Increasing cell size remodels the proteome and promotes senescence

Michael C. Lanz[1,†], Evgeny Zatulovskiy[1,†], Matthew P. Swaffer[1], Lichao Zhang[2], Shuyuan Zhang[1], Dong Shin You[1], Georgi Marinov[3], Patrick McAlpine[2], Josh E. Elias[2], Jan M. Skotheim[1*]

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

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

[3]Department of Genetics, Stanford University, Stanford CA 94305.

[†] These authors contributed equally to this work

[*] Corresponding author. Email: skotheim@stanford.edu

Abstract:

Cell size is tightly controlled in healthy tissues, but it is poorly understood how cell size affects cell physiology. To address this, we measured how the proteome changes with cell size. Protein concentration changes are widespread, depend on the DNA-to-cell size ratio, and are predicted by subcellular localization, size-dependent mRNA concentrations, and protein turnover. As proliferating cells grow larger, concentration changes associated with cellular senescence are increasingly pronounced, suggesting that large size may be a cause rather than just a consequence of cell senescence. Consistent with this hypothesis, larger cells are prone to replicative-, DNA damage-, and CDK4/6i-induced senescence. More broadly, our findings show how cell size could impact many aspects of cell physiology through remodeling the proteome, thereby providing a rationale for cell size control to optimize cell function.

Main Text:

Cells have dedicated mechanisms to control their size, which is one of the most prominent characteristics of distinct cell types (*1-3*). The link between the characteristic cell size and cell function is more obvious at the extremes of the cell size range. Red blood cells and lymphocytes need to be small to squeeze through tight spaces, while macrophages must be larger to engulf a wide range of targets. However, in the middle of the cell size range, including epithelial cells and fibroblasts, the link between cell size and function is unclear. One possibility is that these cells control their size to enhance proliferation, as their rapid turnover is a key part of their physiological function (*4, 5*). Yet, even if these cells are optimized for growth and proliferation, as has been indicated (*4, 6*), it is unclear why there would be an optimal cell size. As cells get larger, it has long been assumed that most proteins and RNA remain at constant concentrations (*7-14*), and organelle volumes, such as the nucleus, increase in direct proportion to cell size (*15-17*). Should protein and RNA concentrations remain constant, larger cells should be capable of proportionally more biosynthesis. However, this is not the case. There is a limit to the size range of efficient biosynthesis (*18*), and excessively large cells exhibit loss of mitochondrial potential (*5*), dilution of the cytoplasm (*6*), and reduced proliferation (*19*). Moreover, recent work has demonstrated the remarkable effect even small variations in cell size can have on hematopoietic stem cell proliferation (*4*).

One possible explanation for why there is an optimal cell size for biosynthesis would be if many key cellular proteins did not remain at constant concentration as cells grew. Then, the further cells got from their target size, the more concentrations of these proteins would change, and the more growth and metabolism would deviate from the optimum. Intriguingly, investigations of the mechanisms cells use to control their size have identified a class of proteins whose concentrations do change with cell size. In budding yeast, human, and plant cells, key cell cycle inhibitors are not synthesized in proportion to cell size so that they are diluted by cell growth, a behavior defined as sub-scaling (*20-22*) (**Fig. 1A**). Larger cells therefore have lower concentrations of cell cycle inhibitors, which promotes their division. In

fission yeast, division is in part promoted by a size-dependent increase in the concentration of a cell cycle activator (*23*), a behavior defined as super-scaling (*24*) (**Fig. 1A**). If this phenomenon of size-dependent protein concentration changes were widespread across the proteome it would provide an explanation for why there is an optimal cell size. This is because the further a cell would be from its target size, the further its intracellular protein concentrations would be from their optimum. To explore this possibility, we measured how the proteome changes as a function of the natural cell size variation in asynchronously proliferating cells.

To measure the scaling behavior of the human proteome (**Fig. 1A**), we developed a method based on triple-SILAC (Stable Isotope Labeling by Amino acids In Cell Culture) proteomics (described in **fig. S1**), which enables the simultaneous measurement of thousands of individual proteins that collectively associate will all major cellular components. Asynchronously proliferating SILAC-labeled primary human lung fibroblasts (HLFs) were gated for G1 DNA content and sorted into three size bins (small, medium, and large) using fluorescence-activated cell sorting (FACS) (**Fig. 1B**). SILAC labeling orientation was swapped for replicate experiments, and the attainment of differentially sized G1 cells was confirmed using a Coulter counter (**Fig. 1C, fig. S2A**). The SILAC channel intensities within a peptide "triplet" represent relative peptide concentrations (**fig. S1A**), so we used the behavior of multiple independent peptide triplets to determine the size scaling relationship for individual proteins (**Fig. 1D**). Rather than measure SILAC ratios, as is typically done, we calculated a slope value to describe the size scaling behavior of each peptide triplet (**Fig. 1E, fig. S1D-G**). Peptides triplets with positive and negative slopes represent peptides (**Fig. 1F**), and therefore proteins (**Fig. 1G**), whose concentrations are super- and sub-scaling, respectively, whereas triplets with a slope value near 0 maintain an approximately constant concentration as function of G1 cell size.

We observed a continuum of size scaling behaviors across the proteome, which spanned slope values from -1 to 1 indicating that a 2-fold increase in cell size can lead to a 2-fold increase or decrease in concentration for a given protein (**Fig. 1H, Data S1**). Several chromatin-associated High Mobility Group

proteins (HMGs), including HMGB1, are diluted in larger cells and so exhibit sub-scaling behavior like what has been previously described for RB/Whi5 and histones (*21, 26, 27*) (**Fig. 1D**). We also identified a diverse set of super-scaling proteins, such as VAT1, whose concentration increases as a function of cell size (**Fig. 1D**). Setting a requirement for multiple independent measurements per protein significantly improved the correspondence between replicate experiments and yielded a high-confidence set of ~1,500 proteins, each having at least 4 distinct peptide measurements (**Fig. 1H, Data S1**). The size-scaling relationship was consistent across all size bins, indicating that similar concentration changes take place when cell size increases from small to medium as when it increases from medium to large (**fig. S3**). Next, we validated candidate super- and sub-scaling behaviors for a subset of proteins using immunofluorescence combined with flow cytometry (**Fig. 1I, fig. S4**). We also confirmed that the process of cell sorting did not affect our proteome measurements (**fig. S5A**) and confirmed that isolation of different sized cells using SSC (Side Scatter) or the total protein dye CFSE (Carboxyfluorescein succinimidyl ester) as proxies for cell size (*14, 28*) yielded similar results (**fig. S5B-D, Data S1**). Since larger G1 cells have, on average, spent more time in G1, it is possible that the protein concentration changes we observe reflect time in G1 rather than cell size. To control for this possibility, we selected different sized G1 cells that were synchronously released from a Thymidine-Nocodazole cell cycle arrest. We found a highly similar size-scaling relationship in our synchronous and asynchronous experiments (**fig. S6A, Data S1**), indicating that cell size, not time in the cell cycle, is driving changes to the proteome. Finally, to ensure that the size-dependent proteome changes we measured in asynchronously dividing cells were not the direct result of replication-related stresses accumulated over many division cycles (*29*), we confirmed that long-term treatment with the CDK4/6 inhibitor Palbociclib, which synchronously enlarges cells over time via G1 arrest (*6, 30*), recapitulated the proteome changes we observed in sorted cells (**fig. S6B, Data S2).**

To examine how cell size may influence different aspects of cellular biology, we asked if different groups of related or interacting proteins exhibited similar size scaling behavior. Indeed, this was the case. For

example, all 17 histone variants sub-scaled and had slope values near -1 (**Fig. 2A**), indicating that histones are proportionally diluted by cell growth in G1 just like the genome (*26, 27*). Moreover, 5 of the 6 detected cathepsin proteases were strongly super-scaling and nearly doubled in concentration from the smallest to the largest cell size bin. Interestingly, we found that all individual protein subunits of the ribosome and proteasome showed small but highly consistent sub- and super-scaling behaviors respectively (**Fig. 2A**). We also re-confirmed the sub-scaling behavior of RB (*21*) **(fig. S5E, S7E).**

We next assessed how size-scaling behavior relates to subcellular location. To do this, we first annotated the proteome based on a strict association with a single, major cellular compartment. Subcellular localization was a strong predictor of size scaling behavior, with ER- and lysosome-resident proteins becoming increasingly concentrated with size, and nucleoplasmic/nucleolar proteins becoming more dilute (**Fig. 2B, fig. S7A**). We observed no clear difference in scaling behavior between luminal- and membrane-associated organelle proteins (**fig. S8**). We confirmed lysosome super-scaling by measuring the lysosome-labelling dye Lysotracker and the lysosomal protein LAMP1 (**fig. S7C**). A less stringent 2D annotation enrichment revealed a range of intermediate scaling behaviors across multiple different cellular compartments (**Fig 2C, Data S1**). Taken together, these analyses refute the commonly held assumption that the protein content of cellular components scales in uniform proportion to cell volume. While total protein concentration may be largely constant within a cell's natural size range, this is not necessarily true for individual proteins.

Next, having identified size-dependent changes in concentrations of individual proteins, we sought to investigate the underlying mechanisms. We first tested whether changes in a protein's concentration are explained by changes in the concentration of the corresponding mRNA. To do this, we performed RNA-seq on size-sorted G1 cells (**Data S3**) and calculated mRNA slope values analogous to the protein slopes described in Figure 1. We found that although there was significant correlation between the protein and RNA slopes, there was large variability in protein scaling slopes for any given RNA slope (**Fig. 2D-E**). Overall, changes in mRNA concentrations explained only a minority of the variation in protein

concentration size-dependency (*31*). We therefore explored additional variables using a series of linear models. In addition to mRNA concentration, we included protein turnover (*32*), mRNA codon affinity, and subcellular localization as variables to predict the protein concentration size-dependence. Iterative incorporation of each parameter significantly improved the model (**Fig. 2F, fig. S9**). Using the correlation between biological replicates as a benchmark, we conclude that our composite model predicts ~60% of the size-dependent variance in protein concentration and strongly supports the conclusion that the size-dependency of protein concentration is regulated both pre- and post-transcriptionally. Importantly, the size-dependency of the proteome is not specific to a single human cell type, as we found a striking degree of similarity in the proteome scaling of primary lung fibroblasts (HLF) and immortalized epithelial cells (*hTERT RPE-1*) (**Fig. 2G, fig. S6**).

Curiously, increasing cell size in proliferating cells is accompanied by proteomic changes normally associated with cell senescence, including the upregulation of beta-galactosidase, lysosomal proteins, and metalloproteases, and the downregulation of Ki67, HMGB proteins, and LaminB (*33*) (**Fig. 3A, fig. S7D**). Senescence-associated secretory phenotype (SASP) proteins (*34*) were also super scaling (**Fig. 3B**). While large cell size is associated with senescence (*33*), it has generally been thought that large size results from a senescent cell's inability to divide while at the same time maintaining cell growth. However, our experiments indicate that increasing cell size itself results in proteome changes that gradually approach those found in senescent cells and support the hypothesis that cell size *per se* promotes senescence (*4, 6*) (**Fig. 3C**). This is consistent with earlier reports that continued cell growth and hypertrophy are required to induce senescence in cell cycle arrested cells (*19*).

To test if large cell size contributes to senescence, we used primary human lung fibroblasts (HLF) that naturally senesce after 10-15 passages. We sorted HLFs into 4 cell size bins at passage #8, re-plated the cells, and then cultured them for an additional 5 passages (**Fig. S10B**). Cells that were larger at the time of sorting started exhibiting high levels of senescence-associated beta-galactosidase (SA-beta-Gal) staining, which is a commonly used marker of senescence, sooner that the cells that were smaller at the time of

sorting (**Fig. 3D,E**). Next, we sought to examine how cell size contributes to senescence in telomerase-immortalized human retinal pigment epithelium (*hTERT RPE-1*) cells exposed to a low dose of a DNA damaging agent (10 ng/ml Doxorubicin), which causes a stress similar to telomere shortening in primary fibroblasts. We note the proteome's size-dependency in untreated G1 *hTERT RPE-1* cells is very similar to that of HLFs (**Fig. 2G, fig. S7**). Consistent with the previous result, larger cells increasingly induced the senescence marker SA-beta-Gal upon Doxorubicin treatment (**Fig. 3F, fig. S10C**). Moreover, co-treatment with the mTOR inhibitor Rapamycin, which reduces cell size, significantly attenuates the SA-beta-Gal staining induced by prolonged treatment with Doxorubicin or the Cdk4/6 inhibitor Palbociclib (**Fig. 3G, fig. S10D,E**), a finding consistent with previous work (*35*).

So far, we have only assayed senescence using indirect markers rather than the durable cell cycle arrest that defines senescence (*36*). We therefore sought to more directly test whether large cell size inhibits cell cycle re-entry following the withdrawal of two treatments previously shown to induce senescence (*6*). We treated *hTERT RPE-1* cells with Palbociclib or a low-dose of Doxorubicin for 4 days to induce senescence in a fraction of the cell population. We then washed out the drugs and imaged these enlarged cells for 4 additional days to determine which cells re-entered the cell cycle and which cells remained durably arrested for the duration of the experiment (**Fig. 3H**). Importantly, Palbociclib or Doxorubicin treatment arrests cells without stopping cell growth, and thus generates a population of cells with a range of large cell sizes that exhibit senescent features (*6*). Since all these cells are exposed to drug for the same duration, we can isolate cell size as a determining factor for durable cell cycle arrest. We found that the cells that were smaller at the time of drug washout are more likely to re-enter the cell cycle (**Fig. 3H)**. This supports the hypothesis that larger size makes cells more prone to senescence and shows that the normal relationship between cell size and division is inverted above a critical cell size, which has also been shown for excessively large budding yeast (*6*). Large cell size may help explain the durable cell cycle arrest in response to transient exposure to a DNA damaging agent. This is because the excessive cell growth that occurs while cells delay cell cycle progression to repair DNA can inhibit cell cycle re-entry.

In addition, it is possible that large cells have difficulty in replicating and repairing DNA since most DNA repair and replication factors subscale with cell size (**fig. S11A**). Crucially, we are not stating that large cell size is the only driver of senescence, but rather that the large size characteristic of senescent cells may itself further promote this state. Thus, while increasing cell size typically drives cell division in small, and intermediate-sized cells, excessively large cell size may inhibit cell division and thereby promotes senescence. Consistent with this hypothesis, mutations to the Rb-family proteins that reduce cell size prevent senescence of mouse embryonic fibroblasts and hematopoietic stem cells (*4, 37, 38*).

