Cell size contributes to single-cell proteome variation

Michael C. Lanz[1,2,*], Lucas Valenzuela[1,*], Joshua E. Elias[2], Jan M. Skotheim[1,2,+]

[1]Department of Biology, Stanford University, Stanford CA 94305.

[2]Chan Zuckerberg Biohub, Stanford, CA 94305.

[*]Contributed equally

[+]Address correspondence to skothiem@stanford.edu

Abstract

Accurate measurements of the molecular composition of single cells will be necessary for understanding the relationship between gene expression and function in diverse cell types. One of the most important phenotypes that differs between cells is their size, which was recently shown to be an important determinant of proteome composition in populations of similarly sized cells. We therefore sought to test if the effects of cell size on protein concentrations were also evident in single cell proteomics data. Using the relative concentrations of a set of reference proteins to estimate a cell's DNA-to-cell volume ratio, we found that differences in cell size explain a significant amount of cell-to-cell variance in two published single cell proteome datasets.

Introduction

Individual cells are the basis of life. It is therefore important to develop techniques that accurately quantify the molecular composition of single cells. Extensive progress examining mRNA composition has been achieved at single cell resolution, helping to catalog diverse cell types in multicellular organisms (1-3). Yet, mRNA sequencing gives an incomplete measurement of the state of the cell because diverse post-transcriptional mechanisms also impact gene expression. For example, the correlation between mRNA and protein amounts is complicated by differing translation and degradation rates (4). Moreover, transcriptomic methods are blind to the diverse set of protein modifications that are often key to activity and function. To address the limitations inherent to measuring only mRNA transcripts, single cell proteomic methods have emerged.

Advances in single cell proteomics are driven by increases in measurement sensitivity from a new generation of mass spectrometers (5). In addition to this increased sensitivity, multiplexed peptide labeling approaches enable the measurement of hundreds and sometimes thousands of proteins from single mammalian cells (6-9). Initial experiments have revealed that the proteomes of single cells are influenced by cell cycle phase (5, 10), though it is unclear which other physiological features underlie cell-to-cell proteome heterogeneity. It is important to measure these and other quantifiable sources of proteome variation to better characterize features that are specific to particular cell types and states.

We recently showed that cell size (*i.e.*, the DNA-to-cell volume ratio) is an important determinant of proteome content (11). Contrary to the assumption that most cellular components would remain at constant concentration in cells of different sizes, we found widespread, size-dependent changes in the concentrations of individual proteins (**Figure 1A**). These changes in protein concentration likely reflect, to a large extent, the size-dependent changes in the cellular growth rate (12-15). Importantly, a recent proteome analysis of the NCI60 cancer lines revealed a similar pattern of size-dependent changes to the proteome (16). Thus, regardless of cell type, cell size has an important influence on proteome composition and therefore should contribute to the cell-to-cell heterogeneity in the proteomes of single cells.

Methods

Data curation

For Brunner *et al.*, protein intensities for the individual G1 cells were obtained from PRIDE (ID: PXD024043). G1-labeled columns were extracted from the file named: "20210919_DIANN_SingleCellOutput.pg_matrix.tsv" (DIANN1.8 cell cycle folder). G1 cells without Histone H4 intensity were excluded from the analysis. Also, G1 cells with the fewest number of protein identifications were excluded until a shared set of ~300 proteins were detected in each single cell. This resulted in the reanalysis of 70 of the 93 G1 cell proteomes (Table S1). For Specht *et al.*, a dataframe containing relative protein concentrations for each single cell was downloaded from https://slavovlab.net ("Proteins-processed.csv"). Mock-treated monocytes were extracted from the "Proteins-processed.csv" dataframe using the "sdrf_scope2.tsv" table (Table S2).

Estimation of cell size

For Brunner *et al.*, we estimated the relative cell size for each of the single G1 cells using the "Histone H4 fraction". To calculate the Histone H4 fraction, we divided the intensity value for Histone H4 (H4_HUMAN) by the summed intensity for all other proteins. To calculate this summed value, we only considered the ion intensity from proteins that were identified in all cells considered for our analysis. We chose a single histone protein, rather than the average of all histone protein, to minimize missing values and therefor maximize the number of cells considered for our analysis. Histone H4 was chosen because a single H4 variant was detected in most cells. The use of other core histone proteins or an averaged value produced similar results (Figure S1). For Specht *et al.*, relative cell size was estimated using the relative concentration of Histone H4 ($\log_2$).

The approach with a single reference protein can be extended to several reference proteins. Namely, we select a small number of reference proteins n_r known to scale significantly with cell size, *i.e.*, the absolute value of the measured slope s_p for those selected proteins is relatively large. We therefore construct a reference dataset of protein fractions $X^r \in \mathrm{R}^{\mathrm{n_r} \times \mathrm{N}}$ and its associated measured slope values $s^r \in \mathrm{R}^{n_r}$ and solve the following regression problem:

$$\min_m ||X^r - s^r m^T||_2^2 = \sum_{c,p} \left(X_{c,p}^r - s_p^r m_c\right)^2$$

where $m \in \mathrm{R}^N$ is the vector of cell sizes we want to estimate, and the subscripts p, c refer to a specific protein or cell, respectively. In other words, the approach aims to find the cell sizes m that, for the reference proteins selected, replicate as closely as possible the previously measured slopes (11). The estimated cell sizes allow the estimation of slopes for all the proteins using standard linear regression.

Estimation of proportion of variance attributable to cell size

Subtracting the contribution of cell sizes for each protein in the dataset yields a second dataset $\tilde{X}$ whose total variance is expected to be lower. Indeed, if cell size is a contributor to cell-to-cell proteome variation, removing its effect should decrease the total amount of variance. Denoting the sample covariance matrices of the original dataset and the new dataset with the effect of cell size removed by Σ and $\tilde{\Sigma}$, the total variances V and $\tilde{V}$ are equal to the sum of the eigenvalues of their respective covariance matrices. Therefore, the amount of leftover variance after removing the effect of slopes is given by $\tilde{V}/V$. If the contribution of cell size is meaningful, we expect this ratio to be smaller than 1. We can compare this ratio to the leftover variance after removing the first principal component in the dataset, *i.e.*, λ_{max}/V, where λ_{max} is the maximum eigenvalue of the sample covariance matrix. The first principal component is the direction in feature space that accounts for maximal variance. Comparing these two metrics reveals the amount of variance contributed by cell size.

Principal component analysis

A dataframe was created that contained individual proteins as rows with columns corresponding to single G1 cell proteomes. PCA analysis was performed in Python using the sklearn package. Results of the PCA analysis were visualized with Seaborn's scatterplot.

Pearson r correlation analysis

For Brunner *et al.*, a dataframe containing intensity values of 295 proteins (row) for 70 single cells (column) was converted to intensity fractions. For each cell, the intensity of each protein was divided by the summed

intensity of all proteins to calculate each protein's proteome fraction, an estimate of a protein's relative concentration. A Pearson r correlation (python's scipy package) was calculated by regressing the relative concentration of each individual protein against a proxy for each cell's size (exemplified in Figure 1E and 1F). For Specht *et al.*, we used the $\log_2$ ratio values published by the authors, so the r value was derived from a regression between the relative protein concentration ($\log_2$) versus the relative Histone H4 concentration ($\log_2$). Only the most abundant ~350 proteins were considered for Figure 1H (filtered by peptide detections in our own dataset). Our analysis of the entire Specht *et al.* dataset can be found in Figure S3 and Table S2.

