A microchemostat array enables the spatio-temporal analysis of the yeast proteome.

Nicolas Dénervaud[1,3], Johannes Becker[1,2], Ricard Delgado-Gonzalo[3], Pascal Damay[4], Arun S. Rajkumar[1,3], Michael Unser[3], David Shore[4], Felix Naef[1,2] & Sebastian J. Maerkl[1,3]

[1] *Institute of Bioengineering, École Polytechnique Fédérale de Lausanne, Switzerland*
[2] *School of Life Science, École Polytechnique Fédérale de Lausanne, Switzerland*
[3] *School of Engineering, École Polytechnique Fédérale de Lausanne, Switzerland*
[4] *Department of Molecular Biology and NCCR program "Frontiers in Genetics," University of Geneva, Switzerland*

Contents

List of Tables

List of Figures

1 Material and methods

1.1 Microfluidic device design and fabrication

1.1.1 Design

The microfluidic device was designed in Clewin (WieWeb software, Netherlands). It is composed of two functional layers: a flow layer and a control layer, which each require the fabrication of a separate mold. The flow mold has a complex layout with three different structure heights for each of its functional components. This mold consists of flow channels (15 μm high), 260x300 μm small microchemostats (5 μm high), and shallow sieve channels (1.7 μm high), which allow for diffusion of nutrients and chemical compounds from the flow channels to the microchemostats. The control layer contains channels necessary for creating microvalves, which control the flow inputs, the chamber button, and a chamber outlet valve. The chamber button pressurizes the chamber and forces cells to grow in a monolayer. The chamber outlet valve permits the partial closing of the chamber to prevent possible contamination during the initial phase of on-chip cell culturing. The control layer has a height of 12 μm (Figure S1.a-c). Each of these structures required the writing of a separate chrome mask and was named as follows: 1. sieve channels, 2. chambers, 3. flow lines and 4. control lines. We placed three devices on one mold, allowing us to fabricate 3 devices at once. Due to PDMS shrinkage, we scaled the flow layer by 100.5% to fit to the spotting pitch of our cell arrays whose pitch is restricted since we used 4 pins for spotting of the arrays (see section 1.5.1). We also scaled the control layer by 101.5% to allow proper alignment with the flow layer.

1.1.2 Mold fabrication

Microfluidic device molds were fabricated using standard photolithography methods [1] in the clean room facility of the center of micronanotechnology (CMI, EPFL, Switzerland). For each of the four channel layers describe above, a chrome mask coated with AZ1518 (Nanofilm, Westlake Village, CA), was written on a DWL200 (Heidelberg Instruments Mikrotechnik GmbH, Germany), with a 4 mm writing head (2 μm resolution). Masks were developed with AZ351 (MicroChemicals GmbH, Germany) in a DV10 robot (Suess MicroTec AG, Germany), etched in a perchloric acid and ceric ammonium nitrate bath (Chromium etchant #1, MicroChemicals GmbH), and finally stripped of the remaining photoresist in a remover bath for 30 min (Remover 1165, Chimie Tech Services, France). These masks where then used to pattern photoresist-coated wafers on a MA6 mask aligner (Suess MicroTec AG). The negative photoresist SU8 (Gersteltec, Switzerland) or the positive resist AZ9260 (MicroChemicals GmbH), were used to generate square or rounded structures, respectively. The sieve channels, microchemostat chambers, and control lines were made of SU8. We used different polymer-solvent ratios to optimize product viscosity and obtain the targeted coating height. Negative re-

sists were coated on a Sawatec LMS200 spin coater (Sawatec AG, Switzerland) and developed in two successive baths of propylene glycol methyl ether acetate (PGMEA; Sigma-Aldrich, Switzerland) followed by active rinsing with 2-propanol. It is important to note that SU8 was not developed directly after the exposure of the first SU8 layer. Instead, layer 1 (sieve channels) and 2 (chambers) were developed together, after layer 2 was exposed and baked. This process prevented the sieve structures from impacting the topology of the second layer, resulting in a flat chamber profile. The flow channels were made of AZ9260, as they required a rounded cross-section to prevent valve leakage. The positive resist was coated on a RC8 THP spin coater (Suess MicroTec AG) and developed with AZ400K (MicroChemicals GmbH) on an automatic resist processing cluster (EVG150, EVGroup, Austria). The flow channels were annealed with a final baking step of the flow wafer at 160°C for 15 min. Table S1 summarizes the different baking times, temperatures, spin speeds, exposure times, and development processes for the four channel layers. Each wafer was profiled with an alpha-step 500 surface profiler (KLA Tencor, Milpitas, CA). The following structure heights were obtained for layer 1 to 4: 1.5-2.0 μm (sieve channels), 4.5-5.5 μm (chambers), 15-16.5 μm (flow lines) and 13-14 μm (control lines). Two different flow wafers (#1'961, #20'527) and two different control wafers (#22'728, #24'206), each with similar heights, were selected and used for the fabrication of all subsequent microfluidic devices.

1.1.3 Microfluidic device fabrication

The microfluidic devices were fabricated essentially as described previously [2]. The flow and control wafers were first treated in a vapour bath of trimethylchlorosilane (TMCS; Sigma-Aldrich) for 15 min. Polydimethylsiloxane (PDMS; Sylgard 184, Dow Corning Corp., Midland, MI) was used with two different elastomer to curing agent ratios: 1:20 for the flow layer and 1:5 for the control layer. 20g of PDMS (1:20) were spin coated on the flow wafer at 1800 rpm with a P6700 spin coater (Specialty coating systems Inc., Indianapolis, IN), to obtain a thin film of PDMS. 42g of PDMS (1:5) were poured onto the control wafer, which was placed in a circular glass dish, and degassed for 15 min in a desiccator. We baked both molds at 80°C for 30 min. Following baking, each of the three devices was cut from the control mold, resulting in three rectangular PDMS pieces of about 4-5 mm thickness. Holes for the control line connections were punched with a manual hole-puncher (Schmidt Technology Corp, Cranberry Twp., PA), using a steal punch with an outside diameter of 0.024 in (Ref. # CR0320245N21R4, Technical innovations Inc., Angleton, TX). The control layer was manually aligned to the flow layer and both layers were bonded at 80°C for 1.5 hour. Finally, each device was cut and peeled from the molds, and holes were punched in the flow layer to allow connection to the flow inlets and outlets.

1.2 Microscope hardware

A fully automated inverted epi-fluorescence microscope (Nikon Eclipse Ti-E, Nikon Instruments Inc., Melville, NY) enclosed by a temperature controlled incubation chamber (iceCube&Box, Life Imaging Service, Basel) was used for all image acquisitions. The microscope consisted of an encoded x-y stage, a motorized objective turret, a motorized filter wheel, and a hardware based auto-focus system. Two automated shutters (35mm Smartshutter TI, Sutter instruments, Novato, CA) were placed in the paths of the trans- and epi-illumination to control bright field and fluorescent exposure times. We used a LED based system for fluorescence illumination (CoolLED pE-2, Custom interconnected Ltd., UK). The LED based system was selected due to its better intensity stability over time (Figure SS2). Three different objectives were used: (i) a 4x plan achromat (Nikon ref. #MRL00042), (ii) a 20x achromat LWD (Nikon ref. #MRP00202), and (iii) an oil immersion 60x plan apochromat (NA 1.4, Nikon ref. #MRD31602). We also used two different filter cubes: (i) GFP-B (Ex 460-500/DM505/BA 510-560, Nikon), (ii) TexasRed filter (Ref. #F36-504, AHF Analysentechnik AG, Germany). Images were acquired with a back-illuminated EMCCD camera with a 1024x1024 pixel sensor array (Ixon DU-888, Andor Technology, UK). The camera was externally cooled with 16°C water (Oasis 160, Solid State Cooling Systems, Wappingers Falls, NY) and further cooled to -70°C using the internal peltier cooler.

1.3 Valve control

Pressure of the flow and the control lines was set by pressure regulators with a range of 1.3 to 25.0 psi (Type 10, Bellofram Corp., Newell, WV), connected to an in-house compressed air supply. To control the flow and the button pressure, we used high precision digital pressure gauges with ceramic sensors (PDC-102N2BFA, KOBOLD Instruments AG, Switzerland). Microfluidic valves were actuated by computer-controlled solenoid valves (Pneumadyne Inc., Plymouth, MN). We custom-designed a printed circuit board to link the solenoid valves to a USB input/output card (USBIO24R, Elexol electronic solutions, Australia), which remotely operated each valve using a Visual Basic interface. Details about the circuit design and its components can be found at cellbase.epfl.ch.

1.4 Microscope software

We wrote a Visual Basic program to control the microscope, its peripherals, the EMCCD camera, and the solenoid valves. The software was manually initialized with three corner points on the device to automatically acquire all positions on the device. Each double chamber position was calculated with simple trigonometry from the reference points. The software defines an optimal path that serpentines through each position and minimizes the distance between adjacent imaging stops (Figure S3). The software also handles two different medium sources and

switches between them for defined periods of time. It acquires time-lapse movies on two different light channels (phase contrast and fluorescence) and supports different camera settings for each channel.

1.5 On chip experiment

1.5.1 Cell preparation, microarraying and aligment

For each experiment, we replicated a different subset of 1152 yeast-GFP strains into 96-well V-shape plates (Nunc Ref. #249662, Fisher scientific, Switzerland). To prevent cross-contamination, plates were sealed with a breathable adhesive membrane (Breathe-Easy, Sigma-Aldrich). After 20-24 hours of incubation at 30°C and 240 rpm in YPD (Ref. #Y1375, Sigma-Aldrich), supplemented with 50μg/ml of ampicillin (Sigma-Aldrich), the cells were sedimented by centrifugation at 2400 rpm for 3min. Each strain was spotted on an epoxysilane coated glass coverslip (Ref. #25X60-1-C50-25, Thermo Fisher scientific specialty glass, Portsmouth, NH) using a standard DNA micro-arrayer (Qarray2, Genetix, UK), and four arraying pins (MP3B, ArrayIt Corp, Sunnyvale, CA), resulting in an array of 1152 (24x48) spots. The humidity within the arraying bed was set to 73% to prevent drying of the cell spots during spotting. Spotting with these settings resulted in an average spot diameter of 96.4 $\pm$9.8 μm (Figure S4). Directly after spotting, a PDMS chip was aligned to the cell array, such that each spot was situated within an individual chamber. During this process, spots quickly dry and did not stick to the chip, allowing multiple alignment trials without causing cross-contamination or disruption of the cells. The time between removal of the cell array from the spotter and priming with media typically ranged between 20 and 30 min.

1.5.2 Chip priming

Control lines were connected to the pressure source via water filled tygon tubing (0.02 in ID, 0.06 in OD) connected to the device with metal pins (0.3 mm ID, 0.65 mm OD). All control lines were primed with dH_2O at 5 psi. Consequently, pressure was increased to 20 psi, such that each valve could be fully switched on or off via computer-controlled solenoid valves (see section 1.4). Additionally, the chamber outlet valve was primed with dH_2O and pressure was maintained at 3 psi, in order to indent the chamber outlet channel without completely closing it. This prevented cells from entering/leaving the chambers, reducing cross-contamination, while at the same time allowing the chambers to fill with medium. As soon as the control lines were primed (time varied between 10-20 min), the flow layer was primed with SD-his at 1.3 psi. A special strategy was employed to maintain cells on the sieve side of the chamber outlet and prevent them from exiting the chambers. First, the flow channel on the sieve side was closed (valves V1 and V3 closed; figure S1d, sequence 1) and medium flowed on the chamber outlet side, progressively flushing each chamber and pushing the cells toward the

sieve channels. Due to surface tension, the medium did not flow through the sieve to fill the opposite flow channel. A short opening of V1 flushed the upstream section of the sieve-side channel, until the first chambers were reached (sequence 2). Then, the medium could flow through the sieve and complete the priming of the flow channel on the sieve side (sequence 3). When flow reached the outlet of the chip, the outlet valve was closed and the remaining air in the device was eliminated by out-gas priming through the PDMS. Once the chip was fully primed with medium, both upstream valves (V1 and V2) were closed and the outlet valve was opened. This released the pressure, which built up during priming and caused chamber ceiling deformation. The chamber returned to its relaxed shape and forced the cells into a well spread monolayer. Finally, both upstream valves were opened simultaneously allowing medium to flow on each side of the chambers efficiently perfusing the cells within the microchemostat chambers (sequence 4).

1.5.3 Initial growth phase control

After priming, chips were continuously perfused with SD-his at 1.7 PSI, at 30°C. The cells were grown for 16-20 hours to let them recover from spotting, enter log-phase, and fully populate each chamber. During that time, low resolution (4x) images of the entire device were acquired every 30min using NIS-element software (Nikon Instruments Inc.). These low magnification image sequences allowed us to systematically check the growth of each strain and identify local problems on the chip. Chambers exhibiting problems were annotated and later discarded from the data set.

1.5.4 Automated time-lapse movie acquisition

The chip and its tubing connections were immobilized on the motorized stage using tape. Images were acquired with a 60x oil immersion objective. The intermediate 1.5x lens was used to obtain a final magnification of 90x. DF 37 immersion oil (Cargille laboratories, Cedar grove, NJ) was selected for its stability over a long period of time; it did not dry out and its refractive index was not affected by a viscosity change, therefore allowing the focal distance to stay constant over time. The two automated shutters were used to switch between phase contrast (Ph3, Nikon instruments Inc.) and fluorescent channels within 50 ms. Focus was maintained at all times by a hardware-based focus system (perfect focus system (PFS); Nikon Instruments Inc.) that constantly measures the distance to the glass-medium or glass-PDMS interface and continuously adapts to vertical variability. As the refractive index of PDMS differs from that of medium, care must be taken that the PFS measures either the glass-medium interface or the glass-PDMS interface. Focus can be set to the cell mono-layer by manually trimming the offset between the glass-medium interface and the cell plane. Camera settings were the same for all experiments. Acquisition time was set to 200ms for the fluorescent channel, with an EM-gain of 100 and an analog gain of 2.4. For the phase contrast channel, acquisition time

was set to 50ms with an EM-gain of 2 and an analog gain of 2.4. The LEDs were set in the $\lambda 2$ channel (470 nm) at 20% intensity. To reduce the acquisition time, pairs of chambers were imaged together. Finally, we acquired time-lapse movies of all 576 positions (1152 chambers) at a sampling frequency of 20min for a total duration of 13 to 24 hours.

1.6 Device characterization

1.6.1 Medium switch

Because by design there was no active flow of medium through our microchemostat, medium switch times depended primarily on the diffusion times of small molecules into our chambers. The medium exchange times in the flow channels impacted the overall switch time to a lesser degree because the medium exchange time was much faster compared to the diffusion times. We connected two different medium sources to each of the three subsections of a device, to measure our medium switch times in our microchemostats. One medium source was SD -his and the other was either SD -his supplemented with sulphorhodamine 101 (Sigma-Aldrich) or a 10 kDa Dextran conjugated Rhodamine B isothiocyanate (Sigma-Aldrich). For each of the three subsections of the device (1. row 1-8, 2. row 9-16, 3. row 17-24), we ran a separate set of medium switches while recording the fluorescence intensity of the chambers. We used a 4x objective, the TexasRed filter and illuminated with our LED system set in the $\lambda 3$ channel (535 nm) at 20%. Acquisition time was set to 200 ms, with an EM-gain of 100 and an analog gain of 2.4. 12 positions, covering the 384 chambers of one subset were visited every 15 s. Medium was switched every 40 min and two repeats were monitored for all subsets. After switching, the new medium rapidly exchanged on each side of the chamber and started diffusing (Figure S5.a). Intensity was calculated as the median intensity in a rectangular area automatically located across the pair of chambers. The equilibration time was defined as the time necessary to reach 95% of the fluorescence at equilibrium (Figure S5.b). We compared two repeats for each subset of one chip and one repeat of one switch taken on another chip and on a different day (Figure S5.c). Equilibrium was reached within 4 to 5 minutes, showing good repeatability between different regions of a same chip and between two different chips. We also measured the equilibration time of 10 kDa Dextran conjugated Rhodamine B isothiocyanate to check the diffusion of a larger compound (Figure S5.d) and we found an equilibration time of 14 min. To measure the impact of cells on diffusion in a packed microchemostat, RPN4-GFP yeast cells were grown at log-phase in 3ml of SD -his, flown through the chambers and captured by the sieve channels. After growing the cells to confluence for 12h another series of sulphorhodamine switches were performed. The results show that the presence of cells did not affect the equilibration time, which was measured to be around 4 min (Figure S5.e).

1.6.2 Lamp calibration and stability assessment

We tested two different light sources: (1.) a mercury source (Intensilight, Nikon instruments Inc.) and (2.) LEDs (CoolLED pE-2, Custom interconnected Ltd., UK). Recordings were taken with a laser power meter (FieldMaxII-TO, Coherant Inc., Santa Clara, CA). The optical sensor was taped directly on top of the 20x objective. To calibrate the mercury source, power was measured for a set of neutral density filters (ND1, ND2, ND4, ND8, ND16, ND32), using the GFP and Texas red filter cubes (Figure S2.a). The LEDs were calibrated in the $\lambda 2$ (470 nm) and $\lambda 3$ (535 nm) channels, using the GFP and Texas red filter cubes respectively. The power was measured for inputs of 0%, 20%, 40%, 60%, 80% and 100% (Figure S2.b). We assessed the relative stability of the two light sources by measuring the respective power of each light source over 12 hours, with an acquisition period of 30 s and a resolution of 1 pW, (Figure S2.c). For the latter experiment, measurements were taken with the GFP filter cube and the LEDs were set in the $\lambda 2$ channel at 10% and the intensilight was used without optical density filter (ND1). We found that the LED light source was impressively more stable than the Intensilight, with maximal variation of 1%. The Intensilight showed bigger variations, a drift over long period of time, and sensitivity to aging. Together, this shows that the LED is a suitable light source for long-term image recordings.

1.7 Screening of the yeast GFP library

1.7.1 Yeast strains

All 4159 strains used in this study are from the yeast GFP clone collection [3] (Invitrogen AG, Switzerland). Strains were grown at 30°C in 96-well U-bottom plates (Nunc Ref. #262162, Fisher scientific, Switzerland) and incubated for 24 hours in yeast synthetic defined medium (Ref. #Y0626, Sigma-Aldrich) complemented with amino-acids without histidine (Ref. #Y1751, Sigma-Aldrich), referred to as SD -his. A working stock in 20% glycerol was kept at -80°C.

1.7.2 Stress conditions

A variety of stress-inducing media additives were used in this study. Medium was connected to the chip just before the start of image acquisition to prevent degradation of the chemicals. The following medium compositions were used:

1. High MMS: SD-his was complemented with methyl methanesulfonate (MMS; Ref. #129925, Sigma-Aldrich) to a final volume of 0.03%
2. Low MMS. SD-his was complemented with MMS to obtain a final volume of 0.0075%
3. Sorbitol. SD-his was complemented with 2M sorbitol (Ref. #S1876, Sigma-Aldrich) to a final concentration of 1M

As an alternative to MMS, UV-induced DNA damage was tested. UVC radiation (254nm) was generated by a mercury lamp (Pen ray 11SC-01, UVP, Upland, CA) equipped with a UVC filter (G-275, UVP). The setup was characterized with a UV radiometer (UVX Radiometer, UVP - courtesy of LESO-PB, EPFL). An irradiance of 11 mW/m^2 was measured just below a PDMS chip, placed at the usual location on the scope. We used three different doses by varying the exposure time (10 min - 6.6 J/m^2, 30 min - 19.8 J/m^2, 1 hour - 39.6 J/m^2).

1.7.3 Titration of MMS concentration and its effect on cell growth

We measured growth curves of 4 yeast strains (Gpd1-GFP, Cdc10-GFP, Rnr4-GFP, Eno1-GFP) for different concentrations of MMS. MMS dilutions were prepared in a flat-bottom 96-well plate filled with 150 μl of SD -his. Strains were grown overnight in 3ml of SD -his and 5 μl of each cell solution was used to inoculate two rows of the plate. 15 μl of mineral oil (Sigma-Aldrich) was added to each well to prevent evaporation. Optical absorbance at 600 nm was measured for 15 hours using a plate reader (Synergy MX, Biotek, Switzerland). Temperature was set to 30°C and the plate was continuously shaken to prevent cell sedimentation. An exponential was fit to the log-phase of each growth curve and doubling times were directly derived from the growth rate (Figure S6).

1.7.4 On-chip doubling time estimation

To estimate the growth rate on chip, we used highways that guide the cells vertically, in the y-dimension. High time resolution sequences were acquired with a time step of 30 seconds, for a total duration of 120 minutes. Single-cells were tracked frame to frame, to determine the vertical movement of the cells Δy as a function of their initial position y (Figure S7). Considering that cells can only escape the chamber on one side, and that the volume of the chamber is constant over time, we established a relationship that defines a cells new position Δy, depending of its initial position y_0, the time step Δt between the two measurements and the average doubling time T_d :

$$y(t_0 + \Delta T) = y_0 \cdot 2^{\frac{\Delta t}{T_d}} \tag{1}$$

Thus, the estimated movement of a cell $\Delta y = y(t_0 + \Delta t) - y_0$ is a linear function of y:

$$\Delta y = y_o \cdot (2^{\frac{\Delta t}{T_d}} - 1) \tag{2}$$

The slope a of this function was measured by averaging single-cell movements over space and time. Knowing a, the doubling time T_d was approximated with the following relationship:

$$T_d = \frac{\Delta t}{log_2(\alpha + 1)} \tag{3}$$

With this method, 12 image sequences from different chambers were analyzed. We found an average doubling time of 129 min, with a standard deviation of 17 min. Batch measurements gave an average of 120 min with a standard deviation of 12 min. The cell doubling time therefore is on average 10 min slower on chip.

1.7.5 Yeast GFP library imaging under standard and MMS conditions

We imaged the full yeast GFP library (4156 strains), 11 plates at a time, or 1056 strains per experiment. A control plate with selected strains was included in each experiment. After 6 hours of recording at steady-state, cells were treated with high MMS and recorded for 7 additional hours. We screened the entire library in duplicate, in a total of 8 experiments. The quality of each movie was manually assessed (see section: 1.9.4). 3860 strains (92.8%) were covered with at least one good quality movie. We manually picked the remaining 299 strains into four 96-well U bottom plates. One additional experiment with these strains allowed us to complete the screen and cover 4085 strains (98.2%) (Figure S2).

1.7.6 Strain subset selection and secondary analysis

Based on the manual annotation and the bulk analysis of the large-scale experiment (see section: 1.9.4, 1.9.3), we selected a subset of 576 strains that showed responses to MMS-treatment, in terms of protein abundance or localization change. We selected all strains that were manually annotated as exhibiting a localization change. For the intensity change, we selected the strains that showed 5 hours post-treatment, an intensity fold-induction above 1.68 ($2^{0.75}$), as measured in bulk. We completed the selection by picking the strains with the highest intensity fold-induction among the remaining strains that were manually annotated as changing in intensity. The 576 strains were automatically cherry picked and re-arrayed using a colony picker (Qpix2 XT, Genetix, UK). Strains were stored in 6, 96-well U bottom plates. 11 experiments were conducted with the selected strains to monitor their response to other sources of stress. The stress conditions and relative experimental timings are summarized in table S2.

1.8 Deletion Strain Construction

1.8.1 Deletion strains

Strains containing C-terminal GFP fusions at 14 genes of interest (genotype: MAT**a** his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 goiX-GFP::HIS3MX; obtained from ATTC and confirmed by PCR of the ORF-GFP junctions) were first crossed to Y9230 (MATα can1Δ::STE2pr-URA3 lyp1Δ ura3Δ0 leu2Δ0 his3,1 met15Δ0). The resulting diploids were then sporulated and haploid segregants of the following genotype were identified: MATα his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 lyp1Δ yfg-GFP::His3MX can1Δ::STE2pr-URA3. These 14 GFP fusion strains were then crossed to a set of 40 different single-gene deletion strains (ORF replacements by kanMX4;

generated by the Saccharomyces Genome Deletion Project) of the following genotype: MAT**a** his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 yfg::KanMX using a liquid-handling robot. The resulting diploids were sporulated and haploids of the following genotype MAT**a** his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 yfg::KanMX lyp1Δ yfg-GFP::His3MX can1Δ::STE2pr-URA3 were selected. Robotic mating, sporulation and haploid selection were done according to Tong et al. [4]. The gene disruptions in strains giving rise to phenotypes in our screen were confirmed by 5 PCR reactions designed to detect both junctions of the kanMX deletion/insertion, the absence of the corresponding junction fragments of the wild-type allele, and a full-length cassette insertion (lack of the wild-type allele). All deletion-GFP strains were imaged in quadruplicate in response to UV-pulses as described above.

1.9 Image processing and single-cell analysis

1.9.1 Overview

The initial analysis of the image sequences required personal intervention to assess movie quality and to make a first round of qualitative observations. However, the single-cell analysis was fully automated and quantitative. The same set of parameters was used to analyze all the movies. Therefore, the analysis pipeline is robust and did not require optimization on an experiment-by-experiment basis. The entire analysis process is the following (with time estimates in brackets):

1. Manual analysis of low magnification movie (see section 1.5.3, 20 minutes)
2. Manual quality analysis (see section 1.9.4, 4 hours for quality only)
3. Updating path settings in the parameter file (see section 1.4, 2 minutes)
4. Submission of jobs to the batch system (5 minutes)
5. Automated computational analysis: chamber separation, cell segmentation, cell sorting, auto-fluorescence deconvolution, and feature extraction (see sections below, 8-12 hours of computation time depending on the number of frames per sequence)
6. Data merging (see section 1.9.9, 20 minutes)

1.9.2 Chamber separation

As we image two chambers at the same time we needed to separate these two chambers into independent image sequences. Also, a small positional or rotational drift could occur during recording due to imperfect taping of the device onto the microscope stage. A MATLAB-based script was written to identify the chamber boundaries, find the spatial shift, the rotation of the chambers, align the position over time, and create two separate stacks of images for each

pair of chambers (Figure S9). To identify chamber boundaries, we successively applied a set of edge detectors and morphological operators to the phase contrast images. First, a Sobel filter was applied to identify cell-containing regions, *i.e.*, the microchemostats. To fill the gaps between the edges, we performed two successive morphological closures: first, with a disk with a radius of 30 pixels and then with a rectangle of 60x10 pixels. This resulted in binary image in which the chambers were white and the rest of the image black. Boundaries were then detected with a Sobel filter. To increase the thickness of the boundaries, we applied a vertical and a horizontal dilation with a line of 5 pixels width. To define chamber rotation, we applied a Hough transform on the boundary image and identified its highest peaks, each representing a line. We found the longest vertical line ($\pm$ 6°) and took its angle as the image rotation. As rotation does not change much over the course of the experiment, we took the mean of the rotation over the full image stack and used this value to rotate each image of the stack with a method of nearest-neighbor interpolation. A binary template was manually drawn and used to find the spatial shift. First, the spatial correlation of the boundary image and the template was calculated. Then, we convoluted the resulting correlation matrix with a trident filter (a horizontal filter with three rectangular shapes). This prevented misalignment of the template with one chamber boundary only, by increasing the middle peak of the correlation. The maximal value of the matrix was identified and the x-y shift derived from its position over time. Limits were set on the vertical and horizontal shifts to prevent aberrant errors. Finally, the shift values for the whole image stack were fitted to a second order polynomial to interpolate a smooth and continuous movement. This was particularly important to prevent "shaky" movies. The interpolated shift values were used to crop the images on the chamber boundaries. All the images of one stack were cropped to the size of its smallest image. The same rotation and cropping was applied to the corresponding stack of fluorescent images. All thus corrected image sequences were saved with a 14 bit depth.