Our data so far are consistent with the hypothesis that large cell size inhibits cell division, and this may be due to widespread changes in the concentrations of individual proteins as cells grow. Thus, the further cells are from their target size, the more protein concentrations deviate from their optimum. However, in apparent contradiction to this model, there are many examples of large animal cells that proceed through the cell cycle and maintain highly efficient cell growth (*39, 40*). In many cases, such large cells are polyploid and thus do not have an aberrant cell size-to-ploidy ratio. We therefore sought to test whether protein scaling is determined by cell size per se or by the cell size-to-ploidy ratio. To do this, we compared the proteomes of diploid, tetraploid, and octoploid (2N, 4N, 8N) G1-phase *hTERT RPE-1* cells (**Fig. 4**). To obtain cells with different ploidies, we induced endoreduplication using a moderate dose of the Aurora kinase B inhibitor barasertib (75nM, 48 hours) (*39*) and sorted 2N, 4N, and 8N G1 cells from a mixed population using FACS (**Fig.4A**). While ploidy increases are accompanied by a corresponding increase in cell size (**Fig.4A**), proteomic analysis showed that the ploidy-sorted cells do not exhibit the same proteome changes as size-sorted cells with 2N ploidy (**Fig.4B**). Taken together, these results indicate that the proteome's size-dependency is largely due to changes in the cell size-to-ploidy ratio. Thus, despite being different sizes, cells of different ploidy maintain similar protein concentrations, which may be why the excessively large size of polyploid cells does not inhibit their cell cycle progression.

Our results here shed some light on the phenomenon of cellular senescence (*33, 36, 41*). While the senescent cell state has been associated with large cell size, this was mostly thought to be a passive

consequence of continuing biosynthesis after cell cycle arrest. However, that such large cell sizes inhibit cell division implies the relationship between cell size and senescence can also be inverted (*4, 6*). Taken together, our work suggests a model where large cell size can drive widespread changes in protein concentrations away from their optimum, which, through a yet unknown mechanism, inhibits cell division to reinforce senescence.

While it has long been thought that most protein concentrations remain constant as cells grow, this paradigm had not previously been tested using a high-throughput quantitative proteomics approach. In contradiction to the previous paradigm, many protein concentrations changed with cell size. Some proteins sub-scaled with cell size, and were diluted in larger cells, while others super-scaled with cell size so that their concentrations increased as cells grew larger. This finding is reflected in the super- and sub-scaling of the mRNA transcripts for various G1/S regulators in budding yeast (*24*). To a large extent, these diverse protein size-scaling behaviors we observed could be predicted from a linear model based on mRNA concentration, protein half-life, and subcellular localization, which indicates the importance of both transcriptional and post-transcriptional size-scaling mechanisms.

Our observation here that most protein concentrations change as cells grow provides a rationale for why many cells control their size. If the proteome content that supports optimal cell growth is only found near the target cell size, then the further a cell deviates from its target size, the further protein concentrations will be from their growth-supporting optimum. While a small change in the concentration of a single protein may not significantly affect cell physiology, the cumulative effect of thousands of small protein concentration changes could account for the drastic drop in the efficiency of biosynthesis in large cells (*4, 6, 19*). Moreover, since proteome concentrations mostly follow the cell size-to-ploidy ratio rather than cell size *per se*, polyploidization is an elegant mechanism for organisms to generate large cells capable of efficient protein synthesis as is commonly found in nature.

Acknowledgements:

We thank Shicong Xie for assistance with python-related data analysis and Marcus Smolka for use of his HPLC for HILIC pre-fractionation. We thank Robert Brooks, Chuck Sherr, Christine Jacobs-Wagner, Marcus Smolka, Jette Lengefeld, and members of the Skotheim lab for helpful feedback on the manuscript. We also thank Frank McCarty and Stanford SUMS for assistance in mass spec related data acquisition. Cell sorting was performed on an instrument in the Shared FACS Facility purchased by Parker Institute for Cancer Immunotherapy.

Funding:

National Institutes of Health R35 grant GM134858 (JMS)

National Institutes of Health F32 grant (ML)

Author contributions:

EZ conceived the project. ML and EZ designed the study. EZ performed and analyzed all cell biology and flow cytometry experiments, prepared and sorted cells for proteomics. ML processed all proteomics samples. ML, LZ, and JE conceived data acquisition strategy. LZ and PM acquired mass spectrometry data. ML analyzed all proteomics data. ML and SZ performed the linear modeling. EZ, MS and GM analyzed the RNAseq data. DY analyzed lysosome scaling with cell size. JMS supervised the study. ML, EZ and JMS interpreted the data and wrote the manuscript.

Competing interests: Authors declare that they have no competing interests.

Data and materials availability: The proteomic datasets produced in this study have been deposited to the PRIDE database (https://www.ebi.ac.uk/pride/archive/) and assigned the identifier PXD-------.

RNA-seq data produced in this study have been deposited at the GEO repository and assigned the accession number GSE------ (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE------).

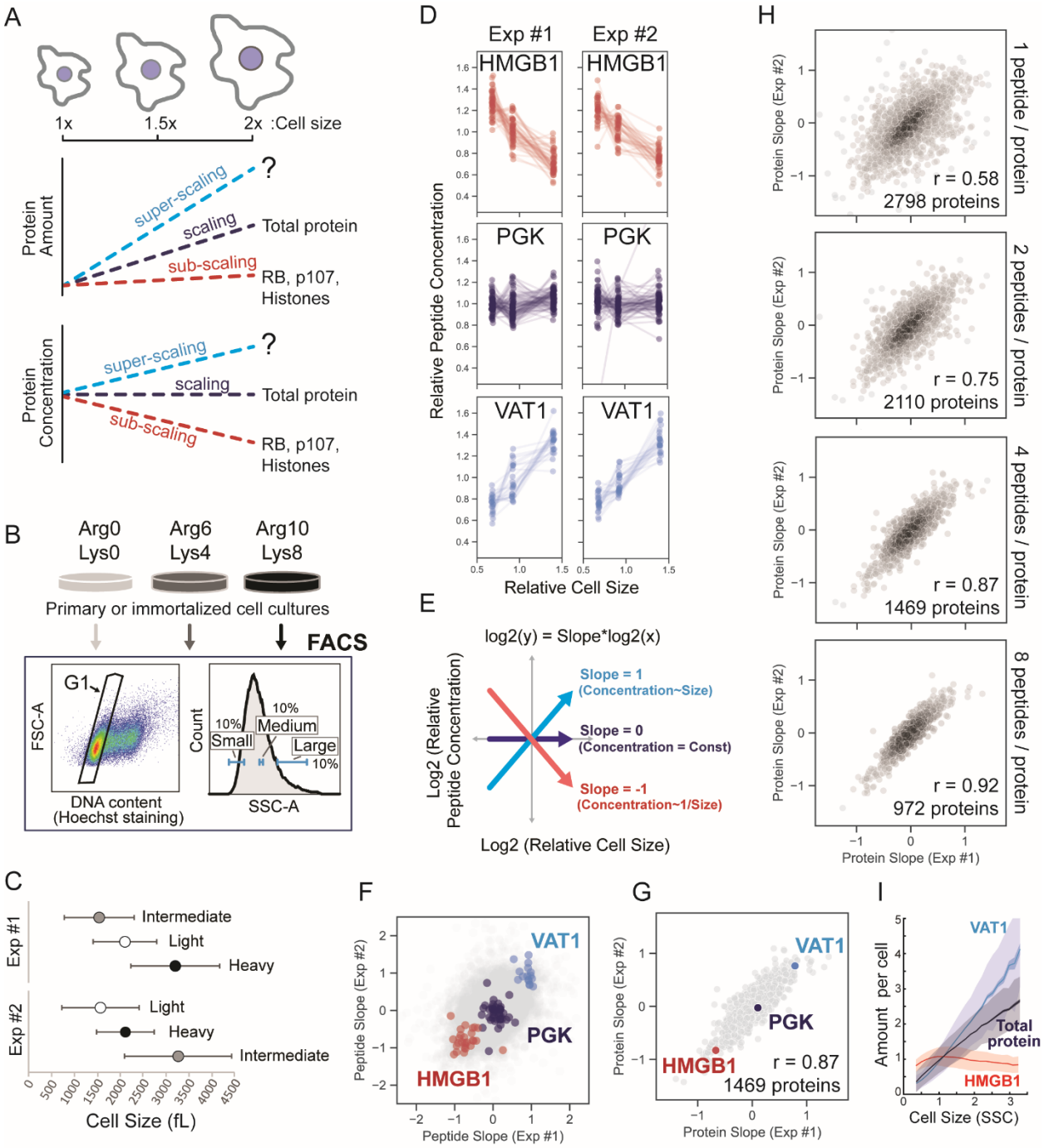


Figure 1. Cell size shapes the human proteome. (A) Schematic illustration of the potential scaling relationships between protein amount, concentration, and cell size. **(B)** Metabolically labeled HLFs cells were gated by G1 DNA content and sorted into three size bins based on the side scatter parameter (SSC) as a proxy for cell size using FACS. **(C)** The attainment of differentially sized G1 cells was confirmed

using a Coulter counter. Central dots represent the mean volume for each size bin and error bars represent the standard deviation. SILAC labeling orientation was swapped for replicate experiments. **(D)** SILAC channel intensities were used to infer changes in relative peptide concentration, which are plotted for three example proteins. Each dotted line represents an independent peptide measurement. **(E)** Derivation of a slope value that describes the scaling behavior of each peptide triplet. A slope value of 1 corresponds to an increase in protein concentration that is proportional to the increase in volume and a slope of -1 corresponds to dilution (1/volume). **(F)** Peptide and **(G)** protein slope values from two replicate experiments. Only proteins with at least 4 independent peptide measurements in both experiments are shown. **(H)** Correlation of protein slope values from two replicate experiments. A threshold for the minimum number of peptide measurements per protein is indicated in each panel. **(I)** Immunofluorescence intensity measured as a function of SSC (cell size) using flow cytometry. The relative change in the total protein amount is inferred from the measurement of carboxyfluorescein succinimidyl ester (CFSE) dye. The data were binned by cell size and plotted as mean protein amounts per cell for each size bin (solid lines). Dark shaded area shows standard error of the mean for each bin, and light shaded area shows the standard deviation. A representative is shown of n=3 biological replicates for each experiment. 100,000 cells were analyzed for each sample.

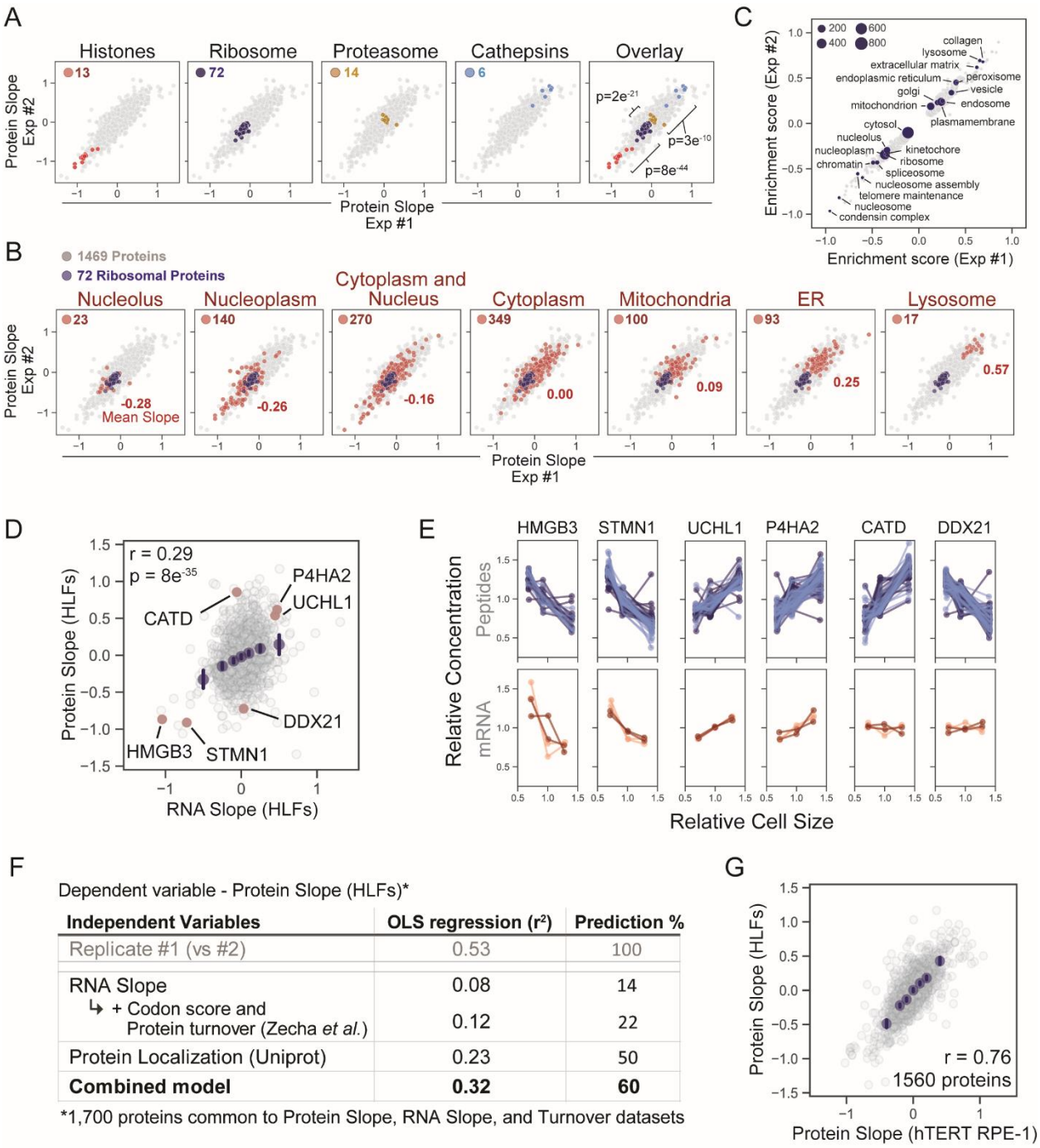


Figure 2. Diverse mechanisms control proteome size-dependency. (A) Scaling behavior of protein groups based on function and **(B)** subcellular localization. Significance is determined by t-test between adjacent protein groups. Ribosomal proteins are plotted in dark blue as a reference and the number of highlighted proteins and their average slope are indicated for each panel. **(C)** 2D annotation enrichment

analysis for replicate experiments depicted in Figures 1 and 2. Data S1 contains a complete list of significantly super- or sub-scaling GO terms. **(D)** Size-dependent concentration changes for a representative set of proteins and their corresponding mRNA transcripts. For proteins, each connected line represents a unique peptide measurement from two biological replicate experiments (Light and dark blue). For RNA, technical replicates are denoted in the same color, while biological replicates are denoted in different colors. **(E)** Correlation between size-dependent proteome and transcriptome changes. Highlighted examples are shown in (D). X-axis bins are shown in dark blue. Error bars represent the 95% confidence interval and r denotes the Pearson correlation coefficient. **(F)** Ordinary least squares regression model predicts the size scaling behavior of 1,700 individual proteins based on their subcellular localization and additional features. The benchmark for predictive accuracy (Prediction %) is set by the correlation between biological replicate experiments. See Figure S9 for a full description of the model. **(E)** Similarity of the cell size-dependent concentration changes between primary HLF and immortalized *hTERT RPE-1* cells. Protein Slope values for each cell type are mean of two biological replicate experiments. Only proteins with at least 3 independent peptide measurements in each biological replicate are depicted. X-axis bins are shown in dark blue. Error bars represent the 95% confidence interval and r denotes the Pearson correlation coefficient.