Results

To investigate whether cell size can explain cell-to-cell variations in proteome content, we reanalyzed data from two recently published single-cell proteome datasets (5, 6) (**Table S1 and S2**). One of these utilized Bruker's ultra-high-sensitivity timsTOF SCP (label-free DIA) to measure the proteomes of single HeLa cells that were proceeding through the cell cycle after being synchronized (5). The authors distinguished the cell cycle phase of single cells based on their measurements. To disentangle cell size and cell cycle-related effects, we only considered the proteomes of G1-enriched single cells for our analysis. An eigenvalue analysis of this set of G1 cell proteomes found that the top eigenvalues of the covariance matrix deviate significantly from the Marcenko-Pastur distribution (17), *i.e.*, the distribution of eigenvalues for datasets with no latent variables (**Figure 1C**). This means there is a significant signal in our data despite the noisy nature of single cell proteomics data. To crudely approximate the relative size of each cell, we used the histone proteins because their amount is proportional to the amount of DNA (11, 18-20). Smaller cells therefore possess proportionally higher concentrations of histone proteins than larger cells (**Figure 1B**). We used the fraction of total ion intensity represented by Histone H4 as representative of the inverse of the cell volume (a proxy for cell size). We performed principal component analysis (PCA) on 70 G1-enriched single-cell proteomes from Brunner *et al.,* reasoning that proteins with cell size-dependent abundances could help explain the variance in these cells. The fraction of total ion signal attributable to histone H4 significantly correlated with the first principal component indicating the importance of cell size in contributing to cell-to-cell proteome variation (**Figure 1D**). Other core histone proteins produced similar results (**Figure S1**). In contrast, substituting a histone protein for a common housekeeping enzyme, PGK1, whose concentration is expected to be independent of cell size (11), did not produce a significant correlation (**Figure S2**).

To further explore the relationship between single cell proteome variation and cell size, we calculated Pearson coefficients (r) for each protein from the correlation between its relative protein concentration and Histone H4, a proxy for cell size (**Figure 1E and 1F**). We then correlated these r values with the protein concentration size-dependence previously reported by Lanz *et al.* (**Figure 1G**). Concentration size-dependence was calculated as a protein slope. In brief, the protein slope is calculated from a linear regression between the $\log_2$ of an individual protein's concentration and the $\log_2$ of the cell volume. Thus, a protein slope value of 0 describes proteins for which concentration does not change with cell volume (scaling), a protein slope value of 1 describes proteins for which concentrations increase proportionally to cell volume (super-scaling), and a protein slope value of -1 describes proteins that are perfectly diluted by cell growth such that their concentration is inversely proportional to cell volume (sub-scaling). The Pearson r value correlating concentration and histone H4 derived from single cells was correlated with the previously published protein slope values (**Figure 1G)**. Having established that cell size influences variation in one single cell proteomics dataset, we next sought to examine the robustness of this result in a second dataset. To do this, we repeated this analysis on a second dataset generated using a different single-cell proteomic platform (6). Using SCoPE2, Specht *et al.* distinguished individual monocytes that were or were not differentiated into macrophages. Like the HeLa cells measured by Brunner *et al.* (**Figure 1G**), Pearson regression analysis of single monocyte proteomes produced r values which significantly correlated with the protein slope values (**Figure 1H)**. Taken together, these correlations strongly support the hypothesis that variations in cell size measurably contribute to single-cell proteome variation.

Since single cell proteomics measurements are noisy, we anticipate that there is significant noise in our estimate of cell size using only histone H4. We therefore sought to derive a more robust method for measuring the size-dependent proteome variation in single cell proteome datasets. To do this, we decided to use more than one reference protein to approximate a cell's size. We selected a subset of reference proteins and

reconstructed cell sizes based on a least-squared-error heuristic (**Figure 2A**). Under this framework, the estimated cell size distribution is the one that most closely reproduces the set of measured protein slopes for this subset of proteins. Assuming that protein-to-protein noise is uncorrelated, using more than one reference protein ensures the impact of measurement noise is reduced. The reference proteins were chosen based on their (i) large absolute measured slope value and (ii) large median correlation coefficients with other reference proteins. These criteria ensure that the reference proteins encode meaningful variations from which a signal can be extracted.

Based on our single reference protein analysis in **Figure** 1, the scaling behavior of proteins is expected to be qualitatively conserved across cell types. Using the reconstructed cell size distribution and the single-cell proteome measurements, we compute *single-cell slopes* for each of the remaining proteins in the dataset, which were well correlated with the previously measured slopes (**Figure 2B**). These results are qualitatively similar to the analysis shown in **Figure 1**, but quantitatively more robust due to the use of five reference proteins to estimate cell size. Indeed, a smaller number of rulers yields larger variations in the estimated single cell slopes and cell size distributions (**Figure S4**). The correlation between estimated and measured slopes does not vary significantly when more than five rulers are used (**Figure S5**).

Having found that cell size contributes to single cell proteome variation, we next sought to quantify the extent of this contribution to the total variation in single cell proteomics data. To explore this contribution, we first subtracted the estimated contribution of cell size to each protein concentration from our data. This resulted in a significant decrease in the covariance and correlation coefficient among a set of proteins whose concentrations varied with cell size (**Figure 2C-D**). To assess whether the estimated cell size distribution is the primary source of variance in the dataset, we compared the first principal component coefficients with the imputed *single-cell* slopes (**Figure 2E**). The strong correlation demonstrates that the estimated slopes closely align with the direction of maximum variance. Consistent with this observation, no correlation is observed with the second PC coefficients (**Figure S6**). Finally, the imputed cell size distribution allowed us to estimate the proportion of variance in the dataset that is attributable to cell size (see Methods; **Figure 2F**). The variance due to cell size differences is comparable to the variance attributed to the first principal component, which sets the upper bound for removable variance by a single linear transformation. While using histone H4 as a single reference protein does remove some variance in the dataset, using more reference proteins increases the amount of variance that can be accounted for. We note that using protein slopes derived from the measurement of another cell type (11) yields generally similar results (**Figure S7**). In contrast, using a set of randomly generated protein slopes does not remove any variance in the dataset.

Discussion

While our data demonstrate that cell size is a major contributor to variation in single cell proteomes extracted from cells of the same type, it is important to note that relationship may be more complex when datasets contain very different types of cells, as was recently demonstrated in a melanoma cell line (10). Nevertheless, even in these complex assemblies of cells, we still anticipate there to be some size-dependent signal because several different cell lines exhibit similar size scaling across their proteomes (11, 16). We also note that differences in ploidy across cell types may have a big effect on the proteome. For example, a comparison of the proteomes of 2N, 4N, and 8N cells revealed that they were highly similar despite the near 4-fold increase in cell size (11). Thus, what we refer to as size scaling is more accurately attributed to changes in the cell volume-to-DNA ratio.

In summary, we re-analyzed the proteome heterogeneity in single cell datasets reported by two independent groups using different single-cell preparation and measurement platforms. In both cases, we found that differences in cell size substantially contribute to variance in single cell proteomes. Remarkably, the effects of cell size trended in agreement with a recent report of cell size-dependent changes to the proteome that were measured in bulk for different types of cells (11). Taken together, these analyses support the conclusion that differences in cell size will account for a significant amount of proteome heterogeneity in single cells. We therefore recommend accounting for differences in cell size in future analyses of single cell proteomes.

