doi:10.1038/nature10098

Supplementary methods

Cell culture and preparation of SILAC media

NIH3T3 (mouse fibroblasts, obtained from ATCC) and MCF7 (human breast cancer cells) cells were used and cultivated in SILAC medium at 37°C with 5% CO_2 and split every second or third day. SILAC media were essentially prepared as described previously [1]. In essence, Dulbecco's Modified Eagle's Medium (DMEM) Glutamax lacking arginine and lysine (a custom preparation from Gibco) supplemented with 10% dialyzed fetal bovine serum (dFBS, Gibco) was used. Heavy (H) and light (L) SILAC media were prepared by adding 84 mg/l $^{13}C_6{}^{15}N_4$ L-arginine plus 146 mg/l $^{13}C_6{}^{15}N_2$ L-lysine or the corresponding non-labeled amino acids (Sigma), respectively. Labeled amino acids were purchased from Sigma Isotec.

RNA isolation

Total RNA was extracted using Trizol Reagent (Invitrogen) following the manufacturer's instructions. In brief, cells were washed once with 1x D-PBS (Gibco), harvested by trypsinization and pelleted by centrifugation for 5 min at 300x g and 4°C with a Heraeus Multifuge 3 S (Heraeus). The supernatant was carefully removed and the cells were lysed by adding 1 ml Trizol Reagent. To ensure homogenization, the cell lysate was passed 8-10 times through a 20G needle. After adding 200 ul of chloroform (Invitrogen), the mixture was transferred to a Phase-Lock-Gel tube (5Prime), vigorously mixed and centrifuged at full speed (20,000 g) for 5 min at 4°C. The upper, RNA containing aqueous phase was removed and transferred to a fresh, RNase-free tube (Ambion). To precipitate RNA, 1/10 reaction volume of 5 M NaCl and an equal volume of 2-propanol were added and incubated for 10 min at RT. Precipitated RNA was collected through centrifugation at full speed (20,000 g) for 30 min at 4°C. The pellet was washed with an equal volume of 70% ethanol and precipitated again at 20,000g for 20 min. RNA pellets were resuspended in 20 µl RNase-free sterile water and RNA quantity was assessed spectrophotometrically using the NanoDrop ND-1000 UV-VIS Spectrophotometer (Thermo Fisher).

Double pulse-labeling of NIH3T3 cells with heavy amino acids and 4-thiouridine

After growing NIH3T3 cells to 60-70 % confluency in light (L) SILAC medium, the fibroblasts were thoroughly washed three times with pre-warmed D-PBS and transferred to heavy (H) SILAC medium. To avoid a possible interference of 4sU with protein turnover (see Fig. 1), only SILAC medium of cells designated for RNA isolation was additionally supplemented with 4-thiouridine (4sU; Sigma-Aldrich) to a final concentration of 400 µM.

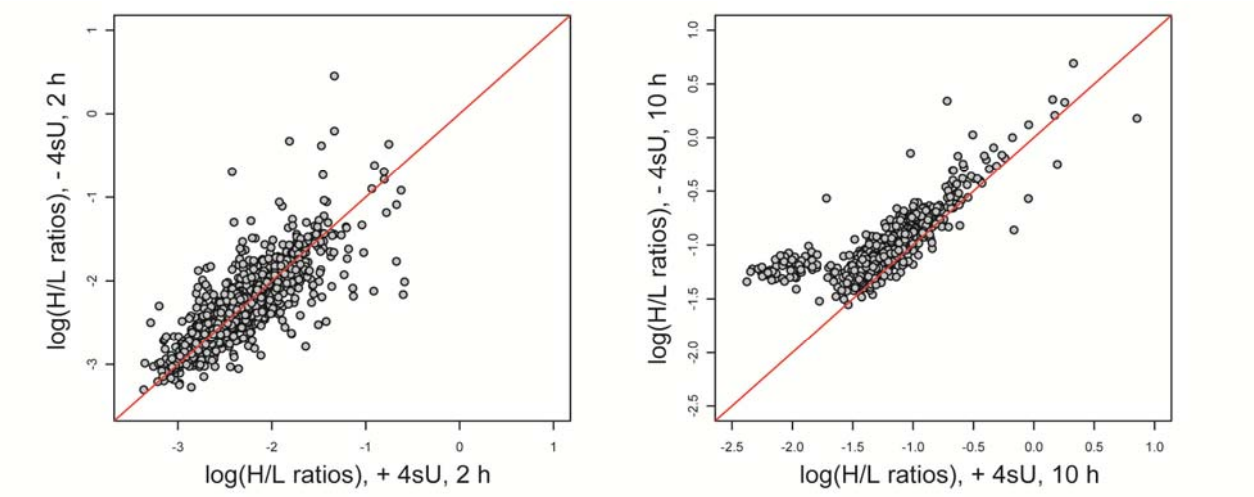

Fig. 1: Effect of 4sU on protein pulse-labeling. NIH3T3 cells were transferred from light (L) to heavy (H) SILAC medium either with or without 4sU (400 µM) for 2h (left panel) or 10h (right panel). 4sU for 2h did not appreciably affect protein turnover given by H/L ratios (left). In contrast, 4sU labeling for 10h resulted in overall slightly decreased turnover compared to untreated control cells. We also observed a population of proteins with significantly lower H/L ratios after 10 h labeling with 4sU. Closer inspection revealed that this protein population is comprised almost exclusively ribosomal proteins. Collectively, these results suggest that pulse labeling with 4sU for extended time periods affects cellular protein turnover. We therefore exclusively added 4sU to the cells used for measuring mRNA half-lives (2 h pulse labeling). For protein turnover measurement we did not use 4sU in any of the three time points.

Pulse-labeling of proteins was performed for 1.5, 4.5 and 13.5 h, while labeling of newly transcribed RNA was conducted for 2 h. At the indicated time points, cells were scraped off, spun down (10 min, 600 g, 4 °C) and used either for protein or RNA isolation.

Preparation of newly-synthesized and pre-existing RNA

Apart from minor changes, all steps described in the following were essentially performed as described previously [2]:

Biotinylation and purification of 4sU-labeled RNA

Biotinylation of 4sU-labeled RNA was performed using EZ-Link Biotin-HPDP (Pierce) dissolved in dimethylformamide (DMF) at a concentration of 1 mg/mL and stored at 4°C. Biotinylation was carried out in RNase-free water (Invitrogen)) with 10 mM Tris-HCl (pH 7.4), 1 mM EDTA and 1 mg/ml Biotin-HPDP (Perbio) at a final RNA concentration of 1 µg/µl RNA for 3 h in the dark at RT. In general 250 µg total RNA were used for the biotinylation reaction. To purify biotinalyted RNA from excess of Biotin-HPDP, a Phenol:Chloroform:Isoamylalkohol (v/v = 25:24:1, Invitrogen) extraction was performed. Phenol:Chloroform:Isoamylalkohol was added to the reaction mixture in a 1:1 ratio, transferred to a Phase-Lock-Gel tube (5Prime), vigorously mixed and centrifuged at full speed (20,000 g) for 5 min at 4°C. The upper, RNA containing aqueous phase was removed and transferred to a fresh, RNase-free tube (Ambion). To precipitate RNA, 1/10 reaction volume of 5 M NaCl and an equal volume of 2-propanol were added and incubated for 10 min at RT. Precipitated RNA was collected through centrifugation at full speed (20,000g) for 30 min at 4°C. The pellet was washed with an equal volume of 75% ethanol and precipitated again at 20,000g for 20 min. Finally, RNA was reconstituted in 25-50 ul of RNase-free water.

