Cell size contributes to single-cell proteome variation

Michael C. Lanz[1,2], Joshua E. Elias[2], Jan M. Skotheim[1]

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

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

Abstract

Accurate measurements of the molecular composition of single cells will be key to elucidating 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 histone proteins to estimate a cell's DNA-to-cell volume ratio, we found that cell size correlated with the cell-to-cell variance in two single cell proteome datasets, each acquired using different preparation and measurement platforms. Moreover, the proteome differences between small and large single cells significantly correlated with how cell size affects the proteomes of cultured cells measured in bulk. We therefore conclude that cell size accounts for a substantial amount of proteome heterogeneity in single cells and should be considered particularly when comparing cells of a similar type.

Main Text

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 and build increasingly complete cell atlases of multicellular organisms (Regev *et al*, 2017; Tabula Sapiens *et al*, 2022; Wu *et al*, 2014). Yet, mRNA sequencing gives an incomplete measurement of the state of the cell because diverse post-transcriptional mechanisms also impact the proteome. For example, the correlation between mRNA and protein amounts is blurred by differing translation and degradation rates (Liu *et al*, 2016). Moreover, transcriptomic methods are blind to the diverse set of protein modifications that are often key to activity and function. To address this gap and more precisely distinguish discrete cell states, single cell proteomic methods have emerged.

Advances in single cell proteomics are driven by increases in measurement sensitivity in the new generation of mass spectrometers (Brunner *et al*, 2022). In addition to this increased sensitivity, multiplexed peptide labeling approaches enable the measurement of hundreds and sometimes thousands of proteins from single mammalian cells (Budnik *et al*, 2018; Derks *et al*, 2022; Petelski *et al*, 2021; Specht *et al*, 2021). Initial experiments have revealed that the proteomes of single cells (of a singular cell type) are influenced by cell cycle phase (Brunner *et al.*, 2022; Leduc *et al*, 2022), though it is unclear which other features underlie cell-to-cell proteome heterogeneity. It is important to measure these and other quantifiable sources of proteome variation in order to better characterize features specific to cell types and states.

In addition to cell cycle phase, recent work demonstrated that cell size (*i.e.*, the DNA-to-cell volume ratio) is an important determinant of proteome content (Lanz *et al*, 2022). Lanz *et al.* used various methods to isolate different populations of cultured mammalian G1 cells by size and measure differences in proteome content with high accuracy. Contrary to the assumption that most cellular components would remain at constant concentration in cells of different sizes, the authors reported 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 (Cadart *et al*, 2018; Liu *et al*, 2022; Zatulovskiy *et al*, 2022).

Importantly, a recent proteome analysis of the NCI60 cancer lines revealed a similar pattern of size-dependent changes to the proteome (Cheng *et al*, 2021). 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.

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 (Brunner *et al.*, 2022; Specht *et al.*, 2021) (**Table S1 and S2**). One of these utilized Bruker's ultra-high-sensitivity timsTOF SCP to measure (label-free DIA) single HeLa cells that were proceeding through the cell cycle after being synchronized (Brunner *et al.*, 2022). The authors were able to distinguish the cell cycle phase of single cells based on their measurements. To disentangle these cell cycle-related effects, we only considered the proteomes of G1-enriched single cells for our analysis. To approximate the relative size of each cell, we used the histone proteins because their amount is proportional to the amount of DNA (Claude *et al*, 2021; Lanz *et al.*, 2022; Swaffer *et al*, 2021; Wisniewski *et al*, 2014). Smaller cells therefore possess proportionally higher concentrations of histone proteins than larger cells (**Figure 1B**), so 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. As expected, the fraction of total ion signal attributable to histone H4 significantly correlated with both principal components one and two (**Figure 1C and 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 (Lanz *et al.*, 2022), did not produce a significant correlation (**Figure S2**).