Acknowledgements

We thank Martin Wühr and members of the Skotheim and Elias labs for feedback on the manuscript. We also thank the authors of Brunner *et al.* and Specht *et al.* for collecting and curating their single cell data in a manner that facilitated our reanalysis.

Funding: Chan Zuckerberg Biohub (collaborative postdoctoral fellowship M.C.L.; investigator award J.M.S.); NIH P01 CA254867.

Data Availability

This study analyzed previously published data.

Author contributions

M.C.L. and L.V. performed data analysis and wrote the manuscript with J.E.E. and J.M.S, who supervised the study.

Declaration of interests

The authors declare no competing interests.

Figures

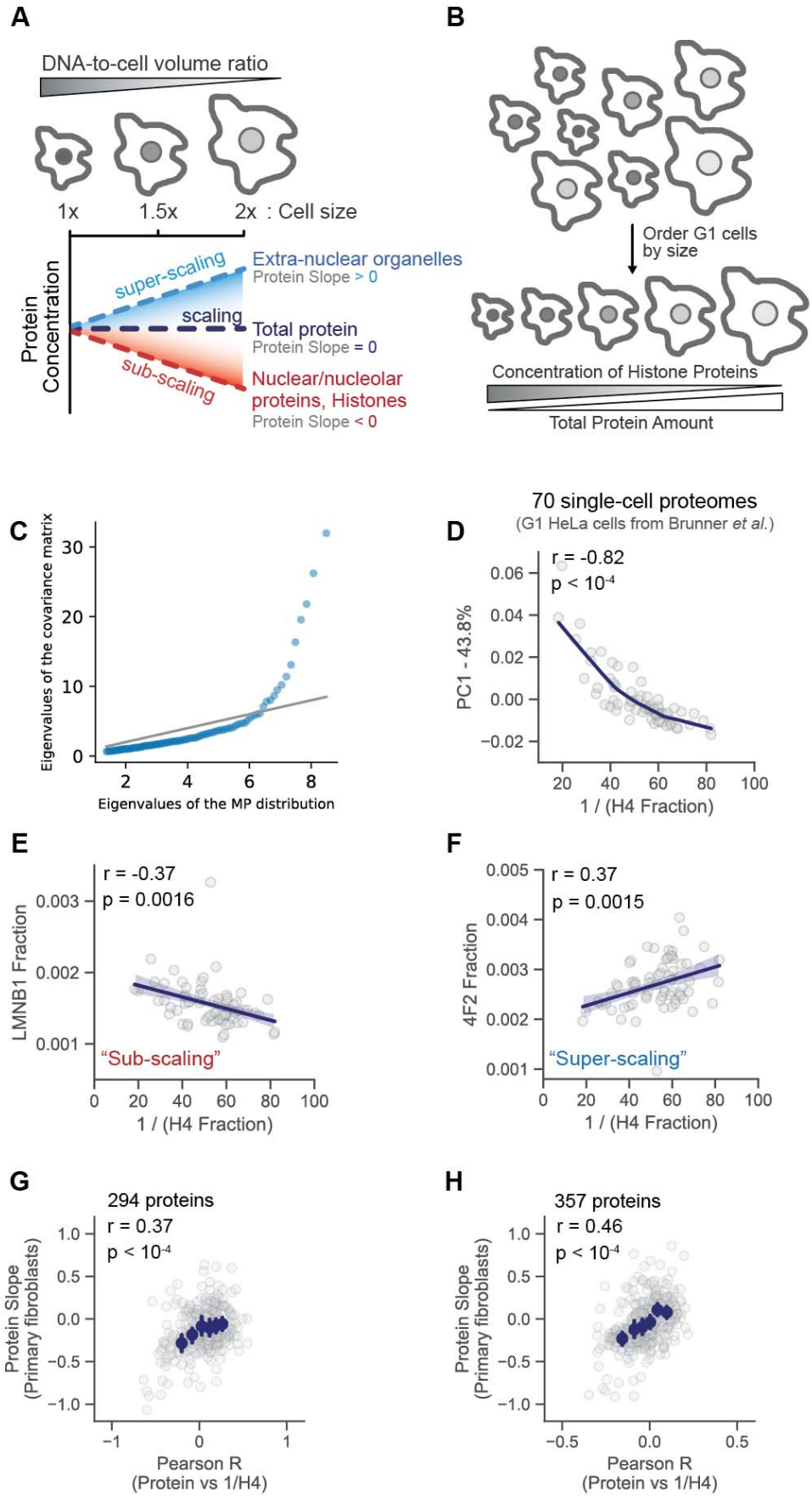
A
DNA-to-cell volume ratio
1x
1.5x
2x : Cell size
Protein Concentration
super-scaling
scaling
sub-scaling
Extra-nuclear organelles
Protein Slope > 0
Total protein
Protein Slope = 0
Nuclear/nucleolar proteins, Histones
Protein Slope < 0
B
Order G1 cells by size
Concentration of Histone Proteins
Total Protein Amount
C
Eigenvalues of the covariance matrix
Eigenvalues of the MP distribution
70 single-cell proteomes
(G1 HeLa cells from Brunner et al.)
D
r = -0.82
p < 10-4
PC1 - 43.8%
1 / (H4 Fraction)
E
r = -0.37
p = 0.0016
LMNB1 Fraction
"Sub-scaling"
1 / (H4 Fraction)
F
r = 0.37
p = 0.0015
4F2 Fraction
"Super-scaling"
1 / (H4 Fraction)
G
294 proteins
r = 0.37
p < 10-4
Protein Slope (Primary fibroblasts)
Pearson R (Protein vs 1/H4)
HeLa cells from Brunner et al.
H
357 proteins
r = 0.46
p < 10-4
Protein Slope (Primary fibroblasts)
Pearson R (Protein vs 1/H4)
Monocytes from Specht et al.

Figure 1 - Cell size contributes to variation in the proteomes of single cells.

A) Proteomes vary with cell size. For example, the amount of histone proteins is maintained in proportion to the genome so that histone concentrations are inversely proportional to cell size. The Protein Slope describes how the concentration of an individual protein scales with cell size (Lanz *et al.*, 2022). Proteins with a slope of 0 maintain a constant cellular concentration regardless of cell volume ("scaling"). A slope value of 1 corresponds to an increase in concentration that is proportional to the increase in volume ("super-scaling"), and a slope of −1 corresponds to dilution (concentration ~ 1/volume; "sub-scaling").

B) Schematic illustrating how relative histone protein concentrations can be used as a proxy for cell size in single cell proteomics datasets in which cell size was not measured.

C) Quantile-quantile plot between the distribution of eigenvalues of the empirical covariance matrix and a sample of the Marcenko-Pastur distribution, which is the distribution expected from uncorrelated, normally distributed random variables. Eigenvalues above the grey identity line indicate the presence of an underlying signal.

D) PCA analysis of 70 single cell proteomes. Each dot represents the proteome of a G1 cell from Brunner *et al.*. The first principal component is plotted against a proxy for G1 cell size (1/H4 Fraction). The fraction of the proteome represented by histone H4 is H4 intensity / summed intensity of all other proteins.