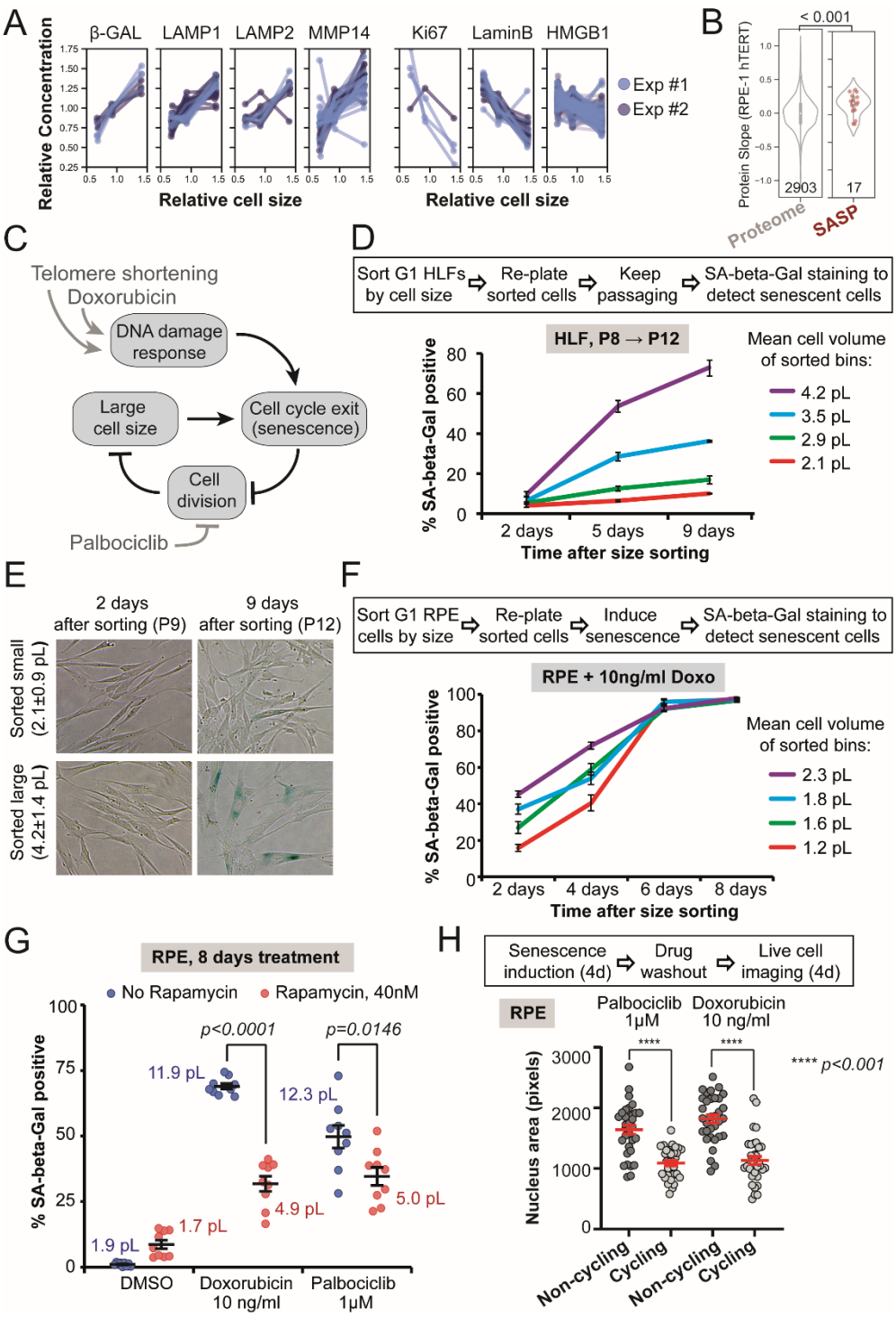


Figure 3. Large cell size promotes cell senescence. (A-B) Examples from our proteomics data set of cell size-dependent intracellular (A) and SASP (senescence-associated secretory phenotype, (B)) protein concentration changes in proliferating cells that are normally associated with senescence. **(C)** Model indicating possible relationships between DNA-related stress, large cell size, and cell senescence. **(D-E)**

Replicative senescence of different-sized primary cells. Asynchronous HLFs were gated for G1 DNA content and sorted into four bins by size using FACS, then replated and stained for the senescence marker SA-beta-Gal at the indicated time points. Percentage of blue-stained SA-beta-Gal positive cells (E) was calculated for each time point and plotted for each sorted size bin as mean ± standard error (D). Cell sizes for each bin are shown as mean ± SD in (E). P denotes the cell passage number. **(F)** Asynchronous *hTERT RPE-1* cells were gated for G1 DNA content and sorted into four bins by size using FACS, replated and cultured in the presence of the DNA damaging agent Doxorubicin (10 ng/ml), and then stained for SA-beta-Gal at the indicated time points. **(G)** Effect of Rapamycin, which reduces cell growth, on the percentage (±standard error) of SA-beta-Gal positive cells in *hTERT RPE-1* cultures treated with Doxorubicin or Palbociclib for 8 days. Values shown next to each condition indicate the mean cell sizes after 8 days of treatment. **(H)** Large cell size inhibits cell cycle entry to promote senescence. *hTERT RPE-1* cells were treated with Palbociclib or a low dose of Doxorubicin for 4 days. Then, the drugs were washed out, and the cells were imaged for 4 days to identify cells that re-enter the cell cycle, and cells that remain arrested in a senescent state. Nuclear area was used as a proxy for cell size and cell cycle re-entry was determined using the fluorescent cell cycle phase reporters Cdt1-mKO2 (G1 reporter) and Geminin-mAG (S/G2/M reporter) (*42*). N = 33 cells for each data point in (H). SA-beta-Gal quantification for every data point in (D), (F), and (G) included 700-1200 cells quantified from 9 different fields of view. All the experiments were performed in biological duplicates.

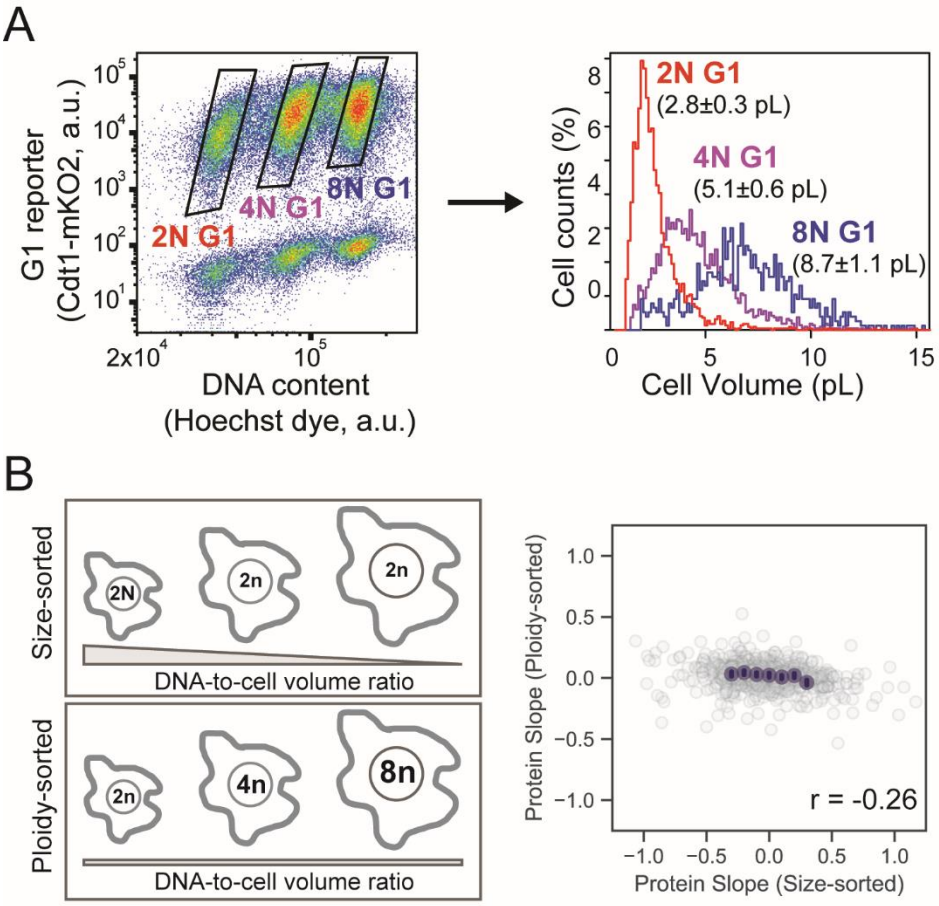


Figure 4. Cell volume-to-ploidy ratio drives size-dependent proteome changes. (A) *hTERT RPE-1* cells expressing fluorescent cell cycle reporters (Cdt1-mKO2, Geminin-mAG) were treated with an aurora kinase inhibitor barasertib (75nM, 48 hours) to partially inhibit cytokinesis. Cells were then sorted based on ploidy and G1 cell cycle phase. The attainment of differentially sized G1 cells was confirmed using a Coulter counter. The histogram shows a representative example of size distributions for sorted cells, and the numbers next to it represent mean cell size ± standard error for n=3 biological replicates. **(B)** Ploidy-sorted Protein Slope values were calculated in manner analogous to the Size-sorted Protein Slope values in Figure 1. For each protein, relative concentration was plotted as against the mean cell size in the 2N, 4N, and 8N bins to calculate a slope value. Despite large increases in cell size from 2N to 8N, concentration changes were minimal.

Supplementary Materials for

Increasing cell size remodels the proteome and promotes senescence

Michael C. Lanz[1†], Evgeny Zatulovskiy[1†], Matthew P. Swaffer[1], Lichao Zhang[2], Shuyuan Zhang[1], Chris You[1], Georgi Marinov[3], Patrick McAlpine[2], Josh E. Elias[2], Jan M. Skotheim[1*]

Correspondence to: skotheim@stanford.edu

This PDF file includes:

Materials and Methods

Figs. S1 to S11

Captions for Data S1 to S5

Other Supplementary Materials for this manuscript include the following:

Data S1 to S5

Materials and Methods

Cell Culture

Recently isolated primary fetal human lung fibroblasts (HLF) were purchased from Cell Applications, telomerase-immortalized retinal pigment epithelium (*hTERT RPE-1*, here also referred to as RPE-1 for brevity) cells were obtained from the Stearns laboratory at Stanford. All cells were cultured at 37ºC with 5% CO_2 in Dulbecco's modification of Eagle's medium (DMEM) with L-glutamine, 4.5 g/l glucose and sodium pyruvate (Corning), supplemented with 10% heat-inactivated fetal bovine serum (FBS, Corning) and 1% penicillin/streptomycin.

Fluorescence-activated cell sorting (FACS)

Fluorescence-activated cell sorting was used to sort live cells by their size and cell cycle phase. To do this, the cells were harvested from dishes by trypsinization, stained with 20 µM Hoechst 33342 DNA dye in PBS for 15 minutes at 37ºC, and then sorted on a BD FACSAria Fusion flow cytometer. Consecutive SSC-A over FSC-A, and FSC-H over FSC-A gates were used to isolate single cells. Then, G1 cells were gated by DNA content (Hoechst staining). Finally, we collected the 10% smallest and 10% largest cells, as well as another 10% of the cells near the average size using the gating based on SSC-A signal. During sorting, all cell samples and collection tubes were kept at 4ºC. To determine the cell size distributions of the collected samples, aliquots were taken from each sorted size bin and measured on a Z2 Coulter counter (Beckman). Sorted cells were used for mRNA or protein isolation, or re-plated for assessing senescence dynamics.

Stable Isotope Labeling In Cell culture (SILAC)

Cells for SILAC were cultured in special Lysine/Arginine-free DMEM for SILAC (Thermo Scientific) with 10% dialyzed heat-inactivated FBS (HyClone) and penicillin/streptomycin. These cultures were

supplemented with “light”, “intermediate”, or “heavy” versions of Lysine (0.8mM) and Arginine (0.4mM) (Cambridge Isotope Laboratories) (*43*). The “light” (Agr0 Lys0) version of the media contained L-Arginine and L-Lysine built with normal ^{12}C and ^{14}N isotopes; the “intermediate" (Arg6 Lys4) version had L-Arginine containing six ^{13}C atoms and L-Lysine containing four deuterium atoms; the "heavy" (Arg10 Lys8) version had L-Arginine containing six ^{13}C and four ^{15}N atoms and L-Lysine containing six ^{13}C and two ^{15}N atoms. Proline (200 mg/l) was added to the media to prevent conversion of isotope-coded Arginine into Proline in cells. We confirmed that cell proliferation is not impaired in our SILAC medium (**Fig. S2B**). To ensure complete labelling, the cells were cultured in SILAC media for 5 passages (approximately 10 doublings) prior to the experiments.

LC-MS/MS sample preparation - SILAC

See Table S3 for a complete list of proteomic experiments. Small, medium and large cells sorted by FACS were pelleted by centrifugation at 500xg for 10 minutes and lysed for 40 minutes on ice in RIPA lysis buffer (Abcam) containing a protease and phosphatase inhibitor cocktail. SILAC-labeled different sized cells were mixed prior to lysis in order to minimize handling error in protein extraction, proteolytic digestion, and peptide desalting. Cell lysates were cleared by centrifugation at 15000xg for 30 minutes at 4°C. The mixed lysates were then denatured in 1% SDS, reduced with 10mM DTT, alkylated with 5mM iodoacetamide, and then precipitated with three volumes of a solution containing 50% acetone and 50% ethanol. Proteins were re-solubilized in 2 M urea, 50 mM Tris-HCl, pH 8.0, and 150 mM NaCl, and then digested with TPCK-treated trypsin (50:1) overnight at 37°C. Trifluoroacetic acid and formic acid were added to the digested peptides for a final concentration of 0.2%. Peptides were desalted with a Sep-Pak 50mg C18 column (Waters). The C18 column was conditioned with 5 column volumes of 80% acetonitrile and 0.1% acetic acid and washed with 5 column volumes of 0.1% trifluoroacetic acid. After samples were loaded, the column was washed with 5 column volumes of 0.1% acetic acid followed by elution with 4 column volumes of 80% acetonitrile and 0.1% acetic acid. The elution was dried in a Concentrator at 45°C.