1.9.3 Bulk background subtraction and fluorescence intensity measurement

Data was initially analyzed in bulk by calculating the average background and foreground intensities over the entire field of view of each image stack. For each cropped and rotated image, a background image was generated. Sliding a window of 50x50 pixels through the image with an increment of 25 pixels, we were able to calculate background values located at each window center. The background value was determined as the 5^{th} percentile of the pixel distribution within the window. Based on this grid of values, we interpolated a background image using a triangle-based cubic interpolation. Finally, fluorescence intensity was calculated as the mean of all the pixel values after background was subtracted.

1.9.4 Manual annotation

To check for image quality and to have a first impression of the data, we watched all movies in the fluorescent channel. Each movie has a duration of 4 to 7 seconds (10 frames/second), and was analyzed frame by frame if necessary. It takes about 12-16 hours for one person to annotate one experiment consisting of 1152 movies, 0.6-0.8 minutes on average per movie. Denervaud and Maerkl annotated the movies based on a defined list of criteria:

1. Empty: tagged if no cells were in the chamber
2. Cross-contamination: tagged if cross-contamination was detected (wrong strain in the chamber, two apparently different stains in one chamber)
3. No growth: tagged if the cells did not recover after spotting and were not growing
4. Late growth: tagged if cells entered log-phase too late
5. Dead cells: tagged if many dead cells were present in the chamber. Dead cells appeared bright in the fluorescent channel.
6. Focus: tagged if cells grew in a multilayer or if the image was out of focus.
7. Washed out: tagged if the cells were partially washed out of the chamber during the experiment.
8. Saturating: tagged if the fluorescent signal saturated the camera.
9. Neighbor leak: tagged if the fluorescence from the neighboring chamber was high and leaked into the chamber, therefore affecting the measured signal.

The two annotators agreed to discard a movie in 63% of the cases. Maerkl agreed with Denervaud 94% of the times, and Denervaud was in general stricter on the quality. To make the quality check stringent, a movie was discarded if either of the two annotators checked it as discarded.
For the large-scale experiment, we also observed the overall cell behavior and annotated intensity changes (up, stable, or down) and localization changes, following MMS treatment. For localization, we annotated changes based on 13 localization patterns:

1. Nothing: no signal above cell background.
2. Cytoplasm
3. Nucleus
4. Nucleolus

5. Nuclear periphery
6. Mitochondrion
7. Structured: combines subcellular structures that are hard to define such as early and late Golgi, and low intensity mitochondria.
8. Bud/bud neck
9. ER
10. Vacuole
11. Punctate: takes together endosome, lipid particles, peroxisomes and other punctate composite patterns.
12. Actin/Tubulin/Spindle pole
13. Unclassified: unclassified low intensity patterns, often due to uneven cell background.

Based on the qualitative observations, we classified each strain in one of these four categories: no change, intensity change, localization change, intensity and localization change. Observations can be made on one or two movies depending on the number of repeats that are good or discarded. Annotations from the two controllers were merged in the following ways:

- For the intensity changes, we denoted a strain as changing if 75% or more of the annotations were positive. By a single annotation, we mean an annotation on one repeat, made by one annotator. Eg: A strain with one repeat has two annotations, therefore both have to agree. A strain with two repeats has four annotations, therefore at least three have to be positive.
- For the localization changes, we denoted a strain as changing if 50% or more of the annotations were positive. Eg: A strain with one repeat has two annotations, therefore at least one must be positive. A strain with two repeats has four annotations, therefore at least two must be positive.

This classification was used to select a small set of strains that showed a response upon MMS induced stress (see section: 1.7.6). However, final classification and analysis was performed using the automated image analysis software.

1.9.5 Cell Segmentation

The segmentation of the cells within the image stack was carried out in a two-step process. The first step, named *detection stage*, consisted in partitioning the images in non-overlapping regions using contextual spatiotemporal information. The set of regions provided a coarse segmentation, each of them representing a cell detection. The second stage, named *outlining stage*, consisted in analyzing each region individually to determine a closed curve that outlined the cell contour within the region. Since the number of images within the dataset was large, it was necessary to limit the access of several images at the same time in order to contain the memory consumption of the algorithm. For this reason, just one image was loaded into memory at a time, including the spatiotemporal analysis of the detection stage.

We began the detection stage by creating a mask within the imaging area that determined the regions where no movement occurred through each image stack. The motivation behind the creation of this mask is to determine the static region within the image sequences without any moving cell at any point in time (*e.g.*, pillars within the chambers or cell carcasses). For this purpose, we computed (pixel-wise) the temporal standard deviation σ_t of the brightness values. This could be achieved efficiently in a single pass through the sequence [5]. The non-moving region (NMR) was then defined as the pixel locations such that

$$\frac{\sigma_t}{\mathrm{Max} - \mathrm{Min}} \tag{4}$$

was below a user-specified threshold τ_{NMR}, and where Max and Min represent the maximum and minimum pixel brightness through the whole sequence respectively. The threshold was determined manually for each experiment, since it is strictly related to the cell dynamics, and to the temporal resolution of the time-lapse sequence. Typical values of τ_{NMR} are in the range of $[0, 0.1]$, that is, between 0-10% of pixel intensity variation though time. We illustrate this process in Figure S10.

Next, for each image of the sequence, a preprocessing step was performed prior to the splitting of the images in non-overlapping regions. We applied a morphological dilation with a 4-connected ball as a structuring element, and a smoothing filter with a Gaussian kernel of variance $\sigma_w^2 = 5^2$. This smoothing proved to be strong enough to eliminate spurious local minima to seed the forthcoming image partitioning algorithm. Finally, the image values were linearly stretched between 0 and 255, and the image was quantized in 8 bits.

The splitting of each image into regions was carried out by the watershed transform [6]. This algorithm is usually described intuitively in terms of flooding simulations. Consider the brightness of the image as a topographic surface, and assume there are local minima. If the surface is immersed in water, the water will flood the catchment basins of the image. Then, dams are raised wherever water coming from different minima confluences. At the end of the process each minimum is surrounded by dams delineating its associated catchment basin. In our case, the cells appeared as dark objects surrounded by brighter halos over a gray background. This

made the watershed algorithm appropriate for separating the cells into local catchment basins, but not suitable for providing accurate cell outlines. The watershed regions that intersected with the non-moving region were discarded, since they corresponded to empty regions or structuring elements of the chamber. We illustrate this process in Figure S11a.

We denoted as *extended watershed sub-images* the regions within the original images that corresponded to the regions within the bounding boxes of the watershed regions plus an extension of $n_f = 3$ pixels in each direction. This parameter provided contextual information of the adjacent cells.

In the outlining stage, each extended watershed sub-image was analyzed independently. The most important techniques used in this stage were the parametric active contours (a.k.a. snakes). Within an image, a snake is a curve that evolves from an initial position toward the boundary of an object [7,8]. The evolution of the curve is usually formulated as a minimization problem, and the associated cost function is the snake energy.

In order to obtain a first estimate of the contour of the cell, we used a minimalistic snake named the ovuscule [9]. This snake takes the shape of an ellipse, and is parametrized by three control points. The elliptic shape of the ovuscule makes it a very robust cell-segmentation algorithm in poor imaging conditions [10]. Moreover, it is fast to compute and its optimization can be carried out by fast gradient-descend-based methods. We initialized each ovuscule within each sub-image such that the bounding box of the ovuscule coincided with the sub-image margins. The optimization of the snake energy was then efficiently carried out by a Powell-like line-search method [11]. Since the initial location of the ovuscule was very close to the cell contour, we limited the maximum number of iterations to 100 without losing accuracy. We illustrate the segmentation accuracy of the ovuscule in Figure S11b when segmenting a cell within a extended watershed sub-image.

Two scenarios were possible depending on eccentricity of the resulting ovuscule after the optimization process. If the eccentricity e was lower than 0.8, the shape provided by ovuscule was considered valid, and a refinement process over the ellipse was carried out. On the contrary, if the eccentricity was higher or equal than 0.8, the ovuscule was not considered valid and a morphological process was applied to segment the cell. Note that $e = 0$ corresponds to a perfect circle, and $e = 1$ to a degenerated ellipse with no area.

For the first scenario, a preprocessing step was performed prior to the refinement of the ellipse. We applied a smoothing filter with a Gaussian kernel of variance $\sigma_s^2 = 4^2$, and we linearly stretched the image values between 0 and 255. Next, we used a more flexible snake variant with variable number of control points named E-snake [12]. This snake was built using exponential splines as basis functions to represent the outline of the shape. While this snake is versatile enough to provide a good approximation of any closed curve in the plane, its most important feature is that it perfectly reproduces circular and elliptical shapes. These features are very appropriate to delineate blob-like objects and allowed for a smooth transition from the ovuscule. Moreover, it has been shown that due to the properties of the basis functions,

this snake minimizes the computational load with the properties stated above [13]. The snake energy was taken to be the addition of the magnitude of the image gradient projected onto the snake, which corresponds to the snake energy with $\alpha = 1$ in [12]. We set the number of control points to $M = 5$. This number of points was high enough to capture small departures from an elliptic shape, and small enough to avoid overfitting to the noise. The optimization of the snake energy was also efficiently carried out by the same Powell-like line-search method than the ovuscule. Since the initial location of the snake corresponded to the exact same shape as the ovuscule, the snake was usually very close to the cell contour. Thus, we also limited the maximum number of iterations to 100. If the snake converged before the 100 iterations, the final cell outline was taken to be the curve provided by the snake, otherwise, the outline of the ovuscule was used. We illustrate the segmentation accuracy of the E-snake in Figure S11c when refining the segmentation of the ovuscule given in Figure S11b.

For the second scenario, that is, when none of the above active contours converged, the sub-image was thresholded using the IsoData algorithm [14]. This classified each pixel into two classes: the cells, and the background. Since the binary class that contained the cells also contained parts of neighboring cells, we labeled all 4-connected components that were contained within the watershed region, and selected the one with the biggest size. Next, we applied a morphological dilation with a 4-connected ball as a structuring element. Finally, we obtained the outline of this region by the Square Tracing Algorithm [15]. The final cell outline was obtained by downsampling by a factor of 5 the result of the tracing algorithm.

1.9.6 Single-cell background subtraction

The background was subtracted for each cell locally. Fluorescence fluctuations due to monolayer issues could be removed, while preserving the relative intensity change for each cell. The background value was calculated with the segmentation plugin. We tested 4 ways of calculating the background:

1. Minimal pixel value of the watershed region.
2. Minimal pixel value of the watershed region excluding the cell region.
3. 5^{th} percentile of the pixel distribution of the watershed region
4. 5^{th} percentile of the pixel distribution of the watershed region excluding the cell region.

We decided to use method 1 for its robustness and its ability to deal with slight inaccuracies in the segmentation. This was particularly important when the fluorescent signal was located mainly in the cell periphery.

1.9.7 Single-cell quality filtering.

We implemented a virtual cell-sorting algorithm to include only well-segmented cells, to gate cell size, and discard dead cells. We implemented a MATLAB script that takes as input the features from the cell segmentation plugin and defines whether a cell is retained in our analysis based on the following criteria:

1. cell distance from the image boundaries
2. cell size
3. cell segmentation quality
4. cell intensity (to detect bright dead cells)

The distance to the image boundary was calculated as the minimal vertical or horizontal distance between the cell center and the boundary. A fixed threshold was manually set to 40 pixels. Because cells that are on the boundaries are generally not well segmented, all the cells closer than this threshold were discarded. To remove aberrant segmented objects such as big empty areas, chamber posts or small artifacts, we set a strict threshold on cell size. Namely, we retained cells that fall within a so-called "safety gate", between 300 and 1500 pixels in size (6.31 to 31.54 μm^2). To further improve size gating and remove cells with aberrant size, we only accepted cells that fit within the 2^{nd} and 98^{th} percentile of the cell size distribution (Figure S12.a) Segmentation quality was assessed using a support vector machine-based classifier (SVM) that takes as input a set of quality features calculated during segmentation. A training set of 24 images (6155 cells) was selected amongst medium to low quality images, covering all possible errors and inaccuracies. Segmented cells in these images were then manually annotated as good or bad quality. For each of the three segmentation methods (ovuscule, snake and region-based thresholding), a classifier was trained based on the following features (see Figure S12.b-d for distribution of the feature values):

1. Cell circularity:
$$c = \frac{4\pi \cdot area}{perimeter^2} \tag{5}$$
2. External intensity (outside the cell) to internal intensity (within cell) ratio
3. Cell area / watershed area
4. Segmentation quality factor for each method:
    - Ovuscule: Minimal energy
    - Snake: Minimal energy
    - Region-based thresholding: cell area / convex area

Cells passing the boundary distance check and size gating were used to train the SVMs. For each segmentation type, a training and a testing set of equal size were randomly chosen from the annotated set of cells. Having more negatives than positives in the training set allowed us to optimize the classifiers for specificity (1-false positive rate). Overall, we correctly identified 83.7% of the cells, with a specificity of 92.3% (Figure S12.e-f).

As a final sorting step, we analyzed the cell's fluorescence intensity distribution in each image and discarded cells that were outliers. For a distribution of intensities $\{I^1, ..., I^N\}$ of N selected cells, we calculated the average ($\bar{I}$), the standard deviation (I_σ), the lower quartile (I_{LQ}), the upper quartile(I_{UQ}) and the inter-quartile range (I_{IQR}). A cell n with intensity I^n was tagged as an outlier if it lies beyond the outer fence and deviates from the mean by more than 3 standard deviations:

1.
$$I^n < I_{LQ} - 3 \cdot I_{IQR} \quad \text{, or} \quad I^n > I_{UQ} + 3 \cdot I_{IQR} \tag{6}$$

2.
$$I^n < \bar{I} - 3 \cdot I_\sigma \quad \text{, or} \quad I^n > \bar{I} + 3 \cdot I_\sigma \tag{7}$$

1.9.8 Auto-fluorescence measurement and deconvolution

To measure the distribution of cell auto-fluorescence, we imaged the GFP library parent strain (BY4741) under the same condition as the GFP strains. A device was bonded to an epoxy-coated coverslip for 20 min at 80°C and primed with SD complete (+his). Cells were flown through the device at 1.5 PSI and forced through the chambers by closing valves V1 and V3 (Figure S1.d). After 12 hours of incubation, we performed time-lapse acquisition of 768 chambers in SD complete medium for 6 hours (18 frames) with the same parameters as previously described (see section 1.5.4). Each movie was manually annotated for quality to obtain a final set of 127 movies. Each image sequence was processed by the same image analysis pipeline described above. For each chamber, we obtained a distribution of fluorescence intensities by pooling all the frames together, resulting in 4897 $\pm$ 467 (mean$\pm$s.d.) cells per chamber. We tested the goodness of fit of four different distributions (normal, gamma, log normal, and logistic) with qq plots (Figure S13a-b) and a Kolmogorov-Smirnov test. The log normal distribution gave the best fit with an average p-value of 0.28$\pm$0.2569 (mean$\pm$s.d.), accepting the null hypothesis (good fit) at the 5% confidence in 99 cases (80%). Averaging the 127 fits together, we obtained a final log normal distribution with the follwing parameters: μ = 4.212$\pm$0.019 (mean$\pm$s.d.) and σ = 0.131$\pm$0.005 (mean$\pm$s.d.).

For each GFP strain distribution, we deconvoluted the auto-fluorescence using a probabilistic approach. For a cell i, the fluorescent signal measured s_i is the sum of the cell auto-fluorescence b_i and the contribution of GFP g_i. Therefore, the probability of the measured

signal $S = \{s_i\}$, considering the auto-fluorescence distribution $B = \{b_i\}$ and the GFP signal $G = \{g_i\}$, is the following:

$$P(S) = \int_0^\infty P(g)\ P(b)\ \delta(s = g + b)\ dg\ db \tag{8}$$

With s=g+b, we get:

$$P(s) = \int_0^s P_g(x)\ P_b(s - x)\ dx \tag{9}$$

Therefore, the probability distribution of the measured signal is the convolution of the GFP distribution and the auto-fluorescence distribution. The auto-fluorescence distribution was represented by the log-normal $\ln\mathcal{N}(\mu,\ \sigma)$, with the parameters measured above. We used a gamma distribution with shape parameter k and scale parameter θ to estimate the GFP signal $P_{gamma}(x, k, \theta)$. The gamma distribution has been shown to be a good approximation of protein distribution [16, 17]. Therefore, the probability of the measured signal can be re-written as:

$$P(s, k, \theta) = \int_0^s P_{gamma}(x, k, \theta) \ln\mathcal{N}(s - x, \mu, \sigma))dx \tag{10}$$

The log-likelihood of this probability for the set of measured cells $\{s_i\}$ is:

$$LogL(k, \theta) = \sum_i \ln P(s_i, k, \theta) \tag{11}$$

We maximized the log likelihood using a Markov Chain Monte Carlo (MCMC) method. Generating the random walk of k and θ over 2000 samples, we obtained the mean and standard deviation of k and θ, μ (=$k\theta$) and η (=1/k). $P(s, k, \theta)$ was calculated using a numerical integration.

1.9.9 Movie selection and data merging

For the first screen, we obtained the following coverage: 2580 strains with double coverage, 1320 strains with single coverage, 185 with 3x or more coverage, and 74 strains with no coverage (see figure S17). We decided to select the best movie, based on the phase contrast image, for each strain in order to obtain a non-redundant final dataset and to avoid any bias between strains with double coverage and single coverage. Thus, we decided to select the best movie for each strain in order to obtain a non-redundant final dataset. To systematically assess movie quality, we measured four criteria, based on our single-cell quality filter:

1. lowest cell count over 50 for the frame with the fewest cells
2. average ratio of good cells over total cell number
3. number of null frames (frames with no good cells, due to transient loss of focus)

4. number of low frames (frames with less than 50 good cells)

To apply the selection, we first set a minimal threshold of 50 on the lowest cell count requiring each frame to contain at least 50 cells (criterion 1). If only one movie fulfilled this condition, it was defined as the best movie. If more than one movie fulfilled this condition, the best movie was defined among those as the one with the highest average rate of good cells (criterion 2). If no movie fulfilled this condition, the best movie was first defined as the one with the lowest number of null frames (criterion 3), and if necessary as the one with the lowest number of low frames (criterion 4). Supplementary list 1 shows a summary of the movie selection process, and the values of the different criteria and the origin of the best movie for each strain.
For the second set of experiments based on a subset of 576 strains, we obtained a higher number of movies per strain (1 to 4, see figure S18a). Instead of selecting the best movie, we took advantage of each good quality movie by averaging the data over all the repeats.

1.10 Analysis of protein abundance

1.10.1 Filtering of abundance data

Average abundance values of the cell population were directly calculated from the gamma fit obtained after the deconvolution of autofluorescence (see section 1.9.8). The minimal possible value was set to 1, to avoid divisions by zero or very small numbers. We first filtered the time-course obtained for each strain. We disregarded every strain for which abundance does not go at any time point above a threshold of 3 times the standard deviation of the autofluorescence distribution. This allows us to concentrate on strains reaching significantly high abundance levels. 1383 strains failed to pass this threshold, showing that many proteins are expressed at very low levels, or simply not expressed. To further improve the quality of our dataset, we disregarded strains for which abundance values varied more than 2-fold during the six hours preceding MMS treatment. 168 strains showed unstable behavior in steady conditions. We obtained a final set of 2534 strains, for which we analyzed the dynamics of protein abundance (see figure 3.a). The same approach was used to filter the data of the second round of experiments.

1.10.2 Fold-change value and significance

Fold-change was calculated in reference to the median of the values before treatment, $\widetilde{x}_{pre}$. Thus, fold-change at each time-point is defined as:

$$f_t = \frac{x_t}{\widetilde{x}_{pre}} \tag{12}$$

We also calculated the significance of the change. The repeatability of fold-change measurement was assessed from 2580 strains, where two measurement repeats are available. Measurement error is calculated for each replicate as the difference of the log2(fold-change). The

errors distribute equally around zero and were well approximated in a given abundance window by a Gaussian. We used a running standard deviation (window size = 250 data points), to approximate the distribution of errors, for a given abundance. Because error decreases with abundance, the running standard deviation was fitted by a negative exponential, dependent on the initial abundance level, $\widetilde{x}_{pre}$ (see figure S15):

$$\sigma^2(\widetilde{x}_{pre}) = 14.07 \cdot \widetilde{x}_{pre}^{-1.09} + 0.25 \tag{13}$$

Thus, the p-value, p_f, of the fold-change, f_t, was directly assessed from the initial abundance level $\widetilde{x}_{pre}$:

$$\begin{aligned} p(\epsilon) &= 1 - \Phi(log_2(f_t), 0, \sigma^2(\widetilde{x}_{pre})), \text{ if } log_2(f_t) > 0 \\ p(\epsilon) &= \Phi(log_2(f_t), 0, \sigma^2(\widetilde{x}_{pre})), \text{ if } log_2(f_t) \leq 0 \end{aligned} \tag{14}$$

where $\Phi(x, 0, \sigma^2)$ is the cumulative distribution function of a Gaussian centered in 0, with a standard deviation of σ^2. This p-value gives us the significance of the change, in regard to the estimated distribution of measurement error.

1.10.3 Periodicity of the abundance signal

To measure the periodicity of abundance variation in the pulse experiments, we calculated a p-value associated with a 6 hour oscillation, using a Fisher test for a specific period, as previously described [18, 19]. In short, the periodic components of the signal were extracted by Fourier analysis. The Fisher test for one component is given by the following equation, where P_k is the power of the k^{th} component of the discrete Fourier transform and P_{6h} is the power for a 6 hours period.

$$S = \frac{2 \cdot P_{6h}}{\sum_{k \neq 0, \infty} P_k} \tag{15}$$

The p-value, p_{per}, was inferred directly from the Fisher score:

$$p_{per} = (1 - S)^{\frac{T}{2} - 2} \tag{16}$$

where T is the length of the signal.

1.10.4 Selection of induced proteins

To select for induced proteins, we applied two thresholds: (i) on the fold-change and (ii) on the significance of the change (see section: 1.10.2). For the experiments with a single stimulus (MMS_{high}, MMS_{low}, Sorbitol), the minimal fold-change was set to 3 and the significance p-value (p_f) had to be below 0.01 (see figure 3.b, S27b and S29b). For the pulse experiments

(MMS_{pulses}, UV_{pulses}), a third threshold was set on the significance of the signal periodicity (see section: 1.10.3). Because protein accumulation is low in response to short stimuli, we set less stringent threshold for the pulse experiments. However, the analysis of the signal periodicity allowed to detect protein response with a high sensitivity and specificity. The fold-change threshold was set to 2, the significance threshold was set to 0.05 and the p-value for periodicity (p_{per}) had to be below $1e^{-4}$ (see figure S28b and S30b).

1.10.5 Analysis of the rate and timing of protein induction

The rate and timing of protein induction was calculated for a set of 124 proteins in MMS_{high} (see figure 3.e), for 10 proteins in MMS_{low} (see figure S25.a-b) and for 24 proteins in MMS_{pulses} (see figure S25.c-d). Rate and timing values were averaged over the number of repeats available for each proteins. For each available movie, abundance time-series were first smoothed, using a smoothing cubic spline (MATLAB function 'csaps'), with a smoothing factor of 0.8. Values were normalized to the median of the last 6 time-points pre-MMS, $\widetilde{y}_{pre}$ (last 2 hours). The time to induction was defined as the first time-point were the abundance variation was significant and remained significant. To assess significance, we first analyzed protein variability in steady conditions, by looking at the distribution of the pre-MMS data points (18 data points, 6 hours). We could find a relationship between the standard deviation and the mean, calculated for the full set of 4085 movies. Thus defined the intra-experiment error, ν, with the following equation (see figure S16d+f):

$$\nu(\widetilde{y}_{ref}) = 0.054 \cdot \widetilde{y}_{ref} + 1.98 \tag{17}$$

To asses the significance of a change in the signal, in regard to the intra-experiment error, we used the p-value, $p(\nu)$, defined as follows:

$$\begin{aligned} p_\nu &= 1 - \Phi(y_t, \widetilde{y}_{pre}, \nu(\widetilde{y}_{pre})), \text{ if } y_t > \widetilde{y}_{pre} \\ p_\nu &= \Phi(y_t, \widetilde{y}_{pre}, \nu(\widetilde{y}_{pre})), \text{ if } y_t \leq \widetilde{y}_{pre} \end{aligned} \tag{18}$$

where $\Phi(y, \mu, \sigma^2)$ is the cumulative distribution function of a Gaussian with mean μ and a standard deviation of σ^2. Finally, time to induction was defined as the first significant data points ($p_\nu < 1e^{-4}$), for which all the following data points where also significant.
Finally, because protein accumulation was well fitted by a linear function, we defined the rate as the slope of a linear fit to the 9 data points following induction (3 hours).

1.11 Classification of dynamical subcellular localization patterns

We aimed at robustly detecting changes of cellular location of proteins following MMS exposure.

1.11.1 Feature extraction

For each cell, we extracted a small rectangular image, surrounding the cell contour. Background was defined as the 5^{th} percentile of the non-cell pixels and was subtracted from the image. Contrast was increased by stretching the image in the full 8-bits range (0-255), between a minimal value, defined as the 5^{th} percentile of the cell pixels, and the maximal pixel value. To capture the texture of the fluorescent signal, we then calculated a set of 97 features (Table S3). We first defined a set of 17 histogram-based features and 3 geometrical features, which we complemented with 7 morphological features [20, 21], 10 granulometry measures [22], and 60 threshold adjacency statistics (TAS) [23].

1.11.2 Supervised Classifier

Our experimental design implies a certain compromise between high-throughput temporal imaging and the level of detail of subcellular localization analysis. Thus, we would like to distinguish between very fine localization patterns as usually defined in cell biology and more objective geometrical shapes. Exploratory analysis indicated that we would be able to robustly distinguish six shapes. As we show below, our probabilistic classification scheme based on assigning probabilities of belonging to one shape to each segmented cell gives us satisfactory spatial resolution both in terms of comparing with previous data [3] and detecting changes in localization. To train our classifier, we built a training set by manually annotating cells extracted from 104 images as representatives of one of the following patterns. Examples can be found in Figure S31A:

- **Periphery:** A fine outline of the cell contour, generally very well distinguishable. Representative strains include membrane proteins uniformly associated with the cell membrane, or in some cases bright dots distributed on the membrane.
- **Structure:** This shape includes filaments, circles and shape-forming dots that are often a direct indication for organelle-related localization of the protein.
- **Punctate:** A number of distinct small dots of sizes smaller than 1 μm ($< 20\%$ of the size of a cell). Typical representatives are *actin*, *lipid particles* and *peroxisomes*.
- **Disk:** One dominant area of GFP signal contained in the interior of the cell. The diameter of these objects is at least around 25% of the diameter of a cell. Typical representatives of this group are strains localized in the *nucleus* and *nucleolus*, but also proteins in the *vacuole* or *vacuolar membrane*.
- **Corona:** Broad ring (donut) around the center that can also be more sickle-shaped. Typical localizations that have a corona-like appearance are *cytoplasm* and in some cases *ER*.

- **Homogeneous:** Cells where the fluorescence is uniformly distributed. In many cases homogeneous cells are of low intensity reflecting background levels.

As the boundaries between these shapes can be fuzzy, we chose to assign to each cell a probability vector reflecting the likelihood to belong to each of the six possibilities. To build such a soft classifier, we first use Reduced-Rank Linear Discriminant analysis (RRLDA) [24] to compress the dimensionality of the feature space. This method is similar to Principal Component Analysis (PCA), but considers the separation into classes as additional information. Instead of considering the directions in parameter space with the largest variance in the data, RRLDA maximizes the between-class variance relative to the within-class variance. We use RRLDA together with our training set and receive a matrix that reduces the dimensions of our feature set from 97 dimensions to five. We verified that increasing the number of features does not improve performance. To assign probabilities to each cell, we use the MATLAB function 'classify', specifying a 'quadratic' discriminant function, which fits multivariate normal densities with a separate covariance estimate for each shape. In addition to the probability vector, the function 'classify' gives us for each cell an estimate of the probability density of the feature set of this cell. This is useful for discarding cells that are atypical (*e.g.*, dead cells) or have ambiguous fluorescence patterns.