E) and **F)** correlation between increasing G1 cell size (1/H4 Fraction) and the relative concentration (*i.e.*, protein intensity / summed intensity of all other proteins) of two proteins previously found to (E) sub-scale and (F) super-scale with cell size (Lanz *et al.*, 2022).

G) and **H)** A Pearson correlation coefficient was calculated by regressing the relative concentration of each individual protein against a proxy for each cell's size (1 / H4 concentration), as exemplified in (E) and (F). The r value for each protein from the (G) Brunner *et al.* and (H) Specht *et al.* datasets are plotted against the previously measured Protein Slope value (11). Histone H4 was excluded from the plot. Error bars represent the 99% confidence interval. The plot in (H) was filtered to display the most abundant proteins. Figure S3B depicts an unfiltered version of this analysis.

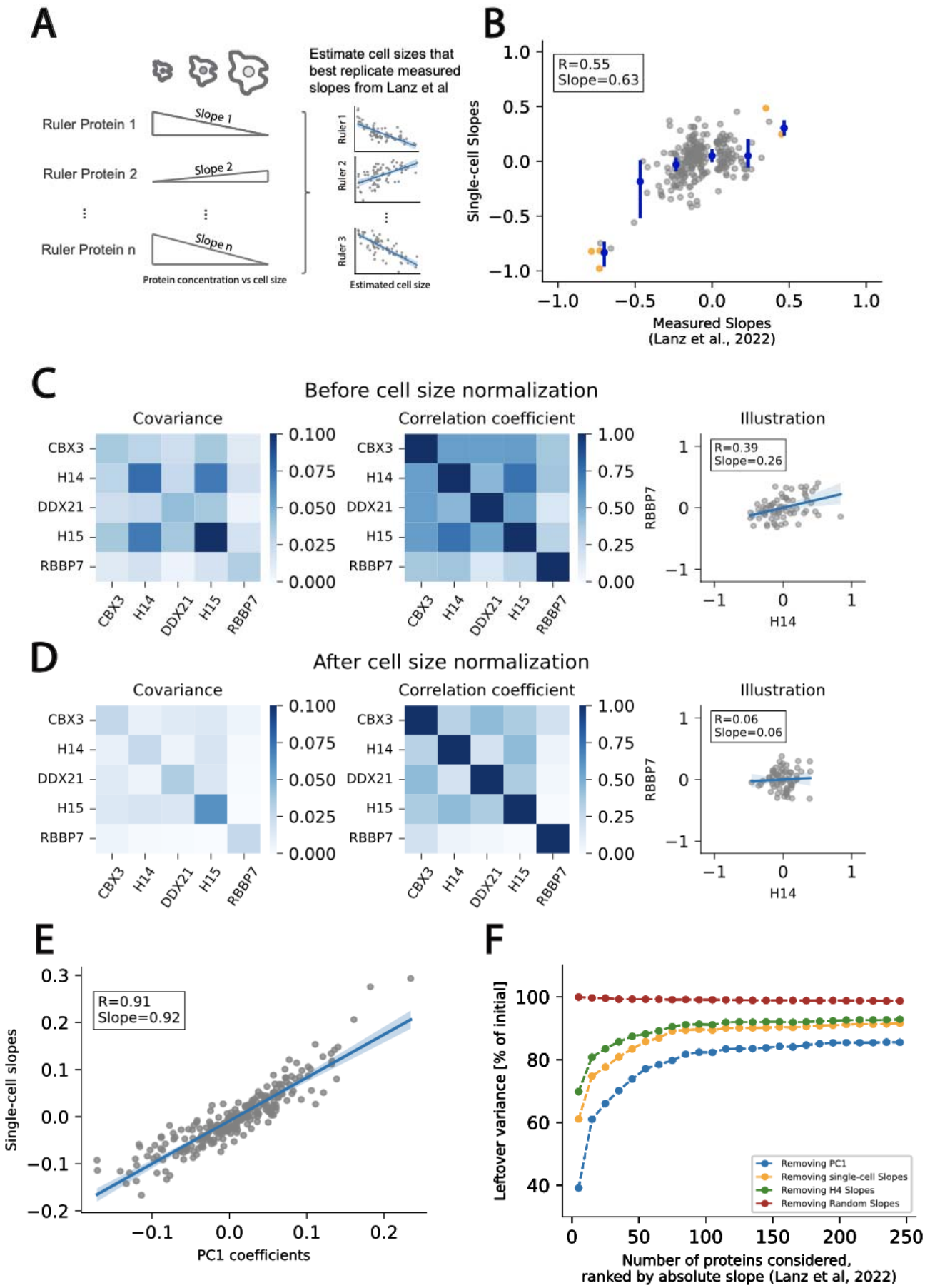
A
Estimate cell sizes that best replicate measured slopes from Lanz et al
Ruler Protein 1
Slope 1
Ruler Protein 2
Slope 2
Ruler Protein n
Slope n
Protein concentration vs cell size
Ruler 1
Ruler 2
Ruler 3
Estimated cell size
B
R=0.55
Slope=0.63
Single-cell Slopes
Measured Slopes
(Lanz et al., 2022)
C
Before cell size normalization
Covariance
Correlation coefficient
Illustration
CBX3
H14
DDX21
H15
RBBP7
R=0.39
Slope=0.26
D
After cell size normalization
R=0.06
Slope=0.06
E
R=0.91
Slope=0.92
Single-cell slopes
PC1 coefficients
F
Leftover variance [% of initial]
Removing PC1
Removing single-cell Slopes
Removing H4 Slopes
Removing Random Slopes
Number of proteins considered,
ranked by absolute slope (Lanz et al, 2022)

Figure 2 – Robust estimation of cell size used to determine the size-dependent variance in single-cell proteomics data.

A) Schematic illustrating the methodology to estimate cell size. We first select a small subset of reference proteins, like histone H4, whose concentrations were shown to be strongly size-dependent (Lanz et al., 2022). Using these reference proteins, and their corresponding size slope values, we performed a least-squares regression to estimate the size of each cell.

B) Having estimated the size of each cell, we then calculate the size slope for each protein in our single cell proteomics data sets. Comparison of slopes estimated via the approach described in this paper and those measured previously (11). Orange dots denote reference proteins and blue dots with error bars denote binned values.

C and D) Comparison of protein concentration covariance and correlation in the initial dataset (**C**) and after removing the estimated effect of cell size (**D**) for a set of 5 proteins with large absolute measured slopes. Removing the estimated effect of cell size reduces the covariance and the correlation coefficient between protein pairs. We illustrate this effect with a given protein pair.

E) Relationship between estimated slopes and the coefficients of the first principal component. Both quantities are very close to each other, indicating the estimated slopes approximate the direction of maximum variance in the dataset.

F) Amount of variance leftover after removing the first principal component (blue), the estimated effect of cell size from measured slopes (orange), the effect of H4 only (green), and the estimated effect of cell size from random slopes (red). The number of proteins included in the analysis (x-axis) was gradually increased based on protein absolute slopes. For example, if 50 proteins were included in the analysis, this set contains the 50 proteins with the highest absolute slopes. The maximum amount of variance removed by cell size is bounded that removed by the first principle component (PC1 blue).