LC-MS/MS sample preparation - TMT

Lysis, denaturation, reduction, and precipitation for SILAC analysis was the same for TMT analysis (working solution of Iodoacetamide was dissolved in HEPES rather than Tris buffer). Our method for TMT labeling was adapted from Zecha et al. (*44*) and the Thermo TMT10plex™ Isobaric Label Reagent Set Protocol. In brief, acetone precipitated samples were resuspended in 100µm TEAB and digested O/N with TPCK trypsin (50:1) in the absence of Tris or Urea. After digestion, peptide concentration was ~1µg/ul in 100µM TEAB for all samples. 20µg of peptide was labeled using 100µg of Thermo TMT10plex™ in a reaction volume of 25µl for 1 hour. The labeling reaction was quenched with 8µL of 5% hydroxylamine for 15 minutes. Labeled peptides were pooled, acidified to a pH of ~2 using drops of 10% formic acid, and desalted with a Sep-Pak 50mg C18 column as described above.

HILIC fractionation - SILAC

Desalted peptide samples were reconstituted in 80% acetonitrile and 1% formic acid and fractionated by hydrophilic interaction liquid chromatography (HILIC) with a TSK gel Amide-80 column (2 mm x 150 mm, 5 µm; Tosoh Bioscience). 90 second fractions were collected between 10 and 25 minutes of the gradient. Three solvents were used for the gradient: buffer A (90% acetonitrile), buffer B (80% acetonitrile and 0.005% trifluoroacetic acid), and buffer C (0.025% trifluoroacetic acid). The gradient used consists of a 100% buffer A at time = 0 min; 88% of buffer B and 12% of buffer C at time = 5 min; 60% of buffer B and 40% of buffer C at time = 30 min; and 5% of buffer B and 95 % of buffer C from time = 35 to 45 min in a flow of 150 µl/min. HILIC fractions were dried in a SpeedVac and reconstituted in 0.1% trifluoroacetic acid. A total of 10 fractions were collected and pooled back into 5 fractions (1-6, 2-7, 3-8, 4-9, 5-10).

High-pH reverse phase fractionation - TMT

TMT-labeled peptides (Experiment from Figure S5D) were fractionated using a Pierce™ High pH Reversed-Phase Peptide Fractionation Kit. The eight default fractions were pooled back to 4 fractions (1-5, 2-6, 3-7, 4-8). Dried peptides were reconstituted in 0.1% formic acid.

LC-MS/MS data acquisition - SILAC

Peptide samples were analyzed using a Fusion Lumos mass spectrometer (Thermo Fisher Scientific, San Jose, CA) equipped with Dionex Ultimate 3000 LC systems (Thermo Fisher Scientific, San Jose, CA). Peptides were separated by capillary reverse phase chromatography on a 25 cm reversed phase column (100 µm inner diameter, packed in-house with ReproSil-Pur C18-AQ 3.0 m resin (Dr. Maisch GmbH)). Liquid chromatography was performed using a two-step linear gradient with 4–25 % buffer B (0.1% (v/v) formic acid in acetonitrile) for 90 min followed by 25-40 % buffer B for 10 min. Data was acquired in top 20 data dependent mode. Full MS scans were acquired in the Orbitrap mass analyzer with a resolution of 120,000 (FWHM) and m/z scan range of 340-1500. Selected precursor ions were subjected to fragmentation using higher-energy collisional dissociation (HCD) with quadrupole isolation, isolation window of 1.6 m/z, and normalized collision energy of 30%. HCD fragments were analyzed in the Orbitrap mass analyzer with a resolution of 15,000 (FWHM). Fragmented ions were dynamically excluded from further selection for a period of 15 seconds. The AGC target was set to 400,000 and 50,000 for full FTMS scans and FTMS2 scans, respectively. The maximum injection time was set to 50 ms for full FTMS scans and dynamic for FTMS2 scans.

LC-MS/MS data acquisition - TMT

Desalted TMT-labeled peptides were analyzed on a Fusion Lumos mass spectrometer (Thermo Fisher Scientific, San Jose, CA) equipped with a Thermo EASY-nLC 1200 LC system (Thermo Fisher Scientific, San Jose, CA). Peptides were separated by capillary reverse phase chromatography on a 25 cm column (75 µm inner diameter, packed with 1.6 µm C18 resin, AUR2-25075C18A, Ionopticks, Victoria

Australia). Electrospray Ionization voltage was set to 1550 volts. Peptides resulting from on-bead digestion were resuspended in 10 µL of 0.1% formic acid. 2 µL was introduced into the Fusion Lumos mass spectrometer using a two-step linear gradient with 6–33% buffer B (0.1% (v/v) formic acid in 80% acetonitrile) for 145 min followed by 33-45% buffer B for 15 min at a flow rate of 300 nL/min. Column temperature was maintained at 40°C throughout the procedure. Xcalibur software (Thermo Fisher Scientific) was used for the data acquisition and the instrument was operated in data-dependent mode. Survey scans were acquired in the Orbitrap mass analyzer over the range of 380 to 1800 m/z with a mass resolution of 70,000 (at m/z 200). Ions were selected for fragmentation from the 10 most abundant ions with a charge state of either 2, 3 or 4 and within an isolation window of 2.0 m/z. Selected ions were fragmented by Higher-energy Collisional Dissociation (HCD) with normalized collision energies of 27 and the tandem mass spectra was acquired in the Orbitrap mass analyzer with a mass resolution of 17,500 (at m/z 200). Repeated sequencing of peptides was kept to a minimum by dynamic exclusion of the sequenced peptides for 30 seconds. For MS/MS, the AGC target was set to 1e5 and max injection time was set to 120ms.

Spectral searches - TMT and SILAC

All raw files were searched using the Andromeda engine (*45*) embedded in MaxQuant (v1.6.7.0) (*46*). See Table S4 for a complete summary of the search parameters used for the SILAC- and TMT-labeled peptide fragments. In brief, 3 label SILAC search was conducted using Maxquant's default Arg6/10 and Lys4/8. For TMT searches, a Reporter ion MS3 search was conducted using 10plex TMT isobaric labels. For both TMT and SILAC searches, variable modifications included oxidation (M) and protein N-terminal acetylation. Carbamidomthyl (C) was a fixed modification. The number of modifications per peptide was capped at five. Digestion was set to tryptic (proline-blocked). Peptides were "Re-quantified", and maxquant's match-between-runs feature was not enabled. Database search was conducted using the Uniprot proteome - Human_UP000005640_9606. Minimum peptide length was 7 amino acids. FDR was determined using a reverse decoy proteome (*47*).

Peptide quantitation - SILAC

Our SILAC analysis pipeline uses the peptide feature information in MaxQuant's "evidence.txt" output file. Each row of the "evidence.txt" file represents an independent peptide triplet measurement. Contaminant and decoy peptide identifications were discarded. Peptides without signal in any of the three SILAC channels were also excluded. Peptide triplets (each row in the "evidence.txt" table) are assigned to a protein based on MaxQuant's "Leading razor protein" designation. For each peptide triplet, the fraction of ion intensity in each SILAC channel was calculated by dividing the "Intensity L/M/H" column by the "Intensity" column. SILAC channels were normalized by adjusting the fraction of ion intensity in each channel by the median for all measured peptides (see **Fig. S1B,C**). After normalization, the relative signal difference between the SILAC channels for each peptide triplet was plotted against the normalized cell size for each of the bins of isolated G1 cells.

For each peptide, we calculated its slope as follows (mean squared error filtering):

$Y_{1,2,3}$ = Relative signal in each SILAC channel (order based on labeling orientation)

Avg. size = (mean volume of small bin + mean volume of medium bin + mean volume of large bin) / 3

x_1 = (Mean volume of small size bin) / Avg. size

x_2 = (Mean volume of medium size bin) / Avg. size

x_3 = (Mean volume of large size bin) / Avg. size

Based on the expectation that our experimental conditions would not result in large, non-linear changes in protein expression, we exclude peptide triplets whose three data points did not loosely fit a linear regression line. Linear regressions on the ~50,000 triplets/experiment were performed using np.polyfit in Python. Regressions with a mean squared error > 0.075 were excluded. Because this step significantly improved the overall data quality (**Fig. S1F,G**), we concluded that our filtering method mostly excludes peptide triplets contaminated by analytical interference or that are near the noise floor.

Individual peptide measurements were consolidated into a protein level measurement using python's groupby.median. Peptides with the same amino acid sequence that were identified as different charge states or in different fractions were considered independent measurements. We summarize the size scaling behavior of individual proteins as a slope value derived from a regression (similar to what is described above for individual peptides), and each protein slope value is based on the behavior of all detected peptides.

For a given protein, we calculate its cell size-dependent slope as follows:

y_i = Relative signal in the ith SILAC channel (median of all corresponding peptides in this channel)

x_i = same normalized cell size x_i as for the peptide slope calculations above

The protein slope value was determined from a linear fit to the log-transformed data using the equation:

$$\mathrm{Log}_2(y) = \mathrm{Slope} * \log_2(x)$$

Variables were log-transformed so that a slope of 1 corresponds to an increase in protein concentration that is proportional to the increase in volume and a slope of -1 corresponds to 1/volume dilution. Pearson r and p values for correlation analyses were calculated using scipy's pearsonr module in python.

Protein annotations

Protein annotations in Figure 2 were sourced from Uniprot columns named "Subcellular location [CC]" or "Protein names" (*48*). For Figure 2B, protein localization was strictly parsed so that each annotated protein belongs to only one of the designated groups. Proteins with 2 or more of the indicated annotations were ignored (with the exception of the "Cytoplasm / Nucleus" category, which required a nuclear and cytoplasmic annotation). 2D annotation enrichment in Figure 2C was performed using Perseus (*49*).

Peptide quantitation - TMT

Our TMT analysis pipeline uses the peptide feature information in MaxQuant's "evidence.txt" output file. Each row of the "evidence.txt" file represents an independent MS^3 TMT measurement. Contaminant and decoy peptide identifications were discarded. Peptides without signal in any of the TMT channels were also excluded. TMT peptide measurements were assigned to protein based on MaxQuant's "Leading razor protein" designation. For each peptide triplet, the fraction of ion intensity in each TMT channel was calculated by dividing the "Reporter ion intensity" column by the sum of all reporter ion intensities. TMT channels were normalized by adjusting the fraction of ion intensity in each channel by the median for all measured peptides (similar to the SILAC normalization in **Fig. S1B,C**). After normalization, the relative signal difference between the TMT channels for each peptide triplet was plotted against the normalized cell size for each of the bins of isolated G1 cells. Slope values in Figure S5 were derived in a manner analogous to the Slope values calculated in the SILAC experiments. Pearson r and p values for correlation analyses were calculated using SciPy's pearsonr module in python.

OLS linear regression model

Multiple linear regression analysis was performed using the statsmodels module in python. The prediction of size scaling behavior was based on the 1,700 proteins shared between the protein turnover (HeLa cells) (*32*), RNA Slope, and Protein Slope datasets (at least 2 peptides / protein). Independent variables for codon affinity, RNA Slope, and Protein turnover (T50%) were each independently standardized by subtracting all values by the dataset's mean and then dividing by the dataset's standard deviation. The codon affinity refers to the binding affinity at the 3rd codon position. The affinity is classified as "low" or "high" based on a prior study (*50*). The affinity score of a gene is the average of the low affinity codon percentage within each amino acid, weighted by the percentage of the amino acid in that gene. The subcellular localization variable was based on Uniprot's "Subcellular location [CC]" annotations and entered as a binary value for each compartment (1 if a protein possessed an annotation

and 0 if it did not). Only subcellular compartments that provided nonredundant predictive power were ultimately included in the model. A constant value was added to the regression equation using the add_constant function in statsmodels. We set the benchmark for predictive accuracy (Prediction %) as the correlation between biological replicate experiments (Protein Slope from Exp #1 vs Exp #2). See Figure S9 for more details on the model and its coefficients.

RNA extraction and sequencing

To compare the transcriptomes of different-sized G1 cells, primary HLFs were arrested in the G1 phase of the cell cycle by a 24-hour treatment with 1μM of the Cdk4/6 inhibitor Palbociclib and then sorted into size bins using FACS. To extract RNA, each sample of sorted HLF cells was split into two technical replicates each of which contained 200,000 HLF cells, which were then mixed with 100,000 *D. melanogaster* S2 cells as a spike-in. Each sample was then pelleted and RNA was extracted using Direct-zol™ RNA Microprep Kit (Zymo Research). mRNA was enriched using the NEBNext Poly(A) mRNA Magnetic Isolation Module (NEB, #E7490). The NEBNext Ultra II RNA Library Prep Kit for Illumina® (NEB, #E7775) was then used to prepare libraries for paired-end (2x150 bp) Illumina sequencing (Novogene). Two independent biological replicates of each sample were collected and for each biological replicate, two technical replicates (*i.e.*, separate lysis, library prep, and sequencing) were processed. Approximately 40 million reads were sequenced per replicate.

RNAseq data processing

RNA samples contained a mixture of *H. sapiens* and *D. melanogaster* spike-in. A combined *H. sapiens* and *D. melanogaster* genome file was created using the hg38 and dm6 versions of the respective genomes and a combined transcriptome annotation was created using the *H. sapiens* gene models from the v29 version of the GENCODE annotation (*51*) and the BDGP6 *D. melanogaster* gene models from ENSEMBL release 90 (*52*). For the purposes of RNA-seq data quality evaluation, genome browser track generation, and calculating the hg38-to-dm6 ratio, reads were aligned against the combined genomes and

combined annotated set of splice junctions using the STAR aligner (version 2.5.3a; settings: --limitSjdbInsertNsj 10000000 --outFilterMultimapNmax 50 --outFilterMismatchNmax 999 --outFilterMismatchNoverReadLmax 0.04 --alignIntronMin 10 --alignIntronMax 1000000 --alignMatesGapMax 1000000 --alignSJoverhangMin 8 --alignSJDBoverhangMin 1 --sjdbScore 1 --twopassMode Basic --twopass1readsN -1) (*53*). Read mapping statistics and genome browser tracks were generated using custom Python scripts. For quantification purposes, reads were aligned as 2x50mers in transcriptome space against an index generated from the combined annotations described above using Bowtie (version 1.0.1; settings: -e 200 -a -X 1000) (*54*). Alignments were then quantified using eXpress (version 1.5.1) (*55*) before effective read count values and TPM (Transcripts Per Million transcripts) were then separated for each genome and renormalized TPMs were calculated with respect to only the *H. sapiens* transcripts.