Separation of biotinylated 4sU-RNA from unlabeled RNA

For specific isolation of biotinylated 4sU-RNA streptavidin coated magnetic beads (µMACS, Miltenyi) were used. To avoid unfavorable secondary RNA structures which potentially impair the binding to the beads, the RNA was first denatured at 65°C for 10 min followed by rapid cooling on ice for 5 min. 4sU-biotinylated RNA was incubated with µMACS streptavidin beads (usually 20 µg labeled RNA with 200 µl beads) for 15 min at RT with rotation. Actual separation of 4sU-RNA from unlabeled RNA was performed by applying the mixture onto a µColumn (Miltenyi) in the magnetic field of the µMACS separator (Miltenyi). Before, µColumns were equilibrated with 100 µl equilibration buffer that is part of the Miltenyi µBeads Kit. To recover unlabeled, pre-existing RNA, the column flow-through (FT) was collected. Then, µColumns were washed three times with 0.5 mL 55°C washing buffer (100 mM Tris-HCl, pH 7.4, 10 mM EDTA, 1 M NaCl, 0.1% Tween20). All washing fractions were discarded. Eluation (E) of labeled RNA from the beads was achieved by adding 100 µl of freshly prepared 100 mM dithiothreitol (DTT) onto the

column followed by a second elution round 3 min later while the columns were retained in the magnetic field. Lastly, RNA from the FT and elution step (E) were recovered using the RNeasy MinElute kit (Qiagen) and reconstituted in 25-50 µl of RNase-free water.

Estimation of the doubling time of NIH3T3 cells

To estimate the duration of the cell cycle of NIH 3T3 mouse fibroblasts under normal cell culture conditions, cells of two culture dishes were pooled and counted six times with haemocytometers (C-Chips™) at 0.75 h, 5 h, 9 h, 13.25 h after the onset of the experiment. The doubling time is estimated assuming exponential growth. Neglecting cell death the doubling time is assumed to be equal to the duration time of one cell cycle. The mean and standard deviation of the measured cell numbers at the indicated time points are used to obtain the doubling time by linear regression. The uncertainty of the parameters was determined by bootstrapping. The 68%-confidence intervals were obtained after resampling the measured cell numbers at each time point with replacement (n = 10,000) and performing the calculation described above. The duration of the cell cycle was determined as 19.9h < 27.5h < 33.6h.

Sample preparation for mass spectrometry analysis

In-gel digestion

Cells harvested after indicated time points were lysed in an appropriate amount of RIPA buffer (50 mM Tris-HCl pH 7.4, 150 mM NaCl, 1% Triton-X100, 1% Sodium deoxycholate and 0.1% SDS) After 20 min on ice, the lysates were cleared by centrifugation at 14,000 rpm in a microcentrifuge for 15 min at 4°C. For proteome analysis approximately 150 µg whole cell lysate were subjected to SDS-PAGE using NuPAGE Novex 4 to 12% gradient gels (Invitrogen) under reducing conditions according to the manufacturer's instructions. Proteins were fixed in fixative solution (50% methanol (v/v), 10% acetic acid (w/v)) and stained afterwards with the Colloidal Blue staining Kit (Invitrogen). Whole gel lanes were cut into 20 gel slices which were individually subjected to reduction, alkylation and in-gel digestion with sequence grade modified trypsin (Promega) according to standard protocols [3]. After in-gel digestion peptides were extracted and desalted using StageTips [4] before analysis by mass spectrometry.

In-solution digestion

Cells were lysed in appropriate amounts of denaturation buffer (6 M urea/2 M thiourea in 10 mM HEPES, pH 8.0) for 20 min on ice. The lysates were cleared by centrifugation for 10 min (14,000 rpm at 4°C) and transferred to fresh tubes. Protein samples were reduced for 30 min at RT in 10 mM dithiothreitol solution followed by alkylation for 20 min by 55 mM iodacetamide in the dark at RT. The endoproteinase LysC (Wako) was added (protein:enzyme ratio of 50:1) and incubated for 4 h at room temperature. After dilution of the sample with four times digestion buffer (50 mM ammonium bi-carbonate (NH_4HCO_3) in water, pH 8.0), sequence grade modified trypsin (Promega; protein:enzyme ratio of 50:1) was used for overnight digestion. Trypsin and Lys-C activity was quenched by adding trifluoroacetic acid to adjust the pH to < 2, and peptides were extracted and desalted using StageTips [4].

HPLC and mass spectrometry

Reversed-phase liquid chromatography (rpHPLC) was performed employing a Eksigent NanoLC – 1D Plus system using self-made fritless C18 microcolumns [5] (75 µm ID packed with ReproSil-Pur C18-AQ 3-µm resin, Dr. Maisch GmbH) connected on-line to the electrospray ion source (Proxeon) of a LTQ-Orbitrap Velos mass spectrometer (Thermo Fisher). Peptide samples were picked up by the autosampler and loaded onto the column with a flow rate of 250 nl/min. Subsequent sample elution was performed at a flow rate of 200 nl/min with a 10 to 60 % acetonitrile gradient over 6 h in 0.5% acetic acid for online MS analysis. The LTQ-Orbitrap Velos instrument was operated in the data dependent mode (DDA) with a full scan in the Orbitrap followed by up to 20 consecutive MS/MS scans in the LTQ. Precursor ion scans (m/z 300–1700) were acquired in the Orbitrap part of the instrument (resolution R = 60,000; target value of 1 x 10^6), while in parallel the 20 most intense ions were isolated (target value of 3,000; monoisotopic precursor selection enabled) and fragmented in the LTQ part of the instrument by collision induced dissociation (CID; normalized collision energy 35 %; wideband activation enabled). Ions with an unassigned charge state and singly charged ions were rejected. Former target ions selected for MS/MS were dynamically excluded for 60 s. Total cycle time for one full scan plus up to 20 MS/MS scans was approximately 2 s.

Processing of mass spectrometry data

Identification and quantification of proteins was carried out with the MaxQuant software package [6] (v. 1.0.13.13) In essence, isotope clusters and SILAC doublets were extracted, re-calibrated

and quantified in the raw data files with Quant.exe (heavy labels: Arg10 and Lys8; maximum of three labeled amino acids per peptide; polymer detection enabled; top 6 MS/MS peaks per 100 Da). The generated peak lists (msm-files) were submitted to a MASCOT search engine (version 2.2, MatrixScience) and searched against an in-house curated concatenated target-decoy database [7] of forward and reversed proteins. The IPI mouse database (v. 3.64) was employed and supplemented with common contaminants (e.g. trypsin, BSA). Databases used in the iBAQ approach for absolute protein quantification were additionally supplemented with UPS2 (Universal Protein Standard 2; Sigma-Aldrich) proteins. Full tryptic specificity was required and a maximum of two missed cleavages and a mass tolerance of 0.5 Da for fragment ions applied. The initial mass accuracy cut-off on the parent ion was 7 ppm but subsequently narrowed down by filtering based on hits to reversed peptides in the target-decoy database. Oxidation of methionine and acetylation of the protein N-terminus were used as variable modifications, carbamidomethylation of cysteine as a fixed modification. Filtering of putative MASCOT peptide identifications, assembly of proteins and re-quantification was performed with Identify.exe. A minimum peptide length of 6 amino acids was required. False discovery rates (FDR) were estimated based on matches to reversed sequences in the concatenated target-decoy database. A maximum false discovery rate of 1% at both the peptide and the protein level was allowed. Peptides were assigned to protein groups (that is a cluster of a base protein plus additional proteins matching to a subset of the same peptides). Protein groups containing matches to proteins from the reversed database or contaminants were discarded. Protein H/L ratios were calculated as the median of all H/L peptide ratios for a specific protein. We required at least three peptide ratios at all three time points combined for calculation of half-lives.

