## Supplementary Methods Online

**Retroviral constructs.** Primers containing splice acceptor and donor sequences[1] and restriction sites were used to amplify YFP from the EYFP-N1 vector (BD Biosciences Clontech, Palo Alto, CA). Three separate constructs, one for each reading frame were created. Each reading frame insert was ligated into a pBabe retroviral vector[2] in reverse orientation to the LTR promoter activity. To suppress the enhancer activity of the viral LTR, the vector was self-inactivated by deletion of the promoter and enhancer sequences in the 3' LTR. This construct was named BabeAE.

**Cells and infection.** H1299 non-small lung cell carcinoma cells (ATCC) were grown in RPMI medium with phenol red, supplemented with 10% FCS and antibiotics. BabeAE was transfected into Phoenix ecotropic packaging cells (ATCC with the generous permission of Gary Nolan) with Fugene transfection reagent (Invitrogen, Carlsbad, CA). To create a permissive human cell line for the murine competent virus, human H1299 were transfected with the viral ecotropic receptor and a cell pool with stable expression of the receptor was selected using Neomycin. These cells were infected with the packaged virus supernatant in the presence of 1ug/ml polybrene (Sigma-Aldrich, St. Louis, MO). In a typical infection, $3x10^6$ cells were infected, resulting in a library with approximately $10^4$ detectably labeled proteins per infection.

**Sorting YFP fluorescent cells by flow cytometry.** Cells were sorted using a Becton Dickinson FACS Vantage machine (BD Biosciences, San Jose, CA) equipped with Cloncyt motorized stage control software. Excitation was at 488 nm with an argon laser and emission was collected at 535 nm (YFP) and 590 nm (auto-fluorescence). Cells were gated for a high 535 nm to 590 nm ratio relative to control. The fluorescence threshold for positive cells was set to give a detectable signal for long period time lapse microscopy using a medium resolution (numerical aperture 0.7) 20X objective. Positive cells were sorted at one cell per well into every second well in 384 well plates (Greiner, Frickenhausen, Germany). The wells into which cells were deflected contained filtered conditioned medium from sub-confluent H1299 cells. Wells in between remained empty to prevent the variability in stream direction from deflecting two cells into the same well. The fraction of cells that formed clones ranged between 20 to 40%.

**Identification of tagged protein by 3' RACE.** RNA was extracted from cells of each clone using RNeasy columns and protocol (Qiagen, Valencia, CA). mRNA was reverse transcribed to cDNA using an oligo dT primer with an adaptor sequence (Invitrogen, Carlsbad, CA) and the superscript III reverse transcriptase (Invitrogen, Carlsbad, CA). The cDNA library was then subjected to one round of PCR with a 5' YFP primer and a 3' primer complementary to the adaptor region inserted after the oligoT. 1 ul of the PCR product was transferred to a new reaction with a nested YFP primer and the same 3' primer. If the nested reaction produced one clear band, it was cleaned by PCRclean (Qiagen, Valencia, CA) and directly sequenced with the nested YFP primer. If two or more bands appeared, each was extracted from the gel and cleaned by Gelclean (Qiagen, Valencia, CA), and sequenced directly as above. Most multiple bands were found to be splice variants. Sequences derived from 3' RACE were used to identify the tagged gene by the Blat genomic alignment tool (http://genome.ucsc.edu/cgi-bin/hgBlat).

**Analysis of DNA content by flow cytometry to estimate relative duration of cell cycle phases.** Since the relative fraction of cells in each phase is determined by how long single cells spend in each cell cycle phase, we used relative fractions of cells $G_1$, S, and $G_2$ to estimate relative durations of these cell cycle phases. The relative fraction of cells in each cell cycle stage was determined by DNA content analysis using flow cytometry. Cells were fixed in a cold solution of 80% ethanol, 20% PBS, then incubated with RNase and labeled with 25 ug/ml of the DNA stain propidium iodide (Sigma-Aldrich, St. Louis, MO). About 20,000 to 50,000 cells were collected using a Becton Dickinson FACS SORT flow cytometer (BD Biosciences, San Jose, CA), and the fraction of cells in each cell cycle stage was estimated by ModFit software (Verity Software House, Topsham, ME) and averaged over 20 different cell clones (**Supplementary Fig. 3**). Measurement of DNA content dynamics per cell by time-lapse microscopy using a vital DNA stain was not performed because of associated toxicity over the span of the movie.

