Edinburgh Research Explorer

Estimating numbers of intracellular molecules through analysing fluctuations in photobleaching

OPEN

Estimating numbers of intracellular molecules through analysing fluctuations in photobleaching

Elco Bakker & Peter S. Swain [iD]*

The impact of fluorescence microscopy has been limited by the difficulties of expressing measurements of fluorescent proteins in numbers of molecules. Absolute numbers enable the integration of results from different laboratories, empower mathematical modelling, and are the bedrock for a quantitative, predictive biology. Here we propose an estimator to infer numbers of molecules from fluctuations in the photobleaching of proteins tagged with Green Fluorescent Protein. Performing experiments in budding yeast, we show that our estimates of numbers agree, within an order of magnitude, with published biochemical measurements, for all six proteins tested. The experiments we require are straightforward and use only a wide-field fluorescence microscope. As such, our approach has the potential to become standard for those practising quantitative fluorescence microscopy.

In fluorescence microscopy, converting measurements of fluorescence into numbers of molecules is a long-standing challenge[1]. This deficit limits our ability both to combine fluorescence measurements from different experiments and to apply quantitative analyses of time series that often must assume known numbers of proteins[2–4].

Although fluorescence standards[1] and the collection of techniques known as fluorescence fluctuation spectroscopy – the most well known of which are fluorescence correlation spectroscopy (FCS)[5,6] and analysis of photon-counting histograms (PCH)[7] – provide a solution, neither are yet commonly adopted to calibrate, for example, time-lapse imaging in cell and systems biology[8].

Instead, fluctuation-based methods have been developed, such as those that measure fluctuations in the distribution of fluorescent proteins between daughter cells at cell division[9–11]. These approaches, however, have been applied mostly to bacteria, are unsuitable for non-dividing cells[12], and do not straightforwardly extend to species that exhibit differences in size between mothers and daughters.

A second approach is to study fluctuations in stochastic processes of decay. Inhibiting translation and transcription has allowed the fluorescence per molecule to be estimated in mammalian cells[13], but the stability of fluorescent proteins can make these experiments time consuming. An alternative technique is to deliberately induce photobleaching[14]: the process by which fluorophores cease to fluoresce when continuously excited. This method has been applied in vitro[15,16] and to bacteria[17], but the analysis relies on photobleaching exhibiting an exponential decay[14], which is expected for single molecules but not necessarily for the fluorescence of cells[18].

Here we develop a method for estimating numbers from deliberately photo-bleached cells that works on the wide-field microscopes used for time-lapse imaging and requires no specialized equipment. We verify our approach using six different proteins tagged with Green Fluorescent Protein (GFP) in budding yeast. In all cases, our estimates of the numbers of molecules are within an order of magnitude of estimates made using biochemical techniques, such as quantitative Western blotting and mass spectrometry.

Before beginning, we remind the reader that fluctuation analyses work because the magnitude of fluctuations in fluorescence are determined not by the concentration of the fluorescent molecules but by their numbers. A fluorescent measurement, Y, is given by the product of the brightness per molecule, ν, and the number of fluorescent molecules, X, and is therefore agnostic to their individual values. If X varies with time, the magnitude of the fluctuations in the underlying biochemical process changing X, however, typically scale with the mean[19]: $\text{Var}[X] \propto E[X]$. Therefore the variance in the fluorescence, scales as $\text{Var}[Y] \propto \nu^2 E[X] = \nu E[Y]$ because $Y = \nu X$. For a given fluorescence (a given $E[Y]$), large fluctuations (large $\text{Var}[Y]$) therefore imply a high value of ν and low numbers of molecules, and vice versa (Fig. 1).

School of Biological Sciences, University of Edinburgh, Edinburgh, United Kingdom. *email: peter.swain@ed.ac.uk
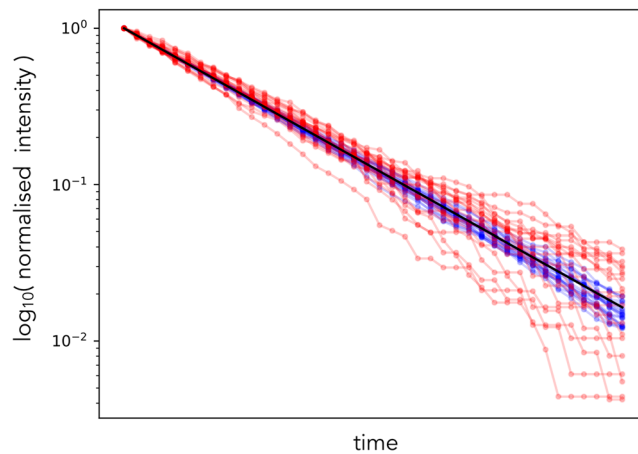
Figure 1. The magnitude of fluctuations is determined by numbers of molecules, which we illustrate by simulating stochastic exponential decay for both a small number of molecules in red and a large number in blue. By normalizing to the same starting value, these simulated time series show fluctuations of a different magnitude depending on their initial numbers of molecules. The mean behaviour is common (black line), but the data for low numbers of molecules (red) shows larger deviations from the mean than the data for high numbers of molecules (blue).
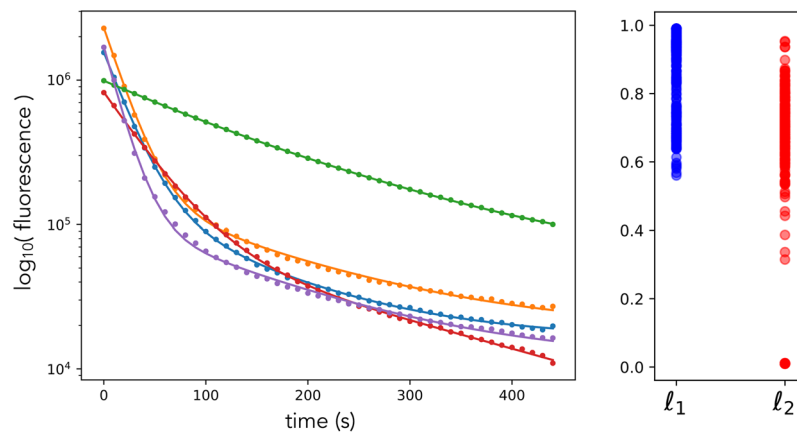
Figure 2. Photobleaching *in vivo* is not described by a homogeneous, single exponential decay. Left: The logarithm of the fluorescence of 5 cells of budding yeast expressing Pgk1-GFP and undergoing photobleaching do not all fall on a straight line as a function of time (c.f. Fig. 1). The data are shown as dots and the fits for double exponential decay as lines, which is modelled for fluorescence at time point i as $f_i = a_0 + a_1 \ell_1^i + a_2 \ell_2^i$ with constants a and $\ell_1 = e^{-\lambda_1 \Delta t}$ and $\ell_2 = e^{-\lambda_2 \Delta t}$ for time scales λ and time Δt between measurements (here 10 s). All data have been corrected for autofluorescence using wild-type cells that do not express GFP. Right: bleaching is heterogeneous across the population of cells. There is substantial cell-to-cell variation in the best-fit values for ℓ_1 and ℓ_2.