Intensity-based absolute quantification (iBAQ) of proteins

The Universal Proteomics Standard (UPS2, Sigma-Aldrich) consists of 48 accurately quantified human proteins formulated into a dynamic range of concentrations spanning six orders of magnitude. The standard was dissolved in denaturation buffer (6 M urea/2 M thiourea in 10 mM HEPES, pH 8.0), mixed with NIH3T3 lysate (4.24 µg UPS2 + 13 µg NIH3T3 ~ 2.8 x 10^5 cells) and subjected to in-solution digest followed by mass-spectrometry analysis.

MaxQuant computes protein intensities as the sum of all identified peptide intensities (maximum detector peak intensities of the peptide elution profile, including all peaks in the isotope cluster). Protein intensities were divided by the number of theoretically observable peptides (calculated by in silico protein digestion with a PERL script, all fully tryptic peptides between 6 and 30 amino

acids were counted while missed cleavages were neglected). The resulting "iBAQ" intensities were log-transformed and plotted against known log-transformed absolute molar amounts of the spiked-in standard proteins (UPS2 standard). Linear regression was used to fit iBAQ intensities to absolute standard protein amounts. The slope and intercept from this calibration curve was then used to convert iBAQ intensities to molar amounts for all identified proteins. To assess the uncertainty of the fit the confidence intervals of the slope and axis intersect of the calibration curve were calculated by bootstrapping. The iBAQ intensities were resampled with replacement (n = 10,000). The slope and intersect were determined from the mean and standard deviation of the resampled data by linear regression. The determined value with the average 68% confidence interval was 3.7 ± 0.31 (8.5%) for the intersect and 1.1 ± 0.08 (7.4%) for the slope. Cellular copy numbers were obtained by calculating the number of molecules using the Avogadro constant followed by division by the number of cells used in the respective experiment. In doing so, absolute cellular copy numbers for almost 2,000 proteins were obtained. Of note, even iBAQ values for abundant NIH3T3 proteins did not reach saturation as they fell within the linear range covered by the UPS2 spike-in standards (data not shown). Finally, cellular copy numbers calculated from the in-solution digest were used to scale (by linear regression) and convert the iBAQ values of the 6,445 proteins identified in the actual in-gel digest protein copies/cell.

Determination of protein half-lives

SILAC enables the determination of protein turnover by measuring the ratio (r) of protein labelled with heavy amino acids (P_H) and light amino acids (P_L) at different time points (1.5h, 4.5h and 13.5h). (Fig. 1 B, main text).

$$r = \frac{P_H}{P_L} \qquad [S1]$$

Proteins labelled with light amino acids (P_L) are assumed to decay exponentially with the degradation rate constant (k_{dp})

$$P_L = P_0 e^{-k_{dp} t} . \qquad [S2]$$

Protein labelled with heavy amino acids (P_H) can be expressed as the difference between total number of a specific protein (P_{total}) and P_L. The total number (P_{total}) is assumed to double during the duration of one cell cycle:

$$P_H(t) = P_{total}(t) - P_L(t) = P_0 2^{t/t_{cc}} - P_L(t), \quad [S3]$$

P_0 is the protein copy number at the onset of the experiment (t = 0) and t_{cc} is the duration of one cell cycle. The rate constant of the protein decay (k_{dp}) can then be obtained by linear regression using Eqs. S1-S3:

$$k_{dp} = \frac{\sum_{i=1}^{m} \log_e(r_{t_i} + 1) t_i}{\sum_{i=1}^{m} t_i^2} - \frac{\log_e 2}{t_{cc}}, \quad [S4]$$

where m is the number of time points (t_i) considered and r_{t_i} the ratio of the light and heavy amino acids containing fraction of a specific protein (Eq.S1) at each time point (Fig. 1C, main text). The half life of a protein ($T_{1/2}$) is then given by

$$T_{1/2} = \frac{\log_e 2}{k_{dp}}. \quad [S5]$$

Note that when k_{dp} approaches 0 half-lives approach infinity. This situation occurs when P_H/P_L ratios are close to the minimal value expected due to cell division (i.e. $\log_e 2/t_{cc}$, Eq. S4). Therefore, long half-lives are sensitive to variations in cell cycle times. Of all identified proteins (n = 6,445) those were taken, which were counted at least three times by the mass spectrometer regardless at how many time points. As a quality test of linear regression, the coefficient of determination (R^2) was calculated from the correlation coefficient between measured and estimated ratios (two or three time point measurements). The half-lives are only calculated for fits that satisfy $R^2 \geq 0.9$ (n = 4,906) or measurements of one time point with minimum three counts (n = 95).

For very unstable proteins we observed that the ratio measurements saturate at the late time point. Therefore, for proteins that failed the quality check ($R^2 < 0.9$) but had a ratio $r \geq 0.5$ at the first time point (1.5 h), only this time point was used to calculate the half-life (n = 27).

Next we estimated the variability of the linear regression slope ($\beta = k_{dp} + \frac{\log_e 2}{t_{cc}}$, cf. Eq. S4) by leave-one-out cross validation. For all two and three time point measurements we determined

the average leave-one-out cross validation sum of squares (CVSQ) for the linear regression slope

$$CVSQ = \frac{1}{n}\sum_{i=1}^{n}(\beta - \beta_i^{(-r_i)})^2, \qquad [S6]$$

where β is the slope of the regression curve using the original number of time points and $\beta_i^{(-r_i)}$ the slope of the regression curve after removing the ratio $r(t_i)$. The number of time points is n. The relative cross validation sum of squares (CVSQ/β)) reveals that 93% of the slopes vary on average less or equal to 1% if one data point is removed (main text, Fig.1D). This is 95% of the linear regression fits that satisfy the condition $R^2 \geq 0.9$.

mRNA sample preparation and sequencing

Total RNA from bulk RNA, pre-existing RNA and newly-synthesized RNA were subjected to Solexa deep-sequencing, following the mRNA sequencing protocol provided by Solexa. Briefly, poly(A) RNA was purified from 10 µg of total RNA with two rounds of hybridization to Dynal oligo(dT) beads (Invitrogen). The poly(A) RNA was first fragmented by using 5x fragmentation buffer (200mM Tris acetate, pH 8.2, 500 mM potassium acetate and 150 mM magnesium acetate) and heating at 94 °C for 3.5 min in a thermocycler. The fragmented RNA was precipitated and used for first- and second- strand cDNA synthesis with random hexamer primers. Both ends of the cDNA fragments were then repaired by T4 polymerase and Klenow DNA polymerase, and T4 polynucleotide kinase. Using Klenow exo (3′ to 5′ exo minus), a protruding “A” base was added to the 3′ end of the DNA fragments for ligation with Solexa adaptors (with a “T” overhang). After ligation, the cDNA with size of 200 +/- 25 bp was selected and followed by 15 cycles of PCR amplification. Finally the adapter-ligated DNA was sequenced for 36 cycles on the Solexa/GAIIX, according to the manufacturer’s instructions.