1.11.3 Similarity of two collections of cells

To obtain an objective measure on how two populations of cells compare in their localizations, we modeled each population (image) as a Dirichlet distribution (DD), which is a natural choice to model a population of probability vectors. We then exploited the property that the Bhattacharyya distance can be easily computed from the parameters of the DD. A straighfroward extension would be to use for each population a mixture of DDs, but no simple formula exists to compute the Bhattacharyya distance in that case. However, we tested that the Bhattacharyya distance computed numerically for three component mixture models actually correlates well ($R^2 > 0.8$) with that of the simple DDs. Thus we adopted the latter for the automatic analysis of localization changes.

1.11.4 Validation and performance of the classifier and the distance measurement

Cross validation. We first assessed the performance of the classifier using 10-fold cross-validation. Left-out data are assigned to just one, the most probable, shape and the procedure is repeated ten times. The confusion matrix (Figure S33A) indicates that, expectedly, most misclassifications happen between groups with the most fuzzy boundaries. Note that assigning hard classes removes information and performance is thus expected to be lower in this assessment.

Comparison of automatic and manual annotation. As another test, we compared manual against automatic annotation. For this, we randomly picked 200 images for which more than

60% of the cells were classified to belong to the same group. Those images then were independently and blindly annotated by ND and JB. For the 182 images where the two manual annotations agreed, there were only two cases of disagreement with the automatic annotation (Figure S33B). In both cases, the difference between automatic and manual annotation can be traced back to a subjective estimation as part of the manual annotation.

Agreement of replicated recordings. A total of 2741 strains, for which we had double coverage, were picked to assess reproducibility. We selected only those 1034 strains with high intensity to avoid the problem that two randomly selected strains that contain only background noise are as well very similar. Those selected strains have an average Bhattacharyya distance of 0.11. In comparison, if we select two of these strains at random, their distance is significantly higher (Figure S32).

Comparison between Bhattacharyya distance and manual annotation To assure that the Bhattacharyya distance produces meaningful values, we calculated the Bhattacharyya distance for the strains before the MMS treatment and at the end of the experiment. JB and ND then independently and blindly annotated for 110 randomly picked images, if there was a change between the last two images before the MMS treatment and the last two images of the experiment. For a Bhattacharyya distance of 0.3 or bigger 75.6% of annotations where changes, while for a Bhattacharyya distance smaller than 0.3, only 25% of strains where annotated as changes. Details can be found in Figure S32D. Together with the finding that more than 90% of replicate experiments have a Bhattacharyya distance smaller than 0.2, we can say with high certainty that our experiment is reliably reproducible.

1.11.5 Comparison with the original yeast GFP library annotations

To test the agreement between the annotations made for the yeast GFP clone collection (UCSF data set [3]) and our assignments, we compared an automatic classification for those strains that had a single subcellular location category in the UCSF data set annotation, similar to previous evaluations of the UCSF data set [25, 26]. We chose those strains with only one annotated subcellular localization in the UCSF data set to better characterize the conversion between our shapes and subcellular components. As many of our strains had low intensity levels, we concentrated on those subcellular components where we had enough high intensity strains. For these strains, there is a reasonable relation between the UCSF annotation and our patterns as Figure S33 shows. Most subcellular localizations showed distinguishable average probability vectors. As our experiments lack the additional information of DAPI staining and we had several groups with only few high intensity representatives, we grouped subcellular localizations into those groups that where in general distinguishable by eye: cytoplasm, nucleus, nucleolus, cell periphery, ER, subcellular periphery, filaments and punctate particle. *Subcellular periphery* contains vacuole, vacuolar membrane and nuclear periphery. *Filament* contains microtubule, mitochondrion, actin and golgi. *Punctate particle* contains endosomes, spindle pole, bud neck,

peroxisomes and lipid particles. Comparing the average probability vectors of each of this group showed a distinct probability density in the six dimensional space. We added an intensity filter to discard strains with low protein expression, for which we could not confidently extract information. We identified two key parameters for the filtering. Firstly, we observed that manual classification of cells was not possible for a mean Top5 pixel intensity below approximately 130 a.u.. Secondly, the homogeneous class represented cells without an observable pattern. Therefore, we discarded all strains that had more than 50% of cells below the Top5 pixel intensity threshold or classified as homogeneous. We used 5-fold cross-validation to separate our data into test and training sets. We did not use a common algorithm like nearest neighbor for our classification, to avoid favoritism of groups with a high number of strains (*e.g.*, cytoplasm). Instead, we used a three component Dirichlet Mixture Model to calculate a density function in probability space for each group in our training set. This allowed us to calculate group membership probabilities for each average probability vector of the test set. We then assigned each strain of the test set to the group for which its average probability vector had the highest group membership probability. Varying the thresholds to discard strains (Top5 pixel intensity, cell probability of pattern homogeneous), we were able to correctly classify between 77% (less stringent thresholds: 989 strains in total) and 90% (more stringent threshold: 637 strains in total) of strains. In the case of the 637 examined strains (Figure S34), around 30% of the misclassified strains showed a localization-unrelated stress pattern. For five strains there is a disagreement between our observation and the UCSF annotation, which could be either a result of cross-contamination or of slightly different growth conditions.

1.11.6 Identification of localization change

Each movie in the first repeat experiment in MMS_{high} was systematically annotated for localization changes (see section 1.9.4), resulting in a first set of candidate proteins. We refined this set by performing another round of manual annotation and kept only the changes that were clearly identified by both annotators. We grouped the proteins in 5 major relocation classes: (i) transition between nucleus and cytoplasm, (ii) punctate aggregation within the nucleus, (iii) focal aggregation of proteins inside the nuclear periphery, (iv) focal aggregation or disaggregation of proteins in the cytoplasm and (v) transition of proteins between the cell membrane and the cell interior. To minimize the number of missed hits, we also used the Bhattacharyya distance to identify strains showing localization probability traces that were similar to the changing strains. This allowed us to find 8 additional strains. Thus, using both manual and quantitative analysis, we were able to robustly select a set of 119 proteins that we observed to relocate in response to MMS treatment. In addition, we found 2 strains, where the change was questionable in MMS, but clearly visible in Sorbitol.

1.11.7 Temporal analysis of localization changes

To further quantify the localization changes and analyze their rate and timing, we focused on the geometrical pattern, which is the most relevant to a given transition (*i.e.*, showing the clearest change). We found that these are (i) disk for transitions between nucleus and cytoplasm, (ii) punctate for proteins that aggregate and (iii) periphery for everything transiting from or towards the cell membrane. For each of those cases we fitted a logistic function to approximate the average probability of the respective relevant pattern, P_t, by minimizing the error between L_t and P_t. L_t is given by:

$$L_t = \frac{\alpha}{1 + e^{(-\lambda(t-\delta))}} + \tilde{P}_{Pre} \tag{19}$$

where $\tilde{P}_{Pre}$ is the median value of P_t for the last two hours before the stress stimulus. α, δ and λ are estimations for the rate, the timing and the slope of a localization change. If the localization change was transient, we fitted the logistic function only for the time until the change reached its peak. Otherwise, we included the time points during the first hour after the peak for robustness. Pulse experiments were automatically calculated as transient. In addition, we manually changed the analysis for those strains with obvious transiency (e.g. Stb3p and Nmd3p in Sorbitol (Figure S39)).

To filter strains for which the change was not quantified robustly, we set a minimal threshold to the score T_P , given by the following equation:

$$T_P = \frac{L_{max} - L_{min}}{\sqrt{P_\sigma^2 + L_{max}}} \tag{20}$$

T_P takes into account the variance of the average probability of the pattern before treatment, P_σ, and the maximum and minimum values for the logistic fit, L_{max} and L_{min}. This way, we were able to sort out strains for which the automatic analysis could not detect a change with certainty. In addition, we required that strains with at least double coverage showed a change in more than one repeat. For the second screen, we discarded strains with single coverage. The threshold of 0.12 was determined empirically. The requirement that all 119 strains in the first experiment show a localization change was used to optimize sensitivity. As a result, 111 of 119 strains (93.3%) passed this filter. Strains that we manually detected as not changing in the second set of experiments were used to maximize specificity. Of all 97 strains that were automatically detected during the four additional experiments of the second screen, we had to discard only 3 strains (3.1%) as misclassification errors. Thus the choice of our threshold was adequate.

To make different stresses comparable, we normalized the data of our relevant pattern by a

modified z-score, Z_t, that is closely related to T_P :

$$Z_t = \mathrm{sgn}_{man} \frac{P_t - \tilde{P}_{Pre}}{\sqrt{P_\sigma^2 + max(L_{max}, P_t)}} \tag{21}$$

sgn_{man} sets the directionality of the function after our annotated localization change (*e.g.*, a protein moving from the nucleus to the cytoplasm has a positive change, although the disk probability decreases). The observed localization changes in MMS_{high} are given in figure 4.e. Figures S37, S38, S39 and S40 summarize the localization changes found in the four additional stress conditions.

1.12 Computing Hardware and Software

All images were stored and processed on an IBM server with 5 compute nodes and 50 TB of disk space. Each compute node consisted of an IBM HS21 XM server with two 2.83GHz Xeon QC E5440 processors and 32 GB of memory. An IBM DS3200 and three EXP3000 storage expansions housing 24 hard drives of 1 TB capacity and 24, 2 TB hard drives for a total of 50TB available storage space. The server ran SUSE Linux Enterprise 10.3, OpenPBS 2.3.16, ImageJ 64 bit version 1.42, Java 1.6.0, MATLAB R2009a, mySQL 5.0.26, Perl 5.8.8, and TWiki 5.0. We wrote custom scripts to automatically submit MATLAB or ImageJ jobs to the OpenPBS queue manager which allowed us to compute up to 40 positions in parallel, each job running on one of the 40 cores of the server. To high loads on the system we generally used 7 CPUs on each of the nodes and 4-6 CPUs on the master node for a total of 32-34 CPUs. All custom written scripts, ImageJ plugins and macros, and MATLAB programs can be downloaded from http://cellbase.epfl.ch.

2 Supplementary references

[1] Whitesides, G. M., Ostuni, E., Takayama, S., Jiang, X. & Ingber, D. E. Soft lithography in biology and biochemistry. *Annu Rev Biomed Eng* **3**, 335–373 (2001).

[2] Thorsen, T., Maerkl, S. J. & Quake, S. R. Microfluidic large-scale integration. *Science* **298**, 580–584 (2002).

[3] Huh, W. K. *et al.* Global analysis of protein localization in budding yeast. *Nature* **425**, 686–691 (2003).

[4] Tong, A. H. Y. & Boone, C. Synthetic genetic array analysis in Saccharomyces cerevisiae. *Methods Mol Biol* **313**, 171–192 (2006).

[5] Welford, B. P. Note on a Method for Calculating Corrected Sums of Squares and Products. *Technometrics* **4**, 419–420 (1962).

[6] Soille, P. *Morphological Image Analysis: Principles and Applications* (Springer-Verlag, 2003).

[7] Kass, M., Witkin, A. & Terzopoulos, D. Snakes - Active contour models. *Int J Comput Vision* **1**, 321–331 (1987).

[8] Delgado-Gonzalo, R. & Unser, M. Spline-based framework for interactive segmentation in biomedical imaging. *Ing Rech Biom / BioMed Eng Res* **34**, 235–243 (2013).

[9] Thévenaz, P., Delgado-Gonzalo, R. & Unser, M. The ovuscule. *IEEE Trans Pattern Anal Mach Intell* **33**, 382–393 (2011).

[10] Delgado-Gonzalo, R., Chenouard, N. & Unser, M. A new hybrid Bayesian-variational particle filter with application to mitotic cell tracking. *Proc IEEE Int Symp Biomed Imaging* 1917–1920 (2011).

[11] Press, W. H., Teukolsky, S. A., Vetterling, W. T. & Flannery, B. P. *Numerical Recipes: The Art of Scientific Computing* (Cambridge University Press, 1986).

[12] Delgado-Gonzalo, R., Thévenaz, P., Seelamantula, C. S. & Unser, M. Snakes with an ellipse-reproducing property. *IEEE Trans Image Process* **21**, 1258–1271 (2012).

[13] Delgado-Gonzalo, R., Thévenaz, P. & Unser, M. Exponential splines and minimal-support bases for curve representation. *Comput Aided Geom D* **29**, 109–128 (2012).

[14] Ridler, T. W. & Calvard, S. Picture thresholding using an iterative selection method. *IEEE Trans Syst Man Cybern Syst* **8**, 630–632 (1978).

[15] Pavlidis, T. Algorithms for graphics and image processing (1982).

[16] Taniguchi, Y. *et al.* Quantifying E. coli proteome and transcriptome with single-molecule sensitivity in single cells. *Science* **329**, 533–538 (2010).

[17] Friedman, N., Cai, L. & Xie, X. S. Linking stochastic dynamics to population distribution: an analytical framework of gene expression. *Phys Rev Lett* **97**, 168302 (2006).

[18] Rey, G. *et al.* Genome-wide and phase-specific DNA-binding rhythms of BMAL1 control circadian output functions in mouse liver. *PLoS Biol* **9**, e1000595 (2011).

[19] Fisher, R. A. Tests of significance in harmonic analysis. *Proceedings of the Royal Society of London. Series A* **125**, 54–59 (1929).

[20] Boland, M. V. M. & Murphy, R. F. R. A neural network classifier capable of recognizing the patterns of all major subcellular structures in fluorescence microscope images of HeLa cells. *Bioinformatics* **17**, 1213–1223 (2001).

[21] Huang, K. & Murphy, R. Automated classification of subcellular patterns in multicell images without segmentation into single cells. *Proc IEEE Int Symp Biomed Imaging* 1139–1142 Vol. 2 (2004).

[22] Walter, T. *et al.* Automatic identification and clustering of chromosome phenotypes in a genome wide RNAi screen by time-lapse imaging. *Journal of Structural Biology* **170**, 1–9 (2010).

[23] Hamilton, N. A., Pantelic, R. S., Hanson, K. & Teasdale, R. D. Fast automated cell phenotype image classification. *BMC Bioinformatics* **8**, 110 (2007).

[24] Hastie, T., Tibshirani, R. & Friedman, J. The elements of statistical learning: data mining, inference, and prediction (2001).

[25] Chen, S.-C., Zhao, T., Gordon, G. J. & Murphy, R. F. Automated image analysis of protein localization in budding yeast. *Bioinformatics* **23**, i66–i71 (2007).

[26] Huh, S., Lee, D. & Murphy, R. F. Efficient framework for automated classification of subcellular patterns in budding yeast. *Cytometry A* **75**, 934–940 (2009).

[27] Picotti, P., Bodenmiller, B., Mueller, L. N., Domon, B. & Aebersold, R. Full dynamic range proteome analysis of S. cerevisiae by targeted proteomics. *Cell* **138**, 795–806 (2009).

[28] Ghaemmaghami, S. *et al.* Global analysis of protein expression in yeast. *Nature* **425**, 737–741 (2003).

[29] Newman, J. R. *et al.* Single-cell proteomic analysis of S. cerevisiae reveals the architecture of biological noise. *Nature* **441**, 840–846 (2006).

[30] Tkach, J. M. J. *et al.* Dissecting DNA damage response pathways by analysing protein localization and abundance changes during DNA replication stress. *Nature Publishing Group* **14**, 966–976 (2012).

[31] Gasch, A. P. *et al.* Genomic expression responses to DNA-damaging agents and the regulatory role of the yeast ATR homolog Mec1p. *Mol Biol Cell* **12**, 2987–3003 (2001).

3 Supplementary tables

Coating						
#	Layer	Resist	Speed	Time	Baking Temp.	Baking time
1	sieve channels	SU8 - GM1040	3000 rpm	40 s	80°C	10 min
2	chambers	SU8 - GM1050	2250 rpm	40 s	80°C	10 min
3	flow	AZ9260	1500 rpm	40 s	115°C	6 min
4	control lines	SU8 - GM1060	1500 rpm	40 s	100°C	10 min

Exposure						
#	Layer	Power	Time	Dose	Baking Temp.	Baking time
1	sieve channels	8 mW/cm^2	5 s	40 mJ/cm2	80°C	10 min
2	chambers	8 mW/cm^2	8 s	64 mJ/cm2	100°C	10 min
3	flow	10 mW/cm^2	5x20 s	1000 mJ/cm2	none	none
4	control lines	8 mW/cm^2	2x11 s	176 mJ/cm2	100°C	10 min

Development						
#	Layer	Developer	Dev. Time	Rinser	Rinsing time	Final height
1	sieve channels	none	none	none	none	1.7 µm
2	chambers	PGMEA	2x5 min	2-propanol	1 min	3.3 µm
3	flow	AZ400 1:3.5	2x115 s	DI water	30 s	15 µm
4	control lines	PGMEA	2x7 min	2-propanol	1 min	12 µm

Table S1: **Summary of the materials and processes used for the microfabrication of the molds.**

	exp. #	# strains (with number of replicates)	stress conditions	timing: pre - imaging	Pre - stress	stress #1 (recovery)	stress #2 (recovery)	stress #3 (recovery)	data size: # frames / movie	# images (phase contrast and epi)	# movies: Total	High quality #	High quality %	# cells: Total # of cells used for analysis	Average / frame
primary screen	1	1056 + 3x control A	0.03% MMS	26h	6h	7h			40	92160	1152	795	69.0	3.1E+06	97
	2	1056 + 3x control A	0.03% MMS	16h	6h	7h			40	92160	1152	943	81.9	5.3E+06	140
	3	1056 + 3x control A	0.03% MMS	24h	6h	7h			40	92160	1152	1013	87.9	8.3E+06	204
	4	1023 + 3x control A	0.03% MMS	18h	6h	7h			40	92160	1152	681	59.1	3.7E+06	134
	5	1023 + 3x control A	0.03% MMS	18h	6h	7h			40	92160	1152	925	80.3	8.3E+06	225
	6	1056 + 4x control B	0.03% MMS	24h	6h	7h			40	92160	1152	736	63.9	5.5E+06	185
	7	1056 + 4x control B	0.03% MMS	21h	6h	7h			40	92160	1152	964	83.7	7.5E+06	195
	8	1056 + 4x control B	0.03% MMS	19h	6h	7h			40	92160	1152	889	77.2	6.6E+06	185
	9	3x 299	0.03% MMS	21h	6h	7h			40	92160	1152	580	50.3	3.6E+06	156
										8.3E+05 802 MB	10368	7526	72.6	5.2E+07	169
secondary screen	10	2x 576	0.03% MMS	20h	6h	7h			40	92160	1152	1045	90.7	9.7E+06	233
	11	2x 576	0.0075% MMS	20h	6h	7h			40	92160	1152	899	78.0	6.4E+06	177
	12	2x 576	0.0075% MMS	16h	6h	7h			40	92160	1152	687	59.6	7.8E+06	284
	15	2x 576	1M Sorbitol	18h	6h	7h			40	92160	1152	848	73.6	5.6E+06	164
	16	2x 576	1M Sorbitol	22h	6h	7h			40	92160	1152	758	65.8	6.9E+06	229
	17	2x 576	UV steps	20h	6h	10 min (6h)	30min (6h)	1h (6h)	75	172800	1152	721	62.6	1.1E+07	211
	18	2x 576	UV steps	24h	6h	10 min (6h)	30min (6h)	1h (6h)	75	172800	1152	759	65.9	1.2E+07	209
	19	2x 576	MMS steps	27h	6h	20 min (6h)	40 min (6h)	1h20 (6h)	75	172800	1152	751	65.2	1.2E+07	217
	20	2x 576	MMS steps	24h	6h	20 min (6h)	40 min (6h)	1h20 (6h)	75	172800	1152	771	66.9	1.0E+07	177
	21	2x 576	0.2M HU	24h	1h	7h			24	55296	1152	1077	93.5	7.2E+06	280
	22	2x 576	0.2M HU	24h	1h	7h			24	55296	1152	889	77.2	5.9E+06	276
										1.3E+06 1221 MB	12672	9205	72.6	9.5E+07	223.4
								Total Size of raw data		2.1E+06 1.98 TB	23040	16731	72.6	1.5E+08	199

Control A: ACE2, SWI4, NDD1, FKH2, MCM1, CDC10, YAP5, SFP1, FHL1, Hog1, SKO1, GPD1, STB5, YAP1, PDR1, CAD1, MSN1, MSN2, CRZ1, RPN4, PHO2, GCN4, MSS1, INO4, HAC1, HSF1, MRC1, RAD9, RAD53, ASF1, RNR4, MAG1

Control B: MCM1, HOG1, GPD1, CRZ1, CDC10, RNR4, WTM1, YAP1, RPN4, PHO2, ACE2, ENO1, ENO2, YNL134C, HDR1, CLB2, DCS2, UBI4, PEX27, SIS1, HSC82, DHH1, WT, empty well

Table S2: **List of on chip experiments.**

Table S3: **List of features.**

#	FEATURE TAG	DESCRIPTION
Histogram-based features		
1	top5vs20	mean(highest 5 pixels) / mean(highest 20 pixels)
2	top5vs50	mean(highest 5 pixels) / mean(highest 50 pixels)
3	top20vs50	mean(highest 20 pixels) / mean(highest 50 pixels)
4	top5vsMed	mean(highest 5 pixels) / median
5	top20vsMed	mean(highest 20 pixels) / median
6	top50vsMed	mean(highest 50 pixels) / median
7	histo1ratio	frequency of highest pixel bin (pixel values 240-255) / frequency of bottom half (pixel values 0-127)
8	histo2ratio	frequency of 2nd highest pixel bin (pixel values 224-239) / frequency of bottom half (pixel values 0-127)
9	histo3ratio	frequency of 3rd highest pixel bin (pixel values 208-223) / frequency of bottom half (pixel values 0-127)
10	histoHLratio	frequency of top half (pixel values 128-255) / frequency of bottom half (pixel values 0-127)
11	bin1vs2	93.75th percentile / 87.5th percentile
12	bin1vs3	93.75th percentile / 81.25th percentile
13	bin2vs3	87.5th percentile / 81.25th percentile
14	bin1vsMed	93.75th percentile / median
15	bin2vsMed	87.5th percentile / median
16	bin3vsMed	81.25th percentile / median
17	binHLratio	Upper quartile / lower quartile
Spatial distribution features		
18	central_signal	mean($P(x \leq x_i, y \leq y_i)$) / total mean, ($x_i$ = cos(t)·3·cell width$/12$, y_i = sin(t)·3·cell height$/12$)
19	middle signal	mean($P(x_i < x \leq x_m, y_i < y \leq y_m)$) / total mean, ($x_m$ = cos(t)·5·cell width$/12$, y_m = sin(t)·3·cell height$/12$)
20	boundary_signal	mean($P(x_m < x \leq x_b, y_m < y \leq y_b)$) / total mean, ($x_b$ = cos(t)·7·cell width$/12$, y_b = sin(t)·7·cell height$/12$)
Morphological features [20, 21]		
21	convex_hull _overlap	SLF 1.14, Convex hull area / cell area (with binary threshold at 0.5·P_{max})
22	convex_hull _roundness	SLF 1.15, The roundness of the convex hull (with binary threshold at 0.5·P_{max})

23	edges_fraction	SLF 1.9, The fraction of the nonzero pixels that are along an edge (with binary threshold at 0.5·P_{max})
24	edges_homogeneity	SLF1.10, Measure of edge gradient intensity homogeneity
25	edges_direction _homogeneity1	SLF1.11, Measure of edge direction homogeneity 1
26	edges_direction _homogeneity2	SLF1.12, Measure of edge direction homogeneity 2
27	edges_direction _difference	SLF1.13, Measure of edge direction difference

Granulometries [22]

28	gray_open_1	mean intensity of $(I - IO^{d1})$, IO^{dr} = grayscale opening of image I with disk of radius r
29	gray_open_2	mean intensity of $(IO^{d1} - IO^{d2})$
30	gray_open_4	mean intensity of $(IO^{d2} - IO^{d4})$
31	gray_open_7	mean intensity of $(IO^{d4} - IO^{d7})$
32	gray_open_12	mean intensity of $(IO^{d7} - IO^{d12})$
33	gray_close_1	mean intensity of $(I - IC^{d1})$, IC^{dr} = grayscale closing of image I with disk of radius r
34	gray_close_2	mean intensity of $(IC^{d1} - IC^{d2})$
35	gray_close_4	mean intensity of $(IC^{d2} - IC^{d4})$
36	gray_close_7	mean intensity of $(IC^{d4} - IC^{d7})$
37	gray_close_12	mean intensity of $(IC^{d7} - IC^{d12})$

Threashold adjacencies statistics (TAS) [23]

38-46	tas_T35_pk	Threshold at 0.35·P_{max}, pixel count with k neighbor above threshold / pixel count above threshold
47	tas_T35_binRatio	Threshold at 0.35·P_{max}, pixel count above threshold / total pixel count
48-56	tas_T50_pk	Threshold at 0.5·P_{max}, pixel count with k neighbor above threshold / pixel count above threshold
57	tas_T50_binRatio	Threshold at 0.5·P_{max}, pixel count above threshold / total pixel count
58-67	tas_T65_pk	Threshold at 0.65·P_{max}, pixel count with k neighbor above threshold / pixel count above threshold
67	tas_T65_binRatio	Threshold at 0.65·P_{max}, pixel count above threshold / total pixel count

Threashold adjacencies statistics (TAS) - inverted image

68-97	tas_inv_Txx_pk	Same as 38-67with inverted image, $P_{inv}(x, y)$ = P_{max}-$P(x, y)$

4 Supplementary figures

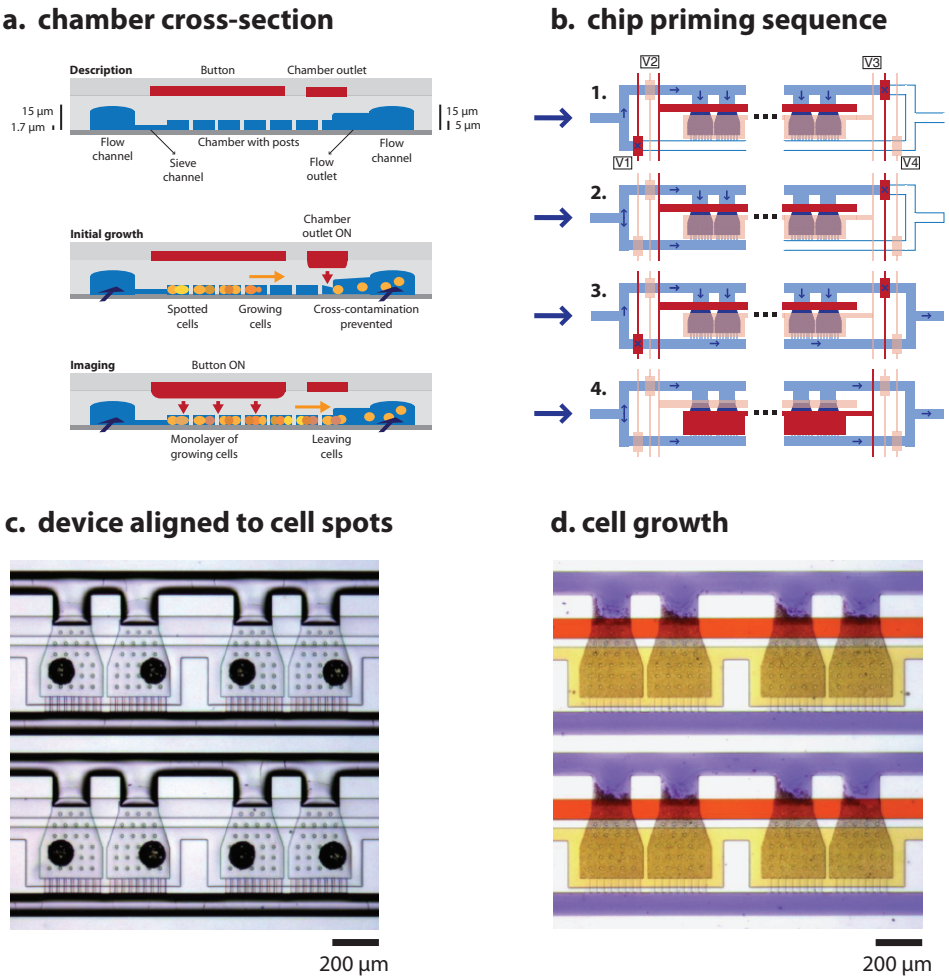

Figure S1: **Additional description of the microfluidic device. a.** Chamber cross-section. Schematic shows the design of the flow layer (in blue) and the control layer (in red). There are three different structures in the flow layer: (i) flow channels (13 μm), (ii) chambers (5 μm), and (iii) sieve channels (1.5 μm). The control layer has two features: (i) a chamber button, which allows to pressurize the chamber roof and thus maintain cells in a monolayer, and (ii) a chamber outlet valve. The middle schematic shows the cross-section during the initial phase of the experiment when cells initiate growth. The chamber outlet is closed to prevent chamber to chamber cross-contamination. The bottom schematic shows a cross-section during imaging. The chamber outlet valve has been released to let the cells exit the chamber while the button is pressurized to prevent the cells from growing in the vertical dimension. **b.** Simplified schematic of the sequence of valve opening/closing, necessary for priming the array with medium. Full flow lines are shown in blue and empty flow lines in white. Arrows show the direction of flow. Pressurized control lines are drawn in dark red and released control lines in light red. Crosses indicate closed valves. **c.** Brightfield micrograph of 8 chambers taken after cell spotting and chip alignment. **d.** Brightfield micrograph of 8 chambers during cell growth. Each chamber is filled with cells. A red and a orange dye was used to stain the chamber outlet and chamber button, respectively. A blue dye was added to the medium to highlight the flow channels.