Results

Photobleaching *in vivo* has more than one time scale. We obtained time series of photobleaching using budding yeast and GFP-tagged proteins[20]. Cells were fixed and photobleached in sustained illumination with fluorescence measurements taken every 10 seconds (Fig. 2 and Methods). We use fixed cells to remove complications arising from cells synthesising and degrading proteins and from fluorescent proteins maturing during the 6 minutes or so of bleaching.

Our data is incompatible with earlier fluctuation-based methodologies (Methods) because it is not well described by a single exponential decay even after being corrected for autofluorescence (Fig. 2; left). First, the data for multiple cells shows systematic deviations from a single exponential and is better fit by a bi-exponential decay. Second, the decay rates of these two exponentials vary substantially between cells (Fig. 2; right). Multi-exponential photobleaching is common[18] and can be caused by, for example, differing intracellular micro-environments[21], molecular rotation[22], and higher order interactions between excited fluorophores[23]. These phenomena can also cause heterogeneity in parameters between cells.

A Bayesian approach systematically underestimates.　　We wish to infer from our data, ν, the factor that converts from fluorescence units to absolute numbers[10]. We follow an established Bayesian methodology for stochastic, chemical systems[24], which uses the linear noise approximation[19] to describe the dynamics of bleaching and a Kalman filter for the inference. Our model is general: each cell contains two pools of fluorescent molecules that bleach with their own time scale and measurement noise has a normal distribution but with a variance that is up to a quadratic function of the total number of fluorescent molecules (Methods). Using a Markov chain Monte Carlo scheme to sample the posterior distribution, the numbers of molecules we infer are, however, too small (Methods).

We conclude that our model of either bleaching or measurement noise is incorrect or that both are incorrect. Shortcomings in the bleaching model, for example, will generate discrepancies between the model's behaviour and the cells' actual behaviour that are correlated over time. We make the common assumption that the measurement noise at each time point is independent, and so the algorithm will interpret any correlated error as coming from bleaching. The magnitude of the fluctuations in bleaching will correspondingly be overestimated and the number of molecules underestimated.

An estimator for the number of molecules.　　We therefore propose instead an estimator that is insensitive to the details of the photobleaching. We motivate the estimator by first considering bleaching with a single rate constant, λ:

$$X \xrightarrow{\lambda} X_{\text{bleached}} \tag{1}$$

Using the linear noise approximation[19] and writing $\ell = e^{-\lambda t}$ for the probability of a fluorophore remaining unbleached at the time t of interest, we can show (Methods) that from x_0 molecules at $t = 0$

$$\mathrm{E}[X_t] = x_0\ell \tag{2}$$

and that

$$\mathrm{Var}[X_t] = x_0(1 - \ell)\ell. \tag{3}$$

Dividing Eq. 3 by x_0^2 implies that the magnitude of these normalized fluctuations in X_t/x_0 scale as $1/x_0$ (Fig. 1). Using Eq. 2, we can replace ℓ in Eq. 3,

$$\mathrm{Var}[X_t] = \frac{\mathrm{E}[X_t]}{x_0}(x_0 - \mathrm{E}[X_t]) \tag{4}$$

which holds for all t. If we ignore measurement noise then fluorescence $Y = \nu X$, and multiplying Eq. 4 by ν^2, assumed to be the same for each molecule, gives

$$\mathrm{Var}[Y_t] = \frac{\mathrm{E}[Y_t]}{x_0}(y_0 - \mathrm{E}[Y_t]) \tag{5}$$

and so our estimator for x_0 is

$$\hat{x}_0 = \frac{\mathrm{E}[Y_t](y_0 - \mathrm{E}[Y_t])}{\mathrm{Var}[Y_t]} \tag{6}$$

for any time $t > 0$. Given y_0, $\mathrm{E}[Y_t]$, $\mathrm{Var}[Y_t]$, and no measurement noise, we can show that $\hat{x}_0$ is a maximum likelihood estimator of x_0 (Methods).