Flow cytometry

For flow cytometry analysis, cells were grown on dishes to ~50% confluence and harvested by trypsinization. The cells were then fixed with 3% formaldehyde for 10 minutes at 37ºC, and then permeabilized with 90% methanol for 30 minutes on ice. Fixed and permeabilized cells were washed once with PBS, blocked with 3% BSA in PBS for 30 minutes at 37ºC, and then stained with primary antibodies for 2 hours at 37ºC. We used the following primary antibodies: HMGB1 (Abcam, ab79823), HMGN2 (CST, 9437S), RPLP0 (Sigma, SAB1402899), Actin (Sigma-Aldrich, A2103), UCHL1 (CST, 13179S), VAT1 (Santa Cruz, sc-515705), LAMP1 (CST, 9091), alpha-Tubulin (Abcam, ab6160). After the primary antibodies, the cells were washed twice with a wash buffer (1% BSA in PBS + 0.05% Tween® 20), stained with the fluorophore-conjugated secondary antibodies Alexa Fluor 488 goat anti-mouse (Life Technologies A11029), Alexa Fluor 594 goat anti-rabbit (Life Technologies A11037), and Alexa Fluor 405 goat anti-rat (Abcam ab175673) at 1:1000 dilution for 1 hour at 37ºC, and then washed twice again. After this treatment, the cells were resuspended in PBS containing 3 μM DAPI for DNA staining, incubated for 10 minutes at room temperature, and then analysed on a Attune NxT flow cytometer

(Thermo Fisher). To compensate for the nonspecific background staining with antibodies, we measured the fluorescence of cells stained with nonspecific Isotype Control antibodies. We then performed a linear regression of this nonspecific background signal with the cell size, and subtracted the background fluorescence corresponding to the cell's size from the actual fluorescence signal measured for each cell. For total protein staining, live cells were resuspended in PBS, then the CellTrace CFSE dye (CarboxyFluorescein Succinimidyl Ester, Thermo Fisher) was added at 5 μM concentration, incubated for 30 minutes at 37ºC. The dye was then washed out with FBS-containing DMEM, and the cells were pelleted and resuspended in PBS for the flow cytometry or for the fixation and antibody staining. For the Lysotracker staining, the cells were harvested and re-suspended in growth media at a concentration of 10^6 cells/mL. Then, Lysotracker Red DND-99 (Thermo Fisher) was added at a concentration of 75nM and incubated at 37ºC for 30 min. Cells were spun down and re-suspended for analysis or additional staining. For plotting the flow cytometry data, all protein amounts and cell size values were normalized to their means. To characterize the degree of size-dependency of protein amounts, we fit a line to the flow cytometry data after normalizing these data to mean values. We performed at least three biological replicates for each experiment that measured 100,000 cells.

Senescence-associated beta-galactosidase (SA-beta-Gal) staining

To detect senescent cells, the RPE-1 or HLF cells on a dish were stained using the Senescence β-Galactosidase Staining Kit (Cell Signaling Technology) following the manufacturer's protocol. Briefly, live cells on a dish were washed once with PBS and fixed with 1x Fixation solution for 10 minutes at room temperature, then rinsed twice with PBS, and stained with β-Galactosidase Staining Solution for 48 hours at 37ºC in a dry incubator (no CO_2). The cells were then imaged on an EVOS M5000 imaging system to obtain a colored brightfield image. The obtained images were quantified manually, by a blinded investigator, to determine the percentage of senescent cells, *i.e.*, the cells that have a pronounced blue staining.

Live cell fluorescence microscopy

In preparation for imaging, cells were seeded on 35-mm glass-bottom dishes (MatTek) at low density and incubated overnight at 37°C and 5% CO_2. Then, the cells were moved to a Zeiss Axio Observer Z1 microscope equipped with an incubation chamber and imaged for 96 hours (*56*). Brightfield and fluorescence images were collected from three dishes at multiple positions every 20 minutes using an automated stage controlled by the Micro-Manager software. We used a Zyla 5.5 sCMOS camera, which has a large field of view allowing us to track motile cells within a field of view for long durations, and an A-plan 10x/0.25NA Ph1 objective. To distinguish G0/G1 and S/G2/M cells in time lapse imaging experiments, we used RPE-1 cells expressing the fluorescent cell cycle reporters mKO2-hCdt1 (G1), and mAG-hGeminin (S/G2) (*42*). These reporters were introduced into RPE-1 cells using a lentivirus vector, and the positive, fluorescent population of cells was isolated using fluorescence-activated cell sorting.

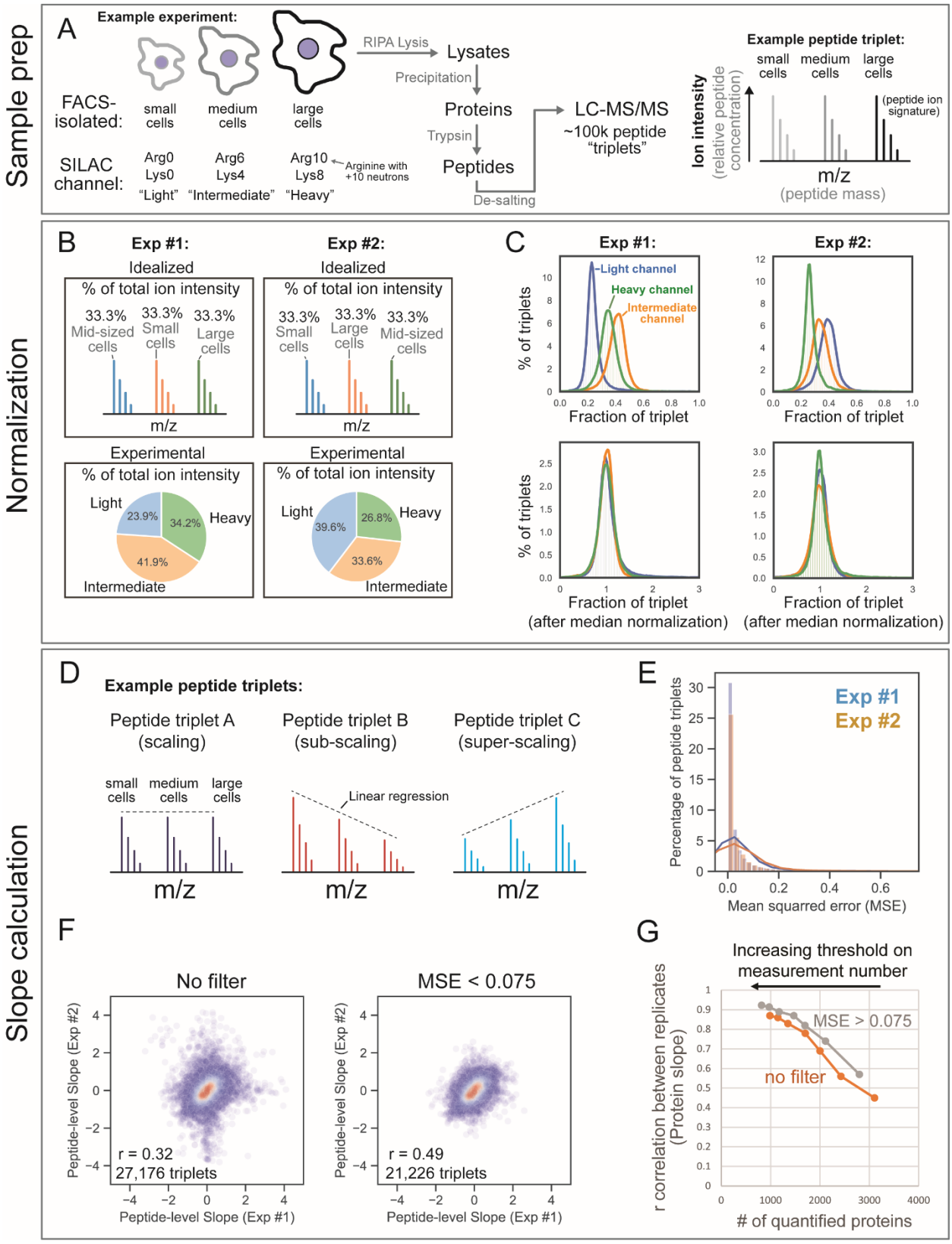


Figure S1. A SILAC proteomics method to measure the proteome as a function of G1 cell size.

(A) Human cells were metabolically labeled in cell culture, sorted by G1 cell size using FACS, and subjected to proteomic analysis. **(B)** Differences in the amount of total protein contributed from the small-, medium-, or large-cell size populations were normalized using the signal proportionality from the light, intermediate, and heavy channel of each peptide triplet. Small, medium, and large cells were mixed prior to lysis, so the amount of protein in each SILAC channel is uneven. Rather than normalize L/H and L/M SILAC ratios separately, we normalize all three channels together so that the values in our dataset represent relative changes to each peptide. **(C)** For each individual peptide triplet, we determined the fraction of the triplet's total ion intensity present in each SILAC channel. The distributions of these fractions were then adjusted by the median (see methods for a complete description of the normalization process). **(D)** Peptide slope values are calculated from a linear regression of the relative ion intensity in each SILAC channel and mean cell size. Mean cell size was determined by Coulter counter prior to mixing and lysis. **(E)** Distribution of mean squared error values for peptide triplet regressions (~50,000 per experiment). The mean squared error was used to track the linear fit of each peptide regression. **(F)** Correlation of peptide slopes calculated from biological replicate experiments before and after applying a filter for mean squared error (MSE). 27,176 unique peptide measurements were identified in both replicate experiments. A unique peptide measurement is defined by the peptide sequence, modification state, charge state, and the fraction number (the fraction number is only considered if the experiment was prefractioned and multiple fractions were analyzed). **(G)** Filtering peptides by mean squared error from a linear fit improves data quality. MSE filtering improves the correlation of protein slope values derived from biological replicates, and the improvement is consistent across different thresholds of measurement confidence (*i.e.*, peptide measurements per protein).

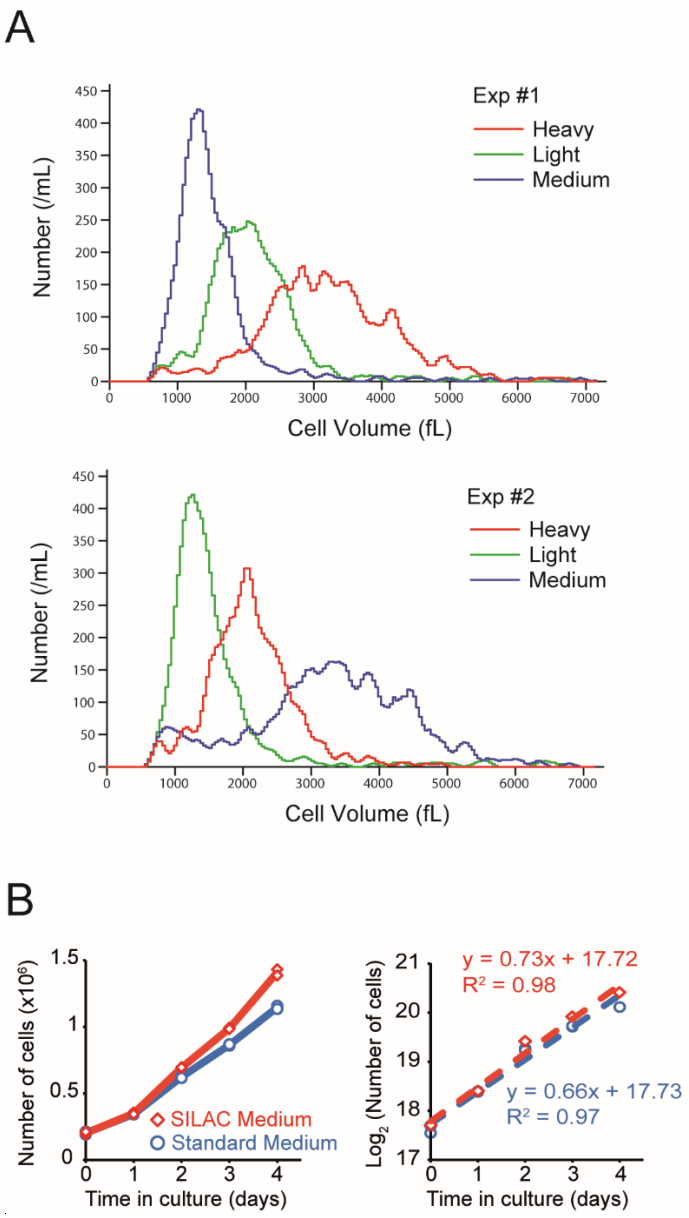


Figure S2. Supporting information for SILAC proteomics data shown in Figure 1: (A) Coulter counter measurement of HLF cells isolated by FACS. Small, medium, and large cell populations are colored based on the SILAC labeling orientation for the two replicate experiments in Figure 1. See Table S3 for cell size measurements for all proteomic experiments. **(B)** HFL primary cell proliferation rates in SILAC vs standard medium.

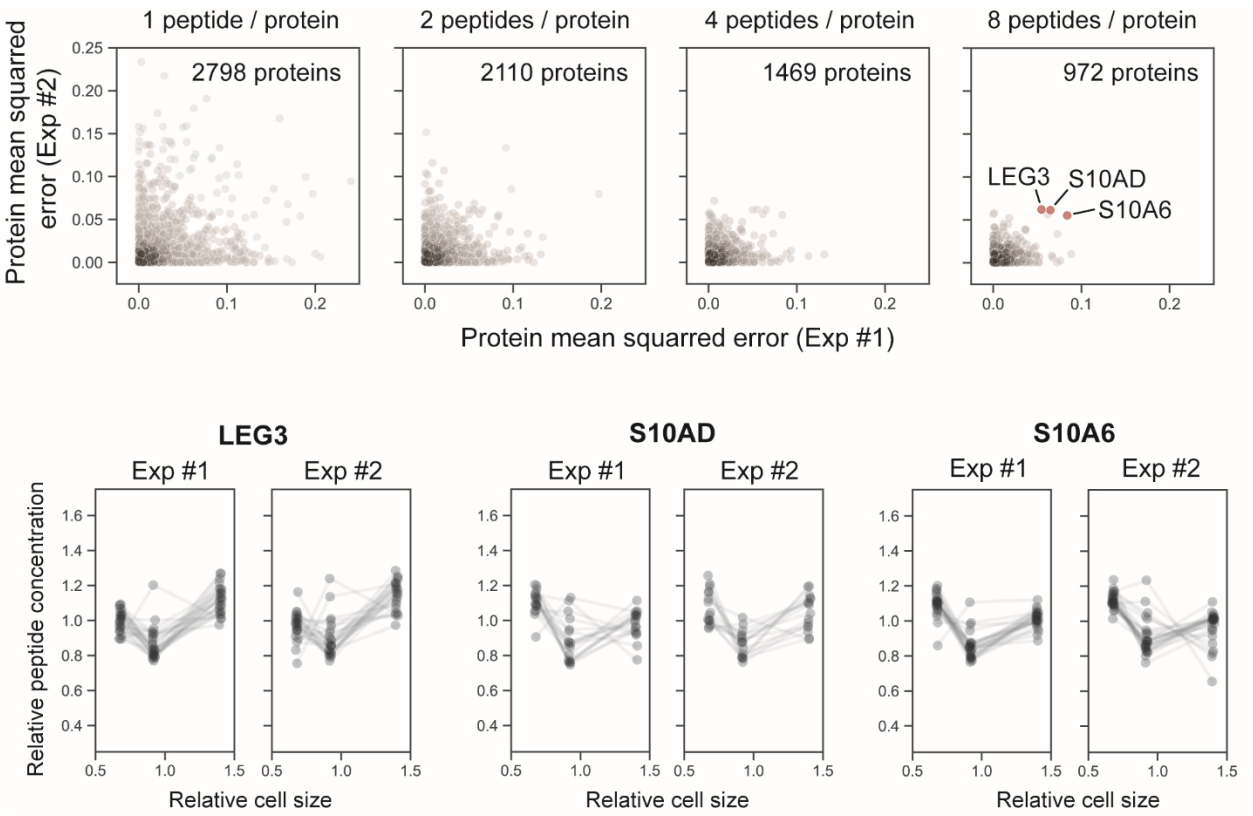


Figure S3. Cell size-dependent changes to concentrations in the proteome are mostly linear: Correlation for the mean squared error (MSE) of the Protein Slope regression from two biological replicates. A threshold for the minimum number of peptide measurements per protein is increased from left to right. Because very few proteins have reproducible large MSE values, we conclude that most proteins scale linearly with G1 cell size. Peptide-level measurements for the few proteins with non-linear scaling are plotted below.