To further explore the relationship between single cell proteome variation and cell size, we plotted the weight coefficients for the individual proteins in the PCA analysis against their corresponding Protein Slope values. Protein Slopes values reflect how a given protein's concentration changes with cell size (Lanz *et al.*, 2022). 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. 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 Protein Slope of -1 describes proteins that are perfectly diluted by cell growth such that their concentration is inversely proportional to cell volume (sub-scaling). The weight coefficients for the individual proteins in the principal components PC1 and PC2 both significantly correlated with their Protein Slope values (**Figure 1E**), and the directionality of each correlation agreed with the expected proteome differences between large and small cells. We also calculated Pearson coefficients (r) for each protein from the correlation between its relative protein concentration and a proxy for each cell's size (as exemplified in **Figure 1F and 1G**). These r values also correlated with the Protein Slope values (**Figure 1H**). These complementary analyses are consistent with the hypothesis that variations in cell size measurably contribute to proteome variation in single cells.

To further test our hypothesis that cell size contributes to proteome variation in single cells, we examined a second dataset generated using a different single-cell proteomic approach. Rather than measure one cell at a time, as in Brunner *et al.*, Specht *et al.* used isobaric mass tags and multiplexed labeling of single cells to boost the MS1-level peptide intensity. They also included a "carrier" channel consisting of ~100 cells to further increase the MS1 peptide signal. Relative changes in peptide concentrations between individual cells were then quantified by comparing the intensities of MS2-level reporter ion fragments. Using this method (referred to as SCoPE2), individual monocytes that were or were not differentiated into macrophages could be distinguished (Specht *et al.*, 2021). We analyzed the mock-treated monocytes from that study to better focus on the influence cell size has on proteome variation. PCA analysis produced a pattern similar to the data from Brunner *et al.*, though histone concentration primarily correlated with PC1 (**Figure S3A and S3B**). Strikingly, the PCA and Pearson r analysis of single monocyte proteomes produced coefficients that significantly correlated with the Protein Slope value (**Figure 1I and S3E**), which was derived from the measurement of primary fibroblast cells (Lanz *et al.*, 2022). This further supports the notion that cell size effects are reflected in single cell proteomic datasets. However, the relationship between cell size and proteome content is likely most evident between cells of a similar type and could be complicated by large deviations in cell identity, as was recently demonstrated in a melanoma cell line (Leduc *et al.*, 2022).

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 the concentration of histone proteins, a proxy for cell size, correlated with the primary axes of variance in the cells we considered for our analysis. 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 (Lanz *et al.*, 2022). Taken together, these analyses support the conclusion that differences in cell size (*i.e.*, the DNA-to-cell-volume ratio) will account for a significant amount of proteome heterogeneity in single cells. We recommend accounting for differences in cell size particularly when comparing cells of a similar type. Our analyses also further validate the feasibility of measuring meaningful biological signal using single-cell proteomics and supports the pioneering work of both Brunner *et al.* and Specht *et al.*

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 reported by Brunner *et al.* (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. For Specht *et al.*, relative cell size was estimated using the relative concentration of Histone H4 ($\log_2$). The same method was used to generate the plots in Figure S1 and S3 that show the other core histones.

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. Proteome fraction is a crude estimation 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 VS a proxy for each cell's size (exemplified in Figure 1F and 1G). 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 ($\log_2$) vs the relative Histone H4 concentration ($\log_2$). Only the most abundant ~350 proteins were considered for Figure 1I (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.

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 manner that facilitated our reanalysis.

Funding: Chan Zuckerberg Biohub collaborative postdoctoral fellowship (M.C.L).

Author contributions

M.C.L. performed all data analysis and wrote the manuscript. J.E.E. and J.M.S. supervised the study and helped revise the manuscript.

Declaration of interests

The authors declare no competing interests.