To gain intuition about when the estimator should be reliable, consider the extreme case where each molecules has its own brightness, ν_i, and bleaches with its own rate. This bleaching should be independent and if ℓ_i is the probability of molecule i remaining unbleached at a time t, then

$$\mathrm{E}[Y_t] = \sum_{i=1}^{x_0} \nu_i \ell_i$$
$$\mathrm{Var}[Y_t] = \sum_{i=1}^{x_0} \nu_i^2 \ell_i (1 - \ell_i)$$
$$y_0 = \sum_{i=1}^{x_0} \nu_i \tag{7}$$

from Eqs 2 and 3. Letting measurement noise have variance, σ_e^2, then the measured variance $\mathrm{Var}[Y_t]$ will be increased by σ_e^2 above the value in Eqs 7, and 6 becomes

$$\hat{x}_0 = \frac{\sum_i \nu_i \ell_i \sum_j (\nu_j - \nu_j \ell_j)}{\sigma_e^2 + \sum_i \nu_i^2 \ell_i (1 - \ell_i)}. \tag{8}$$

The ν_i and ℓ_i for each molecule are unknown, and to proceed we assume that they are samples from a probability distribution, $P(\nu, \ell)$. Then $\sum_i \nu_i \ell_i$ is an empirical estimate for $x_0 \mathrm{E}[\nu \ell]$ because there are x_0 terms in the sum. We can write Eq. 8 as

$$\hat{x}_0 = \frac{x_0}{1 + \epsilon} \tag{9}$$

where

$$\epsilon = \frac{\frac{\sigma_e^2}{x_0} + \text{Cov}[\nu, \nu\ell] - \text{Var}[\nu\ell]}{\text{E}[\nu\ell](\text{E}[\nu] - \text{E}[\nu\ell])} \tag{10}$$

after re-arranging. Equations 9 and 10 imply that $\hat{x}_0$ underestimates x_0 if measurement noise is sufficiently large,

$$\frac{\sigma_e^2}{x_0} + \text{Cov}[\nu, \nu\ell] > \text{Var}[\nu\ell] \tag{11}$$

and overestimates if both measurement noise is sufficiently small and $\text{Var}[\nu\ell] > \text{Cov}[\nu, \nu\ell]$ such that $\epsilon < 0$.

For the estimates to be accurate, $|\epsilon| \ll 1$, which holds if

$$\left| \frac{\sigma_e^2}{x_0} + \text{Cov}[\nu, \nu\ell] - \text{Var}[\nu\ell] \right| \ll \text{E}[\nu\ell](\text{E}[\nu] - \text{E}[\nu\ell]). \tag{12}$$

If σ_e is too large, the estimate of x_0 fails.

If ν is homogeneous so that $\nu = \text{E}[\nu]$ for all molecules, then Eq. 12 simplifies to

$$\frac{\sigma_e^2}{\text{E}[\nu]^2 x_0} \ll \text{E}[\ell] - \text{E}[\ell]^2 + \text{Var}[\ell]. \tag{13}$$

We note that $\text{E}[\ell] > \text{E}[\ell^2]$ because $0 < \ell < 1$. Equation 13 is most easily satisfied midway through the bleaching when $\text{E}[\ell] = 1/2$ and the difference between $\text{E}[\ell]$ and $\text{E}[\ell]^2$ is maximum.

If $\ell = \text{E}[\ell]$ for all molecules, then Eq. 12 simplifies to

$$\frac{\sigma_e^2}{\text{E}[\nu]^2 x_0} \ll (\text{E}[\ell] - \text{E}[\ell]^2)\left(1 - \frac{\text{Var}[\nu]}{\text{E}[\nu]^2}\right). \tag{14}$$

Equation 14 implies that Eq. 6 fails if the variation in ν is so high that $\text{Var}[\nu] > E[\nu]^2$ and again is most easily satisfied midway through the bleaching.

In summary, our estimator, Eq. 6, can work even in the completely heterogenous case where each molecule has its own brightness and rate of bleaching, but is more sensitive to variation in ν than in ℓ. Measurement noise if sufficiently high will cause underestimation and if too high will, as expected, undermine accuracy.

Using the estimator.   After correcting for autofluorescence, flatfield (inhomogeneous illumination), and background (Methods), the data set comprises n_r cells with a time series of n_d fluorescence values for each cell. We initially analyze the cells one at a time. We consider the time series for the first cell as a collection of pairs of measurements – (y_0, y_t) – with one pair for each positive time point and apply Eq. 6 to each pair. To find $\text{E}[Y_t]$ and $\text{Var}[Y_t]$ in Eq. 6, we smooth the time series using a Gaussian process[25]. From the smoothed time series, we estimate $\text{E}[Y_t]$ for all t as the mean of the Gaussian process. Given this mean, we estimate $\text{Var}[Y_t]$ at each t as $(y_t - \text{E}[Y_t])^2$. Equation 6 can then be used directly, and we obtain $n_d - 1$ estimates of $\hat{x}_0$ for that cell. We repeat this process for all cells and therefore obtain $n_r(n_d - 1)$ estimates of $\hat{x}_0$ in total, which we pool. We take the mode of the resulting distribution as the best estimate for the number of molecules.

Comparison with biochemical estimates of protein numbers.   To verify our method, we compare our results with biochemical measurements of the numbers of molecules. We selected six proteins from budding yeast that have a range of absolute numbers – Fus3, Hog1, Guk1, Def1, Gpm1, and Pgk1 – and subjected GFP-fusions of these proteins[20] to a photobleaching analysis (Methods). We compare with a unified data set from 19 separate biochemical experiments[26], involving either mass spectrometry, GFP-tagging and microscopy, or TAP-tagging with immunoblotting.

For all proteins, there is an agreement within an order of magnitude between the two approaches (Fig. 3) showing that the measurement noise is not prohibitively large, at least for yeast.

Discussion

A challenge in quantitative fluorescence microscopy is converting measurements of fluorescence to absolute units. Absolute units enable the pooling of data from different laboratories and are needed both for models to fit single-cell data[4] and to validate the values of fitted parameters[27]. As such, absolute units facilitate the long-term success of systems and synthetic biology[28,29].

We have presented a fluctuation analysis to estimate numbers of fluorescent proteins by deliberately photobleaching cells on a wide-field fluorescence microscope. The method provides a straightforward calibration for time-lapse imaging.