Supplementary information

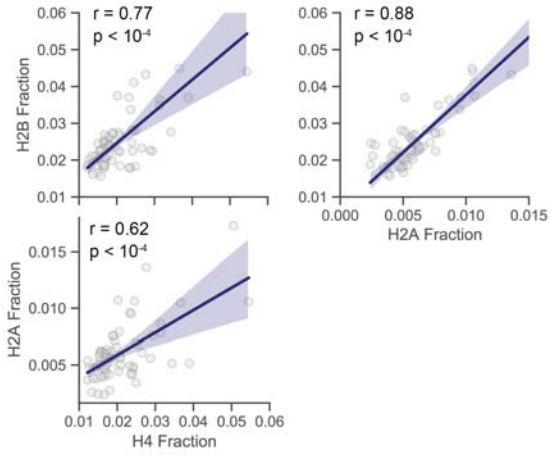

Figure S1: Multiple core histones can be used to estimate cell size.
The concentration of Histone H4 correlated with the concentration of other core histones in single cells. A regression line is plotted in dark blue with 95% confidence intervals. Pearson r value and its associated p-value are shown.

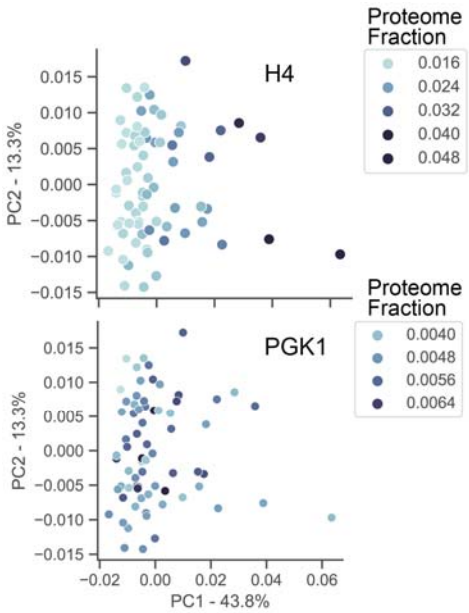

Figure S2: Correlation of principal components with the relative concentration of histone H4 and PGK1, a protein whose concentration is not expected to change with cell size.
Plot of the first two principal components from a PCA analysis of 70 single cell proteomes. Each dot represents the proteome of a G1 cell from Brunner *et al.* and its color indicates the fraction of the proteome represented by either H4 (top) or PGK1 (bottom).

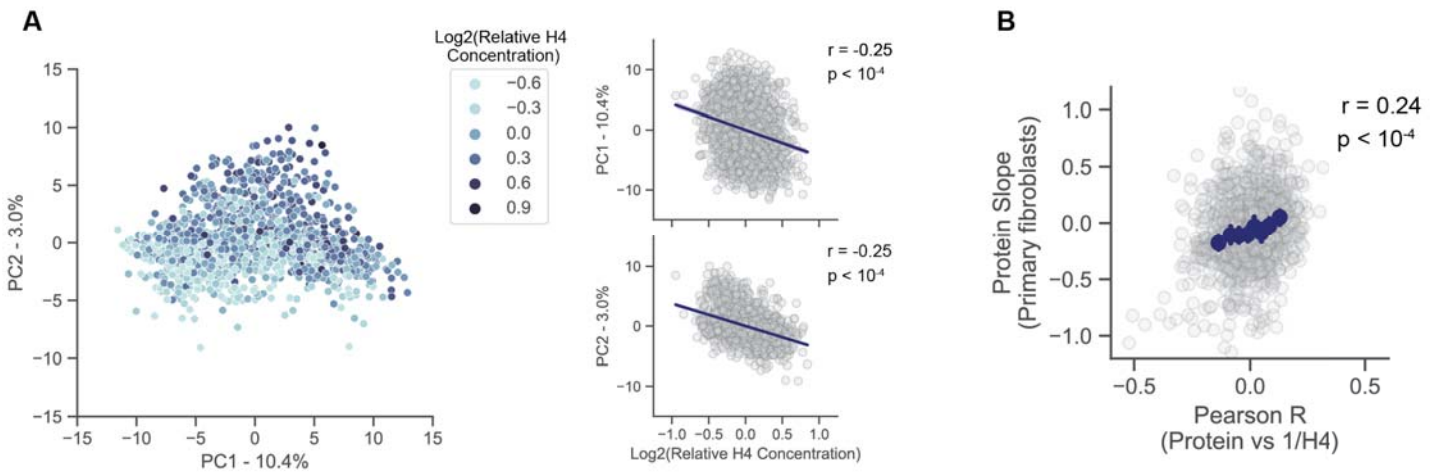

Figure S3: Reanalysis of single cell proteomics data from Specht *et al.*

A) PCA analysis of 1490 single cell proteomes from Specht *et al*. Proteins were quantified using tandem mass tags, so cell size was estimated using the relative concentration of MS2-level reporter ions for Histone H4. Each dot represents a proteome and its color indicates the relative H4 concentration. Correlation between the relative histone concentrations and PC1 and PC2. A regression line is plotted in dark blue with 95% confidence intervals. Pearson r value and its associated p-value are shown.

B) A Pearson correlation coefficient was calculated by regressing the relative concentration of each individual protein against a proxy for each cell's size (histone H4 concentration). The r value for each protein in the Specht *et al.* dataset is plotted against the Protein Slope value (Lanz *et al.*, 2022). Histone H4 was excluded from the plot. Error bars for the binned data represent the 99% confidence interval of the mean. In Figure 1H, only the most abundant proteins are depicted. All proteins identified by Specht *et al.* are included.

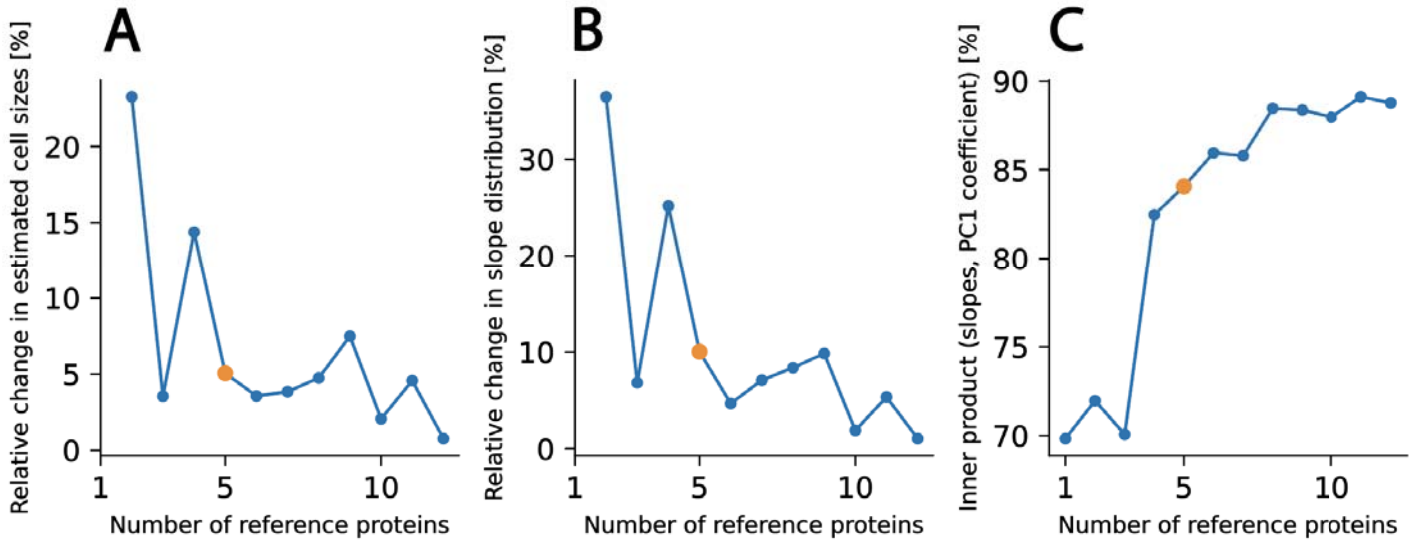

Figure S4 – Metrics used to determine the number of reference proteins

The orange dot represents the number of reference proteins that is presented in the main text. Relative change denotes the value for n reference proteins minus the value for n-1 reference proteins divided by value for n reference proteins.