Determination of mRNA half-lives

The Solexa sequencing reads were mapped to mouse genome reference sequence (mm9, July 2007) by using SOAP2 with a maximum of 2 mismatches allowed. Only uniquely mapped reads were retained. The mRNA abundance was estimated as the number of reads mapped within the exonic regions divided by the total number of reads uniquely mapped on the genome (in Million)

and the cumulated exon length (in Kb) (RPKM). Only the transcripts with the RPKM values > 1 in all 3 samples were retained for further analysis. The newly synthesized transcripts containing more 4sU molecules are more prone to be pulled down and subsequently sequenced. To correct for such a bias in pulldown efficiency of mRNAs containing different numbers of uridines, we applied LOWESS (locally weighted scatter plot smoothing) (Fig. 2).

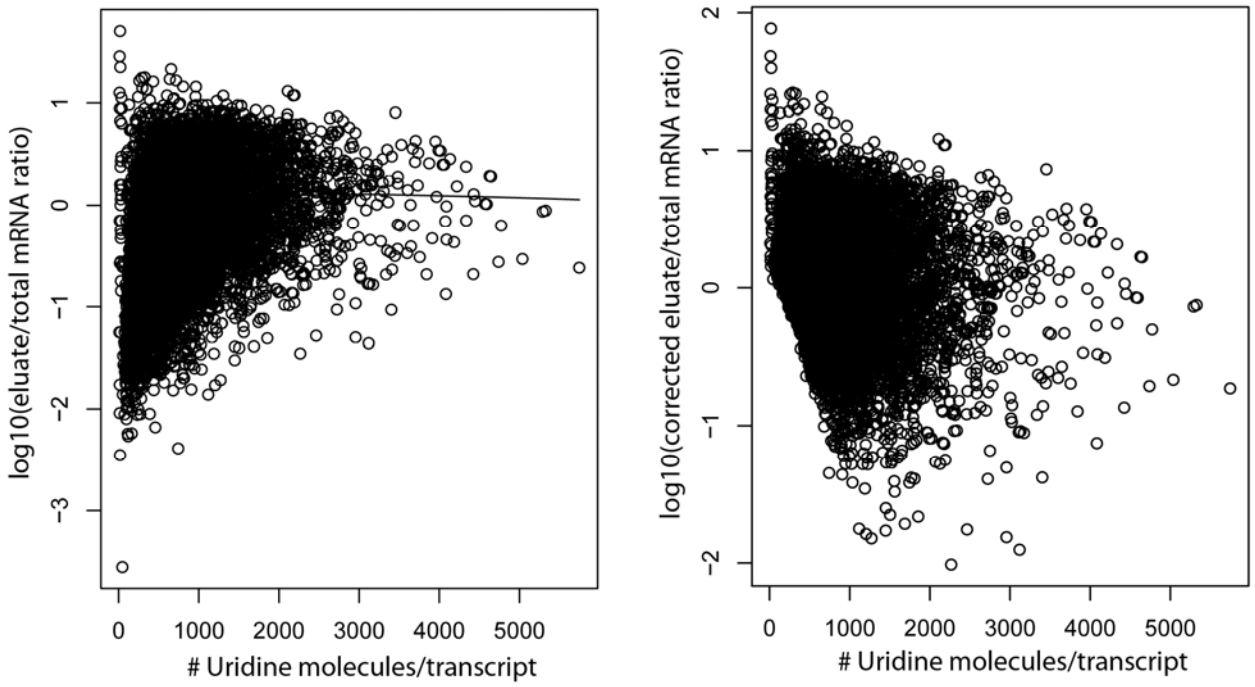

Fig. 2: Correction of 4sU bias by LOWESS. Plotting the number of uridines per transcript against newly-synthesized/total RNA ratios obtained by RNA-Seq reveals the bias in pulldown efficiency (left plot). Correction was done by applying LOWESS (right plot). Note the differently scaled y-axis between the plots.

Afterwards, the normalization of the measurement in the 3 samples (newly-transcribed, pre-existing and total RNA) was performed by linear regression analysis, as described before [2] (Fig 3 and 4):

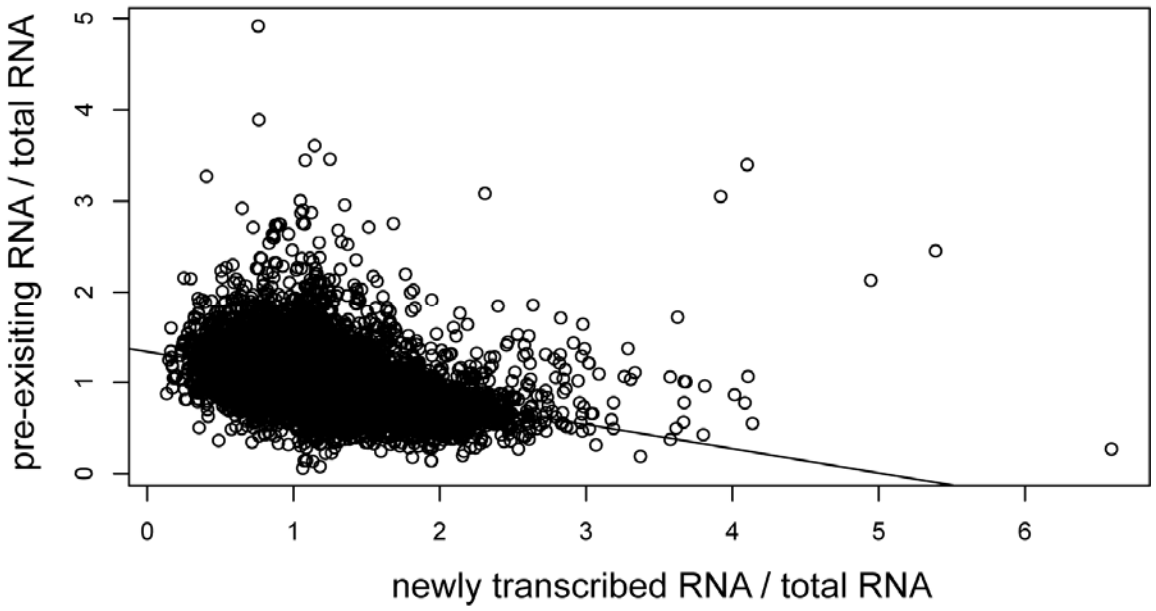

Fig. 3: Correlation of newly transcribed RNA / total RNA and preexisting RNA / total RNA ratios for NIH3T3 cells (2 h labeling) *before* normalization by linear regression analysis.