Although our estimate is within an order of magnitude of the numbers estimated using biochemical methods, it is often an underestimate (Fig. 3). This behaviour is consistent with a sufficiently large measurement noise
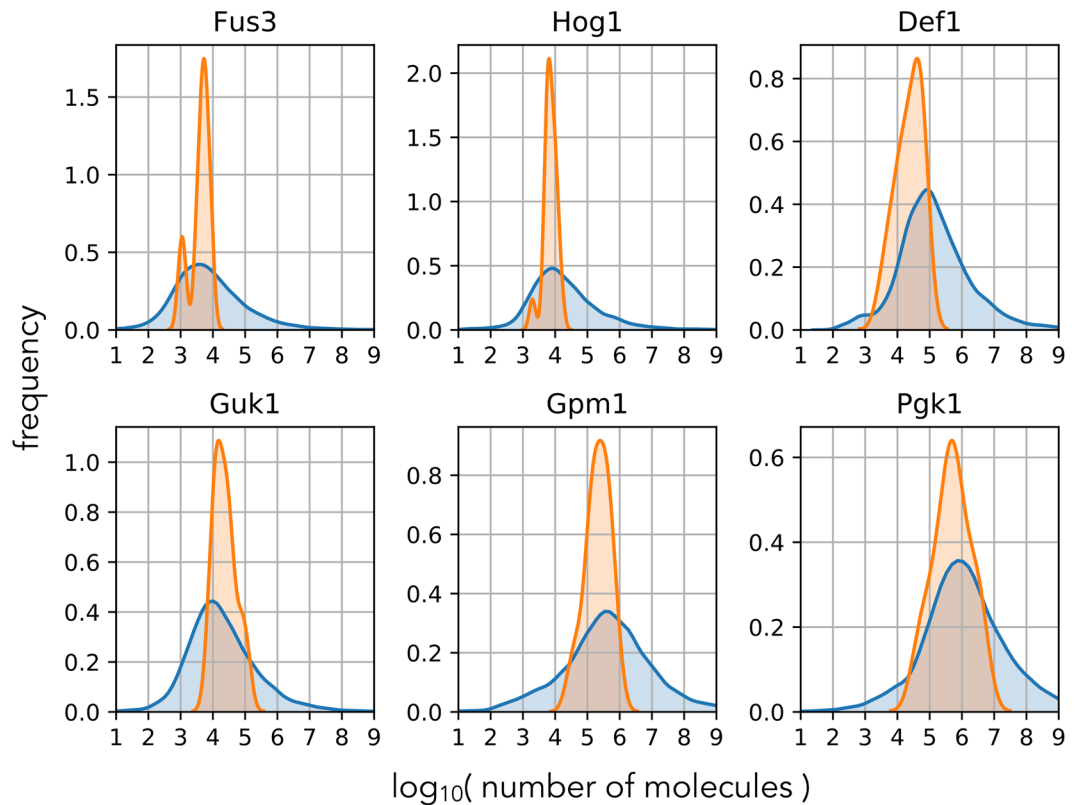
Figure 3. A photobleaching analysis identifies numbers of molecules in agreement with biochemical measurements. We show the distribution comprising the estimates of x_0 for all pairs of y_0 and y_t for $t > 0$ in blue and the distribution of results from 19 biochemical experiments in orange. The numbers of molecules range from approximately 10^3 (Fus3) to 10^6 (Pgk1). For all proteins, the mode of the two distributions are within an order of magnitude. Each cell was imaged 45 times in 10 s intervals, and the numbers of cells are as follows: 355 for Fus3; 114 for Hog1; 239 for Def1; 310 for Guk1; 242 for Gpm1; and 311 for Pgk1.

(Eq. 11). Nevertheless, underestimation is countered by the estimate of $\mathrm{Var}[Y_t]$ in Eq. 6. Using smoothing to find $\mathrm{E}[Y_t]$ and so $\mathrm{Var}[Y_t]$ as a function of time from a single time series can underestimate $\mathrm{Var}[Y_t]$ because the smoothed data – the estimate of $\mathrm{E}[Y_t]$ – will follow a sufficiently large fluctuation whereas the true mean will not. A too small $\mathrm{Var}[Y_t]$ overestimates the numbers of molecules. In practice, the measurement noise appears sufficiently high to disrupt large fluctuations so that our estimates of $\mathrm{E}[Y_t]$ do approximate the true mean, but not too high to undermine using Eq. 6.

We might expect an underestimation too because not all tagged proteins fluoresce. Considering the extreme case of a (long) maturation time of the fluorescent protein of 45 minutes and that its degradation is caused by growth with a (short) doubling time of 100 minutes[30], then we expect the fluorescence at steady-state to be at least $1/45/(1/45 + 1/100) \simeq 0.7$ of the total amount of protein. This fraction is, however, large enough that the order of magnitude of our estimate is unchanged and lies within the expected error (width of the blue distributions in Fig. 3).

It is surprising that more principled methods, such as Bayesian inference, work poorly (Methods). Although a single exponential decay does not describe photobleaching in yeast (Fig. 2), we know neither how many exponential decays do nor even if exponential decay itself is correct. Further, both the brightness and time scales associated with any exponentials may be the same for all cells in the population or specific to each cell. Compounding the uncertainty in the model of bleaching, there is no consensus on a model for the measurement noise for fluorescence microscopes, which has been described as normal[10], log-normal[4], and with a variance that is a function of fluorescence[31] (and Methods).

Measurements both in single cells and in absolute numbers are necessary to bring discoveries from different laboratories into one predictive framework. Our method of calibrating fluorescence measurements will help enable such a quantitative, single-cell biology.

Methods
Selecting proteins to study. When selecting proteins to test, we looked at three whole-proteome datasets of absolute numbers of protein obtained by quantitative Western blot[32], mass spectrometry[33], and fluorescence microscopy[34]. Proteins were selected that appeared in all three data sets. To obtain proteins whose levels would be robust to any stresses from our fixing procedure, we used a protein localisation atlas[35] to select cytoplasmic proteins that showed no significant fold change under starvation and dithiothreitol- and peroxide-induced stress.