Figures

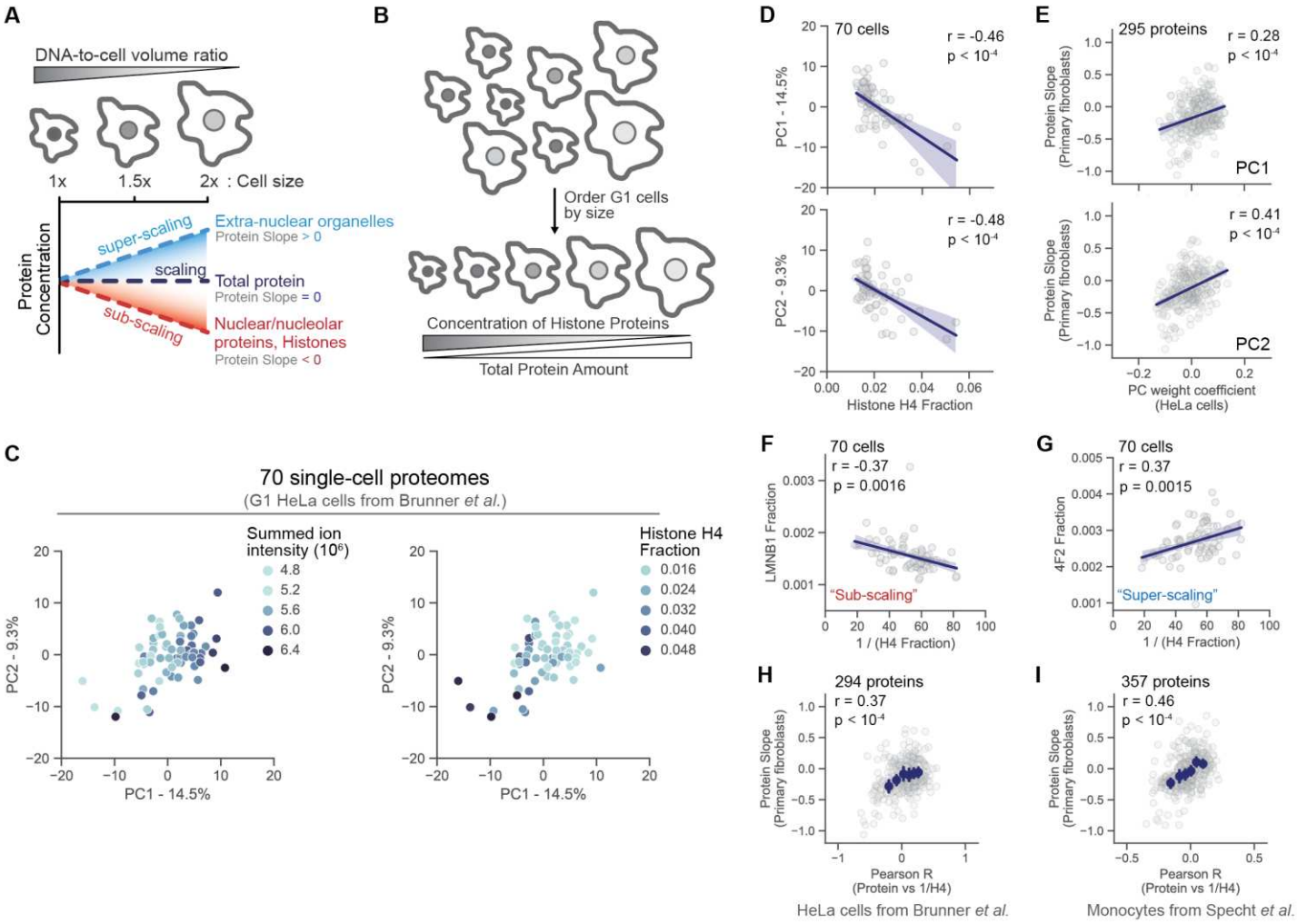

Figure 1 - Cell size contributes to variation in the proteomes of single cells.
A) Proteomes vary with cell size. For example, the amount of histone proteins is maintained in proportion to the genome and thus histone concentrations have a negative slope ("sub-scale") with respect to increasing cell size. The relative histone concentration is therefore 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) Our strategy uses relative histone protein concentrations as a proxy for cell size in single cell proteomics datasets.
C) PCA analysis of 70 single cell proteomes. Each dot represents the proteome of a G1 cell from Brunner *et al.* and its color indicates either (left) the summed ion intensity of all proteins or (right) the fraction of the proteome represented by histone H4 (*i.e.*, H4 intensity / summed intensity of all other proteins).
D) Correlation between the fraction of histone H4 intensity (a proxy for cell size) and PC1 and PC2. A regression line is plotted in dark blue with 95% confidence intervals. Peason r value and its associated p-value are shown.
E) Correlation between the weight coefficients for the individual proteins in PC1 and PC2 with the Protein Slope value as defined in (A). A regression line is plotted in dark blue with 95% confidence intervals. Peason r value and associated p-value are depicted.
F) and **G)** correlation between increasing G1 cell size (1/H4 Fraction) and the relative concentration (*i.e.*, protein intensity / summed intensity of all other proteins) of protein previously found to (F) sub- and (G) super-scale with cell size (Lanz *et al.*, 2022).