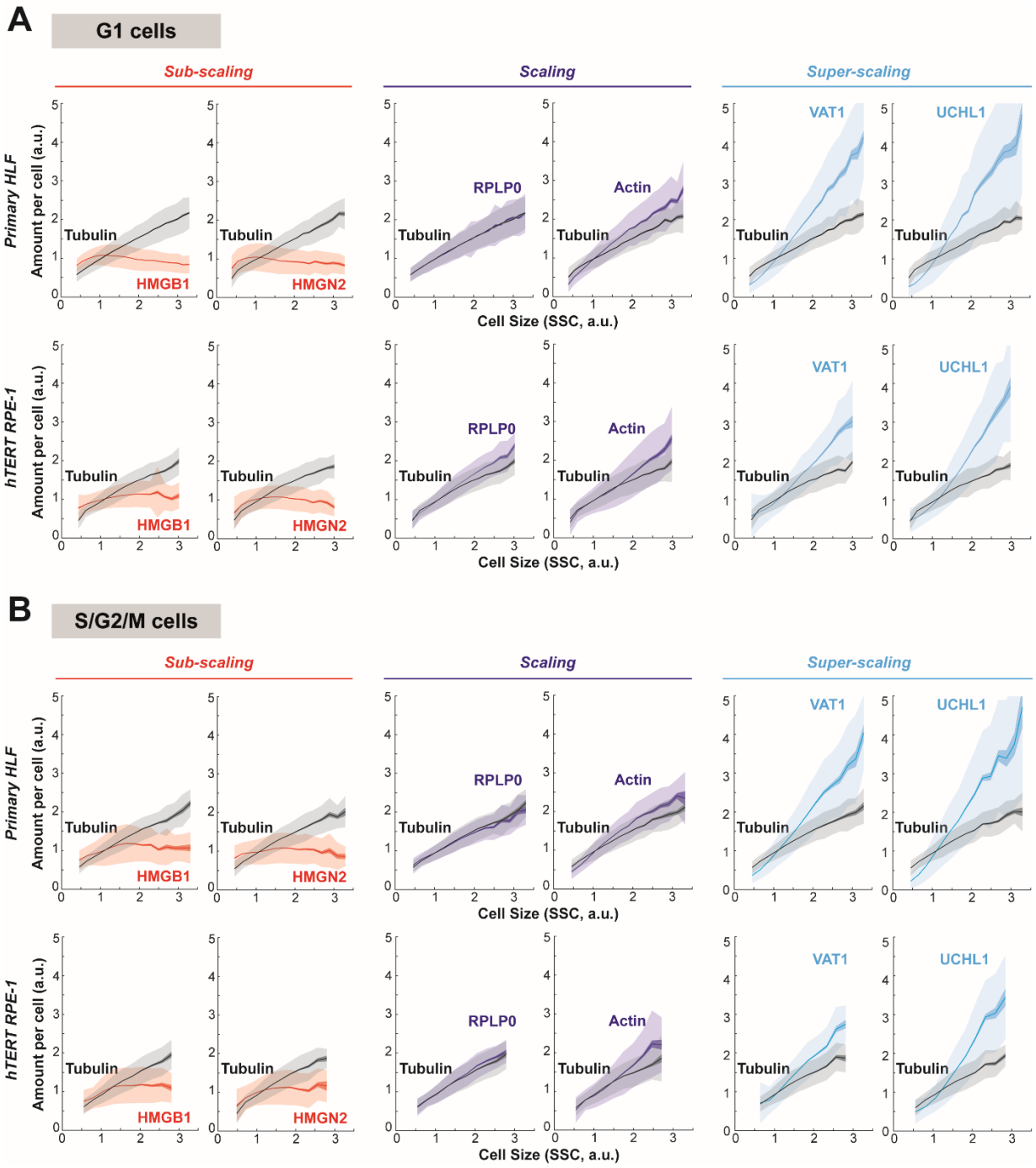
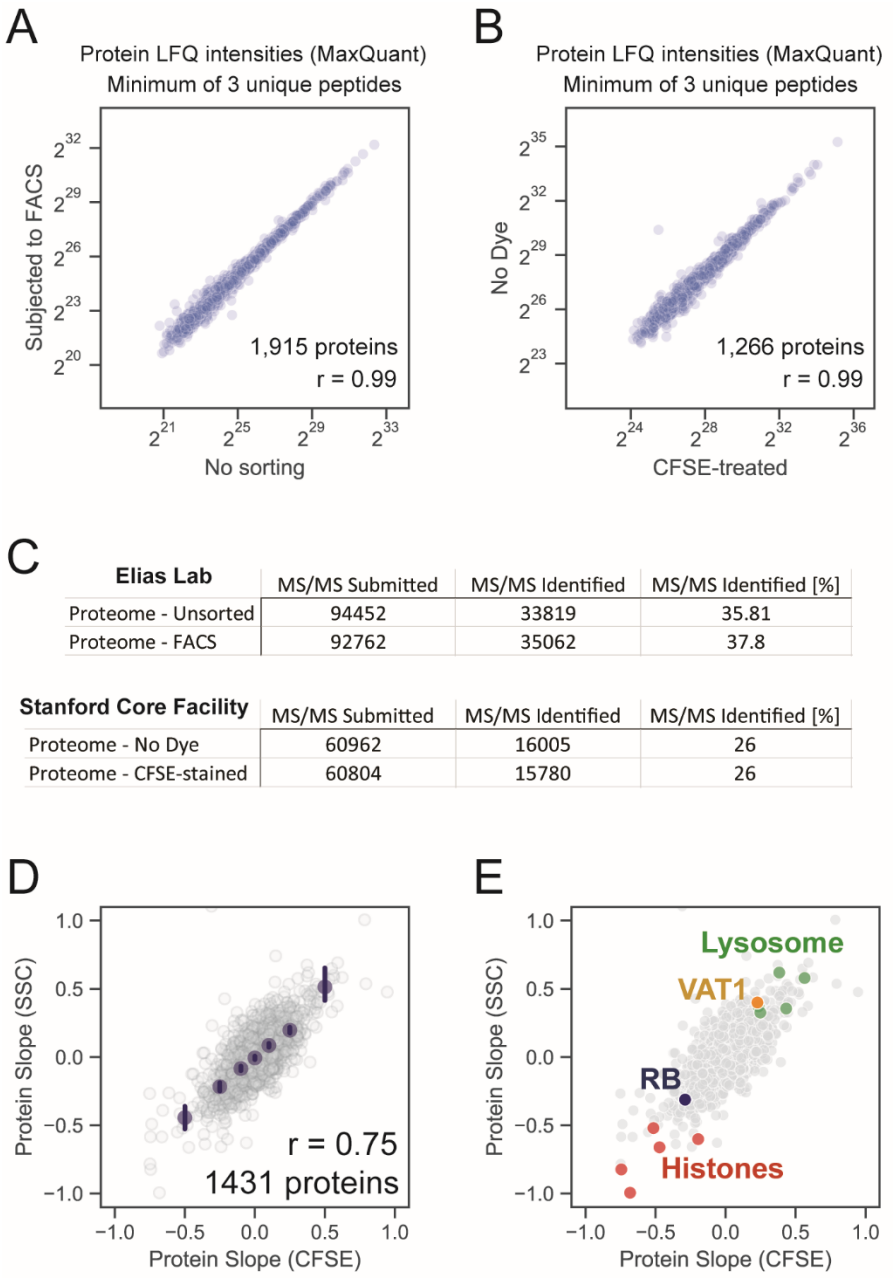


Figure S4. Validation of protein size-scaling behaviors using flow cytometry: Cycling HLF and RPE-1 cells were fixed and stained with antibodies against subscaling (HMGB1, HMGN2), scaling (RPLP0, beta-Actin), and superscaling proteins (VAT1, UCHL1). Alpha-tubulin is used as an internal control for each sample. Using flow cytometry, G1 and non-G1 cells were gated by DNA content (DAPI dye) and

analyzed separately in panels **(A)** and **(B)**, respectively. The data were binned by cell size (SSC, the side scatter parameter) and plotted as mean protein amounts per cell for each size bin (solid lines). Dark shaded area shows standard error of the mean for each bin, and light shaded area shows the standard deviation. A representative is shown of n=3 biological replicates for each experiment. 100,000 cells were analyzed for each sample.



Elias Lab	MS/MS Submitted	MS/MS Identified	MS/MS Identified [%]
Proteome - Unsorted	94452	33819	35.81
Proteome - FACS	92762	35062	37.8

Stanford Core Facility	MS/MS Submitted	MS/MS Identified	MS/MS Identified [%]
Proteome - No Dye	60962	16005	26
Proteome - CFSE-stained	60804	15780	26

Figure S5. Controls indicating that cell sorting does not affect proteomics measurements: (A) A dish of HLF cells were split into two equal parts, and one half was run through the FACS machine while the other half sat on ice. MaxQuant LFQ was used to determine whether cells subjected to FACS exhibited altered proteomes. The strong correlation between the proteomes of sorted and unsorted cells suggests that FACS did not appreciably bias our measurements. **(B)** MaxQuant LFQ was used to compare proteome samples from CFSE-stained and unlabeled cells. The strong correlation between the proteomes

of CFSE-treated and untreated cells suggests that using a total protein dye does not appreciably bias our measurement. **(C)** Peptide discovery is not impacted by FACS or CFSE staining. **(D)** RPE-1 cells with G1 DNA content were sorted using total protein / cell (CFSE stain) or side scatter to achieve 3 bins of different sized cells. Protein Slope values derived from cells sorted by total protein and side scatter are compared. A select set of proteins from the comparison in (D) are highlighted in **(E)**.

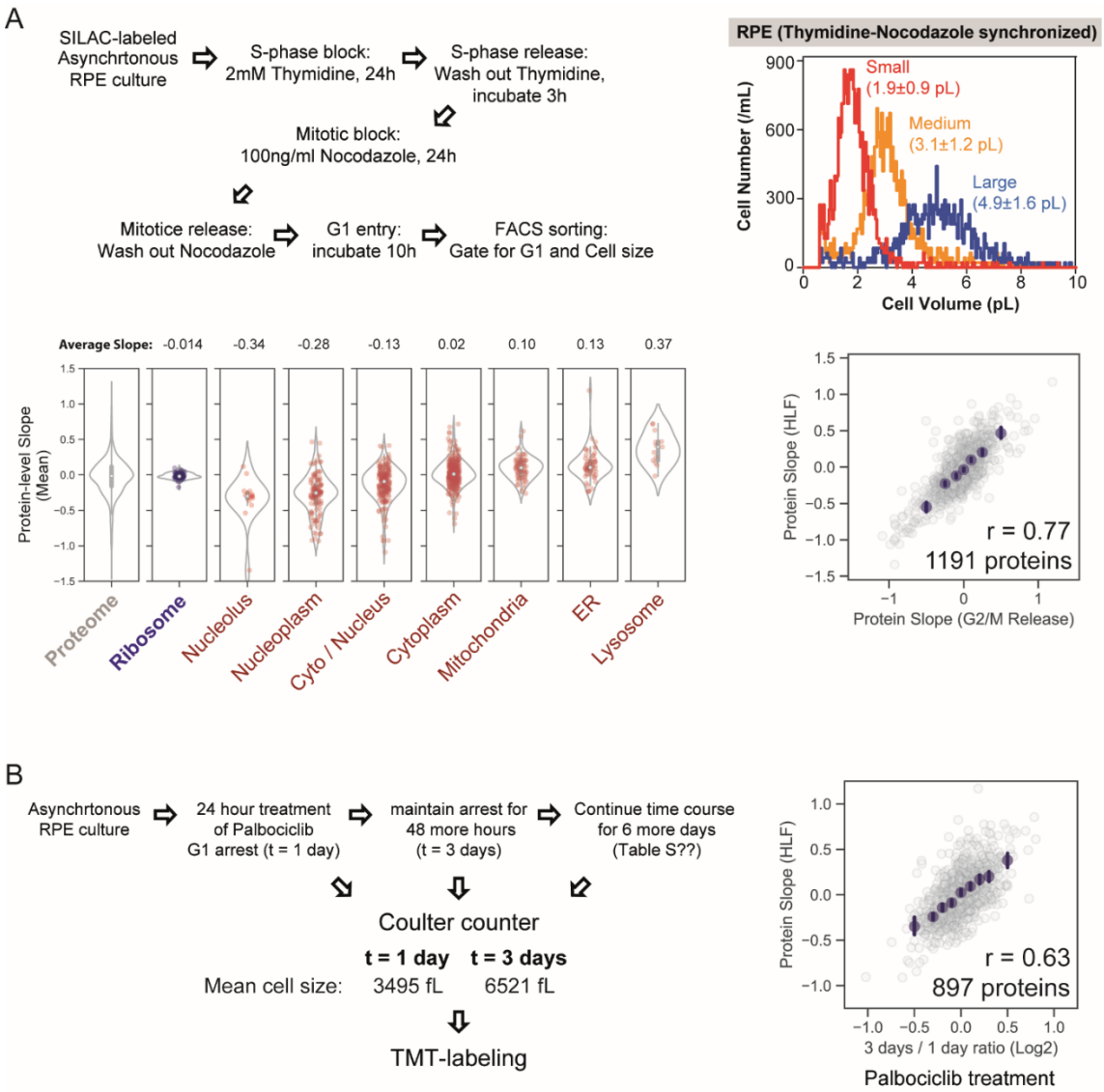


Figure S6. Controls experiments indicating that cell size, and not cell age, drives changes to the proteome changes: (A) Metabolically labeled *hTERT RPE-1* cells were synchronously released into the cell cycle following a Thymidine-Nocodazole cell cycle arrest. Protein slopes were calculated as in Figure 1. Only proteins with at least 4 peptide measurements in both replicate experiments are considered for the violin plots and the correlation to the slope values generated from HLF cells in Figure 1. **(B)** Asynchronous *hTERT RPE-1* cells were treated long-term with Palbociclib. The attainment of differentially sized cells was confirmed by coulter counter analysis. Protein ratios were determined by TMT quantitation.