From these proteins, we selected four giving a broad range of levels: Def1, Guk1, Gpm1, and Pgk1. To these proteins we added Hog1 because of its regular study in our laboratory[36] and Fus3 because of the availability of a measurement by fluorescence correlation spectroscopy (FCS)[37]. Taking a cytoplasmic concentration of 180 nM for *Saccharomyces cerevisiae* cells and a three times higher concentration for the nucleus[37] along with cellular and nuclear volumes of 42 μl and 3 μl[30], we estimate 5,200 molecules of Fus3 per cell from this FCS data.

During the course of our work, however, a comprehensive collection of whole-proteome data sets was published[26], and this data is the data we use for comparison. Our estimate for Fus3 from the FCS data is within the range of values reported in this data set.

Cell preparation.   Cells from the open-reading-frame GFP collection[20] were grown overnight in YEPD (2%) media past the diauxic lag, and 0.5 ml of this culture diluted in 5 ml of fresh media. After 5 hours and at an OD of $\simeq 0.5$, the cells were fixed[38].

Microscopy and image analysis.   All experiments were performed on a Nikon Eclipse Ti inverted microscope, controlled with the Perfect Focus System and custom MATLAB scripts (Mathworks) written for Micromanager[39]. We used a 60X 1.2 NA water immersion objective (Nikon), and images were acquired using an Evolve camera (Photometrics) with a 512×512 sensor in CCD mode. Cells were adhered to slides using concanavalin A.

To photobleach, the GFP excitation LED was kept at full power for the duration of the experiment with cells imaged for fluorescence for 400 ms every 10 s. We repeated this procedure for multiple, isolated fields of view and for wild-type cells, which do not express GFP.

Given that we imaged fixed cells, cells were selected and segmented at the first time point from a bright-field image and whole-image registration used to propagate this outline to other times. Cells in the bright-field image were chosen by eye to be isolated, well focused, present for the whole experiment (i.e. not washed away), and in a region where the illumination intensity (taken from the flat field correction) was at least 80% of the median illumination. Selected cells were outlined based on the out-of-focus bright-field image using custom MATLAB scripts, and these outlines curated by hand.

Cell fluorescence was calculated as the sum of the values of the brightest 80% of the pixels within the cell boundary. This measure reduced the effect of movements of the stage. We discarded cells that displayed large systematic deviations (such as a sudden drop in fluorescence).

Correcting for flat field and background.   Fluorescence images were corrected for both flat field and background. We obtained a flat-field image by flowing 0.001% fluorescein (by mass) through a microfluidic device[40] and imaging multiple positions over a time course. Any microfluidic features were 'blotted out' of the images, and the modified images averaged and normalised to a median of one. A correction was applied to the fluorescence images through an element-wise division by the flat field. To remove any background particular to a slide, we subtracted the average pixel fluorescence for a region of the image containing no cells from each pixel value. Fluorescence images were also registered to correct for drift in the field of view.

Correcting for autofluorescence.   Cells were also corrected for autofluorescence. We corrected bleached wild-type cells for flat field and background, and their average value was subtracted from all fluorescent cells at each time point.

Inference fails assuming mono-exponential decay.   Analysing the data using a previously proposed estimator[14] and following the procedure of Kim *et al.*[17], we find that numbers of proteins are underestimated by several orders of magnitude (the maximum number of molecules over all six proteins is approximately 150). This failure is because that method both uses a poor approximation to the mean behaviour of the photobleaching of individual cells (the means in our data are not well described by exponential decay with one time scale) and its sensitivity to the quality of the fit even when the photobleaching is a decay process with a single time scale.

A maximimum likelihood estimator for x_0.   If there is negligible measurement noise, the data comprises y_0, $\mathrm{E}[Y] = \mu_y$, and $\mathrm{Var}[Y] = \sigma_y^2$, and the likelihood of x_0 is $P(y_0, \mu_y, \sigma_y^2 | x_0)$, which we denote $\mathcal{L}$, then we can write explicit expressions.

Using the rules of probability, and writing $\ell = \mathrm{e}^{-\lambda t}$ for a particular λ and t,

$$\begin{aligned}
\mathcal{L} &= \int d\nu\, d\mu_x\, d\sigma_x^2\, d\ell\, P(y_0, \mu_y, \sigma_y^2 | \nu, \mu_x, \sigma_x^2, \ell, x_0) \\
&\quad \times P(\nu, \mu_x, \sigma_x^2, \ell | x_0) \\
&= \int d\nu\, d\mu_x\, d\sigma_x^2\, d\ell\, P(y_0, \mu_y, \sigma_y^2 | \nu, \mu_x, \sigma_x^2, \ell, x_0) \\
&\quad \times P(\sigma_x^2 | \nu, \mu_x, \ell, x_0) P(\nu, \mu_x, \ell | x_0) \\
&= \int d\nu\, d\mu_x\, d\sigma_x^2\, d\ell\, P(y_0 | \nu, x_0) P(\mu_y | \mu_x, \nu) P(\sigma_y^2 | \sigma_x^2, \nu) \\
&\quad \times P(\sigma_x^2 | \mu_x, \ell, x_0) P(\mu_x | \ell, x_0) P(\ell) P(\nu).
\end{aligned}$$

Assuming $P(\ell)$ is uniform between 0 and 1, we can write

$$\mathcal{L} = \int d\nu \, d\mu_x \, d\sigma_x^2 \, d\ell \, \delta(y_0 - \nu x_0)\delta(\mu_y - \nu\mu_x)\delta(\sigma_y^2 - \nu^2\sigma_x^2)$$
$$\times \, \delta(\sigma_x^2 - \mu_x(1 - \ell))\delta(\mu_x - x_0\ell)P(\nu) \tag{15}$$

following Eqs 2 and 3.