H) and **I)** 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 (F) and (G). The r value for each protein from the (H) Brunner *et al.* and (I) Specht *et al.* datasets are plotted against the Protein Slope value. Histone H4 was excluded from the plot. Error bars represent the 99% confidence interval. The plot in (I) was filtered to display the most abundant proteins. Figure S3F depicts an unfiltered version of this analysis.

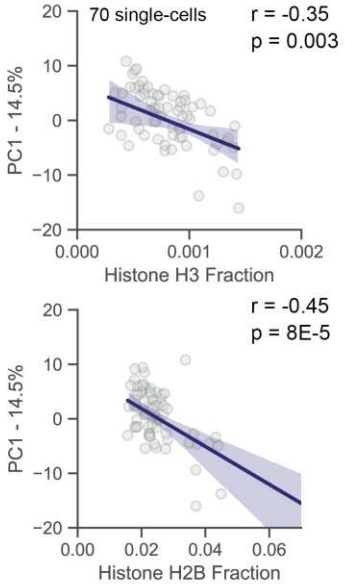

Figure S1: Using other histone proteins as proxies for cell size yields similar results.
In addition to Histone H4, the measurement of other core histone proteins correlates with the variance. A regression line is plotted in dark blue with 95% confidence intervals. Peason r value and its associated p-value are shown.