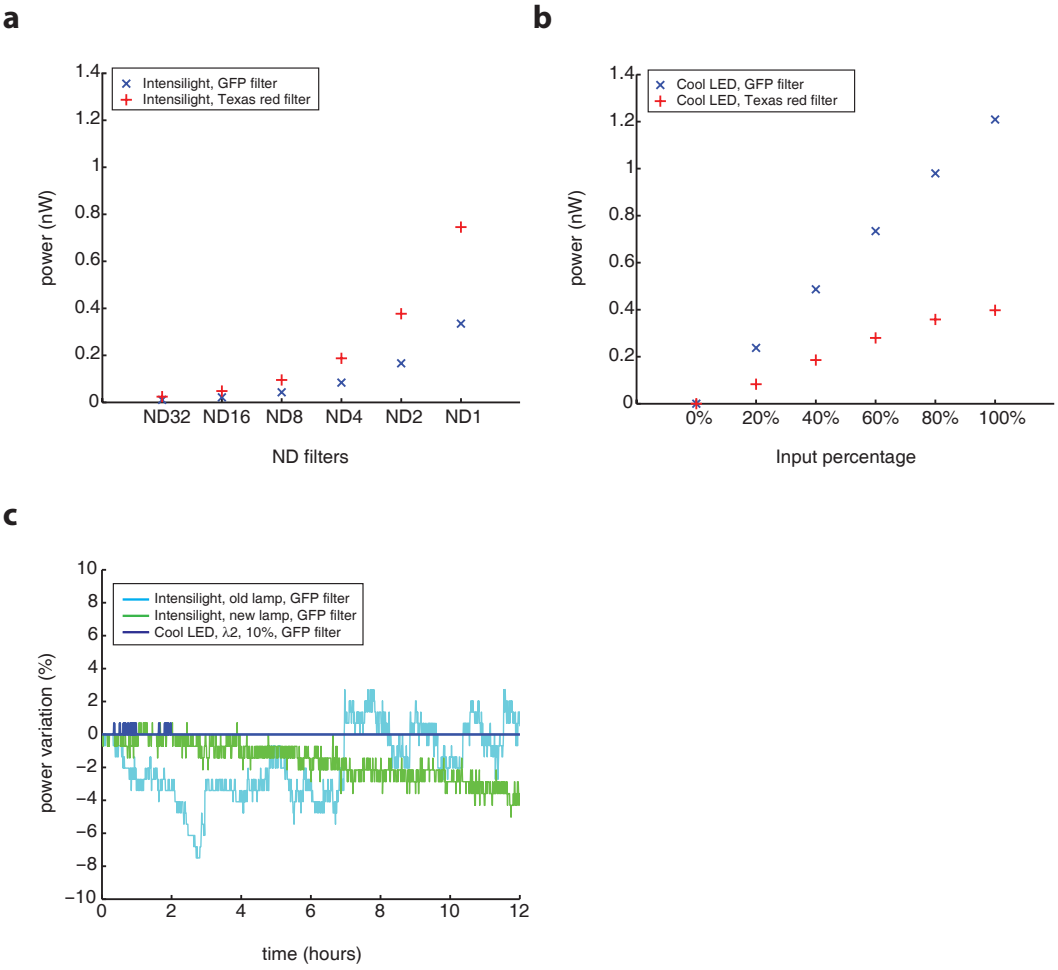

Figure S2: **Lamp calibration and stability test. a.** Calibration curve for the Intensilight showing the lamp power for different neutral density (ND) filters. **b.** Calibration curve for the LED light source showing the power for different input percentage. **c.** Power variation (in percent) of the different light source over time.

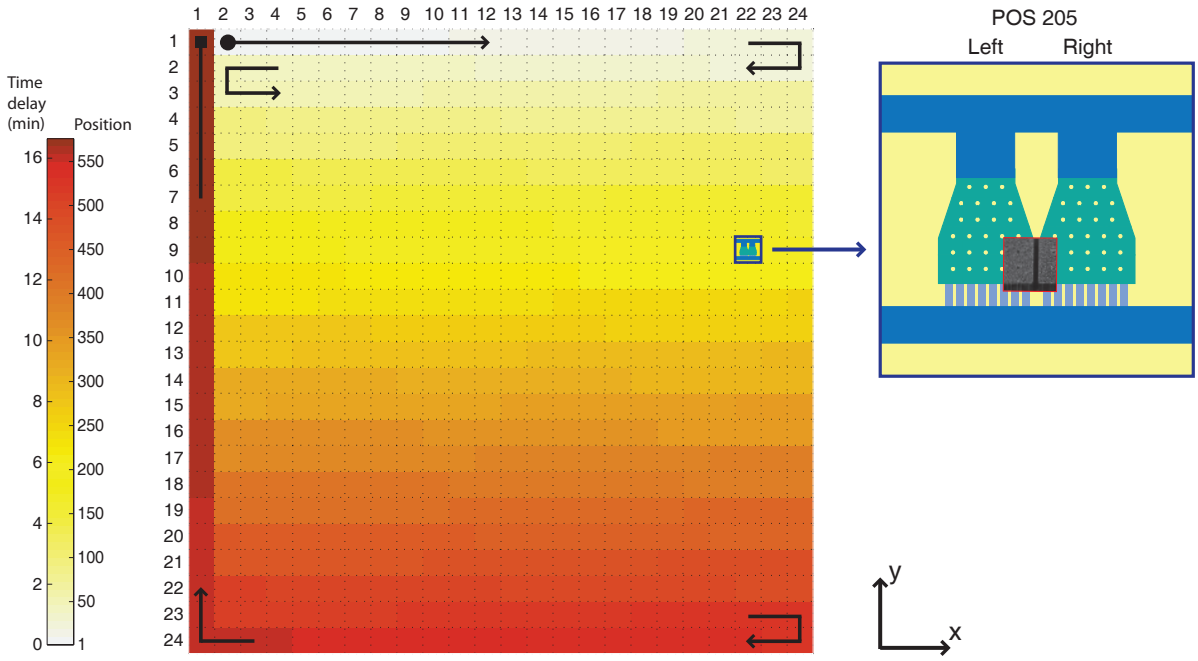

Figure S3: **Microscope path.** Schematic representing the array of positions visited by the microscope during an acquisition iteration. The path, as shown by the arrows and the color gradient, is optimized to reduce the traveling distance of the microscope between positions. The color gradient shows to the time delay from the first position. Each position is located between two chambers, as shown on the inset for position 205.

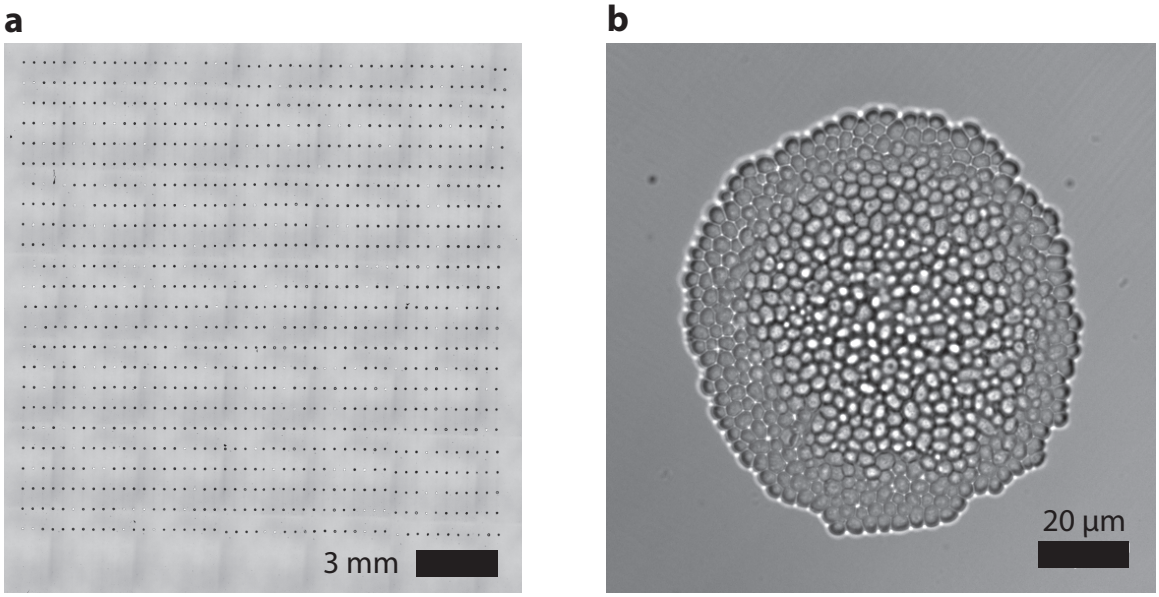

Figure S4: **Cell arraying. a.** Assembly of 7 by 7 images covering the full 24x48 spots of a yeast cell array. **b.** High resolution micrograph of a single cell spot containing hundreds of cells.

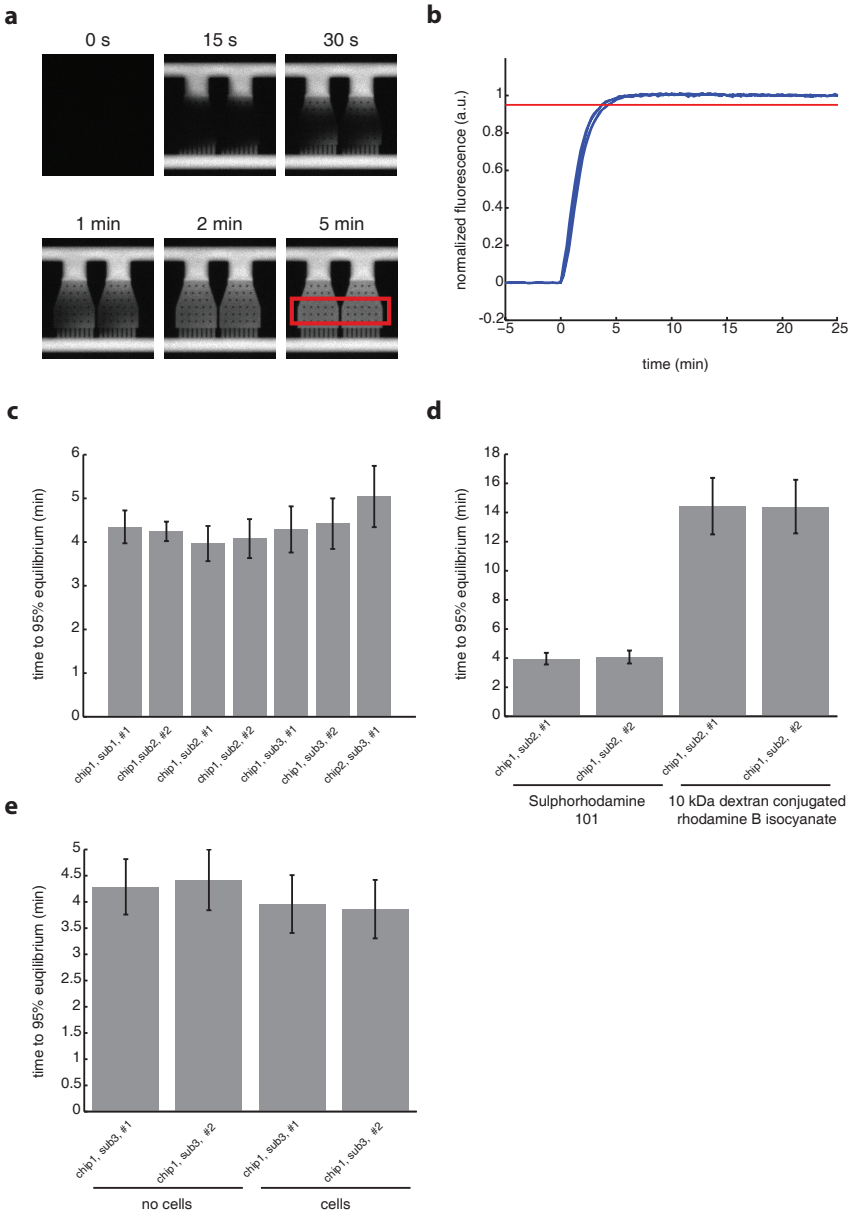

Figure S5: **Characterization of medium switch. a.** Time-series of images showing the diffusion of a fluorescent dye (sulphorhodamine 101), after medium was switched at 0s. The red rectangle defines the area used to calculate the fluorescence intensity. **b.** Normalized fluorescence values over time for 5 pairs of chambers, following a medium switch at t=0s. The red line represents the 95% equilibration threshold, used to define equilibration time. **c.** Time to 95% equilibrium for different set of recordings, from different repeats, different chamber subsets (192 chamber pairs in each subset) and for another chip, recorded on a different day. Equilibration times vary only slightly, ranging between 4 and 5 min. **d.** Comparison of equilibration time between sulphorhodamine and 10kDa dextran conjugated rhodamine B isocyanate. **e.** Equilibration time with or without cells in the chambers.

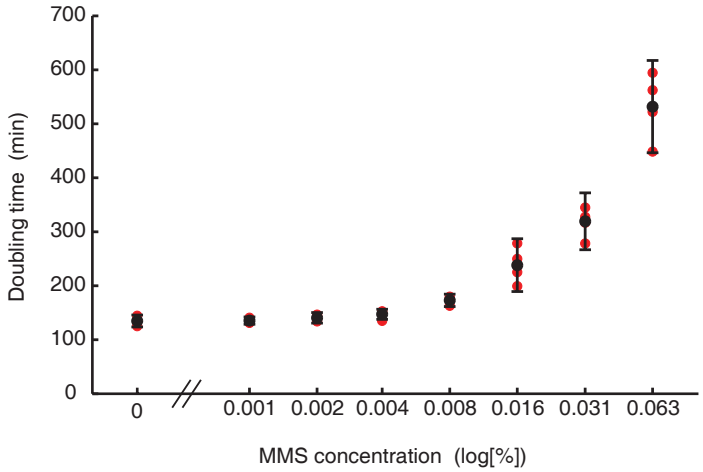

Figure S6: **Effect of MMS concentration on cell growth.** Titration of MMS and its effect on cell growth.

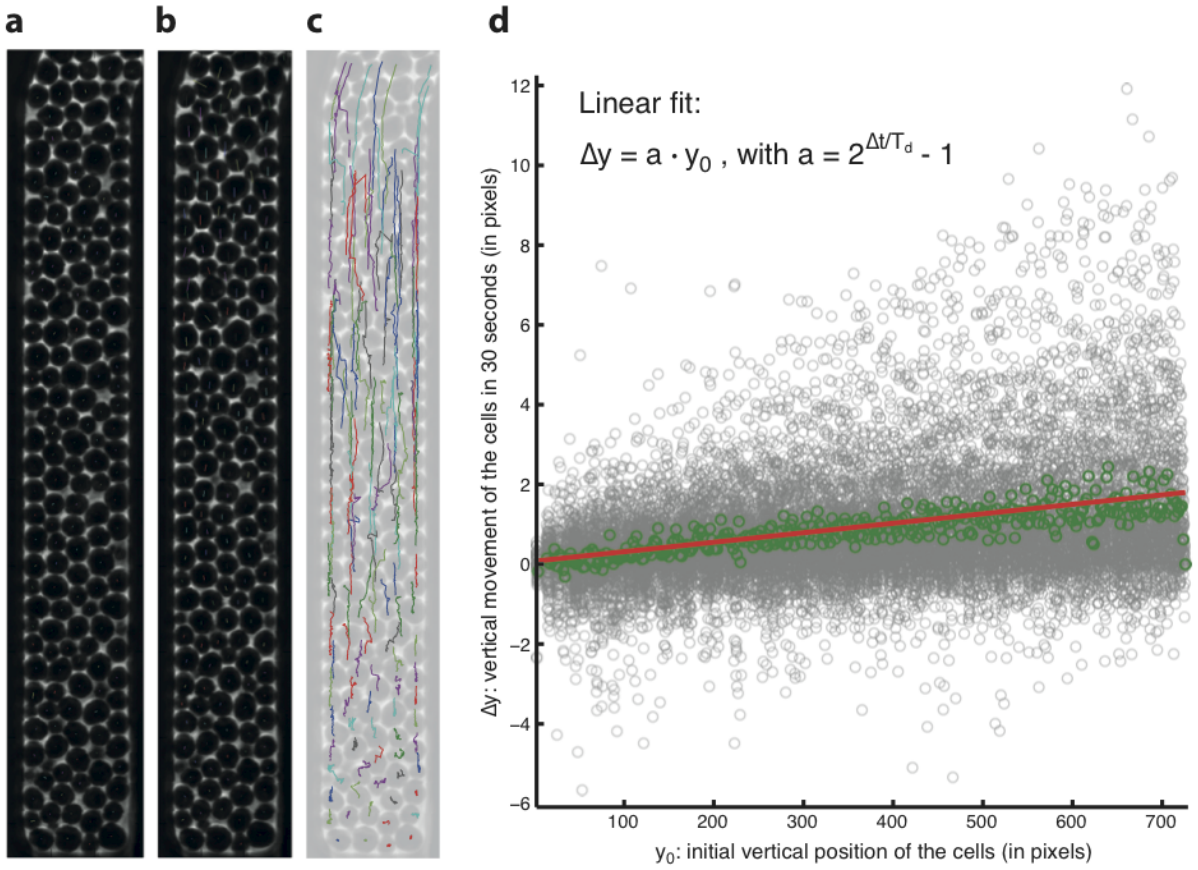

Figure S7: **On-chip doubling time.** By tracking individual cells and their displacement we estimated an average doubling time of 129 minutes, which compares well to bulk doubling time measurements of 120 minutes. Grey data points are individual single cell results (n $\sim$ 15,000). Green date points show the moving average.

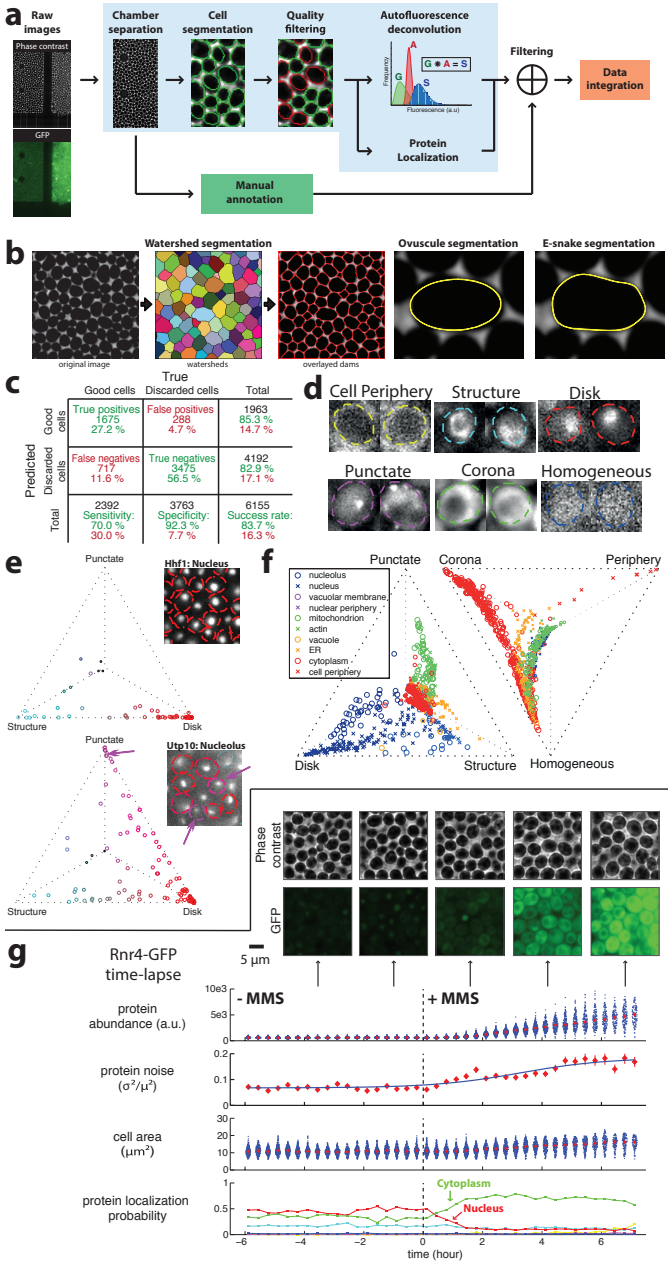

Figure S8: **(a) Image analysis process.** (b) Cell segmentation process. Example of a cell contour determined by an ovuscule (a snake with three nodes) and by an E-snake (with unlimited number of nodes). (c) Summary of cell segmentation quality. (d) The six geometrical patterns used to describe protein localization. (e) A protein localized to the nucleus (Hhf1p) is mainly described by the disk pattern, proteins that were annotated as nucleolus (Utp10) have a punctate component. (f) Distribution of strain averages, annotated with subcellular localizations as defined by Huh et al. [3], showing the relation between geometrical patterns and biological subcellular localizations. (g) Example of a time-lapse sequence for Rnr4-GFP and the data returned by our image analysis pipeline.

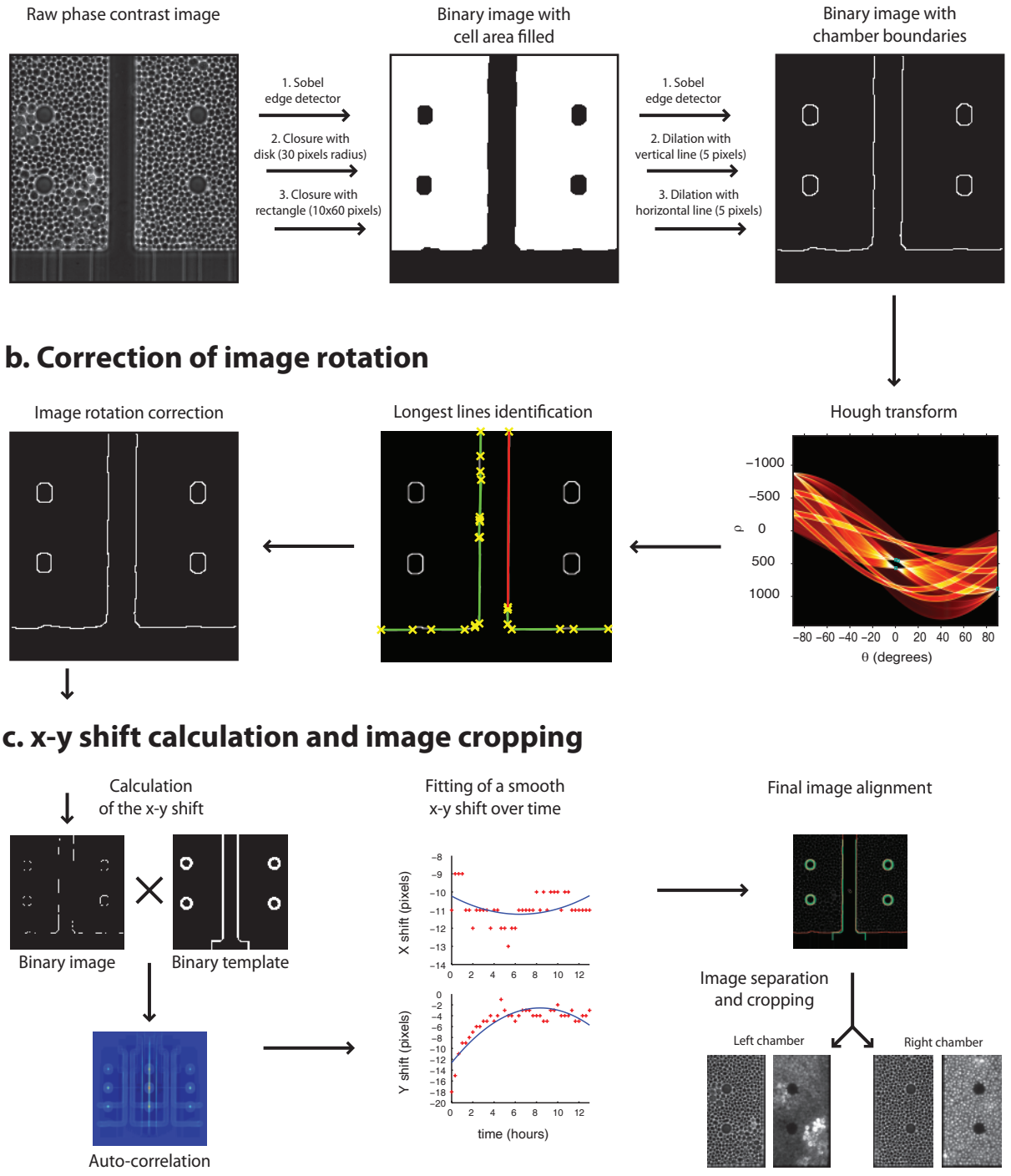

Figure S9: **Illustration of the chamber separation process. a.** Flow chart showing the steps of the chamber identification process. **b.** Illustration of the Hough transform, calculated to determine and correct image rotation. **c.** Alignment of the chamber boundaries to a template to determine the x-y shift and separate the left and right chambers into two distinct set of images.

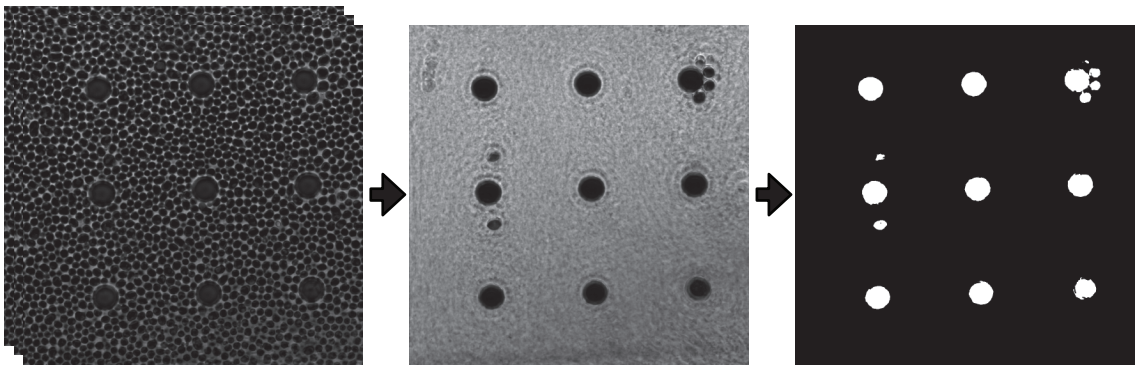

Figure S10: **Cell segmentation - Detection of the non-moving region.** Illustration of the process flow to identify chamber structures, such as post, or other non-moving objects, such as cell carcasses. Non-moving regions have low temporal variation and thus can be identified by thresholding the temporal standard deviation of the image sequence.