**Image analysis and fluorescence quantification**

**Flat field correction:** at the beginning and end of every experiment a set of images was taken of a homogenous medium captured with variable exposure times. The images were used to calculate the offset (grey levels) and gain (grey levels per millisecond of exposure time) of every pixel by performing a linear regression of the grey levels from different exposure times. Exposure times used were either 0, 100, 200, 300, 400, 500, 600 milliseconds when using a solution of 1:1000 GFP (BD Biosciences Clontech, Palo Alto, CA) or 0, 400, 800, 1200, 1600, 2000, 2400 milliseconds when using RPMI culture medium which contained phenol red. A larger weight (X3) was given to the first image in calculating the regression in order to anchor the offset at the correct value. The collection of gains calculated for each pixel separately is the estimate of the non-homogeneity in the illumination pattern. A normalized gain pattern was calculated by dividing the gain pattern by its average over all pixels. Each image in the movie was flat field corrected by subtracting the offset of each pixel and dividing by the gain.

**Background subtraction:** background resulting from auto-fluorescence of the medium as well as other sources was calculated and corrected as follows. The flat field corrected image was divided into blocks of 100 by 100 pixels. The grey levels of all pixels in every such block were ranked. The grey level at the $10^{th}$ percentile was extracted. The set of these extracted grey levels was interpolated to give an image of the background. This image was subtracted from the original image.

**Quantitation of total fluorescence for proteins with nuclear localization:** we used a custom written image analysis tool developed under the Matlab image processing toolbox environment. The main stages were the segmentation of the cells, tracking, and detection of cell division events.

**Segmentation**: Image analysis of general, arbitrary protein localizations is a difficult problem. We focused on the simpler problem of quantifying protein levels for proteins that are localized to the nucleus. We found that using our labeling method for proteins with nuclear localization, we obtained fluorescence images with a high signal to noise ratio and distinct boundaries. Examples of segmentation are shown in **Supplementary Fig. 4a-d**. Segmentation was based on a combination of thresholding and watershed segmentation and was applied to images after flat field correction and background subtraction were performed. The segmentation code uses a preliminary version of an algorithm which is freely available and documented in the open-source

software package CellProfiler (www.cellprofiler.org). The following description applies only to the code used for this work, not to the current version of CellProfiler. A manuscript describing CellProfiler is in preparation (Anne E. Carpenter, Thouis Ray Jones, Michael Lamprecht, Douglas B. Wheeler, Colin Clarke, In Han Kang, Ola Friman, David A. Guertin, Joo Han Chang, Robert Lindquist, Jason Moffat, Polina Golland, and David M. Sabatini, CellProfiler: image analysis for high throughput microscopy). The most comparable module in the current version of CellProfiler is called IdentifyPrimAuto (Intensity/Intensity options).

We used a modular three-step strategy for identifying nuclei in images which is based on a combination of previously published algorithms[3-5]. In step 1, the algorithm determines whether an object is an individual nucleus or two or more clumped nuclei. This determination is accomplished by identifying local maxima in the smoothed intensity image. This works well because the nuclei of interest in this work are bright in the middle and dimmer towards the edges. In step 2, the edges of nuclei are identified. For nuclei within the image that do not touch, the edges are easily determined using thresholding. The threshold is automatically calculated for each image using Otsu's method which was adjusted in this case by an empirically determined factor of 0.7. For finding the dividing lines between touching nuclei, the already identified nuclear markers are used as starting points for a watershed algorithm[6] applied to the original image. This works because the dividing lines between touching nuclei tend to be dimmer than the remainder of the nucleus. In step 3, identified nuclei are discarded if touching the border of the image, or if below a size threshold (250 pixels in this case).

**Tracking:** Tracking of cells throughout the movie was performed by analyzing the movie from end to start and finding for each segmented cell with a given centroid location the cell in the previous image that has the closest centroid. A cell division resulted in two cells converging into the same object when viewed in this end to start manner. For each case where such convergence occurs, the software verifies that this is indeed a cell division and not a case of two cells merged by incorrect segmentation. In order to separate crossing cells from dividing cells when backtracking the movie, we followed every case of cell division/crossing for an extra 5 frames back in the movie. In the case that the cells re-separated, we marked this as a case of cell crossing, whereas if the cells remained "fused", we marked this as a legitimate division event.