A) Relative change in inferred estimated cell size distribution as more reference proteins are added.

B) Relative change in estimated slope distribution as more reference proteins are added.

C) Normalized inner product between the slopes and the PC1 coefficient as a function of the number of

reference proteins . Reference proteins were discarded from the dataset before computing this metric. We chose 5 reference proteins because beyond this number changes produced only marginal differences.

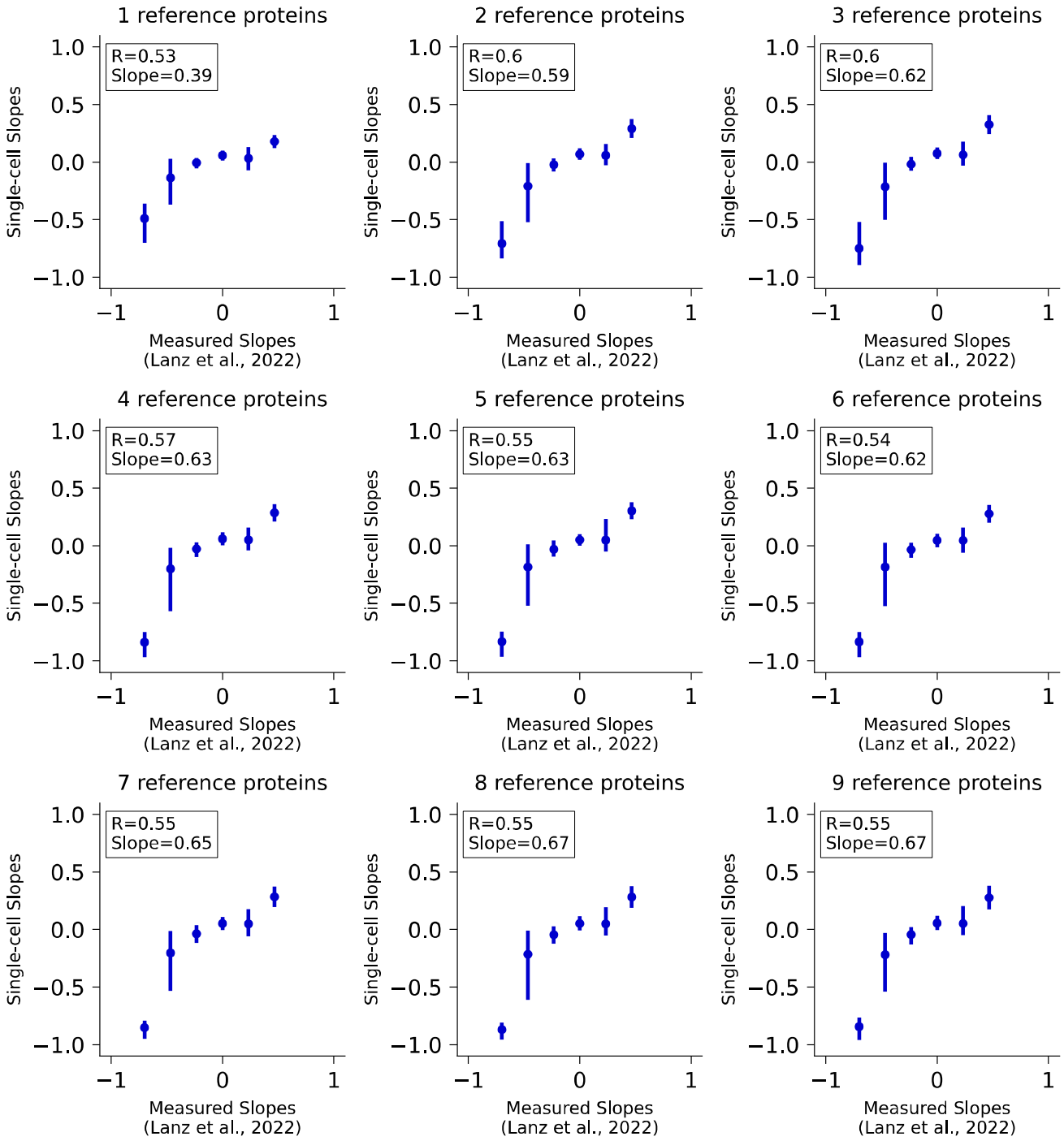

Figure S5 – Impact of increasing number of reference proteins on single cell slopes
Each panel reports the slopes estimated with a given number of reference proteins. Orange dots denote the proteins used as reference and blue dots denote binned data and associated confidence intervals.

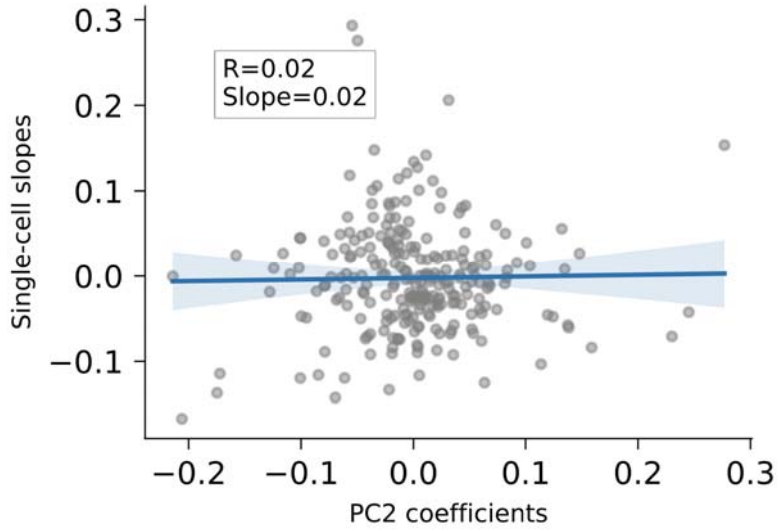

Figure S6 – Correlation of estimated single-cell protein slopes with the coefficients from PC2 from a PCA analysis of *Brunner et al.* as shown for PC1 in Fig. 2E.