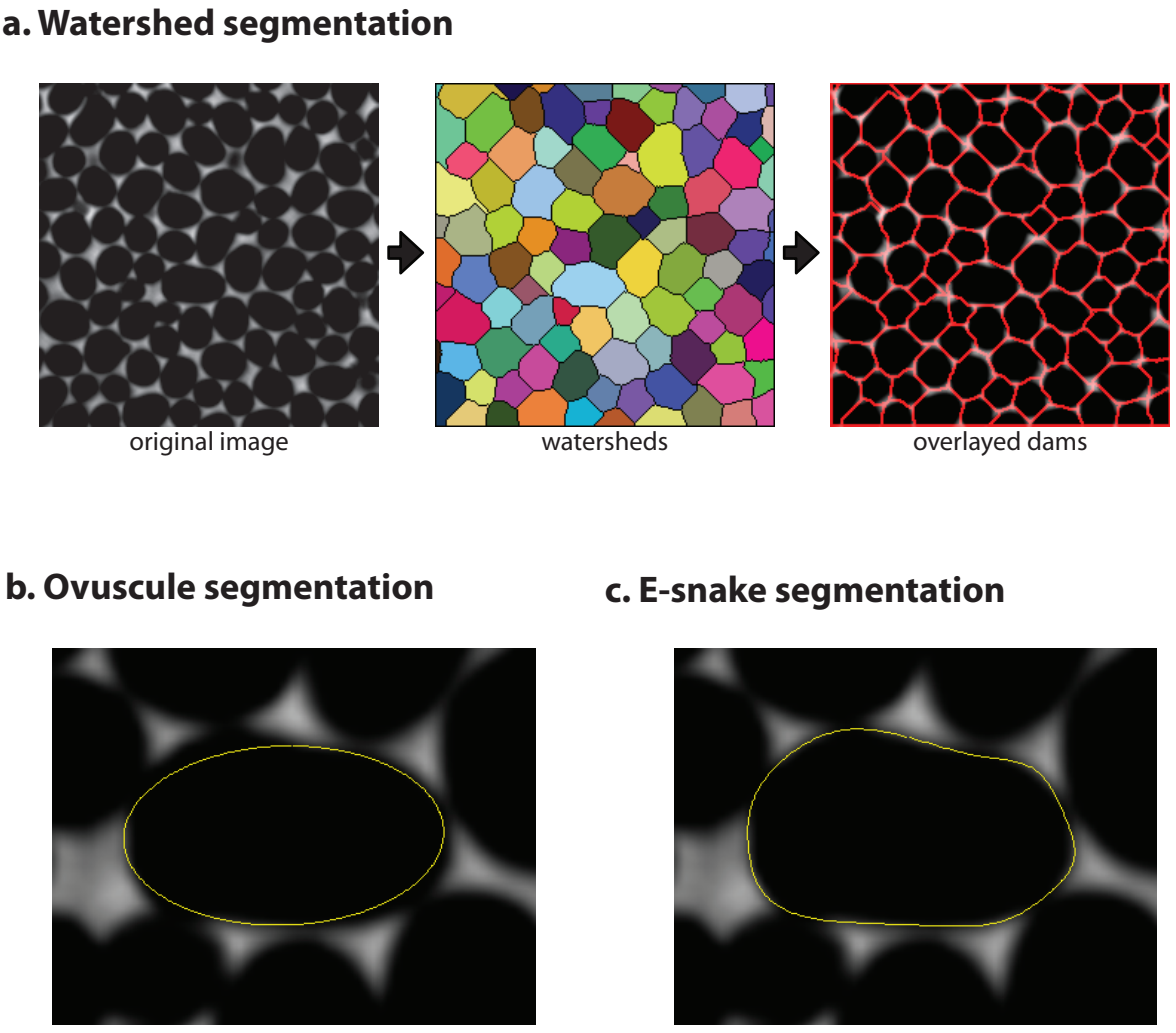

Figure S11: **Cell segmentation - Watershed, ovuscule, E-snake. a.** Process flow of watershed segmentation. **b.** Example of a cell contour determined by an ovuscule (a snake with three nodes) and **c.** by a snake with unlimited number of nodes

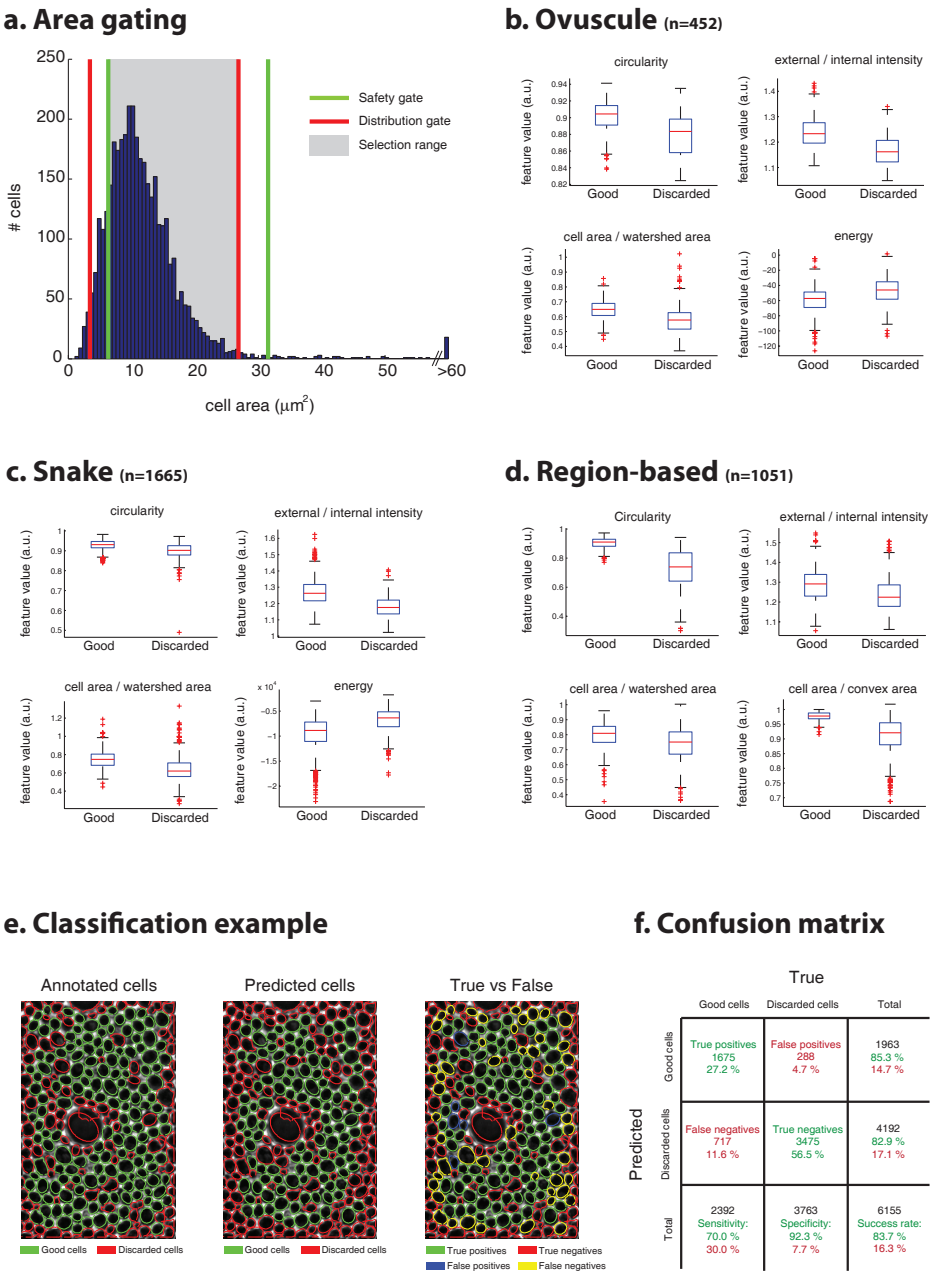

Figure S12: **Cell gating and sorting.** **a.** Histogram illustrating area gating. The selection range is determined by the intersection of the safety gate and the distribution gate. **b-d.** Boxplots showing the distribution of the four features used to discriminate good cells from bad cells for each of the three segmentation methods (ovuscule, snake and region-based). **e.** Training of the classifier illustrated with an example. Predicted classification (central image) is compared to manual annotation (left image) to determine true and false classification (right image). **f.** Overall confusion matrix for the three segmentation methods

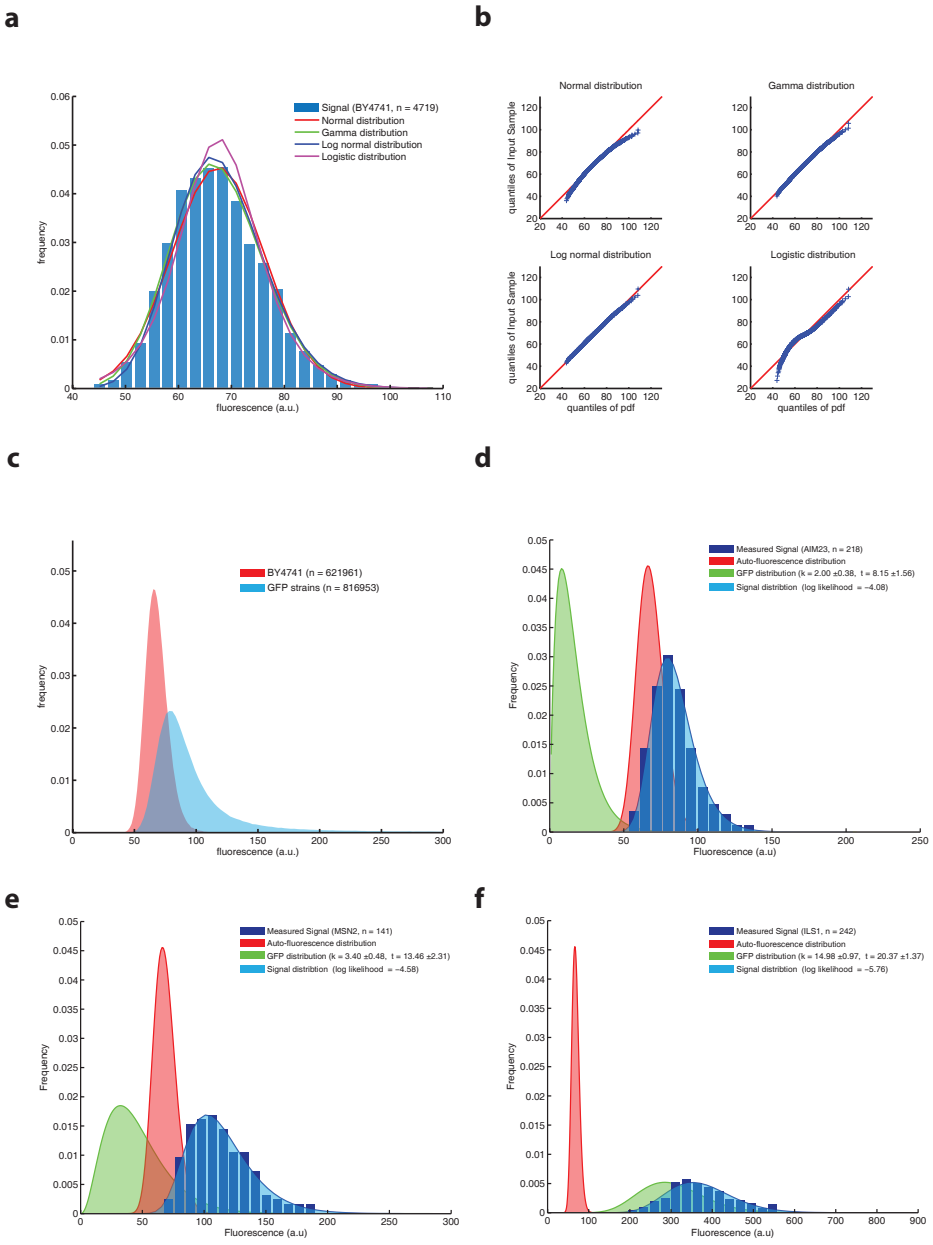

Figure S13: **Auto-fluorescence deconvolution. a.** Distribution of auto-fluorescence values for the wild-type strain *BY4741* and the corresponding fits of four different distribution models. **b.** Quantile-quantile plots for the fits of the four different distribution models shown in **a**. The log normal distribution fits best. **c.** Comparison of single-cell distribution for the wild-type strain *BY4741* (n=621961 cells) and a representative set of cells from the GFP library (n=816'953 cells). **d-f**. Examples of auto-fluorescence deconvolution from three different GFP signals of low (**d**), average (**e**) and high (**f**) intensities. Blue bars represent the distribution of the measured signal. The red curve shows the distribution of the auto-fluorescence. The green curve is the distribution of the GFP contribution. The blue curve represents the convolution of the red and green signal that best fits the measured signal.

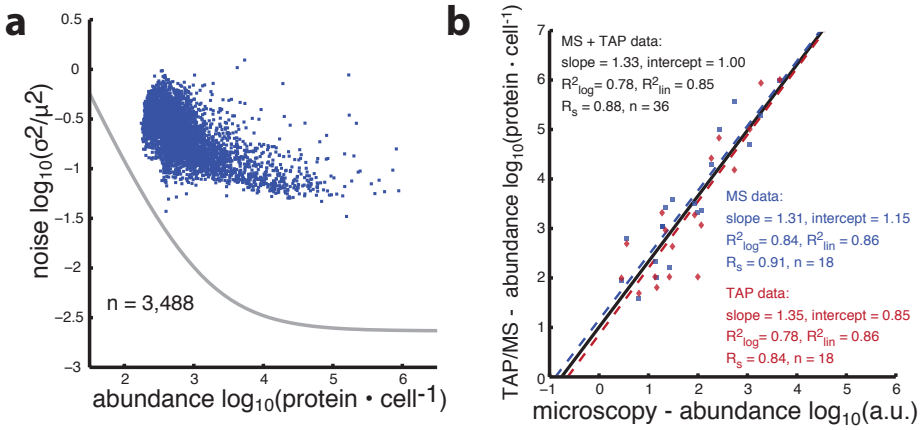

Figure S14: **Protein Noise and Correlation with TAP/MS.**(a) Correspondence between noise and abundance. Instrument noise shown with a gray line. (b) Correlation of our measurements with mass-spectrometry (blue) [27], TAP-western data (red), [28] and both (black).

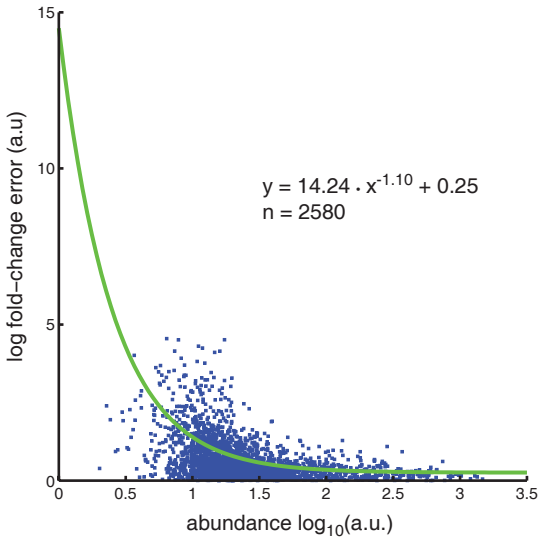

Figure S15: **Fold-change error assessment.** Log fold-change error is shown against abundance for 2580 strains for which two repeats were available (in absolute value). Fold-change is defined as the abundance value after 6 hours of MMS treatment relative to the steady-state baseline. The green line shows a fit through a running standard deviation of the data. Error increases as abundance decreases. This relationship is used to infer the significance of fold-change measurements.

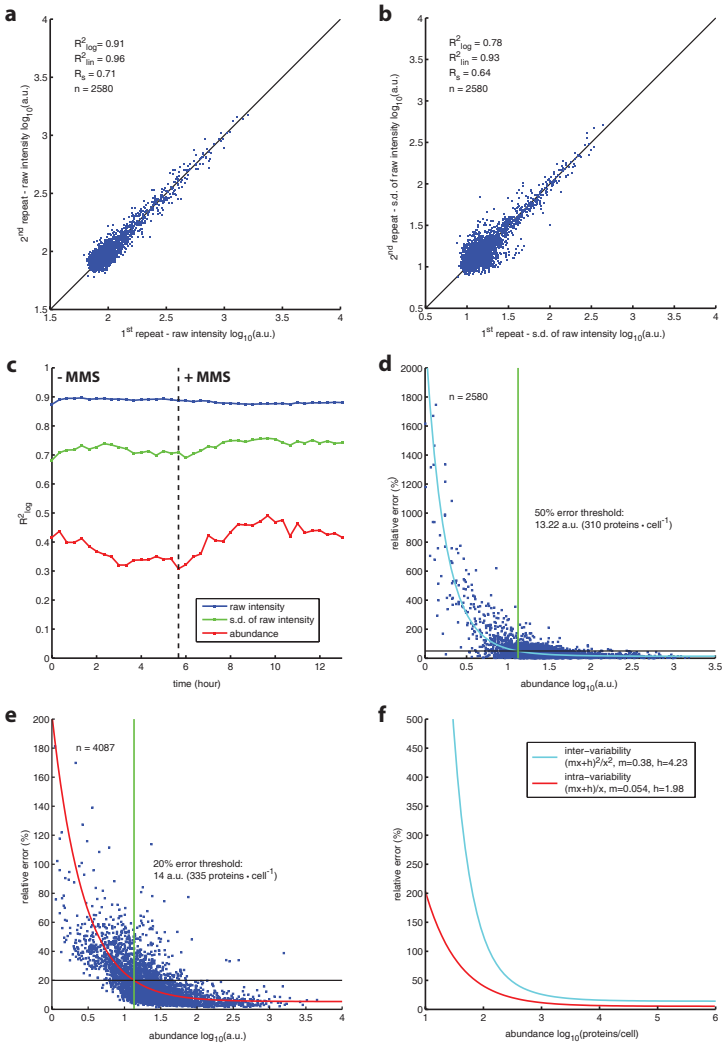

Figure S16: **Repeatability and error estimation.** **a.** Correlation of repeat measurements from two different experiments. Data points show the average intensity of 2580 strains in (SD$_{-his}$). The following correlation coefficient are given: R^2 of the linear data (R^2_{lin}), R^2 for log-transform data (R^2_{log}), Spearman correlation coefficient(R_s). **b.** Same comparison for the standard deviation of single-cell distribution of the same 2580 strains. **c.** R^2_{log} against time at steady-state and during MMS treatment (0.03%) for 2580 repeat data points. Blue line represents the average intensity correlation as shown in **a**. Green line is the standard deviation correlation as shown in **b**. Red line shows the correlation coefficient after auto-fluorescence deconvolution, as shown in the main text. **d.** Relative abundance measurement error against absolute abundance. The blue line shows a fit through the data. Relative error increases as abundance gets smaller. A 50% error threshold is given as indication. **e.** Average abundance variability at steady-state was used to define the measurement error within a time-serie. Relative error is given against absolute abundance. The red line shows a fit through the data. A 20% error threshold is given as indication. **f.** Comparison of the relative error in **d** (inter-variability) and **e** (intra-variability). Protein abundance is given in copies per cell, using the relationship found with other datasets, where absolute protein amount was calculated (see main text figure 2.d)

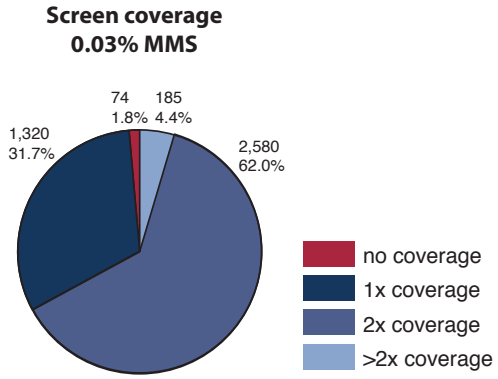

Figure S17: **First screen coverage.** High quality movie coverage of the 4159 strains.

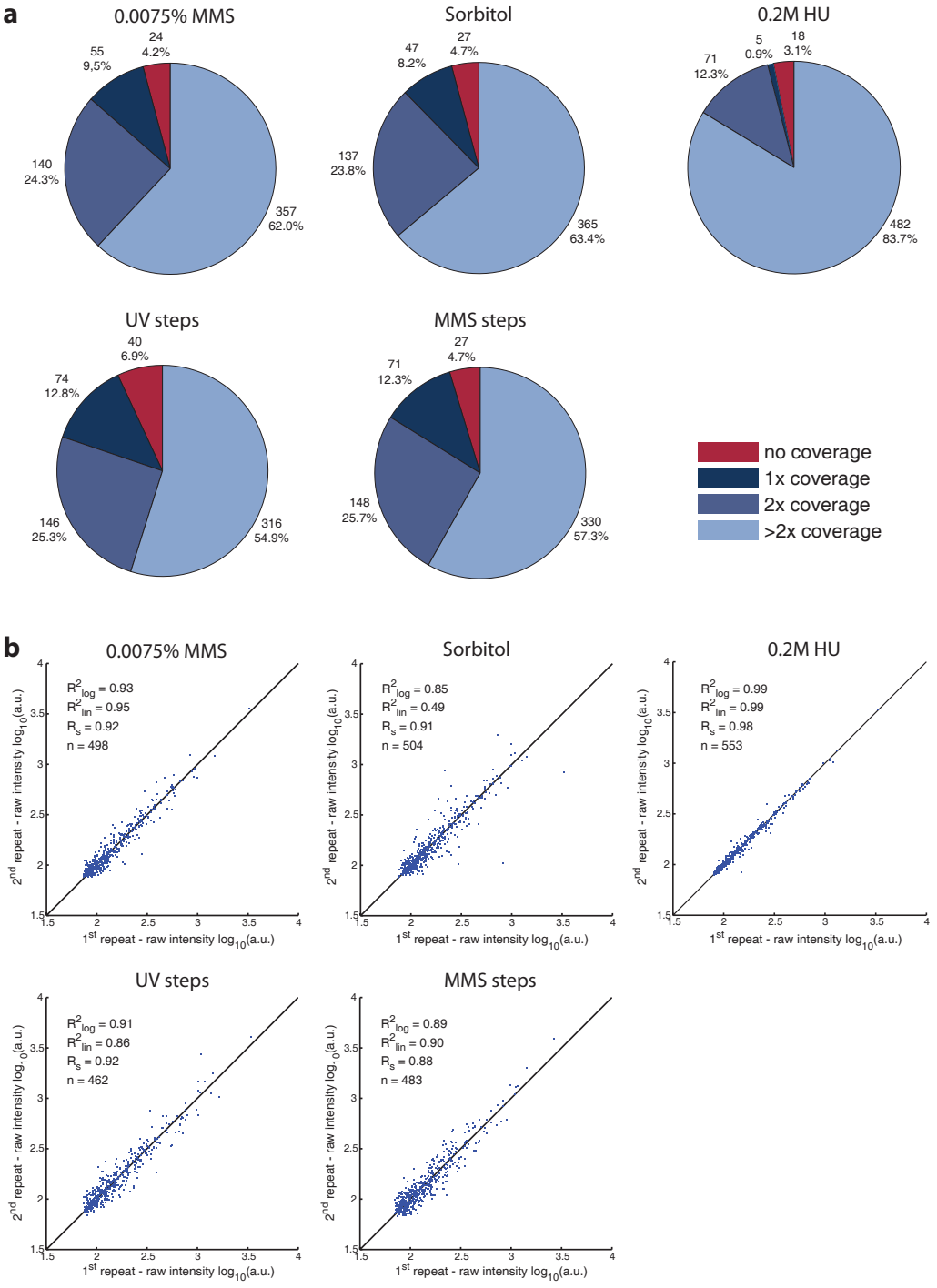

Figure S18: **Second screen coverage and repeatability. a.** High quality movie coverage of the 576 strains present in the second screen is shown for the four different experimental conditions tested (low MMS concentration, osmotic shock with sorbitol, pulses of UV irradiation and pulses of MMS exposure). **b.** Correlation of repeat measurements of average strain intensity for the four experimental conditions.

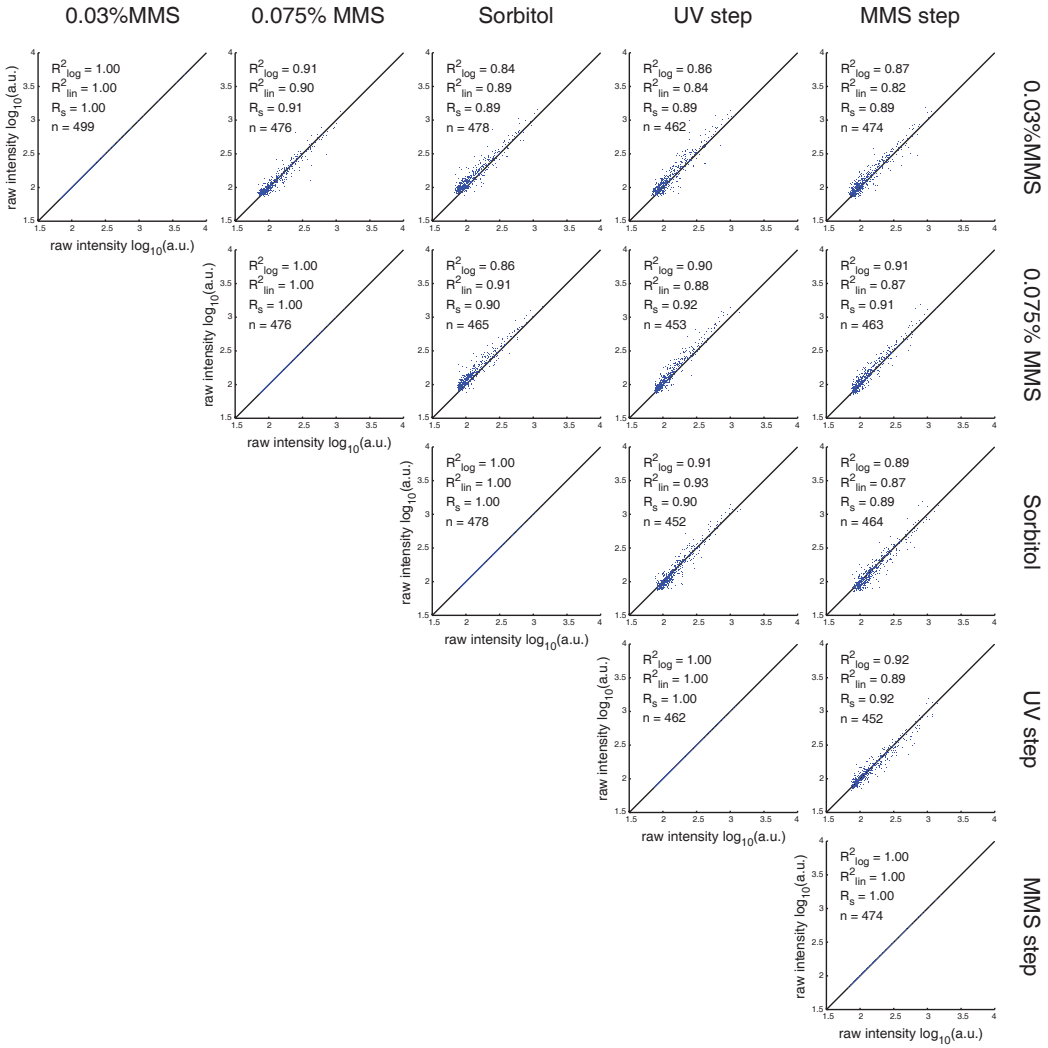

Figure S19: **Second screen experiment to experiment repeatability.** Cross-correlation of average intensity measurements at steady-state for the four different experimental conditions and for the data of the first screen (0.03% MMS).