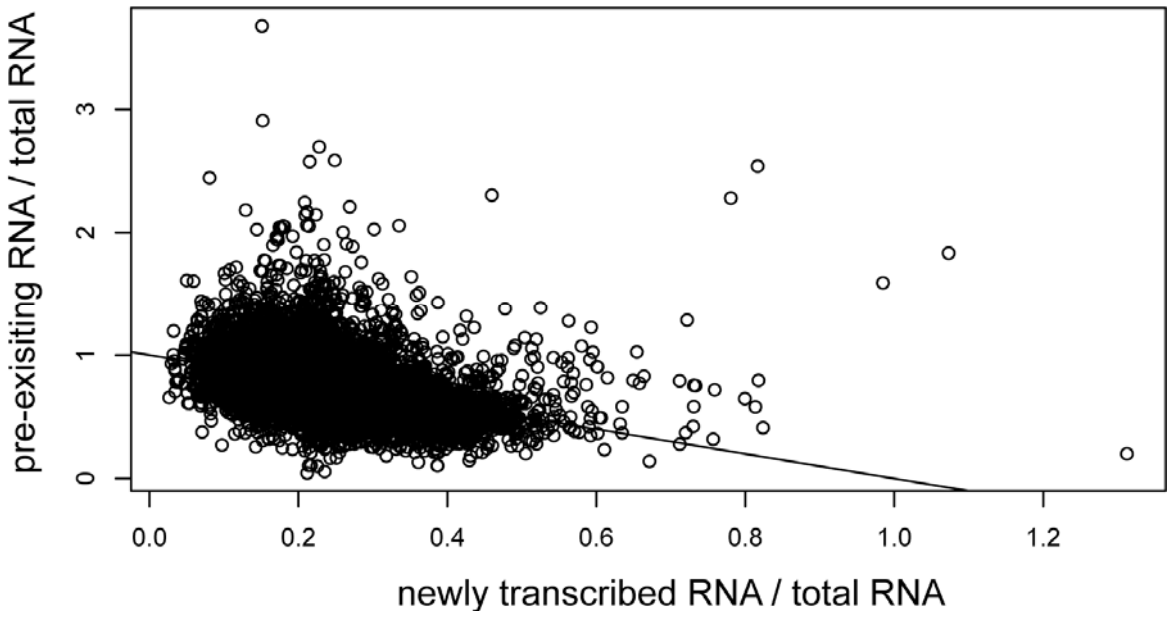

Fig. 4: Correlation of newly transcribed RNA / total RNA and preexisting RNA / total RNA ratios for NIH3T3 cells (2 h labeling) *after* normalization by linear regression analysis.

Of note, the fraction of reads without mismatches was 87.6%, 87.1% and 86.2% for the in the newly synthesized, pre-existing and total mRNA sample, respectively. Therefore, 4sU labeling did not appreciably affect the sequencing error rate.

The mRNA half-lives were then calculated based on the normalized ratios given by newly transcribed RNA/total RNA and preexisting RNA/total RNA, according to the formula:

$$T(1/2)=-t*\ln(2)/\ln(1-1/(1+ \text{Ratio(existing/total)}/\text{Ratio(new/total)})) \quad [S7]$$

The transcripts with the sum of 2 ratios exceeding 1 were excluded for further analysis.

We also compared our mRNA half-lives to previous data. Dolken and co-workers used 1h 4sU labeling combined with microarrays to quantify half-lives in the same cell line (NIH3T3)[2]. The correlation between our data and the Dolken data was high before we corrected for the 4sU bias ($r = 0.64$) and slightly reduced afterwards ($r = 0.59$, note that Dolken et al. did not do the correction). On average, our half-lives were considerably longer (median half-live of all transcripts 7.6 h) than in the Dolken dataset (4.6 h). A recent study on mRNA half-lives in dendritic cells reported even shorter mRNA half-lives after 45 min of 4sU labelling[8]. Previous work based on actinomycin D treatment reported median mRNA half-lives of about 10 h in HepG2 and Bud8 cells[9]. These differences in average mRNA half-lives may be due to the different cell lines and/or cell culture conditions. Alternatively, they could also reflect differences in experimental procedures (4sU or actinomycin D, microarray or RNA-seq, poly(A) enrichment or not, labelling for different length of time) and data analysis (labelling bias correction, computational models). We do not know if the observed differences reflect true differences in average cellular mRNA half-lives, possible inherent biases of the various protocols or a mixture of both. However, despite this uncertainty, the correlation of our mRNA half-lives with the Dolken data was high (see above). Thus, the rank of mRNA half-lives is reproducible across different labs and protocols.

Determination of absolute mRNA copy numbers

We determined absolute cellular mRNA copy numbers based on the measured cellular amount of polyadenylated mRNA and the number of sequencing reads obtained for the unfractionated

mRNA sample (i.e. newly synthesized and preexisting mRNA) similar to the method described previously [10]. First, we measured the yield of total RNA extracted per 3T3 cell as 25.5 pg using a NanoDrop ND-1000 UV-VIS spectrophotometer (Thermo Fisher). In three independent experiments we obtained 331.92, 304.56 and 360 ng of polyadenylated mRNA from 10 µg total RNA. Therefore, the fraction of mRNA in the cellular RNA pool is 3.3 %, corresponding to 8.5 x 10^{-13} g mRNA per cell. The base composition of the 3T3 mRNAs transcriptome was 0.4 : 0.2 : 0.2 : 0.2 (A:C:G:U) according to the reads we obtained. Consequently, the average molecular weight per nucleotide is 323 g/mole. This allows us to calculate the total amount of mRNA nucleotides per cell (T, 2.6x10^{-15} mole/cell). The cellular copy numbers of individual mRNAs (x) can be calculated from the number of sequencing reads mapping uniquely to the transcript (reads_transcript), the total number of reads (reads_total) and the transcript length (L) according to the equation

$$\frac{reads\_transcript}{reads\_total} = \frac{x \times L}{T} \qquad [S8]$$

For example, the popular 'housekeeping' mRNA of GAPDH was identified with 51,569 reads out of 26,761,116 total reads in the total RNA sample. With a transcript length of 1,232 nucleotides we arrive at 2,450 copies. A previous study carefully determined GAPDH copy number in NIH 3T3 cells by single-cell quantitative real-time PCR or bulk measurements [11]. Interestingly, our value falls almost exactly between the previously two published values for bulk and single cell data (1,840 and 2,940, respectively). Note that due to the presence of many GAPDH pseudogenes in the mouse genome we counted all reads mapping to a single GAPDH transcript rather than the corrected uniquely mappable fraction (see below).

For validation of calculated cellular mRNA copy numbers, the nCounter Mouse Inflammation Kit from NanoString Technologies [12] was applied to the total RNA sample (unfractionated). In essence, NanoString uses color-coded molecular barcodes that can hybridize directly to many different types of target molecules. 79 out of 179 genes covered by the Mouse Inflammation Kit matched to our sequencing data. RNA spike-controls of known concentrations spanning 4 orders of magnitude were plotted against obtained NanoString counts on a log-log scale. Linear regression was used to fit NanoString counts to absolute spike control amounts. The slope and intercept was then used to convert NanoString counts for all other mRNAs to arrive at molar

amounts. Cellular copy numbers were derived by dividing through the number of cells used in the respective experiment. mRNA copy numbers for the replicate mRNA-seq experiment were obtained by scaling obtained sequencing reads to mRNA copy numbers in the original experiments (linear fit on log-log scale).

Correcting for different transcript mappabilities

Absolute mRNA quantification is complicated by the fact that some mRNAs have a high degree of sequence identity to other mRNAs or pseudogenes which results in a smaller percentage of uniquely mappable reads. To correct for this effect we determined the fraction f of all the exonic positions on which 36nt sequencing reads that can be uniquely mapped for all mRNAs by in silico simulation. Briefly, 36nt Solexa sequencing reads were first simulated from every refseq transcripts with each position along the transcript (except the last 35 nt) as the first base of one sequencing read. Simulated reads were then mapped to the mouse genome reference sequence and used to estimate transcript abundance using the same procedure as for the real data. The average value of f was 72%. mRNAs with f < 10% were excluded from further analysis. For all other genes, the raw mRNA copy numbers were multiplied by 72 / f to correct for different mappability.