We use the identity

$$\delta(a - \nu x) = \frac{1}{\nu}\delta\left(x - \frac{a}{\nu}\right) \tag{16}$$

to evaluate the integrals.

First, we perform the integral over μ_x using the first delta function in μ_x

$$\mathcal{L} = \int d\nu \, d\sigma_x^2 \, d\ell \frac{1}{\nu}\delta(y_0 - \nu x_0)\delta(\sigma_y^2 - \nu^2\sigma_x^2)$$
$$\times \, \delta\left(\sigma_x^2 - \frac{\mu_y}{\nu}(1 - \ell)\right)\delta\left(\frac{\mu_y}{\nu} - x_0\ell\right)P(\nu) \tag{17}$$

and then over σ_x^2

$$\mathcal{L} = \int d\nu \, d\ell \frac{1}{\nu^3}\delta(y_0 - \nu x_0)\delta\left(\frac{\sigma_y^2}{\nu^2} - \frac{\mu_y}{\nu}(1 - \ell)\right)\delta\left(\frac{\mu_y}{\nu} - x_0\ell\right)P(\nu). \tag{18}$$

Integrating over ℓ, we have

$$\mathcal{L} = \int d\nu \frac{1}{\nu^3 x_0}\delta(y_0 - \nu x_0)\delta\left(\frac{\sigma_y^2}{\nu^2} - \frac{\mu_y}{\nu}\left(1 - \frac{\mu_y}{\nu x_0}\right)\right)P(\nu) \tag{19}$$

and over ν gives

$$\mathcal{L} = \frac{x_0}{y_0^3}\delta\left(\frac{\sigma_y^2 x_0^2}{y_0^2} - \frac{\mu_y x_0}{y_0}\left(1 - \frac{\mu_y}{y_0}\right)\right)P(y_0/x_0)$$
$$= \frac{1}{y_0\sigma_y^2}\delta\left(x_0 - \frac{\mu_y(y_0 - \mu_y)}{\sigma_y^2}\right)P(y_0/x_0) \tag{20}$$

where we have brought x_0/y_0^2 into and σ_y^2 out of the delta function.

For a uniform prior for ν (so that $P(y_0/x_0)$ is constant), the likelihood, Eq. 20, is maximized when

$$x_0 = \frac{\mu_y(y_0 - \mu_y)}{\sigma_y^2} \tag{21}$$

which is Eq. 6.

Estimating E[Y_t]. We smooth each cell's time series, y_t, to estimate its mean, E[Y_t]. To perform the smoothing, we use a Gaussian process with a covariance given by a squared exponential function – $k(x, x') = \theta_0 \exp[-\theta_1(x - x')^2/2]$ – and determine the hyperparameters for each cell by maximizing the marginal likelihood[25]. These hyperparameters are bounded a priori: $10^3 < \theta_0 < 10^{14}$, $10^{-8} < \theta_1 < 1$, and θ_2, which controls the magnitude of the estimated measurement noise, being restricted to $10 < \theta_2 < 10^{10}$. We use an implementation in Python[41].

Using simulated data and normally distributed measurement noise, we find that there is an optimum range for the magnitude of the measurement noise. If measurement noise is too low, the estimated mean follows fluctuations in the data and at times is closer to the data than the true mean. Therefore Var[Y_t] in Eq. 6 is too small, and the estimate of x_0 is too large. If measurement noise is too high, the estimated mean can sometimes be further from and sometimes closer to the data than the true mean, but the estimator in any case then underestimates (Eq. 9). For intermediate magnitudes of the measurement noise, Eq. 6 is accurate, and the measurement noise prevents fluctuations being long-lived and so corrects for bias in the estimate of E[Y_t] and Var[Y_t].

The estimator can perform poorly if the time scales of bleaching are shorter than the the duration of the experiment. Then, all the molecules are bleached at late times, and the resulting noisy data can undermine the estimate of E[Y_t]. The low numbers of molecules also mean that the linear noise approximation and Eqs 2 and 3 are potentially no longer valid.

Inference with an explicit model of bleaching and measurement noise. In general, given data $\mathbf{Y}_t = \{y_1, \ldots, y_t\}$, we wish to infer $\boldsymbol{\theta} = \{\boldsymbol{\theta}_m, \boldsymbol{\theta}_e\}$, where $\boldsymbol{\theta}_m$ are the parameters for the biophysical model of the dynamics of the underlying numbers of proteins, $\mathbf{x}$, and $\boldsymbol{\theta}_e$ are the parameters for the distribution of the measurement noise. The y variables are related to the $\mathbf{x}$ variables only through this measurement noise.

The dynamics of the protein numbers, $\mathbf{x}$, are determined by chemical reactions and can be described by a master equation for $P(\mathbf{x}, t)$, the probability distribution of the state $\mathbf{x}$ at time t[19]. We use the linear noise approximation (first-order terms in an expansion of the master equation in the size of the system – the volume of a cell[19]). This approximation makes $P(\mathbf{x}, t)$ a normal distribution if the initial distribution is either a normal or a delta function.

If we let the stochiometric matrix be $\mathbf{S}$ and the hazards (propensities) be $\mathbf{h}$ and noting that $\mathbf{x}$ describes numbers of molecules not concentrations, then

$$P(\mathbf{x}, t) = \mathcal{N}(\mathbf{x}; \mu(t), \mathbf{\Sigma}(t)) \tag{22}$$

where $\mathcal{N}$ denotes a normal distribution and $\mu(t)$, its mean, and $\mathbf{\Sigma}(t)$, its covariance matrix, obey[24]

$$\frac{d}{dt}\mu = \mathbf{S}^T\mathbf{h}(\mu, t) \tag{23}$$

and

$$\frac{d}{dt}\mathbf{\Sigma} = \mathbf{J}\mathbf{\Sigma} + \mathbf{\Sigma}\mathbf{J}^T + \mathbf{S}^T\mathbf{H}\mathbf{S} \tag{24}$$

with $\mathbf{J}$ as the Jacobian:

$$J_{ij} = \frac{\partial}{\partial x_j}\sum_k S_{ik}h_k(\mu, t) \tag{25}$$

and $\mathbf{H}$ as a matrix of zeros with $\mathbf{h}(\mu, t)$ on the diagonal.