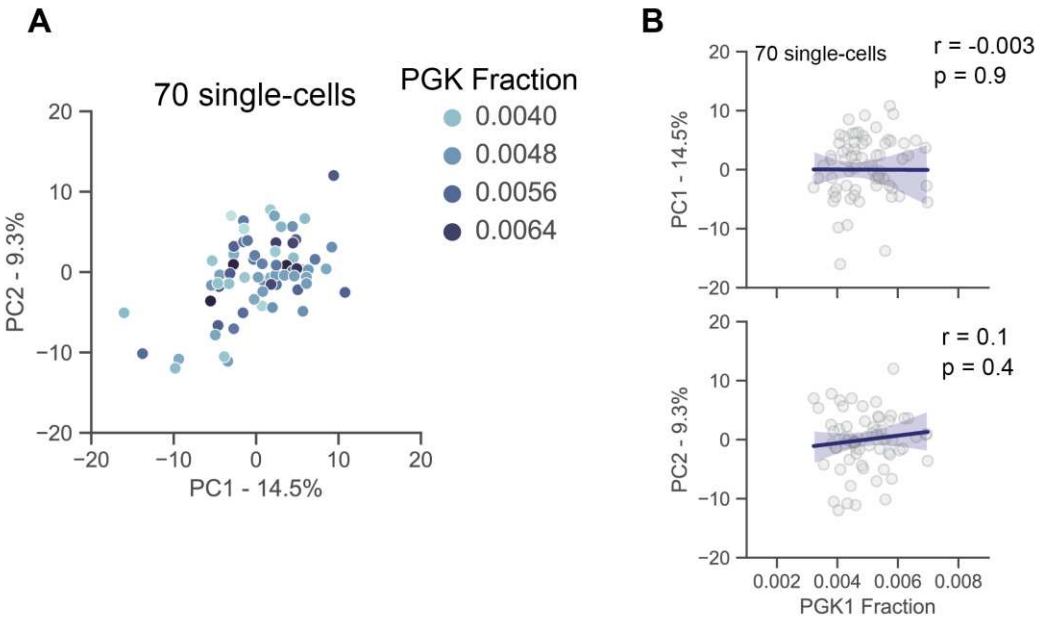

Figure S2: Correlation of principal components with the relative concentration of PGK1, a protein whose concentration is not expected to change with cell size.
A) PCA analysis of 70 single cell proteomes. Each dot represents the proteome of a G1 cell from Brunner *et al.* and its color indicates the fraction of the proteome represented by PGK1 (PGK1 intensity / summed intensity of all other proteins). **B)** Correlation between the fraction of PGK1 intensity and PC1 and PC2. A regression line is plotted in dark blue with 95% confidence intervals. Peason r value and its associated p-value are shown.

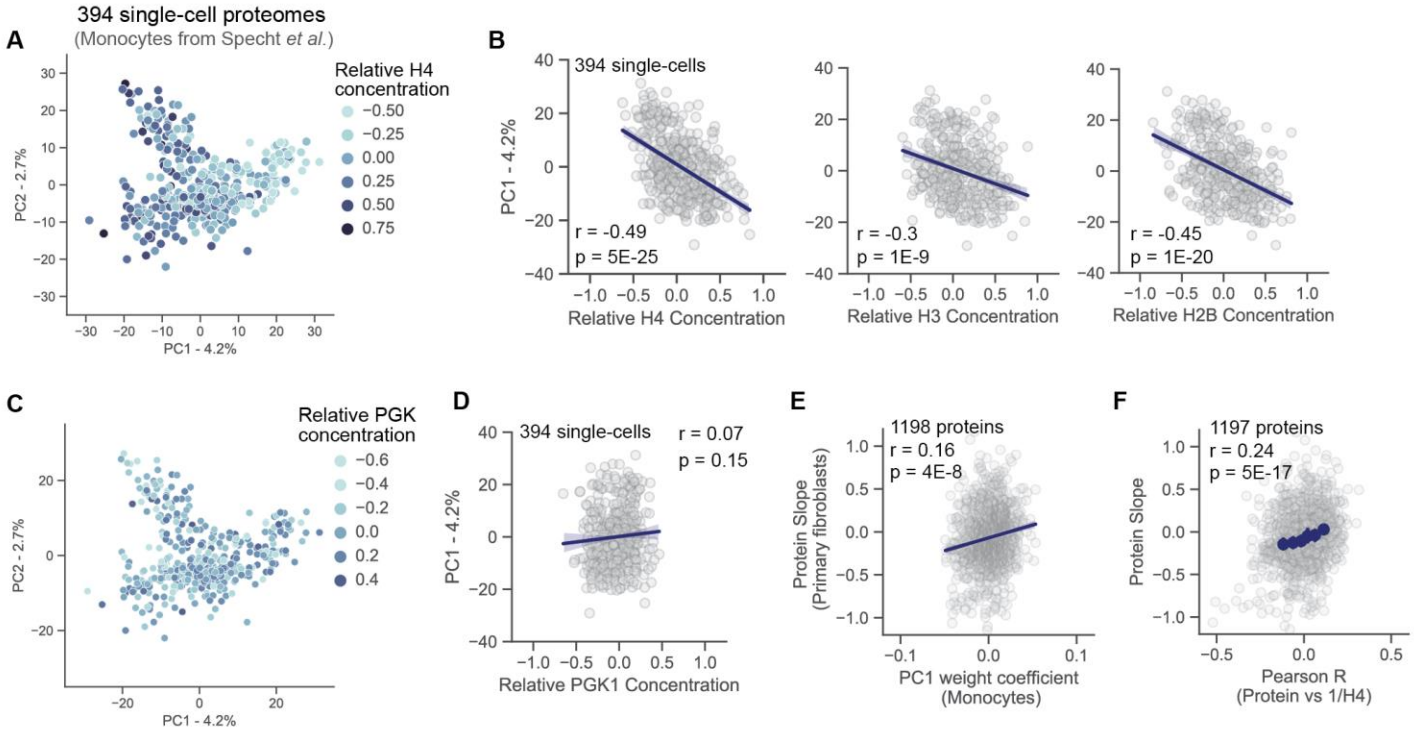

Figure S3: Reanalysis of single cell proteomics data from Specht *et al.*

A) PCA analysis of 394 single cell proteomes. Only the uninduced monocytes from Specht *et al.* are depicted. 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 the proteome of a G1 cell from Brunner *et al.* and its color indicates the relative H4 concentration.