**Detection of cell division events and data collection:** Cell division was detected by having two cell tracks converge into a single cell (viewed from end to start), or by a sharp ~2 fold change in total fluorescence level between consecutive images. Following the detection of division events cell data were stored in a binary tree like database using pointers from mother cell to daughter cells. This data structure enabled easy analysis of cell cycle based properties of the cells. The main parameters measured for each cell nucleus were the x and y centroid locations, area, total, mean and variance of the fluorescence.

**Automatic quality tests for cells tracks:** Automatic quality tests were used to eliminate cells where the location or fluorescence tracks showed unfeasible jumps in values that stem from errors in the segmentation and tracking algorithms. An incorrect tracking was detected by a movement of the centroid of a cell by more than 45 pixels (more than 2 nuclei radii) in any direction from one image to the next. Such tracks were marked as erroneous and were not used. For division events this value was increased to 60 pixels.

Mistakes in measurements of a cell's fluorescent levels were detected by large jumps (log ratio of two consecutive fluorescent values being less than 0.75 or larger than 1.25 when not the result of cell division). Moreover, the ratio of a cell fluorescence level at the end of the cell cycle divided by the fluorescence level at the beginning was calculated. The expected ratio should be around 2. Cells with a ratio<1.4 or >3 were omitted. These usually indicate erroneous segmentation (over-segmentation or aggregation of cells).

Cells with cell cycle length >170 frames (28.3 hours) or <80 frames (13.3 hours) were omitted. These usually indicate erroneous segmentation (aggregation of cells, or over-segmentation). The distribution of cell cycle lengths is shown in **Supplementary Fig. 2.**

**Manual quality control for cells tracks:** We visualized 10-20 individual cells throughout the cell-cycle using the synchrogram. This was done for randomly selected cells in approximately half of the proteins tested. For each cell, we verified that the cell division is evident at the beginning and end of the cell cycle, and that the same cell is followed throughout the track.

**Bleaching correction:** Bleaching effects were corrected both for the quantitation of protein amounts and for visualization of the cells. The average total fluorescence of cells decays over time because of bleaching. Bleaching effects were at most about 30-40% of the fluorescent intensity over a 60 hour movie. A bleaching function was calculated separately for each movie. We used the measured average cell total fluorescence to quantify this effect. We empirically found the parameters for the correction by fitting a decaying exponential with an offset to the average total fluorescence level time course. The bleaching correction function was of the form $B(t)=a1+a2*\exp(-alpha*t)$. The measured fluorescence was divided by this function to yield a corrected fluorescence profile. A typical bleaching correction for a movie lasting 60 hours, had values of a1=0.85, a2=0.3, alpha=1/1500, where t is measured in minutes.

**Accumulation rate calculation:** To measure the accumulation rate at time t we calculated the derivative of the protein fluorescence level: $(f(t+\tau/2)-f(t-\tau/2))/\tau$. Here f(t) is the total fluorescence at time t. Both t and $\tau$ are in units of percent of elapsed cell-cycle following the automatic cell cycle synchronization. We used a value of $\tau$=10%. Similar results were obtained for values of $\tau$=[6, 20, 30] %. Accumulation rates at times t<15% or t>85% of the cell-cycle were not calculated because of fluctuations in the segmentation near the division event. The nuclear accumulation rate was defined as the temporal derivative of the nuclear protein fluorescence level, divided by the nuclear area. The calculation proceeded as follows: the fluorescent levels of each cell were measured as a function of time. To reduce measurement noise, protein levels were filtered by a hybrid Gaussian-median filter. In the filter each measurement is replaced by taking the 5 adjacent time points, delete the maximal and minimal level and calculate the weighted average of the rest of the points with weights of 25%, 50% and 25%. Based on the times of division events, the fluorescent levels were interpolated at 1% of the cell cycle time intervals. For each cell we calculated the accumulation rate of the protein into the nucleus by the difference equation depicted above. The accumulation rate in the nucleus of each cell was normalized by dividing by the average accumulation rate over the whole cell cycle so as to give an average accumulation rate of 1. The accumulation rate calculated in this manner for each cell separately was averaged to find the mean accumulation rate as a function of the cell cycle. The standard error was calculated by dividing the standard deviation by the root mean square of the number of cells.