We let the biophysical model have two decay processes (Fig. 2). For each cell, indexed by j, there are two pools of fluorescence proteins, $x_{1,j}$ and $x_{2,j}$, which bleach at constant, but cell-dependent, rates, $\lambda_{1,j}$ and $\lambda_{2,j}$:

$$x_{1,j} \xrightarrow{\lambda_{1,j}} \varnothing \quad x_{2,j} \xrightarrow{\lambda_{2,j}} \varnothing \tag{26}$$

Equations 23 and 24 then give

$$\mu_{i,j} = x_{i,j}^{(0)}e^{-\lambda_{i,j}t} \tag{27}$$

and

$$\begin{aligned}
\Sigma_{ii,j} &= \Sigma_{ii,j}^{(0)}e^{-2\lambda_{i,j}t} + e^{-\lambda_{i,j}t}(1 - e^{-\lambda_{i,j}t})x_{i,j}^{(0)} \\
\Sigma_{12,j} &= \Sigma_{12,j}^{(0)}e^{-(\lambda_{1,j}+\lambda_{2,j})t}
\end{aligned} \tag{28}$$

where initial values have superscripts of zero.

We use a measurement noise that depends on the numbers of fluorescence proteins giving a standard deviation that scales with the mean. If $y_j(t)$ is the measured fluorescence then

$$\begin{aligned}
y_j(t) \sim \ &\mathcal{N}(y_j(t); \nu[x_{1,j}(t) + x_{2,j}(t)] + f_j, \\
&\sigma_{e,0}^2 + \nu[x_{1,j}(t) + x_{2,j}(t)]\sigma_{e,1}^2 + \nu^2[x_{1,j}(t) + x_{2,j}(t)]^2\sigma_{e,2}^2)
\end{aligned} \tag{29}$$

for constant $\sigma_{e,i}$ and any residual autofluorescence f_j. This model does not have the positive skewness of log-normal noise and has support for negative fluorescence values, which we observe after correcting images for background fluorescence.

Linear noise and sequential data – a Kalman filter.

We wish to infer $\boldsymbol{\theta}$ given $\mathbf{Y}_t$. Bayes's rule states:

$$P(\boldsymbol{\theta}|\mathbf{Y}_t) \propto P(\mathbf{Y}_t|\boldsymbol{\theta})P(\boldsymbol{\theta}) \tag{30}$$

or, by using the rules of probability to factorize the likelihood,

$$P(\boldsymbol{\theta}|\mathbf{Y}_t) \propto P(\boldsymbol{\theta}) \prod_{i=1}^{t} P(y_i|\mathbf{Y}_{i-1}, \boldsymbol{\theta}). \tag{31}$$

We sequentially find each term in Eq. 31 by considering the dynamics of $\mathbf{x}$ from one time point to the next and then correcting that dynamics given the observed data[24]. Assume that at time point $i - 1$ the distribution $P(\mathbf{x}_{i-1}|\mathbf{Y}_{i-1}, \boldsymbol{\theta})$ is normal with a known mean μ_{i-1}^* and covariance matrix $\mathbf{\Sigma}_{i-1}^*$:

$$P(\mathbf{x}_{i-1}|\mathbf{Y}_{i-1}, \boldsymbol{\theta}) = \mathcal{N}(\mathbf{x}_{i-1}; \mu_{i-1}^*, \mathbf{\Sigma}_{i-1}^*). \tag{32}$$

Using the linear noise approximation for the dynamics of $\mathbf{x}$, we can, with Eq. 32 providing the initial condition, integrate Eqs 23 and 24 over one time interval to time point i to find μ_i and $\mathbf{\Sigma}_i$ and that

$$P(\mathbf{x}_i | \mathbf{Y}_{i-1}, \boldsymbol{\theta}) = \mathcal{N}(\mathbf{x}_i; \mu_i, \boldsymbol{\Sigma}_i). \tag{33}$$

We next wish to extend the conditioning in Eq. 33 to include the data point, y_i, at time point i. To do so, note that

$$
\begin{aligned}
P(\mathbf{x}_i | \mathbf{Y}_i, \boldsymbol{\theta}) &= P(\mathbf{x}_i | y_i, \mathbf{Y}_{i-1}, \boldsymbol{\theta}) \\
&= \frac{P(y_i | \mathbf{x}_i, \mathbf{Y}_{i-1}, \boldsymbol{\theta}) P(\mathbf{x}_i | \mathbf{Y}_{i-1}, \boldsymbol{\theta})}{P(y_i | \mathbf{Y}_{i-1}, \boldsymbol{\theta})} \\
&\propto P(y_i | \mathbf{x}_i, \boldsymbol{\theta}) P(\mathbf{x}_i | \mathbf{Y}_{i-1}, \boldsymbol{\theta})
\end{aligned}
\tag{34}
$$

using Bayes's rule, conditioning on $\mathbf{Y}_{i-1}$ and $\boldsymbol{\theta}$, and assuming that the measurement noise only depends on the current value of $\mathbf{x}$. If the measurement noise too has a normal distribution

$$P(y | \mathbf{x}, \boldsymbol{\theta}) = \mathcal{N}(y; \mathbf{U}\mathbf{x}, \mathbf{V}) \tag{35}$$