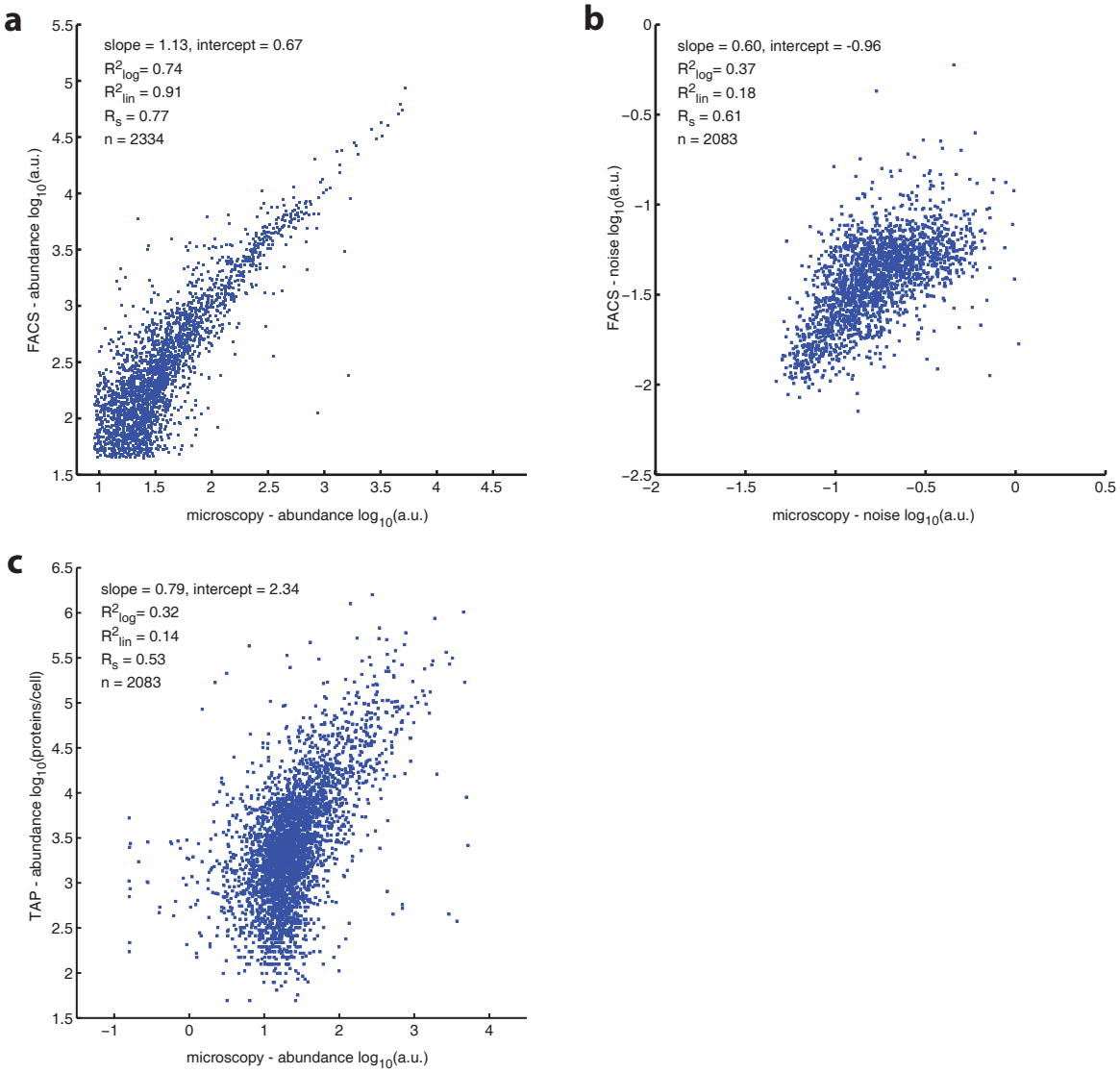

Figure S20: **Correlation of microscopy measurement with FACS and TAP-western measurement. a.** Correlation of abundance values from our data with FACS data [29]. **b.** Correlation of noise values from our data with FACS data. **c.** Correlation of abundance values from our data with TAP-western data [28].

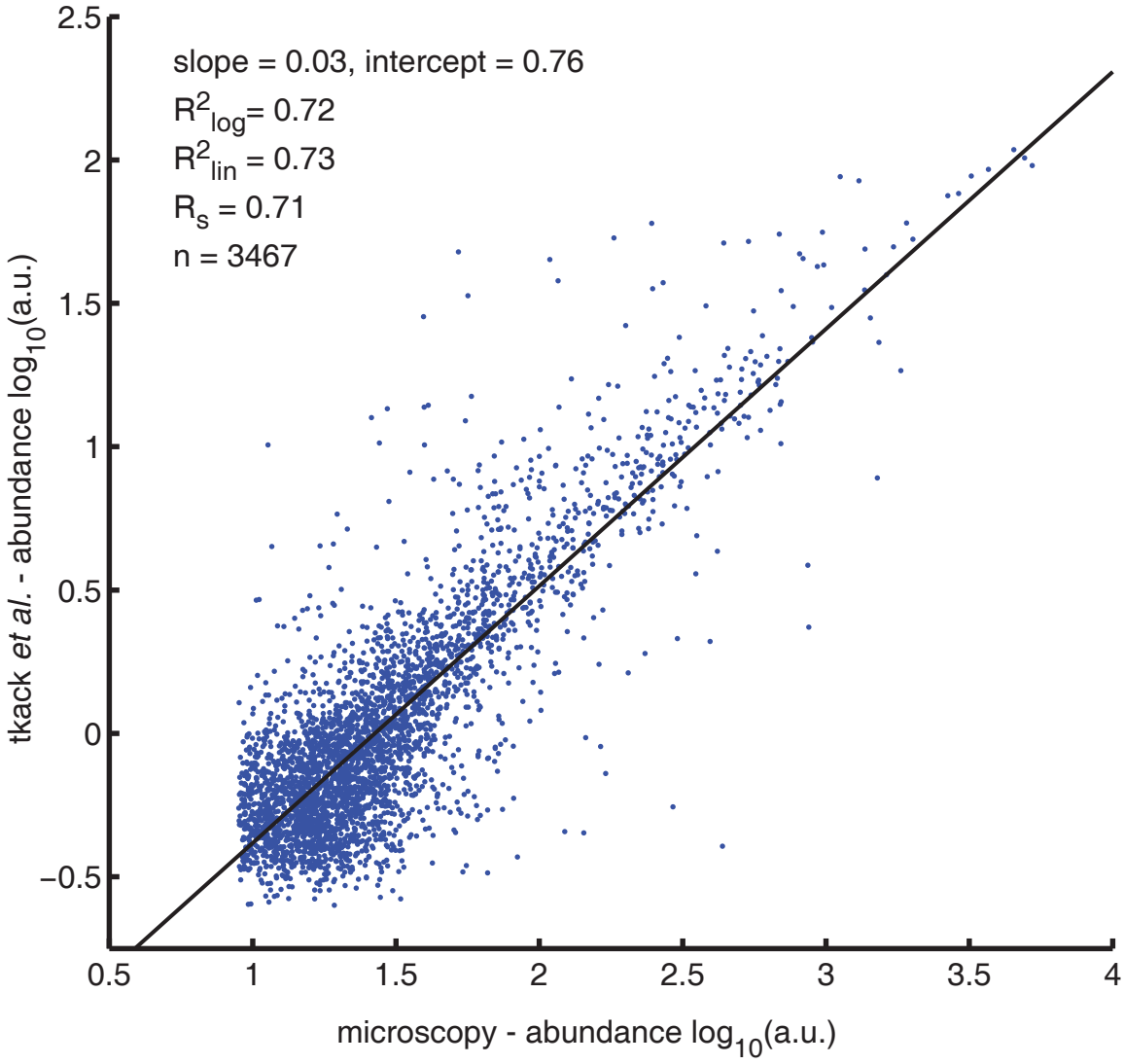

Figure S21: **Correlation of HU abundance values determined by Tkach et al. [30].**

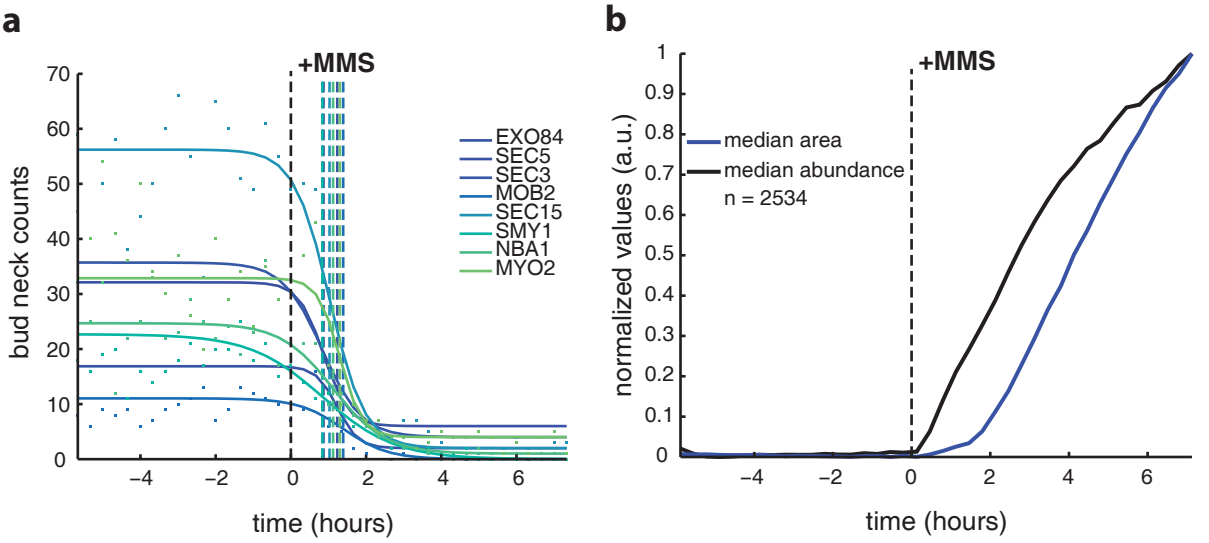

Figure S22: **Assessment of cell arrest following MMS treatment. a.** Assessment of cell division events following MMS treatment. 8 transient bud neck markers were used to count the number of cell division events. The number of cell division events decreases drastically 1 hour after MMS treatment. **b.** Median abundance variation and median area variation is shown for 2534 strains. Abundance increases directly after MMS treatment, whereas cell size increases with a 1.5 hour delay.

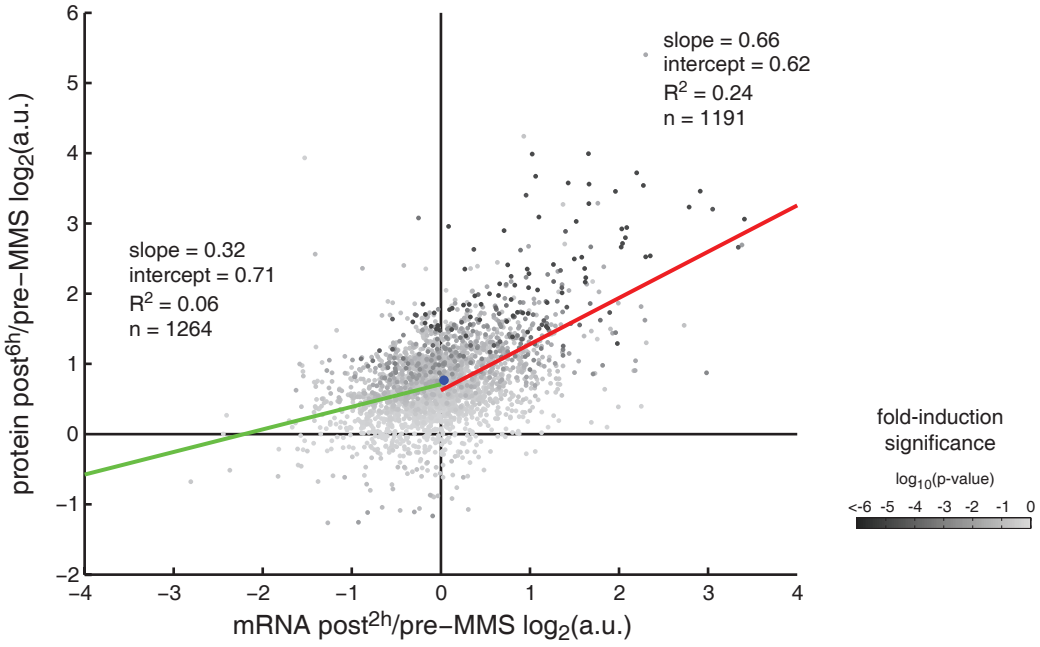

Figure S23: **Correlation of protein fold-change with mRNA fold-change.** 2455 strains that reach abundance level significantly higher than the auto-fluorescence are represented. mRNA data was taken from Gasch *et al.* [31]. The green line shows the correlation for down-regulated genes and the red line for up-regulated genes. The gray scale shows the significance of the protein fold-change.

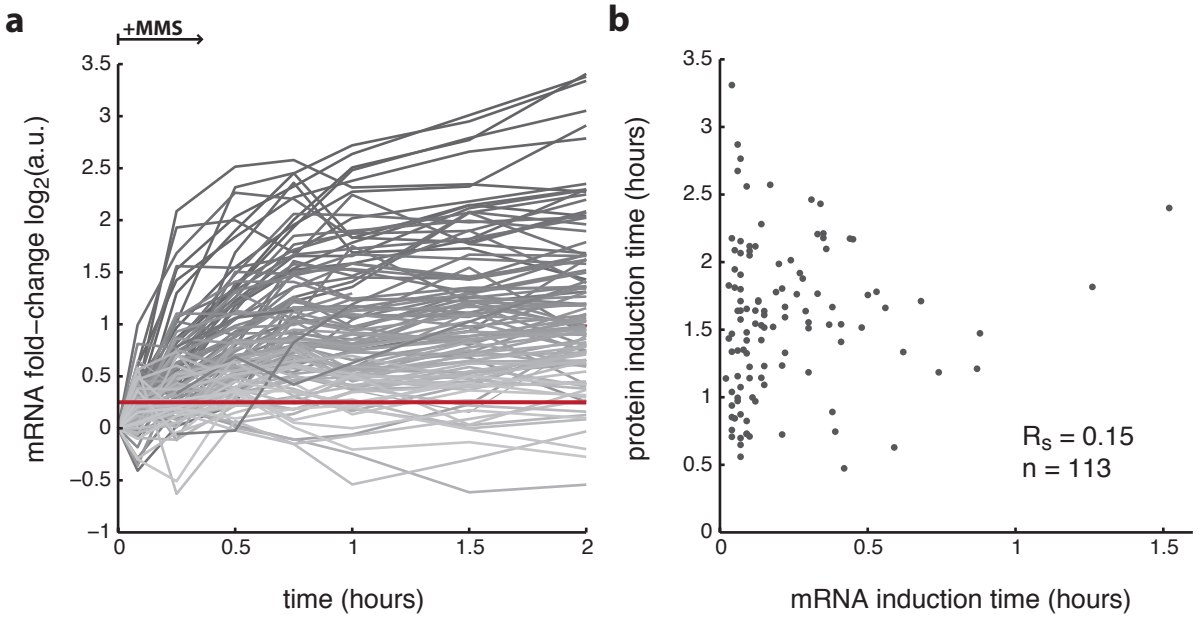

Figure S24: **Dynamics of mRNA levels after MMS treatment.** mRNA data was taken from Gasch *et al.* [31]. Data was available for 113 genes of the 124 for which we observed protein induction. **a.** Time-course of gene expression over 2 hours of MMS treatment. **b.** Comparison of protein and mRNA induction time. mRNA induction times were defined as the first time-point, after data interpolation, when mRNA fold-change is above 0.25.

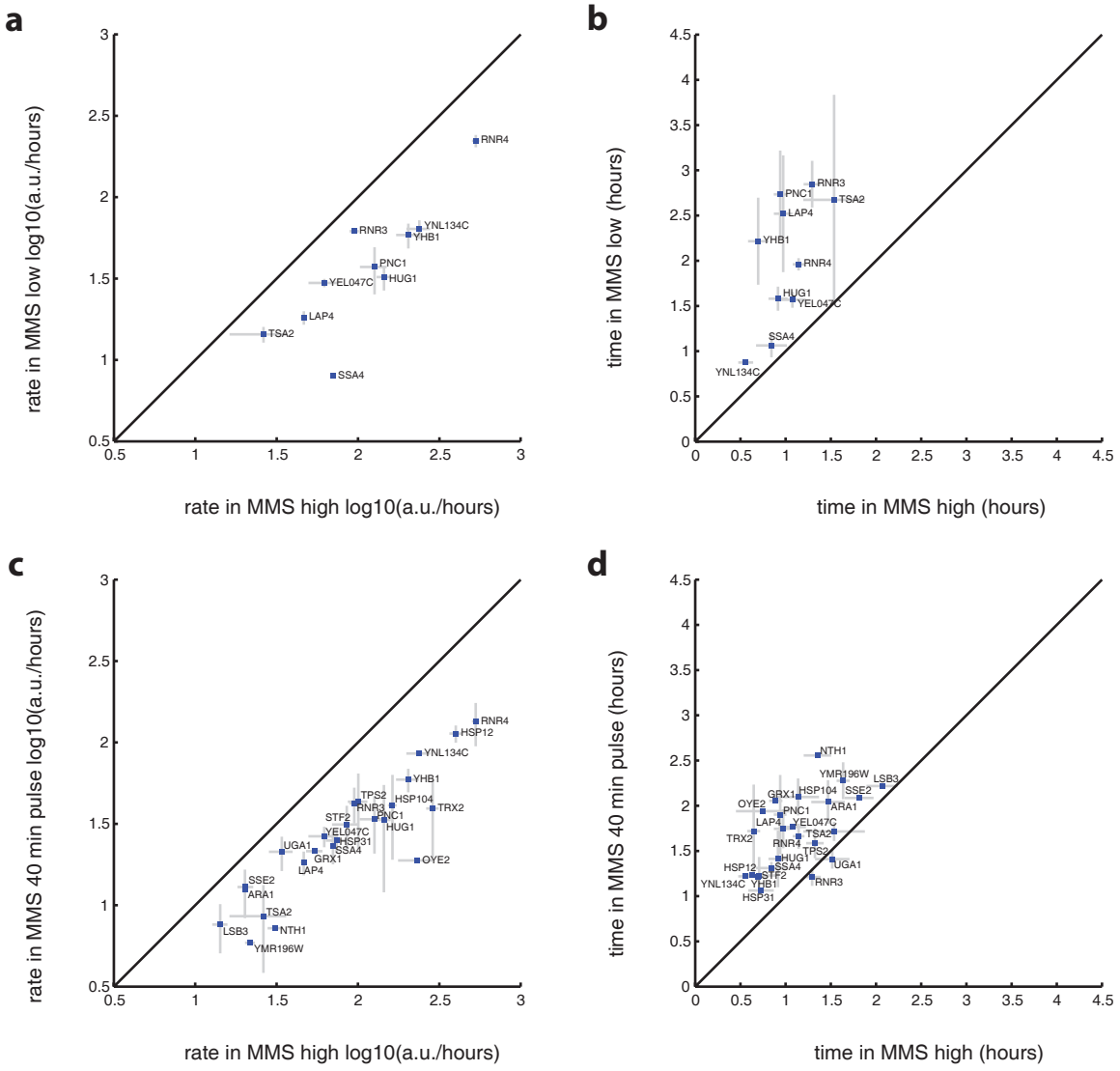

Figure S25: **Rate and timing of protein induction for different MMS conditions.** Correlation of protein induction rate **(a)** and time **(b)** for proteins induced in MMS_{low} and MMS_{high}. Protein induction rate **(c)** and time **(d)** for the second stimulus in MMS (40 min exposure) compare to MMS_{high}. Error bars represent the standard error of the mean (s.e.m). The number of repeats for each data points varies between 2 and 4.

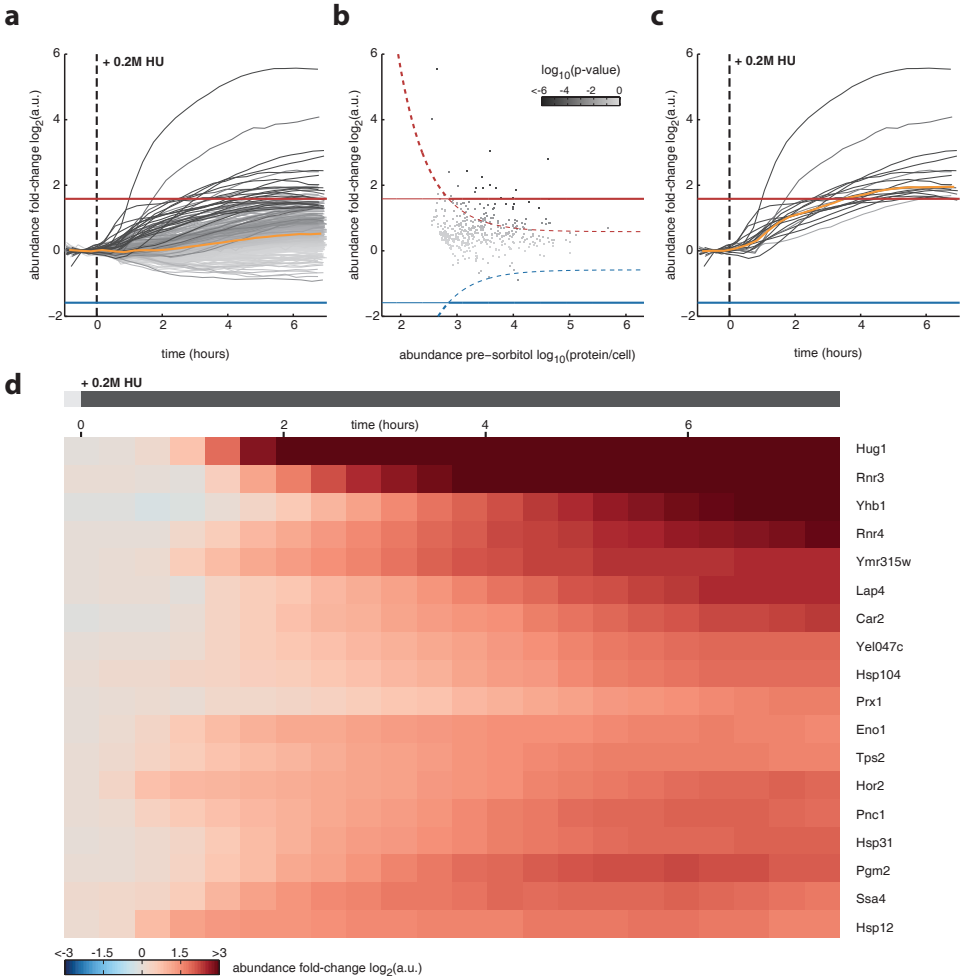

Figure S26: **Protein abundance dynamics for HU treatment. a.** Mean abundance variation against time is shown for 439 strains. Strains that never reach higher abundance levels from the autofluorescence are not shown. For each strain, abundance is normalized to median values within 1 hour before HU treatment. The orange line shows the median of all abundance traces. The red line represents a 3-fold increase threshold and the blue line a 3-fold decrease threshold. The gray scale shows the significance of the fold-change for each protein. **b.** Abundance fold-change against initial pre-HU abundance. The dashed lines show a threshold of significance (p = 0.01). The continuous line show a 3-fold change threshold. 18 proteins are above the red and bold line (p-value < 0.01 and fold-change > 3). **c.** Time-series as in **a** for those 18 proteins. The orange line shows the median of all traces. **d.** Clustergram of the 18 up-regulated proteins. MMS treatment starts after the first time point.

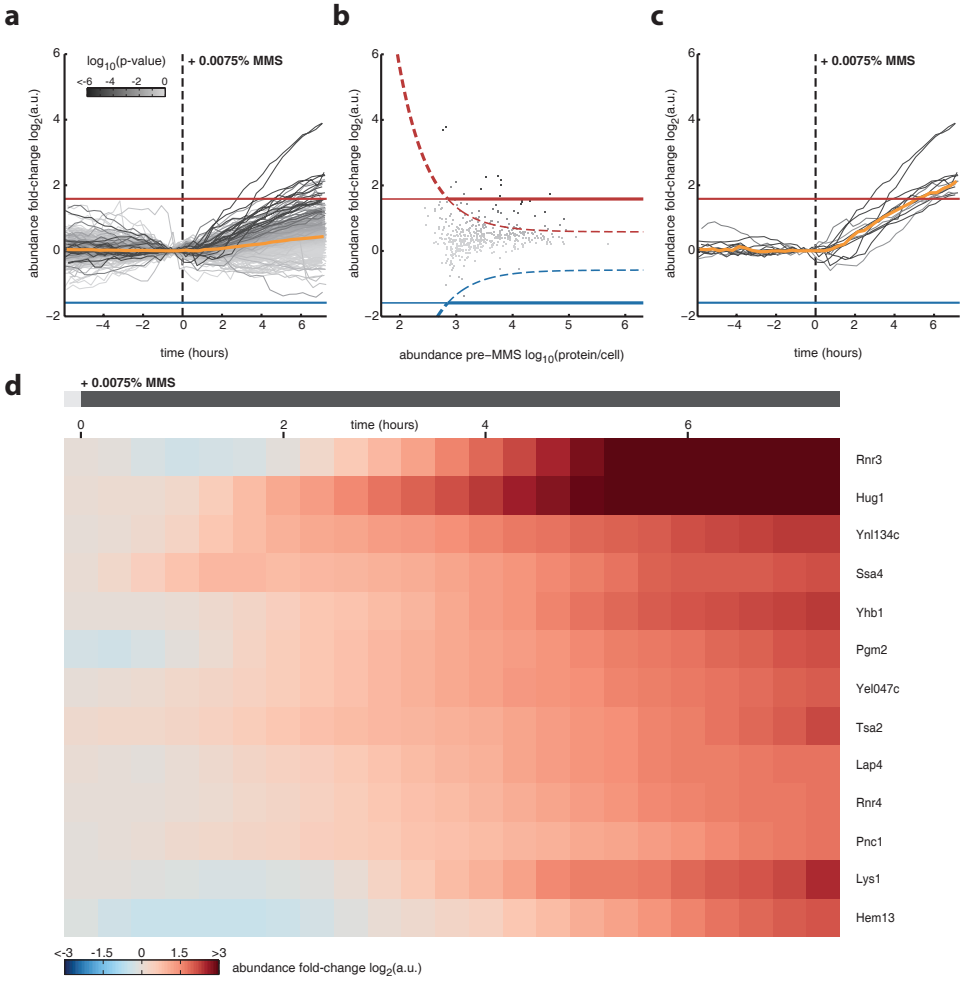

Figure S27: **Protein abundance dynamics for low MMS treatment. a.** Mean abundance variation against time is shown for 439 strains. Strains that never reach higher abundance levels from the autofluorescence are not shown. For each strain, abundance is normalized to median values within 1 hour before MMS treatment. The orange line shows the median of all abundance traces. The red line represents a 3-fold increase threshold and the blue line a 3-fold decrease threshold. The gray scale shows the significance of the fold-change for each protein. **b.** Abundance fold-change against initial pre-MMS abundance. The dashed lines show a threshold of significance (p = 0.01). The continuous line show a 3-fold change threshold. 13 proteins are above the red and bold line (p-value $<$ 0.01 and fold-change $>$ 3). **c.** Time-series as in **a** for those 13 proteins. The orange line shows the median of all traces. **d.** Clustergram of the 13 up-regulated proteins. MMS treatment starts after the first time point.