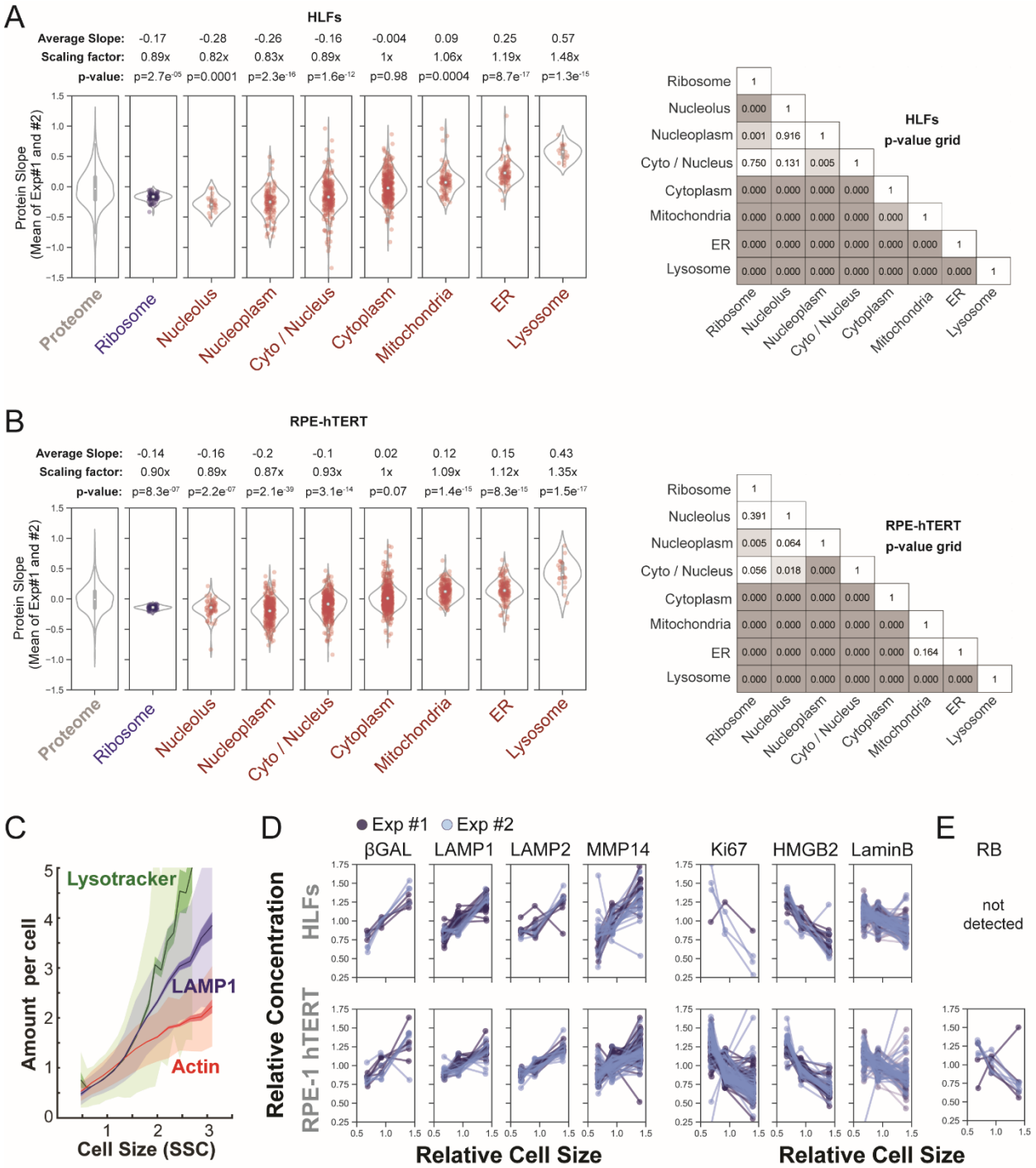


Figure S7. Differential scaling in the proteome content of distinct sub-cellular compartments: (A) Distribution of slopes derived from HLF cells for proteins associated with the indicated compartments. Violin plots depict the average slopes for the proteins highlighted in Figure 2B. P-values above the violin plots are derived from a t-test between the indicated protein group and the rest of the dataset. t-tests

comparing the slopes for each group of proteins are visualized in a grid format. **(B)** Replicate experiment using the immortalized *hTERT RPE-1* cell line was performed as in (A). **(C)** Validation of lysosome super-scaling with cell size using flow cytometry. Both the lysosomal protein LAMP1 and the Lysotracker dye amount increase with cell size faster than Actin, which is a proxy for total protein. The data for G1 *hTERT RPE-1* cells were binned by cell size (SSC, the side scatter parameter) and plotted as mean protein amounts per cell for each size bin (solid lines). Dark shaded area shows standard error of the mean for each bin, and light shaded area shows the standard deviation. A representative is shown of n=5 biological replicates for each experiment. About 100,000 cells were analyzed for each sample. **(D)** Examples from our proteomics data set of cell-size-dependent protein concentration changes in proliferating cells that are normally associated with senescence. **(E)** RB is diluted with G1 cell size.

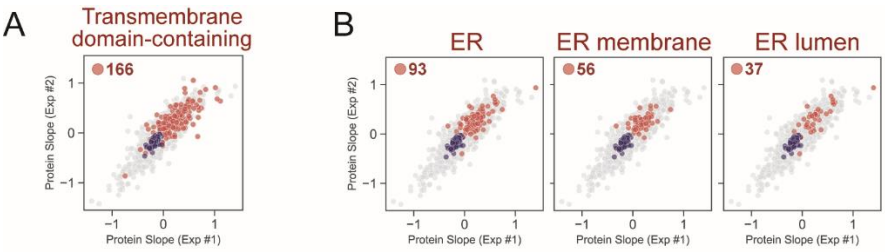


Figure S8. Scaling behavior of lumenal and membrane-associated proteins: (A) Scaling behavior of proteins predicted to contain a transmembrane domain. **(B)** Scaling behavior of ER proteins that are annotated to be either membrane-associated or not (*i.e.*, lumenal).

OLS Regression Results - Summarized
1,700 proteins common to Protein Slope, RNA Slope, and Protein Turnover (T-50%) datasets

Dependent variable - Protein Slope (HLF)

Independent variables	coef	std_err	t-stat	P>|t|	[0.025]	[0.975]			Prediction %
constant	-0.0082	0.005	-1.586	0.113	-0.018	0.002	R-squared	0.528	
Replicate #1 (vs #2)	0.689	0.016	43.552	0	0.658	0.72	Adj. R-squared	0.527	100%
							Prob (F-statistic)	7.53E-279	
constant	-0.0209	0.02	-1.025	0.306	-0.061	0.019	R-squared	0.076	
RNA Slope	0.2835	0.024	11.848	0	0.237	0.33	Adj. R-squared	0.076	14%
							Prob (F-statistic)	3.59E-31	
constant	-0.0802	0.022	-3.591	0	-0.124	-0.036	R-squared	0.115	
Protein Turnover (T-50%) - Zecha *et al.*, 2018	0.1473	0.025	5.99	0	0.099	0.195	Adj. R-squared	0.113	22%
Codon Score	-0.1198	0.02	-5.964	0	-0.159	-0.08	Prob (F-statistic)	1.42E-44	
RNA Slope	0.2852	0.023	12.153	0	0.239	0.331			
constant	0.1439	0.028	5.167	0	0.089	0.199	R-squared	0.266	
Mitochondria_Localization	0.1449	0.057	2.543	0.011	0.033	0.257	Adj. R-squared	0.264	50%
Lysosome_Localization	1.2597	0.123	10.274	0	1.019	1.5	Prob (F-statistic)	4.07E-111	
ER_Localization	0.5075	0.06	8.437	0	0.39	0.626			
Nucleus_Localization	-0.6631	0.039	-16.887	0	-0.74	-0.586			
Ribosome_Localization	-0.5632	0.099	-5.678	0	-0.758	-0.369			
constant	0.0619	0.029	2.151	0.032	0.005	0.118	R-squared	0.315	
Protein Turnover (T-50%) - - Zecha *et al.*, 2018	0.0966	0.022	4.407	0	0.054	0.14	Adj. R-squared	0.312	60%
Codon Score	-0.0811	0.018	-4.534	0	-0.116	-0.046	Prob (F-statistic)	3.86E-133	
RNA Slope	0.1971	0.021	9.242	0	0.155	0.239			
Mitochondria_Localization	0.1771	0.055	3.207	0.001	0.069	0.285			
Lysosome_Localization	1.2251	0.119	10.324	0	0.992	1.458			
ER_Localization	0.4829	0.058	8.255	0	0.368	0.598			
Nucleus_Localization	-0.545	0.039	-13.804	0	-0.622	-0.468			
Ribosome_Localization	-0.5665	0.096	-5.897	0	-0.755	-0.378			

Figure S9. Linear regression analysis predicts size scaling behavior of individual proteins: The prediction of size scaling behavior was based on the 1,700 proteins that are in the published protein turnover dataset (HeLa cells) (*32*), as well as our RNA Slope, and Protein Slope datasets (at least 2 peptides / protein) that we report here. Independent variables for codon affinity, RNA Slope, and Protein turnover (time to replace 50% of a given protein species) were each independently standardized by subtracting all values by the dataset's mean and then dividing by the dataset's standard deviation. The subcellular localization variable was based on Uniprot's "Subcellular location [CC]" annotations and entered as a binary value for each compartment (1 if a protein possessed an annotation and 0 if it did not). Only subcellular compartments that provided nonredundant predictive power were ultimately included in the model. A constant value was added to the regression equation using the add_constant function in statsmodels. We set the benchmark for predictive accuracy (Prediction %) as the correlation between biological replicate experiments, *i.e.*, Protein Slope from Exp #1 vs Exp #2.

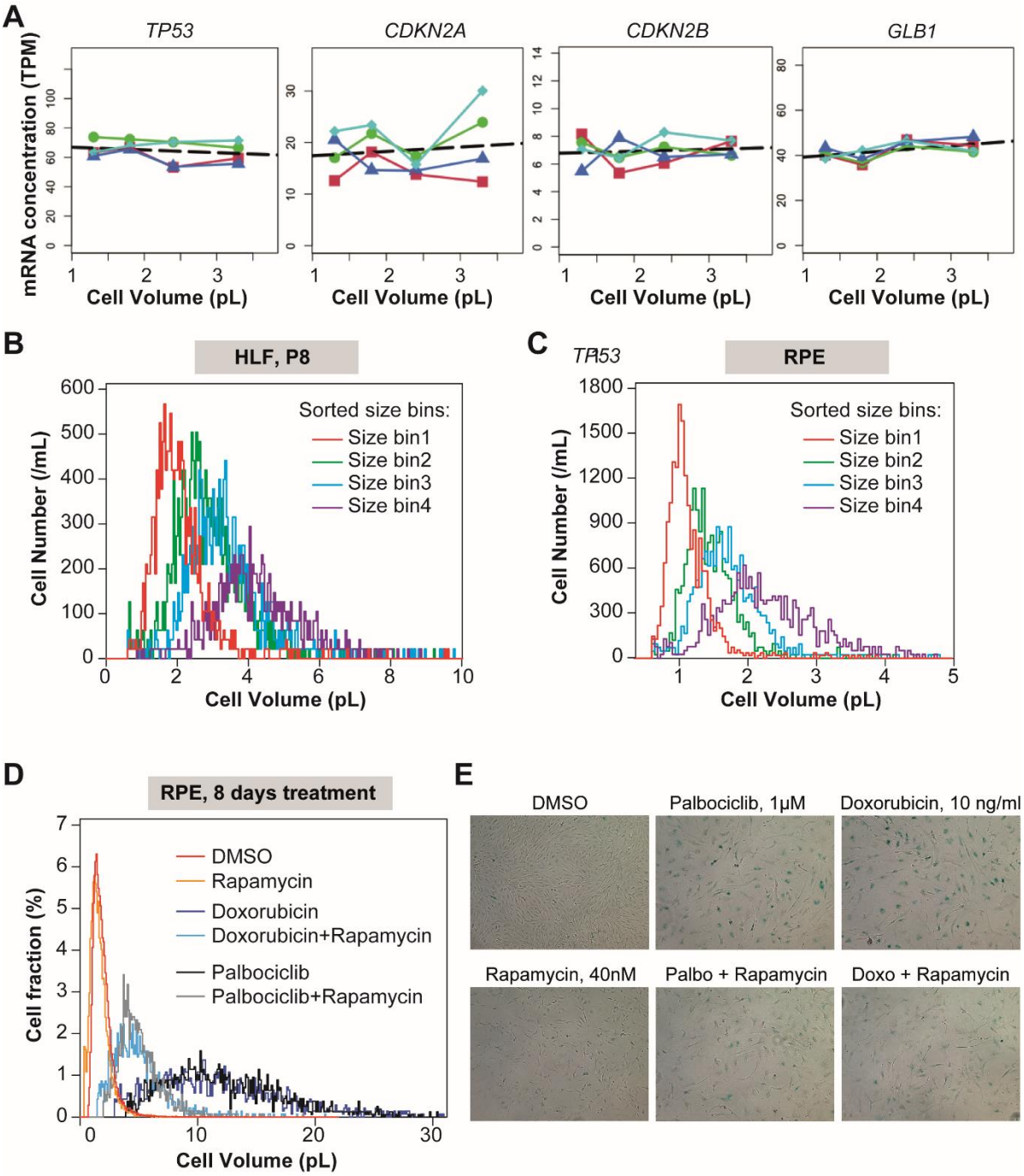


Figure S10. Analysis of senescence markers in different-sized cells: (A) Transcript levels of key senescence reporter genes in size-sorted HLF cells. G1 HLF cells were sorted into four size bins using FACS. The concentrations of *TP53, CDKN2A, CDKN2B,* and *GLB1* mRNAs were then determined by RNAseq and plotted against the mean cell size for each bin. Each colored line represents one of four

replicates. **(B)** Cell size distributions for the four bins of FACS-sorted HLF cells that were then re-plated and passaged to determine replicative senescence dynamics (see Fig. 3C-D), as measured with a Coulter counter. **(C)** Cell size distributions for the four bins of FACS-sorted RPE-1 cells that were then re-plated and passaged in the presence of Doxorubicin to determine the DNA-damage-induced senescence dynamics (see Fig. 3F), as measured with a Coulter counter. **(D-E)** Cell size distributions (D) and SA-beta-Gal staining images (E) of the RPE-1 cells treated for 8 days with DMSO, Palbociclib or Doxorubicin, in the presence or absence of Rapamycin to determine the effect of cell size reduction on senescence dynamics (see Fig. 3G). Each of the experiments in (B)-(E) was performed in two biological replicates.