Cluster analysis of gene ontology (GO) terms

In order to test whether specific genes with certain combinations of mRNA and protein stability have distinct biological functions, regimes were defined as follows: First, genes were separately sorted according to their mRNA and protein half-lives. Second, the top third of mRNA half-lives were combined with either the upper or lower third of protein half-lives, and *vice versa* for protein half-lives. This resulted in four regions comprising distinct combinations of mRNA and protein half-lives.

Next, gene ontology analysis using the DAVID Bioinformatics Database (DAVID Bioinformatics Resources, http://david.abcc.ncifcrf.gov/) [13] was performed with the IPI protein identifier of genes in each region to find enriched biological processes. Calculation of over-represented GO terms was done using the entire list of identified proteins as background (threshold count = 2; EASE score = 1). Terms with a p-value <0.01 in at least one regime were selected, log- and z-transformed, hierarchically clustered and plotted as a heatmap using an in-house perl and R script.

Statistical analysis and model analysis

Statistical data analysis was done using the R project for Statistical Computing (R Foundation for Statistical Computing, Vienna, Austria) or Prism 4.0 (GraphPad Software, San Diego, CA, USA). Linear regression and model anylsis was performed using MATLAB 2009 (MathWorks, Natick, MA, USA).

Energy calculations

Using our quantitative model we were able to estimate the theoretical energy costs (in high energy phosphate bonds) to synthesize mRNAs (i.e. primary transcripts) and proteins from their building blocks (nucleotide monophosphates and amino acids, respectively). Transcription and translation rates required to maintain steady-state mRNA and protein levels were directly calculated from half-lives (rate = ln(2) / half-life). For transcription, two ATP molecules are needed for adding a single nucleotide (elongation) plus another two ATPs for the initiation step [14]. For protein synthesis, four ATP molecules are need for the addition of single amino acids to the nascent chain (2 ATPs for tRNA loading and 2 ATPs for elongation) plus another 2 ATPs for initiation and one for termination [15]. By multiplication of the above mentioned costs per building block (nucleotide or amino acid) with transcription/translation rates and mRNA/protein lengths overall energy expenses for mRNA and protein synthesis were obtained with:

mRNA synthesis = (transcription rate * primary transcript lengths * 2 ATP) + 2 ATP
protein synthesis = (translation rate * protein length * 4 ATP) + 3 ATP

Applying the quantitative model to predict protein concentrations in MCF7 cells

In order to test to what extent the obtained quantitative parameters in NIH3T3 cells also apply to a different organism, mRNA levels obtained for MCF7 cells were used to predict corresponding protein levels according to our quantitative model. mRNA levels were obtained by deep-sequencing of MCF7 RNA (see section above on ´mRNA sample preparation and sequencing´) followed by calculation of absolute mRNA levels (see also section above on ´Determination of absolute mRNA copy numbers´). Predictions of protein abundances were compared to measurements of absolute protein levels in MCF7 cells employing iBAQ (see above) on an in-

house protein dataset. Mapping between mouse and human genes was done based on official gene symbols.

Correlation of mRNA and protein half-lives with sequence features

Correlation of 3'UTR, 5´UTR, CDS sequence lengths and AU-rich elements and Pum2 binding sites with mRNA stability

Mouse 3'UTR, 5´UTR and CDS sequences based on the mouse the reference sequence (NCBI Build 37/mm9) were extracted from UCSC Genome Browser (http://genome.ucsc.edu) and the lengths were counted. 3´UTR sequences were searched for AU-rich core motifs ("ATTTA"; Barreau et al., 2006) and Pumilio2 binding sites ("TGTA(A|T|C|G)ATA" with "I" referring to "or"; Hafner et al., 2010) using an in-house Perl script. The average number of either motif in bins of 100 genes were plotted against averaged mRNA half-lives of the same bins. To account for stochastic motif occurrences due to vastly different 3´UTR lengths, motif densities (motifs/nucleotide) were calculated by dividing the number of either motif counted per 3´UTR by 3´UTR sequence length.

Correlation of amino acid sequence frequency and protein stability

A possible relationship between the presence of certain amino acids and protein stability was examined as follows: First, relative fractions of amino acids per protein were calculated. Second, relative amino acid fractions were plotted against protein half-lives in bins of 100 genes. Third, Spearman Rank correlation coefficients (R_s) were used to assess the statistical significance of possible correlations.

Correlation of the degree of protein disorder with protein stability

As a proxy for a protein´s "unstructuredness", a publicly available mouse protein database [16] (http://bioinfadmin.cs.ucl.ac.uk/disodb/) was used that predicts potentially disordered amino acid residues in a context-specifc manner and also assigns confidence values accordingly. The overall percentage of potentially disordered amino acids was extracted from the database for individual proteins. A possible correlation was investigated by plotting averaged "unstructuredness" in percent in bins of 100 genes against averaged protein half-lives of the same bins.

Quantitative model of gene expression

Calculating synthesis rates of mRNA and protein

A widely used minimal description of the dynamics of transcription and translation (Fig. 5) includes the synthesis and degradation of mRNA and protein, respectively [17-19].

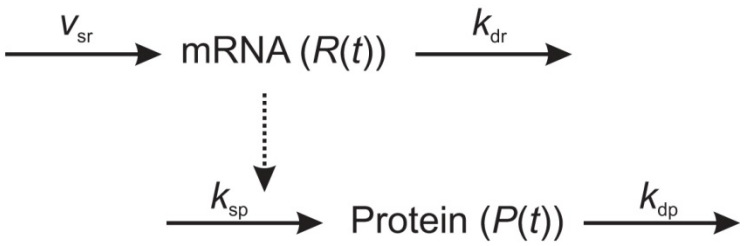

Fig. 5: Minimal model of gene expression. The mRNA (R) is synthesized with rate v_{sr} and degraded according to their numbers with rate constant k_{dr}. The rate constant of protein (P) synthesis is k_{sp}. Protein degradation is characterized by the rate constant k_{dp}.

The scheme can be translated into a system of ordinary differential equations:

$$\frac{dR}{dt} = v_{sr} - k_{dr}R \qquad [S9]$$

$$\frac{dP}{dt} = k_{sp}R - k_{dp}P \qquad [S10]$$

The mRNA (R) is synthesized with a constant rate v_{sr} and degraded proportional to their numbers with rate constant k_{dr}. The protein level (P) depends on the number of mRNAs, which are translated with rate constant k_{sp}. Protein degradation is characterized by the rate constant k_{dp}. This model can be solved analytically [20]. Here it is used to calculate the synthesis rates of mRNA and protein from their measured half lives and levels.