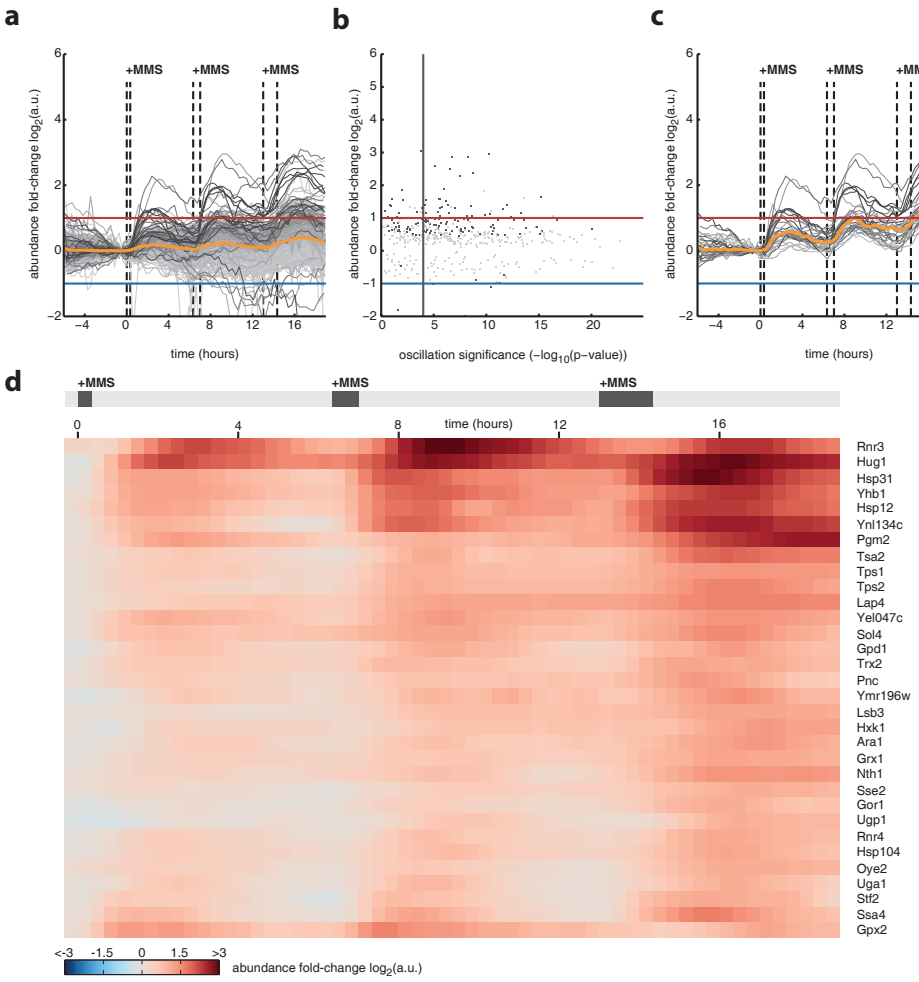

Figure S28: **Protein abundance dynamics for MMS pulses. a.** Mean abundance variation against time is shown for 445 strains. Strains that never reach higher abundance levels from the autofluorescence are not shown. For each strain, abundance is normalized to median value within 1 hour before MMS treatment, which starts at t=0 and lasts 20 min. Following MMs exposure last 40min and 1h30. The orange line shows the median of all traces. The red line represents a 3-fold increase threshold and the blue line a 3-fold decrease threshold. **b.** Abundance fold-change against oscillation significance. Fold-change is defined as the maximal fold-change (increase or decrease) at any time after t=0. Oscillation significance represents the enrichment of a 6 hour period in the fourrier spectrum of each time-course. The red and blue line show a 3-fold change threshold. The vertical gray line shows a 6 hour oscillation significance threshold of 1e-4. Dark data points have a significant abundance fold-change (p-value $<$ 0.05). 32 proteins have a significance response to oscillating stimulus (p-value $<$ 1e-4) and a significant fold-change (p-value $<$ 0.05 and fold-change $>$ 2). **c.** Time-series as in **a** for those 32 proteins. The orange line shows the median of all traces. **d.** Clustergram of the 32 pulse-sensitive proteins. Time line shows the timing and duration of the MMS pulses.

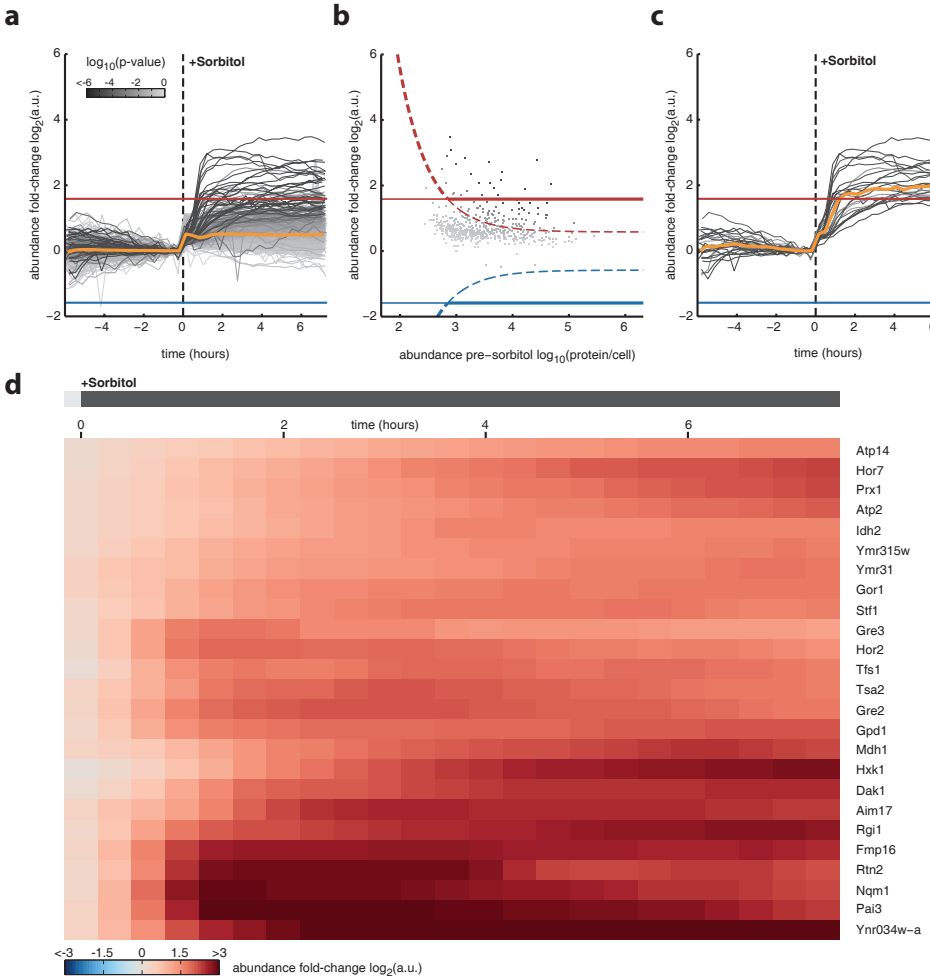

Figure S29: **Protein abundance dynamics for sorbitol treatment.** **a.** Mean abundance variation against time is shown for 547 strains. Strains that never reach higher abundance levels from the autofluorescence are not shown. For each strain, abundance is normalized to median value within 1 hour before Sorbitol treatment, which starts at t=0. The orange line shows the median of all traces. The red line represents a 3-fold increase threshold and the blue line a 3-fold decrease threshold. The gray scale shows the significance of the fold-change for each protein. **b.** Abundance fold-change against initial pre-sorbitol abundance. AFold-change is defined as the maximal fold-change at any time after t=0. The dashed lines show a threshold of significance (p = 0.01). The continuous line show a 3-fold change threshold. 25 proteins are above the red and bold line (p-value < 0.01 and fold-change > 3). **c.** Time-series as in **a** for those 25 proteins. The orange line shows the median of all traces. **d.** Clustergram of the 25 up-regulated proteins. Sorbitol treatment starts after the first time point and last 6.5 hours.

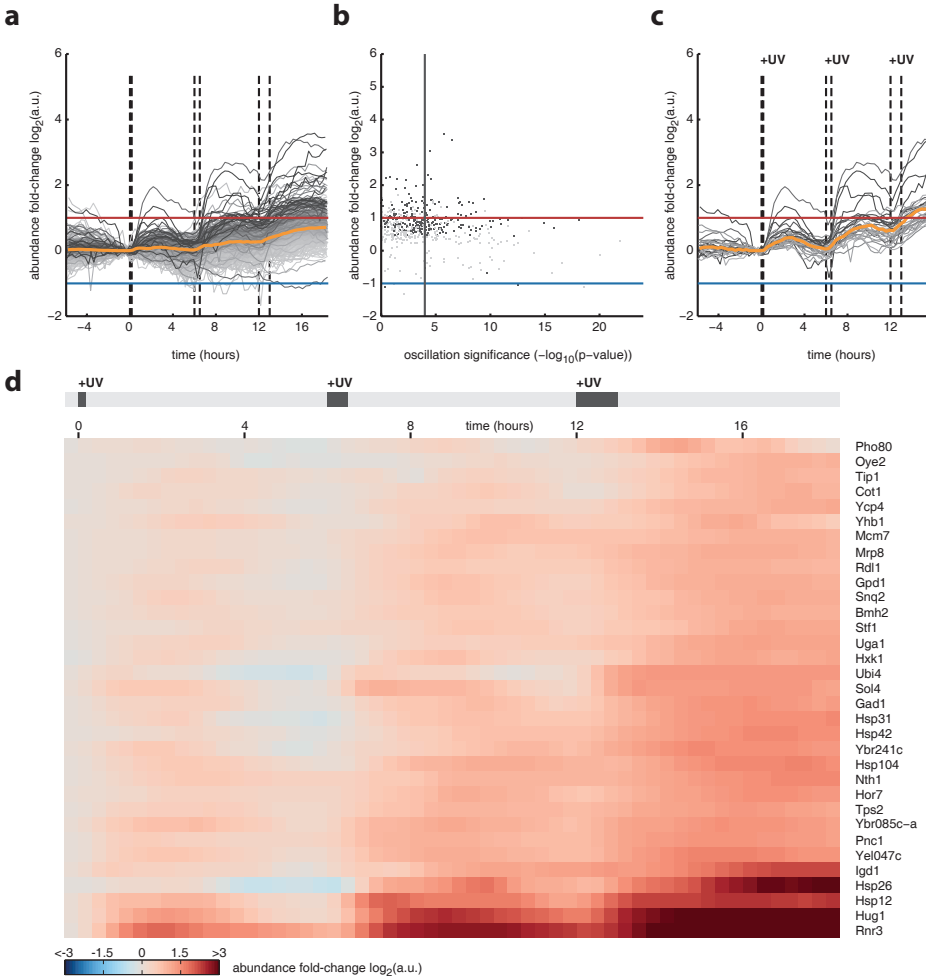

Figure S30: **Protein abundance dynamics for pulsing UV irradiation. a.** Mean abundance variation against time is shown for 507 strains. Strains that never reach higher abundance levels from the autofluorescence are not shown. For each strain, abundance is normalized to median value within 1 hour before the first UV irradiation, which starts at t=0 and last 10 min. Following irradiation last 30min and 1 hour. The orange line shows the median of all traces. The red line represents a 3-fold increase threshold and the blue line a 3-fold decrease threshold. **b.** Abundance fold-change against oscillation significance. Fold-change is defined as the maximal fold-change (increase or decrease) at any time after t=0. Oscillation significance represents the enrichment of a 6 hour period in the fourrier spectrum of each time-course. The red and blue line show a 3-fold change threshold. The vertical gray line shows a 6 hour oscillation significance threshold of 1e-4. Dark data points have a significant abundance fold-change (p-value $<$ 0.05). 33 proteins have a significance response to oscillating stimulus (p-value $<$ 1e-4) and a significant fold-change (p-value $<$ 0.05 and fold-change $>$ 2). **c.** Time-series as in **a** for those 33 proteins. The orange line shows the median of all traces. **d.** Clustergram of the 33 pulse-sensitive proteins. Time line shows the timing and duration of the UV pulses.

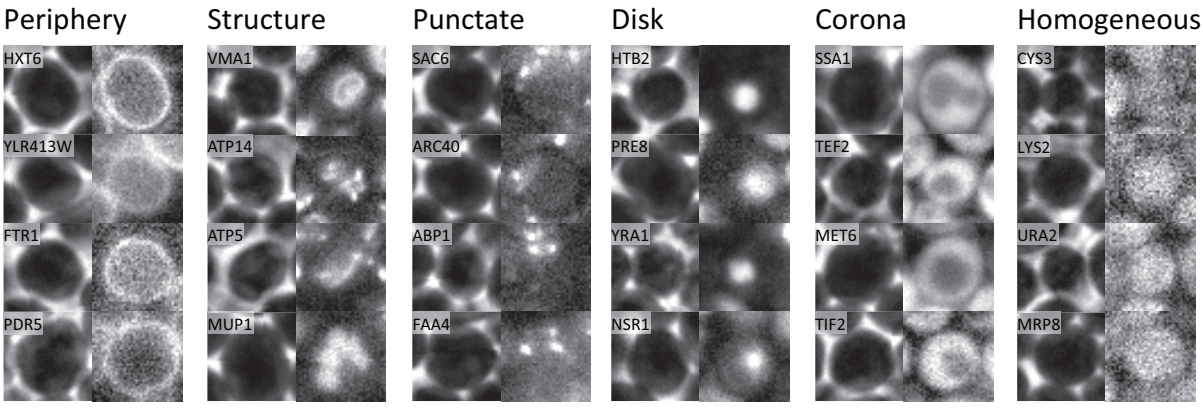

Figure S31: **Single cell bright field and epifluorescence examples for our geometrical shape classes.** Each image has a size of 5.8 micrometers. A manually annotated set containing 6982 cells was used to build a classifier.

a

Actual \ Predicted	Periphery	Structure	Punctate	Disk	Corona	Homogeneous
Periphery	92.6%	0.7%	0.7%	0.0%	3.0%	3.0%
Structure	1.2%	65.0%	10.9%	11.9%	7.1%	4.0%
Punctate	2.4%	12.7%	60.5%	7.5%	1.3%	15.7%
Disk	0.0%	25.7%	2.5%	67.4%	2.3%	2.2%
Corona	0.9%	11.4%	2.5%	0.7%	48.5%	36.1%
Homogeneous	0.0%	0.8%	1.7%	0.2%	9.0%	88.3%

b

Manual \ Computer	Periphery	Structure	Punctate	Disk	Corona	Homogeneous
Periphery	13	0	0	0	0	0
Structure	0	22	1	0	0	0
Punctate	0	0	4	0	0	1
Disk	0	0	0	50	0	0
Corona	0	0	0	0	69	0
Homogeneous	0	0	0	0	0	22

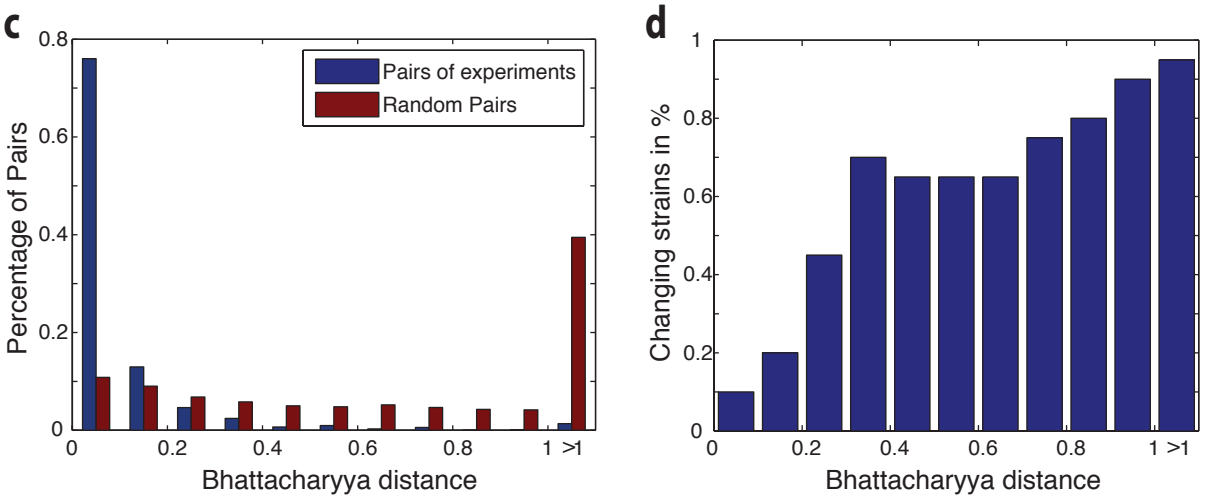

Figure S32: **Validation of the classifier.** (a) Confusion matrix to compare the manual annotation of single cells with the predicted geometrical shape using 10-fold cross validation. (b) Comparison of automatic and manual annotation: 200 images for which the classifier found one characteristic group, where manually and independently predicted by ND and JB. (c) Agreement of replicated recordings: Histogram of the Bhattacharyya distance of high intensity strains with double coverage, in comparison to the distance of randomly selected strains. (d) Relation between Bhattacharyya distance and the visual perception of a change between the images.

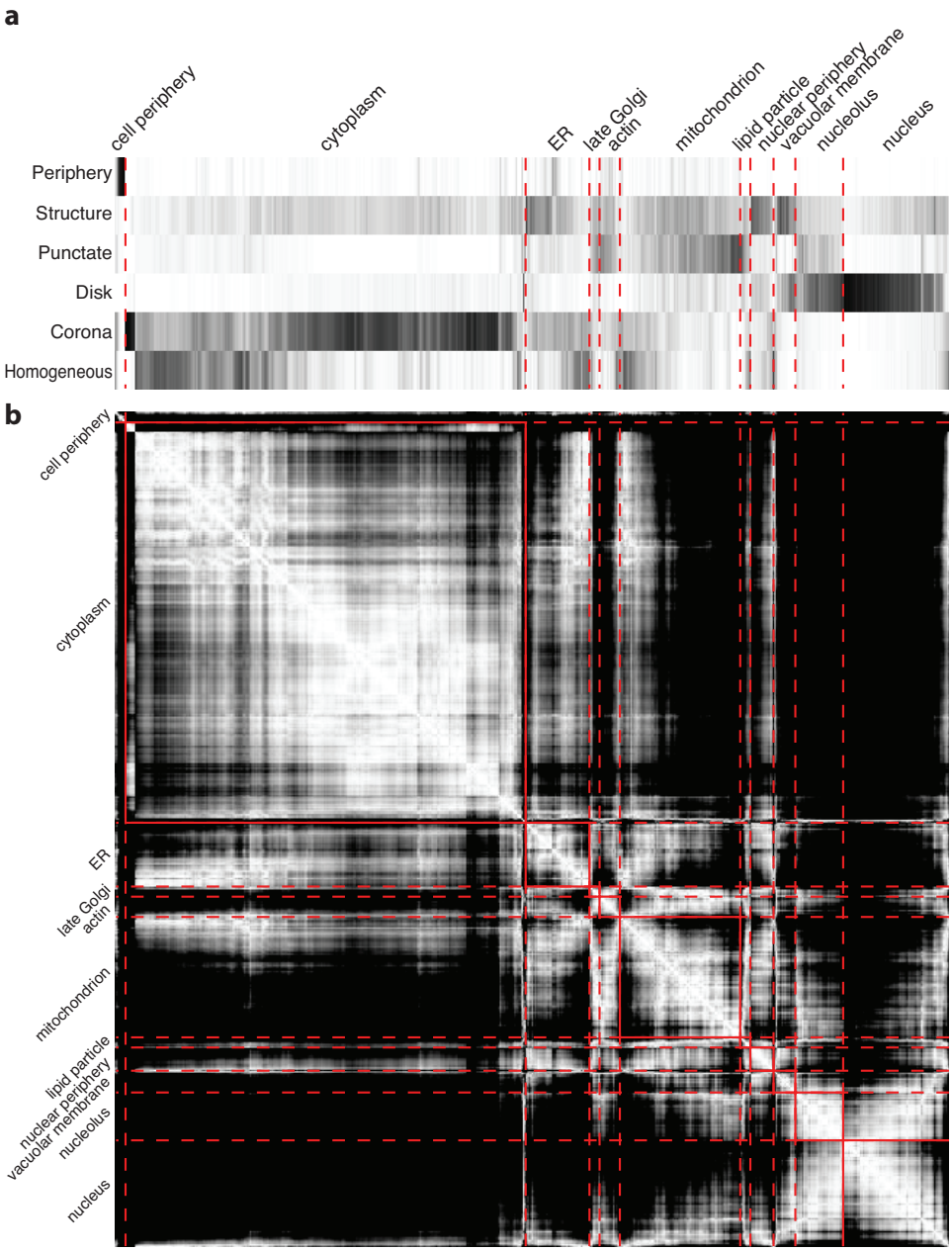

Figure S33: **Comparison with the UCSF data.** Evaluation of strains with high intensity and well defined location patterns. The strains are clustered within their groups, as defined by Huh *et al.* [3] . (a) Average probability of the six geometrical classes for each strain. White means 0% probability, black 100%.(b) Heat map that shows the Bhattacharyya distances between the strains. The distance goes from zero (white) to one or above (black).

UCSF \ Predicted	Cell Periphery	Cytoplasm	ER	Innercell membranes	Nucleus	Nucleolus	Filaments	Punctate
Cell Periphery	8	0	0	0	0	0	0	0
Cytoplasm	0	181	5	2	2	1	1	1
ER	0	1	38	1	0	0	1	0
Innercell membranes	0	0	3	40	4	0	0	0
Nucleus	0	0	0	9	142	2	0	2
Nucleolus	0	0	0	0	1	41	1	3
Filaments	0	1	4	0	2	2	104	8
Punctate	0	0	1	0	1	1	4	19
Number of strains	637							
Correct Classification	90.0%							

Figure S34: **Comparison with the UCSF data using cross-validation.** This table shows the number of strains in each category that were classified by our algorithm (columns) and by Huh *et al.* [3] (rows).

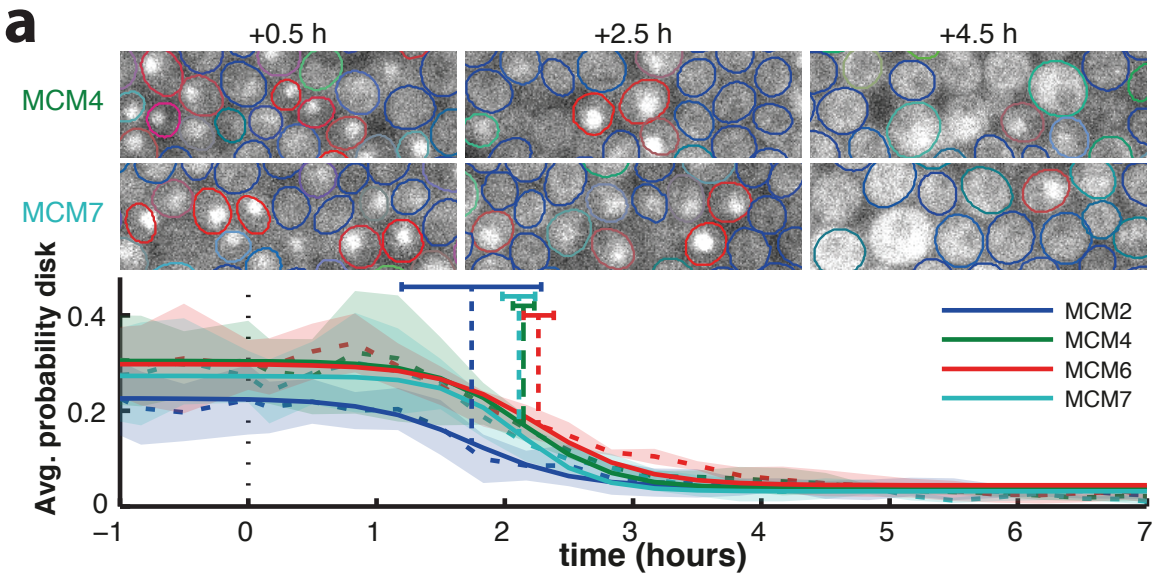

Figure S35: **Mcm timing analysis.** (a) Average disk probability for Mcm proteins that translocate from the nucleus to the cytoplasm. The dashed line represents the average and the transparent area shows the error (±s.d.). Traces were fitted with a sigmoid (solid line). The vertical dashed lines show the transition times and their corresponding error bars (±s.d.).

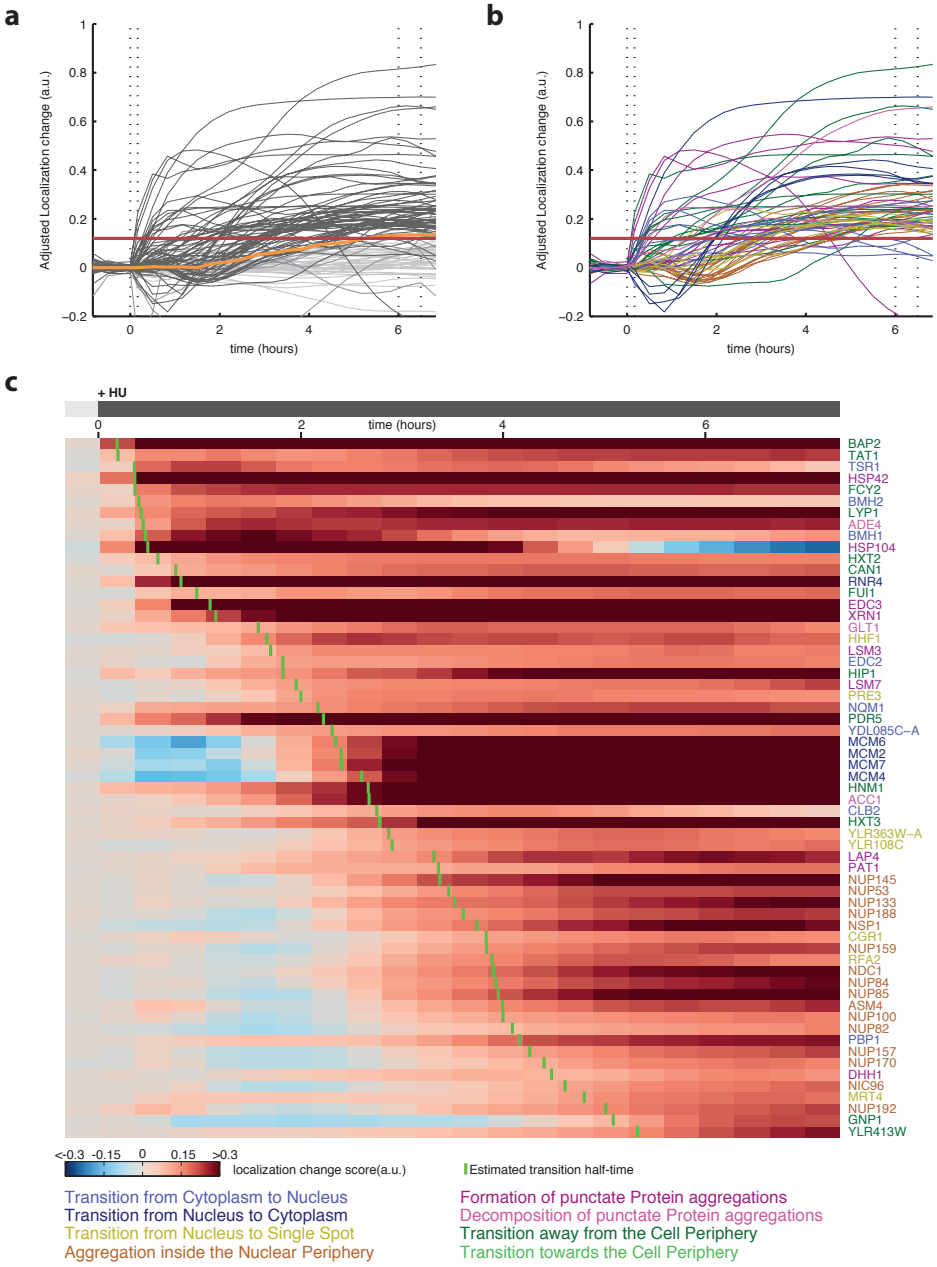

Figure S36: **Protein localization for HU treatment. a.** Localization change is shown for proteins in response to 0.2M HU. The red line shows the minimal threshold to select proteins that change. The orange line shows the median. **b.** Localization change is shown for the proteins above the threshold. **c.** Same data as above, shown in a heatmap, ordered by timing. The green bar represents the time of the change. The color code in **(b)** and **(c)** indicates the different transition classes, given by the legend.