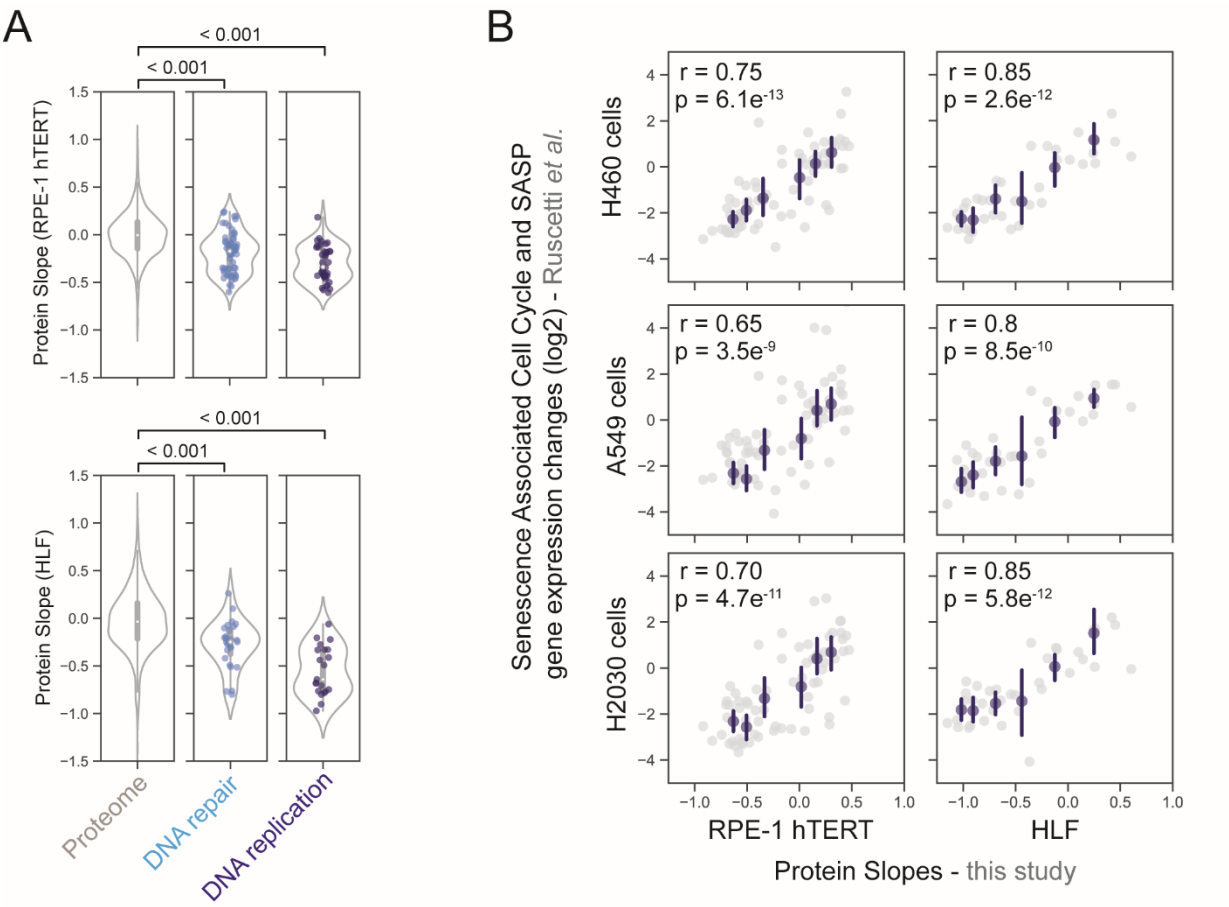


Figure S11. Classes of sub- and super-scaling proteins in HLF and hTERT RPE-1 cells: (A) Slopes for proteins grouped by Uniprot GO annotations for DNA repair (GO:0006281) and DNA replication (GO:0006260). **(B)** Size-dependent proteome changes from this study correlate with the senescence-associated SASP and cell cycle gene expression changes defined by Ruscetti, *et al.* (*34*).

Data S1. Protein and Peptide Slope values for size-sorted G1 cells

Data S2. Long-term Palbociclib time course

Data S3. mRNA concentrations in G1 HLF cells sorted into three size bins

Data S4. Ploidy-sorted Protein Slopes

Data S5. MaxQuant spectral search parameter file

References

1. E. Zatulovskiy, J. M. Skotheim, On the Molecular Mechanisms Regulating Animal Cell Size Homeostasis. *Trends Genet* **36**, 360-372 (2020).
2. M. B. Ginzberg, R. Kafri, M. Kirschner, Cell biology. On being the right (cell) size. *Science* **348**, 1245075 (2015).
3. A. C. Lloyd, The regulation of cell size. *Cell* **154**, 1194-1205 (2013).
4. J. Lengefeld *et al.*, Cell size is a determinant of stem cell potential during aging. *bioRxiv*, 2020.2010.2027.355388 (2020).
5. T. P. Miettinen, M. Bjorklund, Cellular Allometry of Mitochondrial Functionality Establishes the Optimal Cell Size. *Dev Cell* **39**, 370-382 (2016).
6. G. E. Neurohr *et al.*, Excessive Cell Growth Causes Cytoplasm Dilution And Contributes to Senescence. *Cell* **176**, 1083-1097 e1018 (2019).
7. O. Padovan-Merhar *et al.*, Single mammalian cells compensate for differences in cellular volume and DNA copy number through independent global transcriptional mechanisms. *Mol Cell* **58**, 339-352 (2015).
8. H. A. Crissman, J. A. Steinkamp, Rapid, simultaneous measurement of DNA, protein, and cell volume in single cells from large mammalian cell populations. *J Cell Biol* **59**, 766-771 (1973).
9. H. Kempe, A. Schwabe, F. Cremazy, P. J. Verschure, F. J. Bruggeman, The volumes and transcript counts of single cells reveal concentration homeostasis and capture biological noise. *Mol Biol Cell* **26**, 797-804 (2015).
10. J. Lin, A. Amir, Homeostasis of protein and mRNA concentrations in growing cells. *Nat Commun* **9**, 4496 (2018).
11. D. H. Williamson, A. W. Scopes, The distribution of nucleic acids and protein between different sized yeast cells. *Exp Cell Res* **24**, 151-153 (1961).
12. E. Zlotek-Zlotkiewicz, S. Monnier, G. Cappello, M. Le Berre, M. Piel, Optical volume and mass measurements show that mammalian cells swell during mitosis. *J Cell Biol* **211**, 765-774 (2015).
13. S. Son *et al.*, Resonant microchannel volume and mass measurements show that suspended cells swell during mitosis. *J Cell Biol* **211**, 757-763 (2015).
14. D. F. Berenson, E. Zatulovskiy, S. Xie, J. M. Skotheim, Constitutive expression of a fluorescent protein reports the size of live human cells. *Mol Biol Cell* **30**, 2985-2995 (2019).
15. Y. H. Chan, W. F. Marshall, Scaling properties of cell and organelle size. *Organogenesis* **6**, 88-96 (2010).
16. W. F. Marshall, Scaling of Subcellular Structures. *Annu Rev Cell Dev Biol* **36**, 219-236 (2020).
17. H. Cantwell, P. Nurse, Unravelling nuclear size control. *Curr Genet* **65**, 1281-1285 (2019).
18. J. Zhurinsky *et al.*, A coordinated global control over cellular transcription. *Curr Biol* **20**, 2010-2015 (2010).
19. Z. N. Demidenko, M. V. Blagosklonny, Growth stimulation leads to cellular senescence when the cell cycle is blocked. *Cell Cycle* **7**, 3355-3361 (2008).
20. K. M. Schmoller, J. J. Turner, M. Koivomagi, J. M. Skotheim, Dilution of the cell cycle inhibitor Whi5 controls budding-yeast cell size. *Nature* **526**, 268-272 (2015).
21. E. Zatulovskiy, S. Zhang, D. F. Berenson, B. R. Topacio, J. M. Skotheim, Cell growth dilutes the cell cycle inhibitor Rb to trigger cell division. *Science* **369**, 466-471 (2020).
22. M. D'Ario *et al.*, Cell size controlled in plants using DNA content as an internal scale. *Science* **372**, 1176-1181 (2021).

23. D. Keifenheim *et al.*, Size-Dependent Expression of the Mitotic Activator Cdc25 Suggests a Mechanism of Size Control in Fission Yeast. *Curr Biol* **27**, 1491-1497 e1494 (2017).
24. Y. Chen, G. Zhao, J. Zahumensky, S. Honey, B. Futcher, Differential Scaling of Gene Expression with Cell Size May Explain Size Control in Budding Yeast. *Mol Cell* **78**, 359-370 e356 (2020).
25. M. Mann, Functional and quantitative proteomics using SILAC. *Nat Rev Mol Cell Biol* **7**, 952-958 (2006).
26. K.-L. Claude, D. Bureik, P. Adarska, A. Singh, K. M. Schmoller, Transcription coordinates histone amounts and genome content. *bioRxiv*, 2020.2008.2028.272492 (2020).
27. M. P. Swaffer *et al.*, Size-independent mRNA synthesis and chromatin-based partitioning mechanisms generate and maintain constant amounts of protein per cell. *bioRxiv*, 2020.2008.2028.272690 (2020).
28. A. Tzur, J. K. Moore, P. Jorgensen, H. M. Shapiro, M. W. Kirschner, Optimizing optical flow cytometry for cell volume-based sorting and analysis. *PLoS One* **6**, e16053 (2011).
29. M. S. Nassrally *et al.*, Cell cycle arrest in replicative senescence is not an immediate consequence of telomere dysfunction. *Mech Ageing Dev* **179**, 11-22 (2019).
30. C. Tan *et al.*, Cell size homeostasis is maintained by CDK4-dependent activation of p38 MAPK. *Dev Cell* **56**, 1756-1769 e1757 (2021).
31. Y. Liu, A. Beyer, R. Aebersold, On the Dependency of Cellular Protein Levels on mRNA Abundance. *Cell* **165**, 535-550 (2016).
32. J. Zecha *et al.*, Peptide Level Turnover Measurements Enable the Study of Proteoform Dynamics. *Mol Cell Proteomics* **17**, 974-992 (2018).
33. A. Hernandez-Segura, J. Nehme, M. Demaria, Hallmarks of Cellular Senescence. *Trends Cell Biol* **28**, 436-453 (2018).
34. M. Ruscetti *et al.*, NK cell-mediated cytotoxicity contributes to tumor control by a cytostatic drug combination. *Science* **362**, 1416-1422 (2018).
35. Z. N. Demidenko *et al.*, Rapamycin decelerates cellular senescence. *Cell Cycle* **8**, 1888-1895 (2009).
36. N. E. Sharpless, C. J. Sherr, Forging a signature of in vivo senescence. *Nat Rev Cancer* **15**, 397-408 (2015).
37. J. Sage *et al.*, Targeted disruption of the three Rb-related genes leads to loss of G(1) control and immortalization. *Genes Dev* **14**, 3037-3050 (2000).
38. J. H. Dannenberg, A. van Rossum, L. Schuijff, H. te Riele, Ablation of the retinoblastoma gene family deregulates G(1) control causing immortalization and increased cell turnover under growth-restricting conditions. *Genes Dev* **14**, 3051-3064 (2000).
39. L. Mu *et al.*, Mass measurements during lymphocytic leukemia cell polyploidization decouple cell cycle- and cell size-dependent growth. *Proc Natl Acad Sci U S A* **117**, 15659-15665 (2020).
40. H. O. Lee, J. M. Davidson, R. J. Duronio, Endoreplication: polyploidy with purpose. *Genes Dev* **23**, 2461-2477 (2009).
41. R. Kumari, P. Jat, Mechanisms of Cellular Senescence: Cell Cycle Arrest and Senescence Associated Secretory Phenotype. *Front Cell Dev Biol* **9**, 645593 (2021).
42. A. Sakaue-Sawano *et al.*, Visualizing spatiotemporal dynamics of multicellular cell-cycle progression. *Cell* **132**, 487-498 (2008).
43. J. Deng, H. Erdjument-Bromage, T. A. Neubert, Quantitative Comparison of Proteomes Using SILAC. *Curr Protoc Protein Sci* **95**, e74 (2019).
44. J. Zecha *et al.*, TMT Labeling for the Masses: A Robust and Cost-efficient, In-solution Labeling Approach. *Mol Cell Proteomics* **18**, 1468-1478 (2019).
45. J. Cox *et al.*, Andromeda: a peptide search engine integrated into the MaxQuant environment. *J Proteome Res* **10**, 1794-1805 (2011).

46. J. Cox, M. Mann, MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol* **26**, 1367-1372 (2008).
47. J. E. Elias, S. P. Gygi, Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat Methods* **4**, 207-214 (2007).
48. C. UniProt, UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res* **47**, D506-D515 (2019).
49. J. Cox, M. Mann, 1D and 2D annotation enrichment: a statistical method integrating quantitative proteomics with complementary high-throughput data. *BMC Bioinformatics* **13 Suppl 16**, S12 (2012).
50. M. Frenkel-Morgenstern *et al.*, Genes adopt non-optimal codon usage to generate cell cycle-dependent oscillations in protein levels. *Mol Syst Biol* **8**, 572 (2012).
51. A. Frankish *et al.*, GENCODE reference annotation for the human and mouse genomes. *Nucleic Acids Res* **47**, D766-D773 (2019).
52. A. D. Yates *et al.*, Ensembl 2020. *Nucleic Acids Res* **48**, D682-D688 (2020).
53. A. Dobin *et al.*, STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21 (2013).
54. B. Langmead, C. Trapnell, M. Pop, S. L. Salzberg, Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10**, R25 (2009).
55. A. Roberts, L. Pachter, Streaming fragment assignment for real-time analysis of sequencing experiments. *Nat Methods* **10**, 71-73 (2013).
56. C. Schwarz *et al.*, A Precise Cdk Activity Threshold Determines Passage through the Restriction Point. *Mol Cell* **69**, 253-264 e255 (2018).