The determined half lives of many proteins in NIH 3T3 cells are longer than the cell cycle duration (t_{cc}) of 27.5h. For proteins with half lives comparable or longer than the cell cycle, the steady state is not reached before cell division. Consequently, levels per cell are not only defined by synthesis and degradation, but also by dilution due to cell division [21]. In a growing population the cells are distributed in their cell cycle phase. Assuming a homogenous distribution of non-synchronized cells over the duration of one cell cycle, the measured population mean of mRNA ($\langle R \rangle$) and protein ($\langle P \rangle$) is set equal to the averaged copy number of one cell cycle

$$\langle X \rangle = \frac{1}{t_{cc}} \int_{t_{c0}}^{t_{cc}} X(t)dt \,, \qquad \text{[S11]}$$

where $\langle X \rangle$ is the mean population level of either mRNA ($\langle R \rangle$) or protein ($\langle P \rangle$) and X(t) their solution of the differential equation system (Eq. S9 and S10) with the initial conditions $R(0) = R(t_{c0})$ and $P(0) = P(t_{c0})$, respectively. The mRNA and protein copy numbers at the beginning of each cell cycle (t_{c0}) is half the numbers at the time of cell division (t_{cc}), therefore $2R(t_{c0}) = R(t_{cc})$ and $2P(t_{c0}) = P(t_{cc})$. Hence, the average mRNA copy numbers are given by

$$\langle R \rangle = \frac{v_{sr}}{k_{dr}} \left[1 + \frac{1}{k_{dr} t_{cc}} \left(1 - \frac{1 - e^{-k_{dr} t_{cc}}}{2 - e^{-k_{dr} t_{cc}}}\right)\left(e^{-k_{dr} t_{cc}} - 1\right) \right], \qquad \text{[S12]}$$

while the average protein numbers are determined by

$$\langle P \rangle = k_{sp} \left[\frac{v_{sr}}{k_{dr} k_{dp}} + \frac{1}{k_{sp} t_{cc} k_{dp}} \left(P_{ss} - P_{t_{c0}}\right)\left(e^{-k_{dp} t_{cc}} - 1\right) \right] - \frac{k_{sp}}{t_{cc}} \left(R_{ss} - R_{t_{c0}}\right) \frac{\frac{1}{k_{dr}}\left(e^{-k_{dr} t_{cc}} - 1\right) - \frac{1}{k_{dp}}\left(e^{-k_{dp} t_{cc}} - 1\right)}{k_{dr} - k_{dp}} \qquad \text{[S13]}$$

with

$$R_{t_{c0}} = R_{ss} \frac{1 - e^{-k_{dr} t_{cc}}}{2 - e^{-k_{dr} t_{cc}}}$$

and

$$P_{t_{c0}} = \left(2 - e^{-k_{dp} t_{cc}}\right)^{-1} \left[P_{ss}\left(1 - e^{-k_{dp} t_{cc}}\right) - k_{sp}\left(R_{ss} - R_{t_{c0}}\right) \frac{e^{-k_{dp} t} - e^{-k_{dr} t}}{k_{dr} - k_{dp}} \right].$$

The average mRNA copy numbers (Eq. S12) is used to calculate the transcription rate (v_{sr}). The translation rate constant (k_{sp}) is determined from the mean protein numbers (Eq. S13). Since the assumption, that measured protein levels in cell culture represent steady-state values, is commonly used for the analysis of related questions [22] a comparison between synthesis rates determined by our analysis and obtain by the steady-state equations for mRNA ($R_{ss} = v_{sr}/k_{dr}$) and protein ($P_{ss} = R_{ss}*k_{sp}/k_{dp}$) may be instructive (Fig. 6). For the synthesis rates of mRNA (Fig. 6, light gray), the deviation between the two approaches is small, because mRNA half lives are

mostly smaller than the cell cycle time. For protein synthesis (Fig. 6, dark gray), the differences are substantial; they can differ for more than one order of magnitude. Note, that calculating synthesis rates from measured mean mRNA and protein levels as described above is still a simplification and gives mean synthesis rates over time and cell population.

They do not take into account that gene expression in mammalian cells is non-continuous. In addition, the non-uniform age distribution of cells in culture as described in [19, 23] is neglected, since this effect is expected to be small compared to the deviation obtained by neglecting the cell cycle.

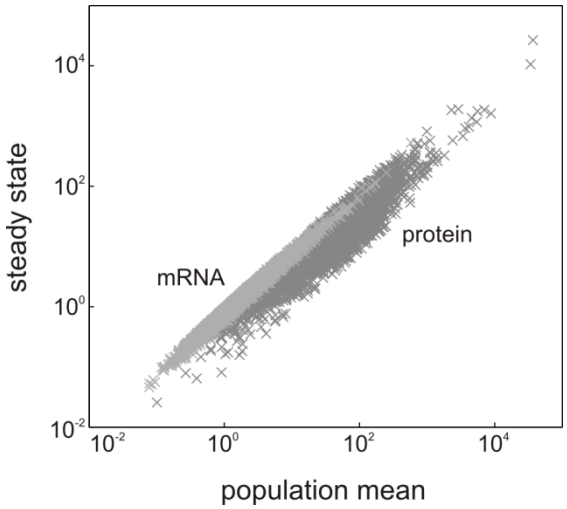

Fig. 6: Comparison of synthesis rates of mRNA and proteins assuming the measured levels reflect averages over one cell cycle or steady-state values. For the mRNA synthesis rates (molecules/h, light gray) both assumptions give comparable results, whereas the translation rate constant (h^{-1}, dark gray) differ partly more than one order of magnitude.

Validation of protein half-lives by cylcoheximide chase analysis and immunoblotting

For protein half-lives validation, cellular protein degradation was monitored after blockage of *de novo* protein synthesis via cycloheximide treatment. Cycloheximide chase analysis was performed with seven proteins spanning a reasonable range of protein half-lives (1.2 - >60 h). For this purpose, protein synthesis of NIH3T3 cells was blocked by adding cycloheximide directly to the culture medium at a final concentration of 50 µg/ml. Cells were harvested after cycloheximide treatment at indicated time points followed by protein extraction, SDS-PAGE and immunoblotting to visualize protein degradation of respective proteins.

Primary antibodies against CEBPβ (sc-150), JNK (sc-7345), Hells (sc-46665), JunB (sc-46) were purchased from Santa Cruz Biotechnologies or Cell Signaling (GSK3β, 9315) or Sigma-Aldrich (β-Actin, A5441) or BD Biosciences (556470, CCND1).

Harvested cells were lysed in appropriate amounts of RIPA buffer (50 mM Tris-HCl pH 7.4, 150 mM NaCl, 1% Triton-X100, 1% Sodium deoxycholate and 0.1% SDS) for 20 min on ice. The lysates were cleared by centrifugation for 10 min (14,000 rpm at 4°C) and transferred to fresh tubes. SDS-PAGE was performed with whole-cell lysates using NuPAGE Novex 4 to 12% gradient gels (Invitrogen) under reducing conditions according to the manufacturer's instructions. Gels were blotted onto PVDF (Perkin-Elmer) membranes using the XCell II Blot Module (Invitrogen). 1 mA of current per cm^2 of membrane for 2 h was applied and the membrane was activated prior to use with 100% methanol and washed once in transfer buffer. Unspecific binding sites were blocked with 5% dry milk powder in 1x TBST at RT for 30 min followed by washing three times for 10 min with 1x TBST. In order to visualize specific protein bands, the membrane was incubated with the respective primary antibody at a 1:1,000 dilution in blocking solution at 4°C overnight with shaking. Blots were washed 3x in TBST and incubated either with an anti-mouse or an an anti-rabbit secondary antibody (all from Amersham) conjugated to horseradish peroxidase diluted 1:3,000 in TBST for 1 h at RT. After three more washing steps in TBST for 10 min each the signal of the bound secondary antibodies were detected by applying Western Blot Chemiluminescence Reagent Plus for ECL immunostaining (PerkinElmer) to the membrane followed by exposition to X-ray films (GE Healthcare). For sequential detection of different proteins on the same membrane, the PVDF membrane was stripped in stripping buffer (50 °C, 20 min), washed once with isopropanol and thoroughly with distilled H_2O. Subsequently, the membrane was blocked again and treated as described above.