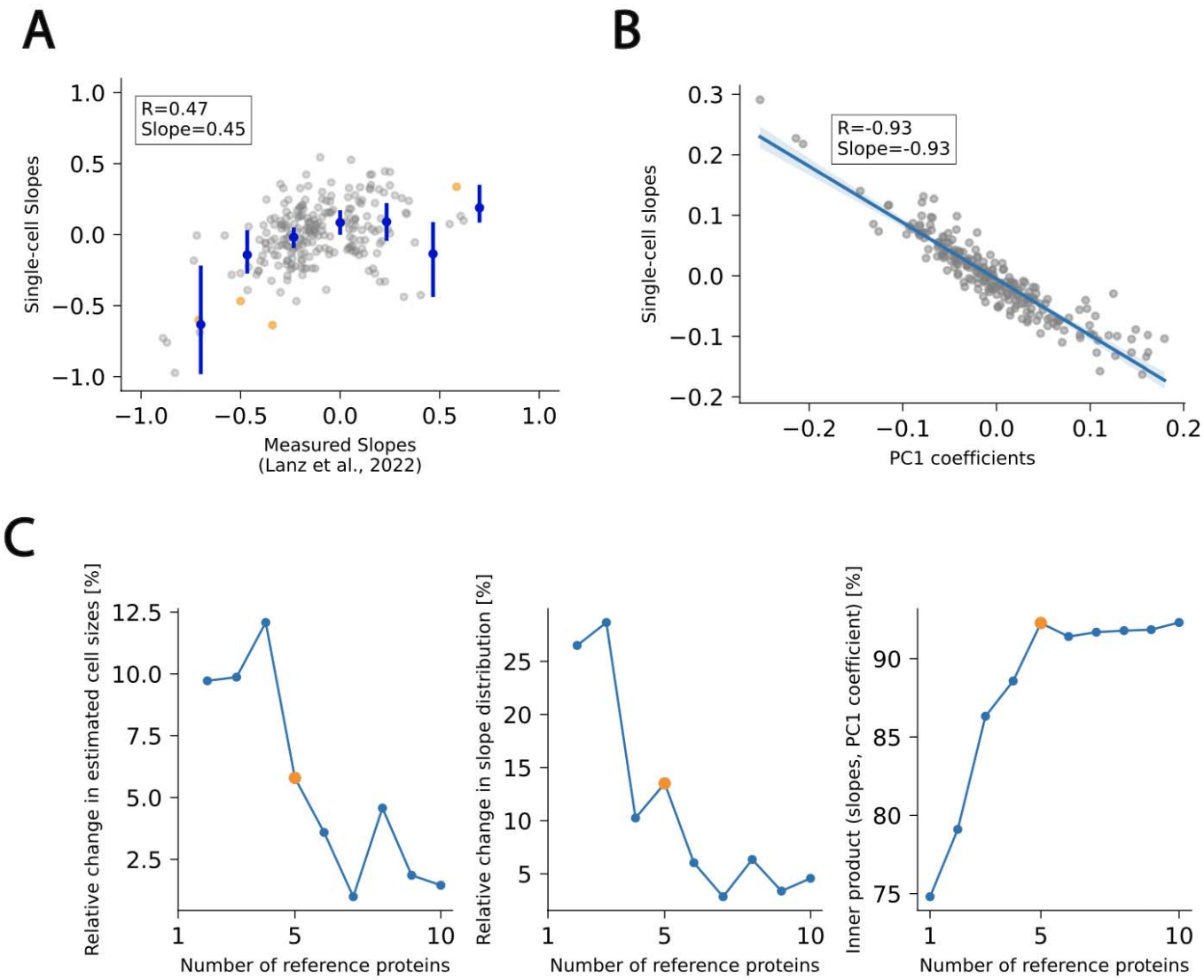

Figure S7 – Analysis of the *Brunner et al.* dataset using HLF protein slopes from *Lanz et al.* rather than RPE1 protein slopes.

A) Comparison of protein slopes estimated via the approach described in this paper and those measured in *Lanz et al*. Orange dots denote the proteins used as reference and blue dots denote binned data and associated confidence intervals.

B) Relationship between estimated slopes and the coefficients of the first principal component.

C) Relevant metrics with increasing number of reference proteins as in **Figure S4**.