B) Correlation between the relative histone concentrations and PC1 and PC2. A regression line is plotted in dark blue with 95% confidence intervals. Peason r value and its associated p-value are shown.

C) PCA analysis of 394 single cell proteomes.

D) Correlation between the fraction of PGK1 intensity and PC1.

E) Correlation between the weight coefficients for the individual proteins in PC1 and PC2 with the Protein Slope value (Lanz *et al.*, 2022).

F) 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 Spect *et al.* datasets are plotted against the Protein Slope value (Lanz *et al.*, 2022). Histone H4 was excluded from the plot. Error bars represent the 99% confidence interval.

References

Brunner AD, Thielert M, Vasilopoulou C, Ammar C, Coscia F, Mund A, Hoerning OB, Bache N, Apalategui A, Lubeck M *et al* (2022) Ultra-high sensitivity mass spectrometry quantifies single-cell proteome changes upon perturbation. *Mol Syst Biol* 18: e10798

Budnik B, Levy E, Harmange G, Slavov N (2018) SCoPE-MS: mass spectrometry of single mammalian cells quantifies proteome heterogeneity during cell differentiation. *Genome Biol* 19: 161

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

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

Claude KL, Bureik D, Chatzitheodoridou D, Adarska P, Singh A, Schmoller KM (2021) Transcription coordinates histone amounts and genome content. *Nat Commun* 12: 4202

Derks J, Leduc A, Wallmann G, Huffman RG, Willetts M, Khan S, Specht H, Ralser M, Demichev V, Slavov N (2022) Increasing the throughput of sensitive proteomics by plexDIA. *Nat Biotechnol*

Lanz MC, Zatulovskiy E, Swaffer MP, Zhang L, Ilerten I, Zhang S, You DS, Marinov G, McAlpine P, Elias JE *et al* (2022) Increasing cell size remodels the proteome and promotes senescence. *Mol Cell* 82: 3255-3269 e3258

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

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

Liu Y, Beyer A, Aebersold R (2016) On the Dependency of Cellular Protein Levels on mRNA Abundance. *Cell* 165: 535-550

Petelski AA, Emmott E, Leduc A, Huffman RG, Specht H, Perlman DH, Slavov N (2021) Multiplexed single-cell proteomics using SCoPE2. *Nat Protoc* 16: 5398-5425

Regev A, Teichmann SA, Lander ES, Amit I, Benoist C, Birney E, Bodenmiller B, Campbell P, Carninci P, Clatworthy M *et al* (2017) The Human Cell Atlas. *Elife* 6

Specht H, Emmott E, Petelski AA, Huffman RG, Perlman DH, Serra M, Kharchenko P, Koller A, Slavov N (2021) Single-cell proteomic and transcriptomic analysis of macrophage heterogeneity using SCoPE2. *Genome Biol* 22: 50

Swaffer MP, Kim J, Chandler-Brown D, Langhinrichs M, Marinov GK, Greenleaf WJ, Kundaje A, Schmoller KM, Skotheim JM (2021) Transcriptional and chromatin-based partitioning mechanisms uncouple protein scaling from cell size. *Mol Cell* 81: 4861-4875 e4867

Tabula Sapiens C, Jones RC, Karkanias J, Krasnow MA, Pisco AO, Quake SR, Salzman J, Yosef N, Bulthaup B, Brown P *et al* (2022) The Tabula Sapiens: A multiple-organ, single-cell transcriptomic atlas of humans. *Science* 376: eabl4896

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

Wu AR, Neff NF, Kalisky T, Dalerba P, Treutlein B, Rothenberg ME, Mburu FM, Mantalas GL, Sim S, Clarke MF *et al* (2014) Quantitative assessment of single-cell RNA-sequencing methods. *Nat Methods* 11: 41-46

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