For each protein the nuclear accumulation rate calculation is based on tracks of 50 to 400 cells. Each track comprises of ~100 frames covering the complete cell-cycle. Altogether >5000 full cell-cycles were obtained using >500,000 cell segments. To correct for systematic measurement errors on the order of 20%, we normalized all curves by a standard curve: the nuclear accumulation rate was normalized by the median accumulation profile of the 25% least variable proteins. Very similar scoring of cell-cycle profiles is obtained if this normalization is omitted.

When using movies of the same clone taken on separate dates, the average fluorescent level over all cells over all times was calculated and used to normalize all intensities such that differences in lamp intensity on different days will not create a difference in the accumulation rates.

The accumulation rate is a combination of the production rate minus the degradation rate plus any net translocations into or out of the nucleus. For the 20 proteins presented in this study, translocation effects are seen only for USP7 and RBBP7. In these proteins, the import of the proteins into the nucleus after nuclear envelope reformation following mitosis is markedly slower than for other proteins. For the times where accumulation rates are measured (15%-85% of the cell cycle), this results in an initially high and rapidly decreasing rate of accumulation. No other translocation effects were evident for any of the proteins in any of the stages of the cell cycle other than the beginning of the cell cycle.

We define a root mean squared distance between two cell cycle accumulation profiles x1(t), x2(t) as: D(x1,x2)=Sqrt(Mean((x1(t)-x2(t))^2)). Similarly, we define the cell cycle dependence of an accumulation profile as its distance from a constant accumulation rate profile (one having a constant value of 1). Therefore, the cell cycle dependence score (CCS) of a protein y is: $CCS_y$= Sqrt(Mean((y(t)-1)^2)). We arrived at an estimate of the experimental noise and clonal differences by analyzing the distance between accumulation profiles of different clones of the same protein. The average distance $mu_{clones}$ between all pairs of clones of the same protein was calculated, as well as their standard deviation $sigma_{clones}$. A CCS of a protein y was considered statistically significant if it was larger than the average distance between clones by more than two standard deviations, that is $(CCS_y-mu_{clones})/sigma_{clones}>2$.

**Synchrogram visualization**: In order to show localization events related to mitosis, we used a time axis that is not linear, but rather has more time points in the beginning and end of the cell-cycle. Notice throughout that intensity is scaled, so that localization can cause apparent dimming of the image. For example, when USP7 nuclear bodies appear, the total intensity appears to decrease. Similarly, in the case of RPL4, the 97% time point shows the concentrated fluorescence of the nucleoli, whereas the 99% time point is after the nucleoli breakdown, and thus the fluorescence is more spread and the maximal intensity used for the image scaling is lower. This results in the apparent change in total fluorescence.

References


1. Smith, D. J. Mini-exon epitope tagging for analysis of the protein coding potential of genomic sequence. *Biotechniques* **23**, 116-20 (1997).
2. Morgenstern, J. P. & Land, H. Advanced mammalian gene transfer: high titre retroviral vectors with multiple drug selection markers and a complementary helper-free packaging cell line. *Nucleic Acids Res* **18**, 3587-96 (1990).

3. Wahlby, C., Sintorn, I. M., Erlandsson, F., Borgefors, G. & Bengtsson, E. Combining intensity, edge and shape information for 2D and 3D segmentation of cell nuclei in tissue sections. *J Microsc* **215**, 67-76 (2004).
4. De Solorzano, C. O., Malladi, R., Lelievre, S. A. & Lockett, S. J. Segmentation of nuclei and cells using membrane related protein markers. *J Microsc* **201**, 404-15 (2001).
5. Malpica, N. et al. Applying watershed algorithms to the segmentation of clustered nuclei. *Cytometry* **28**, 289-97 (1997).
6. Vincent, L. & Soille, P. Watersheds in Digital Spaces - an Efficient Algorithm Based on Immersion Simulations. *Ieee Transactions on Pattern Analysis and Machine Intelligence* **13**, 583-598 (1991).