38

39

References

1. Tabula Sapiens, C., Jones, R. C., Karkanias, J., Krasnow, M. A., Pisco, A. O., Quake, S. R., Salzman, J., Yosef, N., Bulthaup, B., Brown, P., Harper, W., Hemenez, M., Ponnusamy, R., Salehi, A., Sanagavarapu, B. A., Spallino, E., Aaron, K. A., Concepcion, W., Gardner, J. M., Kelly, B., Neidlinger, N., Wang, Z., Crasta, S., Kolluru, S., Morri, M., Pisco, A. O., Tan, S. Y., Travaglini, K. J., Xu, C., Alcantara-Hernandez, M., Almanzar, N., Antony, J., Beyersdorf, B., Burhan, D., Calcuttawala, K., Carter, M. M., Chan, C. K. F., Chang, C. A., Chang, S., Colville, A., Crasta, S., Culver, R. N., Cvijovic, I., D'Amato, G., Ezran, C., Galdos, F. X., Gillich, A., Goodyer, W. R., Hang, Y., Hayashi, A., Houshdaran, S., Huang, X., Irwin, J. C., Jang, S., Juanico, J. V., Kershner, A. M., Kim, S., Kiss, B., Kolluru, S., Kong, W., Kumar, M. E., Kuo, A. H., Leylek, R., Li, B., Loeb, G. B., Lu, W. J., Mantri, S., Markovic, M., McAlpine, P. L., de Morree, A., Morri, M., Mrouj, K., Mukherjee, S., Muser, T., Neuhofer, P., Nguyen, T. D., Perez, K., Phansalkar, R., Pisco, A. O., Puluca, N., Qi, Z., Rao, P., Raquer-McKay, H., Schaum, N., Scott, B., Seddighzadeh, B., Segal, J., Sen, S., Sikandar, S., Spencer, S. P., Steffes, L. C., Subramaniam, V. R., Swarup, A., Swift, M., Travaglini, K. J., Van Treuren, W., Trimm, E., Veizades, S., Vijayakumar, S., Vo, K. C., Vorperian, S. K., Wang, W., Weinstein, H. N. W., Winkler, J., Wu, T. T. H., Xie, J., Yung, A. R., Zhang, Y., Detweiler, A. M., Mekonen, H., Neff, N. F., Sit, R. V., Tan, M., Yan, J., Bean, G. R., Charu, V., Forgo, E., Martin, B. A., Ozawa, M. G., Silva, O., Tan, S. Y., Toland, A., Vemuri, V. N. P., Afik, S., Awayan, K., Botvinnik, O. B., Byrne, A., Chen, M., Dehghannasiri, R., Detweiler, A. M., Gayoso, A., Granados, A. A., Li, Q., Mahmoudabadi, G., McGeever, A., de Morree, A., Olivieri, J. E., Park, M., Pisco, A. O., Ravikumar, N., Salzman, J., Stanley, G., Swift, M., Tan, M., Tan, W., Tarashansky, A. J., Vanheusden, R., Vorperian, S. K., Wang, P., Wang, S., Xing, G., Xu, C., Yosef, N., Alcantara-Hernandez, M., Antony, J., Chan, C. K. F., Chang, C. A., Colville, A., Crasta, S., Culver, R., Dethlefsen, L., Ezran, C., Gillich, A., Hang, Y., Ho, P. Y., Irwin, J. C., Jang, S., Kershner, A. M., Kong, W., Kumar, M. E., Kuo, A. H., Leylek, R., Liu, S., Loeb, G. B., Lu, W. J., Maltzman, J. S., Metzger, R. J., de Morree, A., Neuhofer, P., Perez, K., Phansalkar, R., Qi, Z., Rao, P., Raquer-McKay, H., Sasagawa, K., Scott, B., Sinha, R., Song, H., Spencer, S. P., Swarup, A., Swift, M., Travaglini, K. J., Trimm, E., Veizades, S., Vijayakumar, S., Wang, B., Wang, W., Winkler, J., Xie, J., Yung, A. R., Artandi, S. E., Beachy, P. A., Clarke, M. F., Giudice, L. C., Huang, F. W., Huang, K. C., Idoyaga, J., Kim, S. K., Krasnow, M., Kuo, C. S., Nguyen, P., Quake, S. R., Rando, T. A., Red-Horse, K., Reiter, J., Relman, D. A., Sonnenburg, J. L., Wang, B., Wu, A., Wu, S. M., and Wyss-Coray, T. (2022) The Tabula Sapiens: A multiple-organ, single-cell transcriptomic atlas of humans. *Science* 376, eabl4896
2. Regev, A., Teichmann, S. A., Lander, E. S., Amit, I., Benoist, C., Birney, E., Bodenmiller, B., Campbell, P., Carninci, P., Clatworthy, M., Clevers, H., Deplancke, B., Dunham, I., Eberwine, J., Eils, R., Enard, W., Farmer, A., Fugger, L., Gottgens, B., Hacohen, N., Haniffa, M., Hemberg, M., Kim, S., Klenerman, P., Kriegstein, A., Lein, E., Linnarsson, S., Lundberg, E., Lundeberg, J., Majumder, P., Marioni, J. C., Merad, M., Mhlanga, M., Nawijn, M., Netea, M., Nolan, G., Pe'er, D., Phillipakis, A., Ponting, C. P., Quake, S., Reik, W., Rozenblatt-Rosen, O., Sanes, J., Satija, R., Schumacher, T. N., Shalek, A., Shapiro, E., Sharma, P., Shin, J. W., Stegle, O., Stratton, M., Stubbington, M. J. T., Theis, F. J., Uhlen, M., van Oudenaarden, A., Wagner, A., Watt, F., Weissman, J., Wold, B., Xavier, R., Yosef, N., and Human Cell Atlas Meeting, P. (2017) The Human Cell Atlas. *Elife* 6
3. Wu, A. R., Neff, N. F., Kalisky, T., Dalerba, P., Treutlein, B., Rothenberg, M. E., Mburu, F. M., Mantalas, G. L., Sim, S., Clarke, M. F., and Quake, S. R. (2014) Quantitative assessment of single-cell RNA-sequencing methods. *Nat Methods* 11, 41-46
4. Liu, Y., Beyer, A., and Aebersold, R. (2016) On the Dependency of Cellular Protein Levels on mRNA Abundance. *Cell* 165, 535-550
5. Brunner, A. D., Thielert, M., Vasilopoulou, C., Ammar, C., Coscia, F., Mund, A., Hoerning, O. B., Bache, N., Apalategui, A., Lubeck, M., Richter, S., Fischer, D. S., Raether, O., Park, M. A., Meier, F., Theis, F. J., and Mann, M. (2022) Ultra-high sensitivity mass spectrometry quantifies single-cell proteome changes upon perturbation. *Mol Syst Biol* 18, e10798
6. Specht, H., Emmott, E., Petelski, A. A., Huffman, R. G., Perlman, D. H., Serra, M., Kharchenko, P., Koller, A., and Slavov, N. (2021) Single-cell proteomic and transcriptomic analysis of macrophage heterogeneity using SCoPE2. *Genome Biol* 22, 50
7. Budnik, B., Levy, E., Harmange, G., and Slavov, N. (2018) SCoPE-MS: mass spectrometry of single mammalian cells quantifies proteome heterogeneity during cell differentiation. *Genome Biol* 19, 161
8. Petelski, A. A., Emmott, E., Leduc, A., Huffman, R. G., Specht, H., Perlman, D. H., and Slavov, N. (2021) Multiplexed single-cell proteomics using SCoPE2. *Nat Protoc* 16, 5398-5425
9. Derks, J., Leduc, A., Wallmann, G., Huffman, R. G., Willetts, M., Khan, S., Specht, H., Ralser, M., Demichev, V., and Slavov, N. (2022) Increasing the throughput of sensitive proteomics by plexDIA. *Nat Biotechnol*

10. Leduc, A., Huffman, R. G., Cantlon, J., Khan, S., and Slavov, N. (2022) Exploring functional protein covariation across single cells using nPOP. *bioRxiv*, 2021.2004.2024.441211

11. Lanz, M. C., Zatulovskiy, E., Swaffer, M. P., Zhang, L., Ilerten, I., Zhang, S., You, D. S., Marinov, G., McAlpine, P., Elias, J. E., and Skotheim, J. M. (2022) Increasing cell size remodels the proteome and promotes senescence. *Mol Cell* 82, 3255-3269 e3258

12. Zatulovskiy, E., Lanz, M. C., Zhang, S., McCarthy, F., Elias, J. E., and Skotheim, J. M. (2022) Delineation of proteome changes driven by cell size and growth rate. *Front Cell Dev Biol* 10, 980721

13. Liu, X., Yan, J., and Kirschner, M. W. (2022) Beyond G1/S regulation: how cell size homeostasis is tightly controlled throughout the cell cycle? *bioRxiv*, 2022.2002.2003.478996

14. Cadart, C., Monnier, S., Grilli, J., Saez, P. J., Srivastava, N., Attia, R., Terriac, E., Baum, B., Cosentino-Lagomarsino, M., and Piel, M. (2018) Size control in mammalian cells involves modulation of both growth rate and cell cycle duration. *Nat Commun* 9, 3275

15. Ginzberg, M. B., Chang, N., D'Souza, H., Patel, N., Kafri, R., and Kirschner, M. W. (2018) Cell size sensing in animal cells coordinates anabolic growth rates and cell cycle progression to maintain cell size uniformity. *Elife* 7

16. Cheng, L., Chen, J., Kong, Y., Tan, C., Kafri, R., and Björklund, M. (2021) Size-scaling promotes senescence-like changes in proteome and organelle content. *bioRxiv*, 2021.2008.2005.455193

17. Johnstone, I. M. (2001) On the distribution of the largest eigenvalue in principal components analysis. *Ann Stat* 29, 295-327

18. Wisniewski, J. R., Hein, M. Y., Cox, J., and Mann, M. (2014) A "proteomic ruler" for protein copy number and concentration estimation without spike-in standards. *Mol Cell Proteomics* 13, 3497-3506

19. Swaffer, M. P., Kim, J., Chandler-Brown, D., Langhinrichs, M., Marinov, G. K., Greenleaf, W. J., Kundaje, A., Schmoller, K. M., and Skotheim, J. M. (2021) Transcriptional and chromatin-based partitioning mechanisms uncouple protein scaling from cell size. *Mol Cell* 81, 4861-4875 e4867

20. Claude, K. L., Bureik, D., Chatzitheodoridou, D., Adarska, P., Singh, A., and Schmoller, K. M. (2021) Transcription coordinates histone amounts and genome content. *Nat Commun* 12, 4202