Validation of protein degradation and synthesis by cycloheximide chase analysis with SILAC

Protein degradation and synthesis were validated using cycloheximide chase analysis to block *de novo* protein synthesis. Note that on one hand, drug treatment should be as short as possible to miminize toxic effects. On the other, the time span of drug treatment should be close to the half-life of proteins to allow for accurate protein quantification. Due to this intrinsic trade-off the analysis is inherently limited to high turnover proteins. NIH3T3 cells grown in light (L) SILAC medium were treated for 2 h with 50 µg/ml cycloheximide while control cells cultivated in heavy

(H) SILAC medium remained untreated. Cells were subsequently mixed in a 1:1 ratio and subjected to in-gel digestion prior to mass spec analysis. Assuming that drug treatment does not affect protein degradation, relative changes in protein levels ($R(t)$) given by H/L ratios can be used to calculate decay rates constants (k_{dp}). Proteins labeled with light (L) medium decay exponentially, whereas for heavy (H) labeled proteins the measurement at the onset of the experiment is used, therefore $H = P_0 = \text{const.}$:

$$R(t) = \frac{H}{L(t)} = \frac{P_0}{P_0 e^{-k_{dp} t}} \qquad [S14]$$

$$k_{dp} = \frac{\log_e R}{t}$$

Assuming that during cycloheximide treatment for 2 h the net cell growth can be ignored, measured mean protein levels do not differ significantly from the steady state levels and. Protein synthesis rate constants (k_{sp}) can be calculated using the steady-state equation ($P_{ss} = R_{ss} * k_{sp}/k_{dp}$):

$$k_{sp} = \frac{k_{dp} P_{ss}}{R_{ss}} \qquad [S15]$$

For mRNA (R_{ss}) and protein (P_{ss}) steady-levels we used the concentrations (protein copies/cell) derived from our absolute quantifications that have been validated independently by UPS2 spike-in experiments (see above).

Validation of RNA degradation and synthesis by actinomycin-D chase analysis with NanoString

mRNA decay and production were validated using actinomycin-D chase analysis to block *de novo* mRNA synthesis. As already mentioned for cycloheximide and proteins, this analysis is inherently limited to short-lived transcripts (see above). Transcription of NIH3T3 cells was inhibited by applying actinomycin-D at a final concentration of 5 µg/ml. Cells were harvested for RNA isolation 0h, 1h, 2h, 3h, 4h and 5h after treatment. For quantification of absolute mRNA

levels in each sample, we used the nCounter Mouse Inflammation Kit from NanoString Technologies [12] (see also section above ´Determination of absolute mRNA copy numbers´ for details on the method). To account for slight differences in the RNA amount used as input, raw NanoString counts for each time point were normalized based on the ten most stable mRNAs according to our half-life measurements. Assuming that drug treatment does not affect mRNA degradation, mRNA levels of actinomycin-D treated (*R(t)*) and untreated control cells (*R(0)*) were used to compute decay rates constants (k_{dr}) according to:

$$R(t) = R_0 e^{-k_{dr} t} \qquad \text{[S16]}$$

$$k_{dr} = \frac{\log_e(\frac{R_0}{R(t)})}{t}$$

As shown previously, decay rates obtained by actinomycin-D treatment can vastly differ depending on the length of drug treatment [2]. To take this into account, we used the average of all degradation rate constants obtained for the different length of actinomycin-D treatment. For calculation of mRNA synthesis (v_{sr}), we assumed that during actinomycin-D treatment the net cell growth can be ignored. We further assume that mean population level does not differ from the steady-state level. Hence, using the steady-state equation ($R_{ss} = v_{sr}/k_{dr}$), mRNA synthesis can be computed according to:

$$v_{sr} = k_{dr} R_{ss} \qquad \text{[S17]}$$

For steady-levels mRNA levels (R_{ss}) we used the concentrations (mRNA copies/cell) derived from our absolute quantification that have been validated independently using NanoString (see above).

References

1. Ong, S.E. & Mann, M. A practical recipe for stable isotope labeling by amino acids in cell culture (SILAC). *Nat Protoc* **1**, 2650-2660 (2006).
2. Dolken, L. et al. High-resolution gene expression profiling for simultaneous kinetic parameter analysis of RNA synthesis and decay. *RNA* **14**, 1959-1972 (2008).

3. Shevchenko, A., Tomas, H., Havlis, J., Olsen, J.V. & Mann, M. In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nat Protoc* **1**, 2856-2860 (2006).
4. Rappsilber, J., Mann, M. & Ishihama, Y. Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nat Protoc* **2**, 1896-1906 (2007).
5. Ishihama, Y., Rappsilber, J., Andersen, J.S. & Mann, M. Microcolumns with self-assembled particle frits for proteomics. *J Chromatogr A* **979**, 233-239 (2002).
6. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol* **26**, 1367-1372 (2008).
7. Elias, J.E. & Gygi, S.P. Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat Methods* **4**, 207-214 (2007).
8. Rabani, M. et al. Metabolic labeling of RNA uncovers principles of RNA production and degradation dynamics in mammalian cells. *Nat Biotechnol* (2011).
9. Yang, E. et al. Decay rates of human mRNAs: correlation with functional characteristics and sequence attributes. *Genome Res* **13**, 1863-1872 (2003).
10. Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L. & Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods* **5**, 621-628 (2008).
11. Marcus, J.S., Anderson, W.F. & Quake, S.R. Microfluidic single-cell mRNA isolation and analysis. *Anal Chem* **78**, 3084-3089 (2006).
12. Geiss, G.K. et al. Direct multiplexed measurement of gene expression with color-coded probe pairs. *Nat Biotechnol* **26**, 317-325 (2008).
13. Huang da, W., Sherman, B.T. & Lempicki, R.A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4**, 44-57 (2009).
14. Voet, D.V.a.J.G. Biochemistry. (John Wiley and Sons, New York; 1995).
15. Moldave, K. Eukaryotic protein synthesis. *Annu Rev Biochem* **54**, 1109-1149 (1985).
16. Pentony, M.M. & Jones, D.T. Modularity of intrinsic disorder in the human proteome. *Proteins* **78**, 212-221 (2010).
17. Ben-Tabou de-Leon, S. & Davidson, E.H. Modeling the dynamics of transcriptional gene regulatory networks for animal development. *Dev Biol* **325**, 317-328 (2009).
18. Cohen, A.A. et al. Protein dynamics in individual human cells: experiment and theory. *PLoS One* **4**, e4901 (2009).
19. Powell, E.O. Growth rate and generation time of bacteria, with special reference to continuous culture. *J Gen Microbiol* **15**, 492-511 (1956).
20. Hargrove, J.L. & Schmidt, F.H. The role of mRNA and protein stability in gene expression. *FASEB J* **3**, 2360-2370 (1989).
21. Alon, U. An Introduction to Systems Biology. (Chapman & Hall, 2007).
22. de Sousa Abreu, R., Penalva, L.O., Marcotte, E.M. & Vogel, C. Global signatures of protein and mRNA expression levels. *Mol Biosyst* **5**, 1512-1526 (2009).
23. Berg, O.G. A model for the statistical fluctuations of protein numbers in a microbial population. *J Theor Biol* **71**, 587-603 (1978).