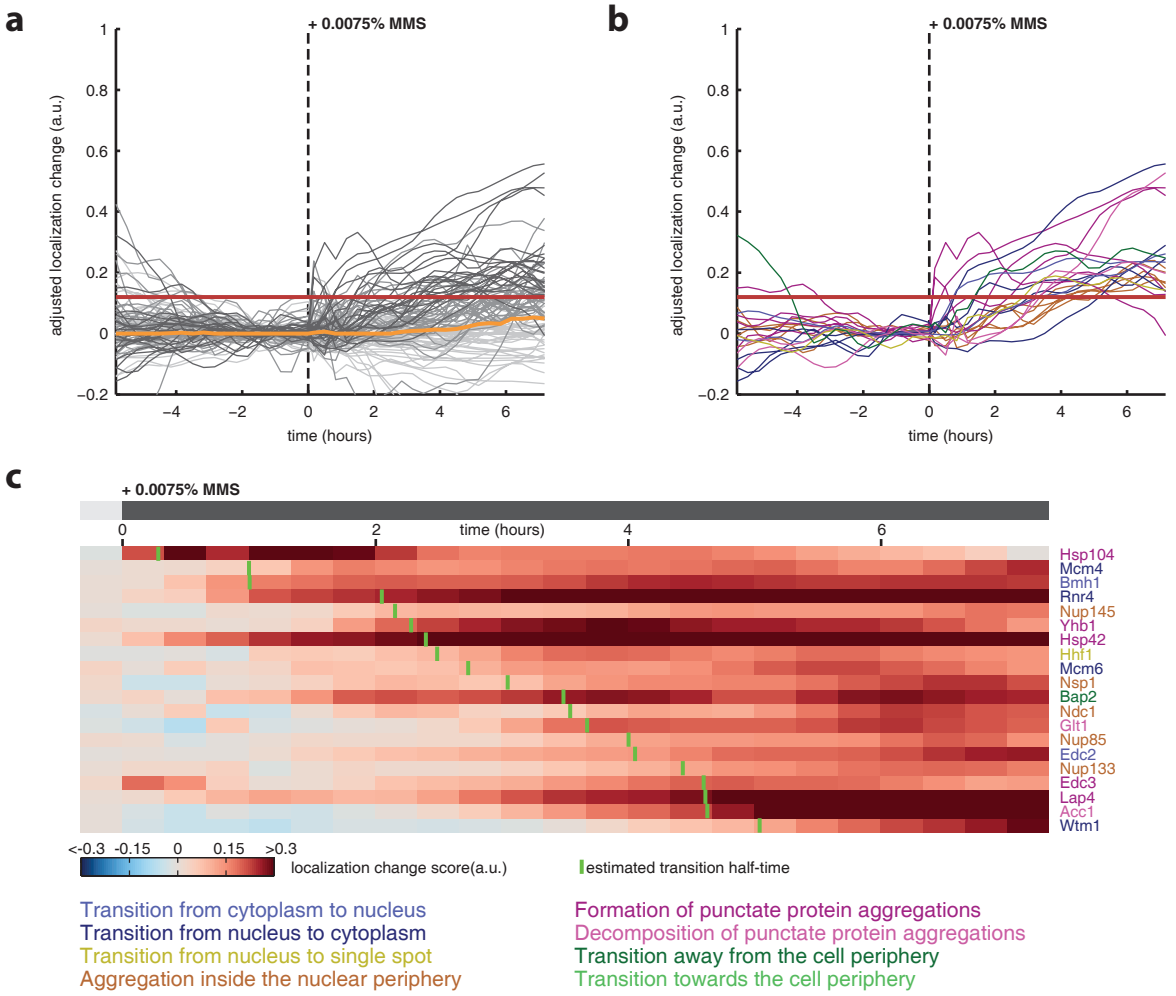

Figure S37: **Protein localization for low MMS treatment. a.** Localization change is shown for proteins in response to 0.0075% MMS. The red line shows the minimal threshold to select proteins that change. The orange line shows the median. **b.** Localization change is shown for the proteins above the threshold. **c.** Same data as above, shown in a heatmap, ordered by timing. The green bar represents the time of the change. The color code in **(b)** and **(c)** indicates the different transition classes, given by the legend.

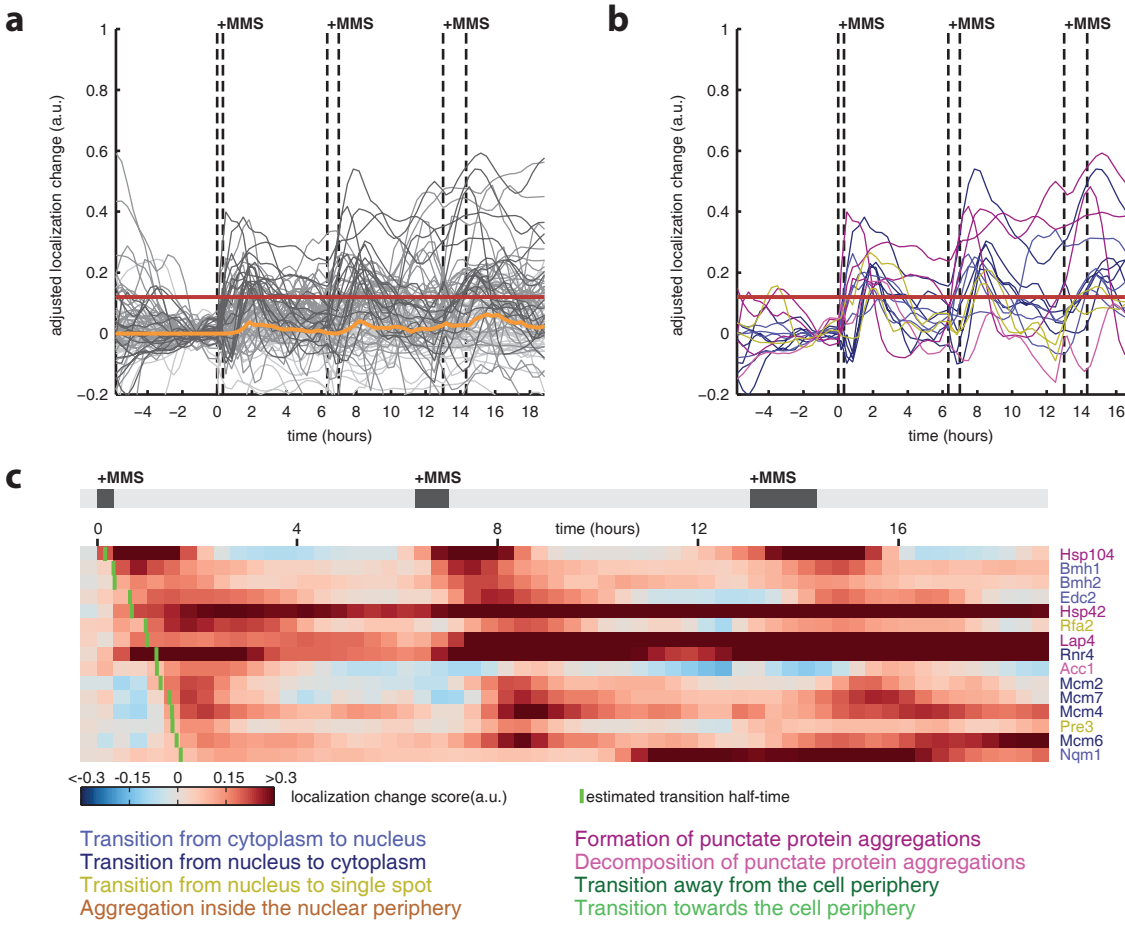

Figure S38: **Protein localization for MMS pulses. a.** Localization change is shown for proteins in response MMS pulses of 20min, 40min and 1h20. The red line shows the minimal threshold to select proteins that change. The orange line shows the median. **b.** Localization change is shown for the proteins above the threshold. **c.** Same data as above, shown in a heatmap, ordered by timing. The green bar represents the time of the change. The color code in **(b)** and **(c)** indicates the different transition classes, given by the legend.

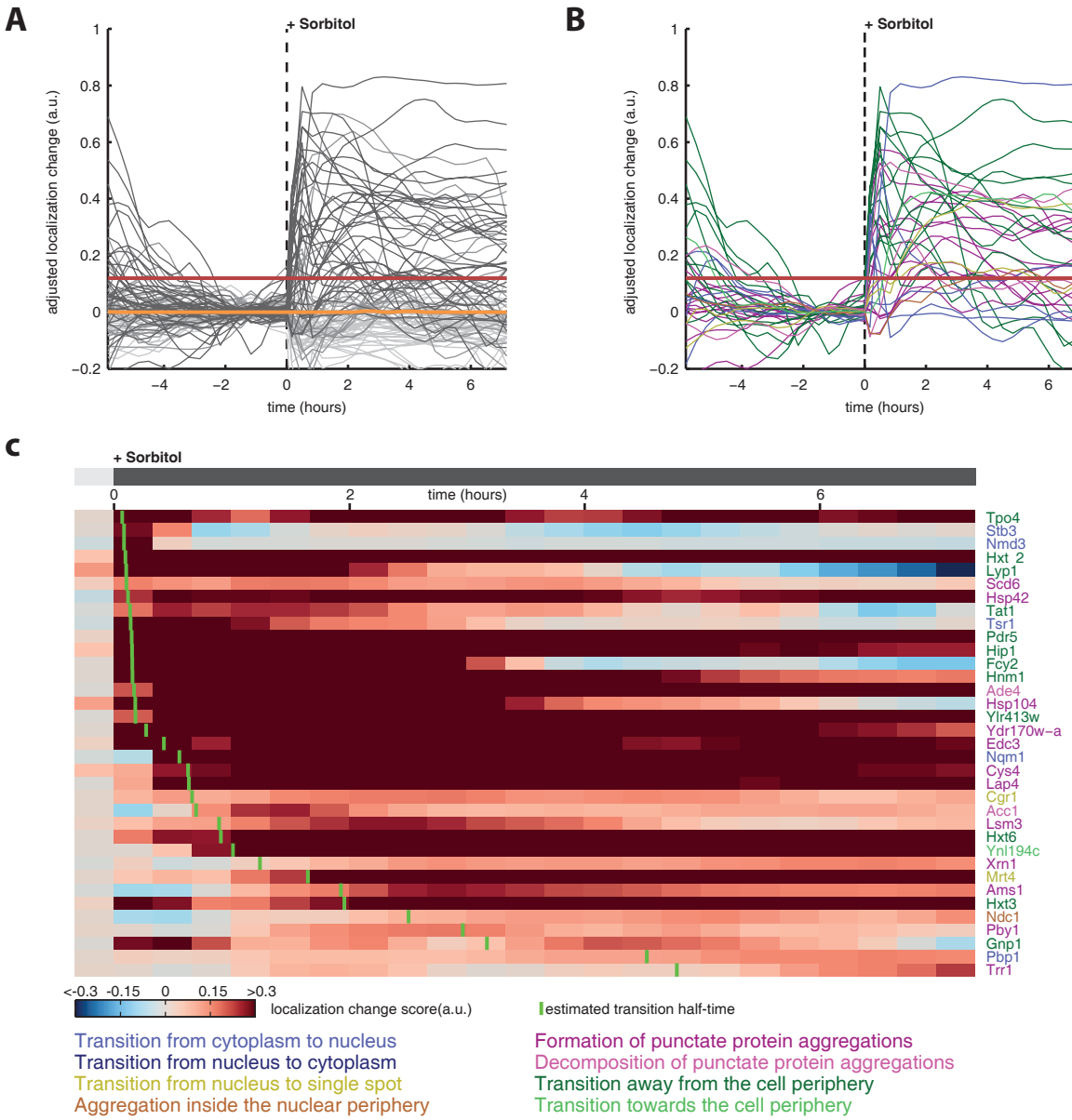

Figure S39: **Protein localization for sorbitol treatment. a.** Localization change is shown for proteins in response to 1M sorbitol. The red line shows the minimal threshold to select proteins that change. The orange line shows the median. **b.** Localization change is shown for the proteins above the threshold. **c.** Same data as above, shown in a heatmap, ordered by timing. The green bar represents the time of the change. The color code in **(b)** and **(c)** indicates the different transition classes, given by the legend.

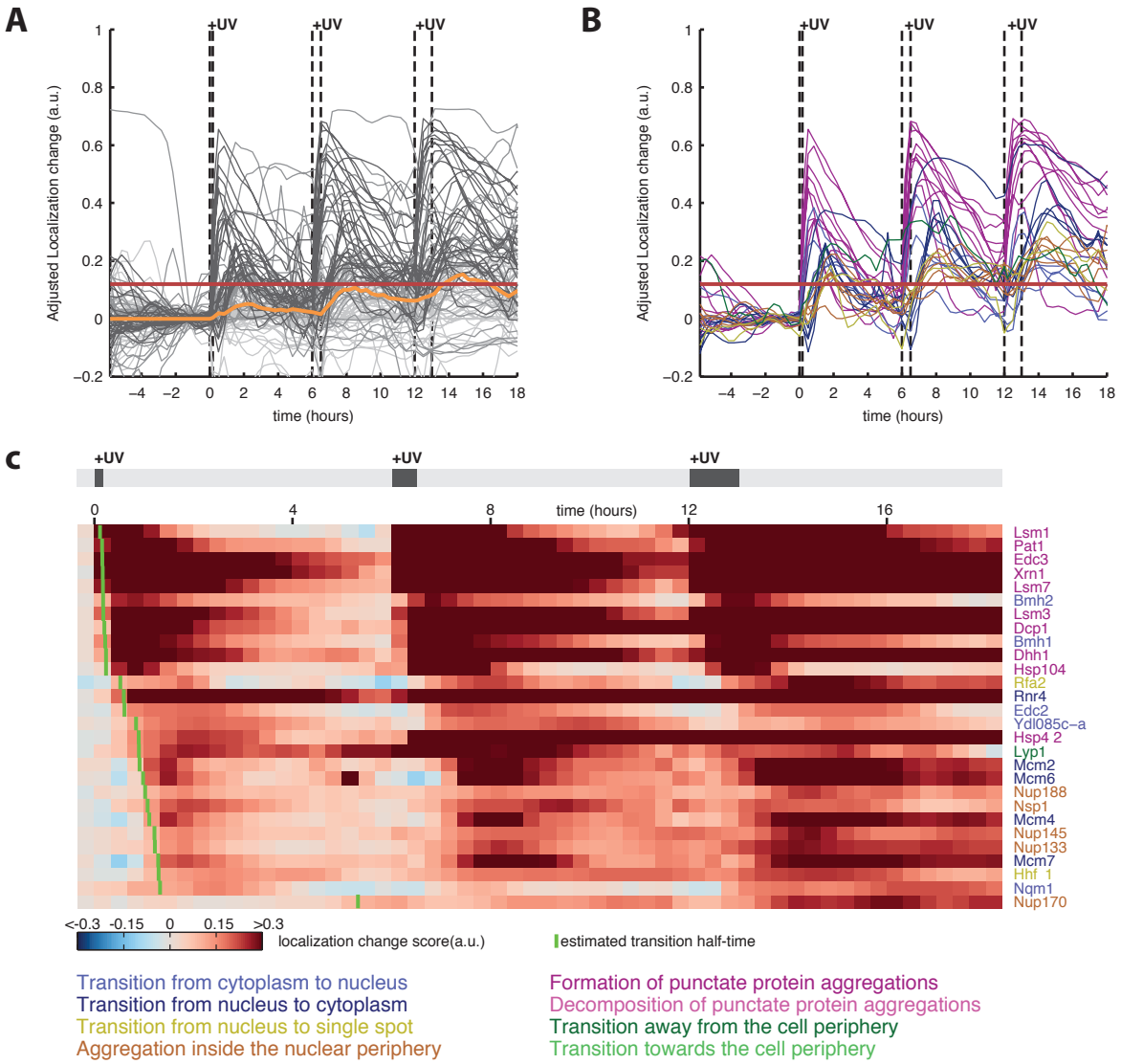

Figure S40: **Protein localization for pulsing UV irradiation. a.** Localization change is shown for proteins in response UV pulses of 10min, 30min and 1h. The red line shows the minimal threshold to select proteins that change. The orange line shows the median. **b.** Localization change is shown for the proteins above the threshold. **c.** Same data as above, shown in a heatmap, ordered by timing. The green bar represents the time of the change. The color code in **(b)** and **(c)** indicates the different transition classes, given by the legend.

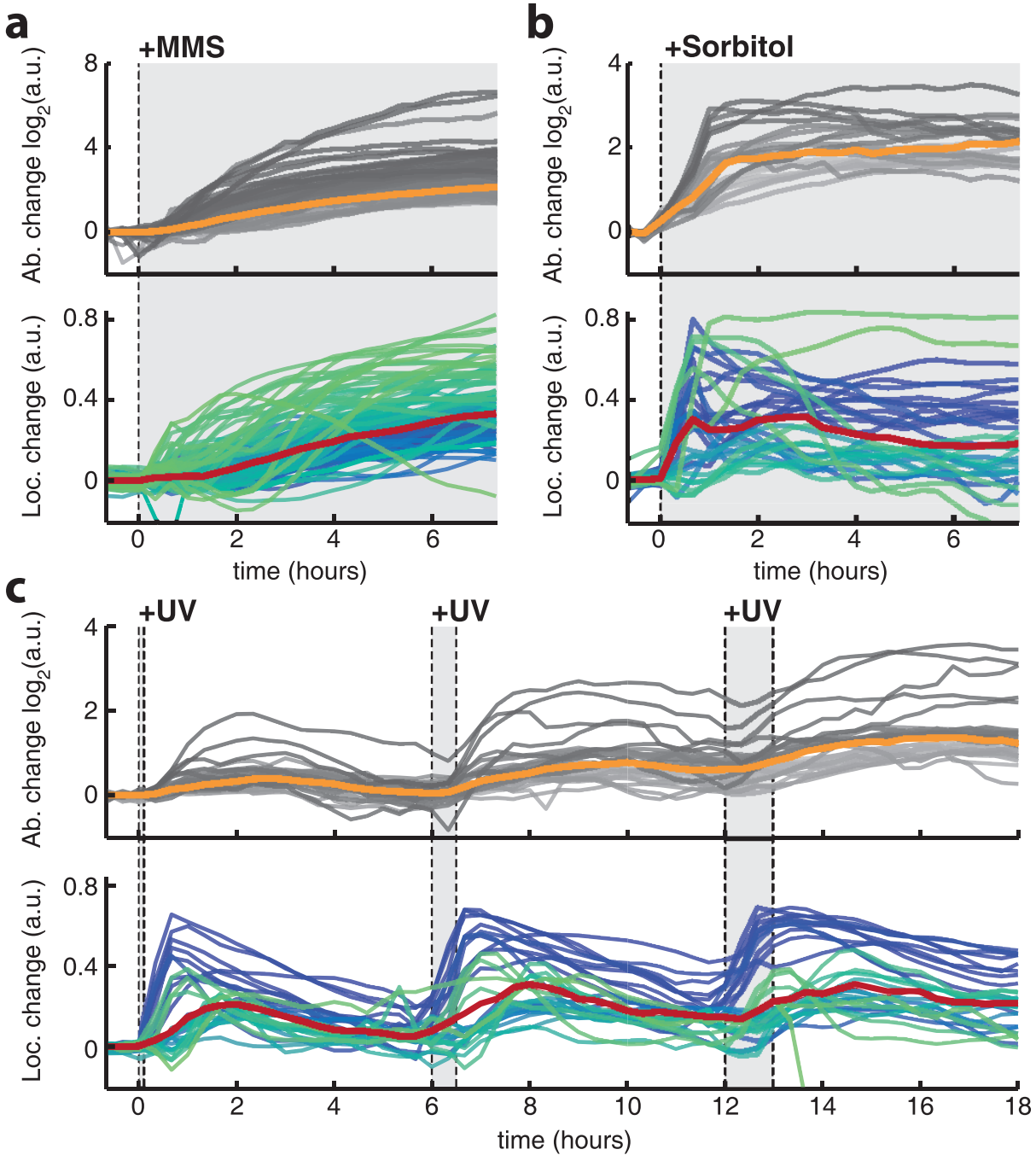

Figure S41: **Comparison of abundance and localization dynamics.** Abundance fold-change (top) and localization change (bottom) are given as a function of time for every protein that responded in **(a)** MMS high, **(b)** hyperosmotic shock, **(c)** and UV irradiation. The gray areas represent the duration of the stimuli. The three UV pulses lasted 10, 30, and 60 min.

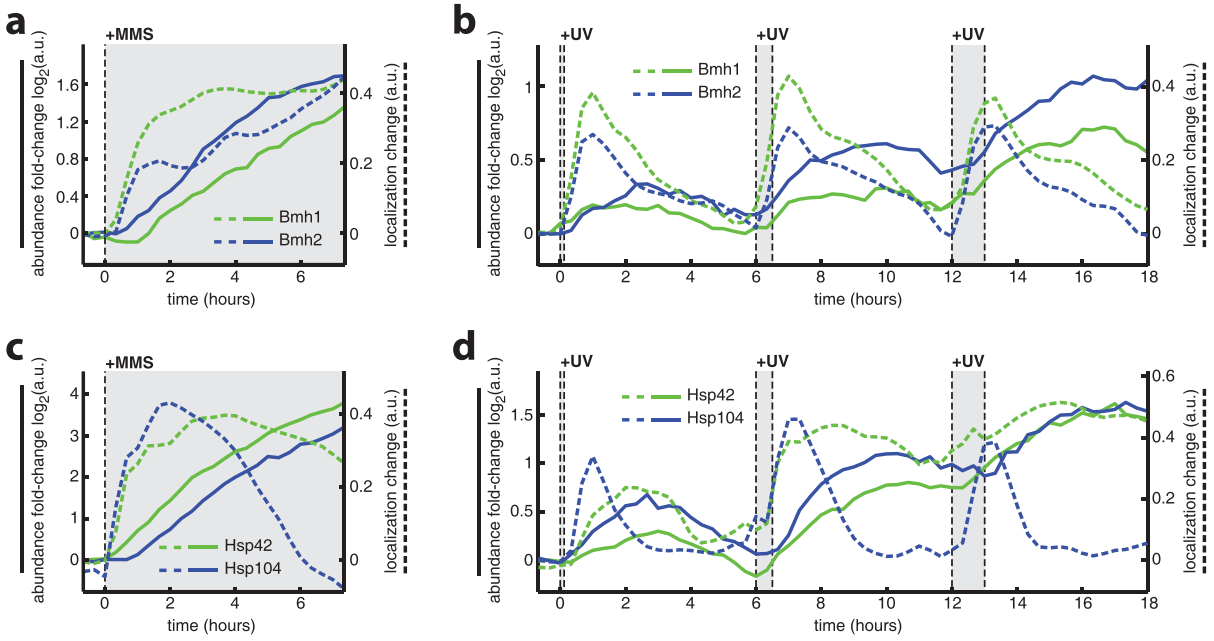

Figure S42: **Concomitant abundance and localization changes.** Abundance (solid line) and localization (dashed line) changes of Bmh1 and Bmh2 in response to **(a)** MMS treatment and **(b)** UV pulses. Abundance and localization change of Hsp42p and Hsp104p in response to **(c)** MMS treatment and **(d)** UV pulses.

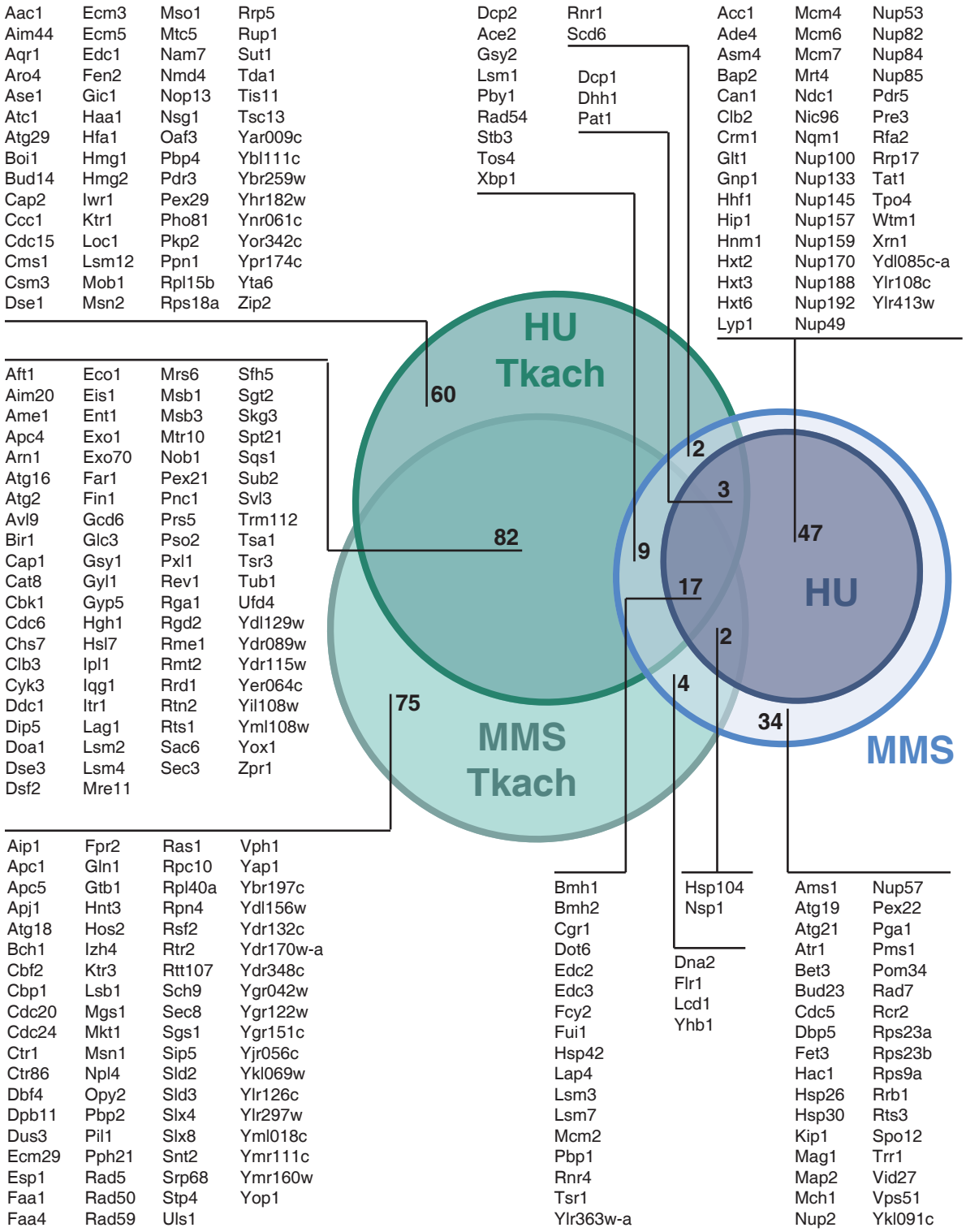

Figure S43: **Comparison of proteins defined as changing in localization in our dataset and by Tkach et al. [30]**