with $\mathbf{U}$ being a constant projection matrix and $\mathbf{V}$ being a covariance matrix, then we can simplify Eq. 34 using the properties of normal distributions. We find that $P(\mathbf{x}_i | \mathbf{Y}_i, \boldsymbol{\theta})$ is also normal with a mean μ_i^* and a covariance $\boldsymbol{\Sigma}_i^*$ that satisfy[24]

$$
\begin{aligned}
\mu_i^* &= \mu_i + \boldsymbol{\Sigma}_i \mathbf{U}^T (\mathbf{U}\boldsymbol{\Sigma}_i \mathbf{U}^T + \mathbf{V})^{-1}(y_i - \mathbf{U}\mu_i) \\
\boldsymbol{\Sigma}_i^* &= \boldsymbol{\Sigma}_i + \boldsymbol{\Sigma}_i \mathbf{U}^T (\mathbf{U}\boldsymbol{\Sigma}_i \mathbf{U}^T + \mathbf{V})^{-1}(-\mathbf{U}\boldsymbol{\Sigma}_i).
\end{aligned}
\tag{36}
$$

The predictions of μ_i and $\boldsymbol{\Sigma}_i$ found from μ_{i-1}^* and $\boldsymbol{\Sigma}_{i-1}^*$ using the linear noise approximation are corrected to μ_i^* and $\boldsymbol{\Sigma}_i^*$ because of the new data point and the measurement noise.

The factors in Eq. 31, $P(y_i | \mathbf{Y}_{i-1}, \boldsymbol{\theta})$ obey

$$P(y_i | \mathbf{Y}_{i-1}, \boldsymbol{\theta}) = \int d\mathbf{x}_i P(y_i | \mathbf{x}_i, \boldsymbol{\theta}) P(\mathbf{x}_i | \mathbf{Y}_{i-1}, \boldsymbol{\theta}) \tag{37}$$

and are therefore the normalizing factors for Eq. 34, satisfying

$$P(y_i | \mathbf{Y}_{i-1}, \theta) = \mathcal{N}(y_i; \mathbf{U}\boldsymbol{\mu}_i, \mathbf{U}\boldsymbol{\Sigma}_i\mathbf{U}^T + \mathbf{V}). \tag{38}$$

Hence from a normal prior distribution for $\mathbf{x}_1$, $P(\mathbf{x}_1 | \boldsymbol{\theta})$, we use Eq. 38 to find $P(y_1 | \boldsymbol{\theta})$, the first term in the factorization of the likelihood (Eq. 31) and Eq. 36 to find $P(\mathbf{x}_1 | y_1, \boldsymbol{\theta})$, the starting normal distribution in Eq. 32 for the sequential inference.

To specialize the algorithm to photobleaching, the matrices in Eq. 35 are $\mathbf{U} = [\nu \ \ \nu]$ and $\mathbf{V} = \sigma_e^2$, a constant. We subtract the autofluorescence, f_j, from each data point before applying the Kalman filter. The Kalman update, Eq. 36, can result in unphysical, negative components of μ, which we set to zero.

We must specify a prior $P(\mathbf{x}_1 | \boldsymbol{\theta}_m)$ to begin the inference scheme. To do so, we introduce two parameters: x_0, which is the total amount of fluorescent protein at $t = 0$, and α, which is the partitioning of this fluorescent protein between the two pools. We infer both these parameters. For the Kalman filter, we require that $P(\mathbf{x}_1 | \boldsymbol{\theta}_m)$ be a normal distribution, but we wish to start $\mathbf{x}_1$ at a known value and so use

$$\mu_1^* = x_0 \begin{bmatrix} 1 - \alpha \\ \alpha \end{bmatrix} \text{ and } \Sigma_1^* = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}. \tag{39}$$

so that the normal distribution approximates a delta function.

Extending to state-dependent measurement noise. The variance of the measurement noise of Eq. 29 depends on $\mathbf{x}$, but our inference scheme assumes a constant variance (Eq. 35). We therefore approximate Eq. 29 by replacing the explicit dependence of the variance on $\mathbf{x}$ by its expected value given $\mathbf{Y}_{j-1}$:

$$
\begin{aligned}
y_j | \mathbf{x}_j &\sim \mathcal{N}(y_j; \mathbf{U}\mathbf{x}_j + f_j, \sigma_{e,0}^2 + \mathbf{U}\mathbf{x}_j \sigma_{e,1}^2 + (\mathbf{U}\mathbf{x}_j \sigma_{e,2})^2) \\
&\simeq \mathcal{N}(y_j; \mathbf{U}\mathbf{x}_j + f_j, \mathrm{E}_{\mathbf{x}_j | \mathbf{Y}_{j-1}}[\sigma_{e,0}^2 + \mathbf{U}\mathbf{x}_j \sigma_{e,1}^2 + (\mathbf{U}\mathbf{x}_j \sigma_{e,2})^2]) \\
&= \mathcal{N}(y_j; \mathbf{U}\mathbf{x}_j + f_j, \sigma_{e,0}^2 + \mathbf{U}\mu_j \sigma_{e,1}^2 + \mathbf{U}(\mu_j \mu_j^{\mathrm{T}} + \boldsymbol{\Sigma}_j)\mathbf{U}^{\mathrm{T}}\sigma_{e,2}^2)
\end{aligned}
\tag{40}
$$

The usual update in the Kalman filter, Eq. 36, can then be used.

Sampling the posterior probability. To sample from $P(\boldsymbol{\theta} | \mathbf{Y}_t)$ in Eq. 30 we use both optimization and a Markov chain Monte Carlo method.

We distinguish between heterogenous parameters, which are specific to each cell (λ_1, λ_2, f, x_0 and α), and homogeneous parameters, which have the same values for all cells (the $\sigma_{e,i}$ and ν). The homogenous parameters and x_0 have scale-free priors: for example, $P(\nu) = 1/\nu$. The heterogenous parameters other than x_0 have flat priors. All priors are proper and bounded to physical values. To ensure the expected behaviour is dependent only on
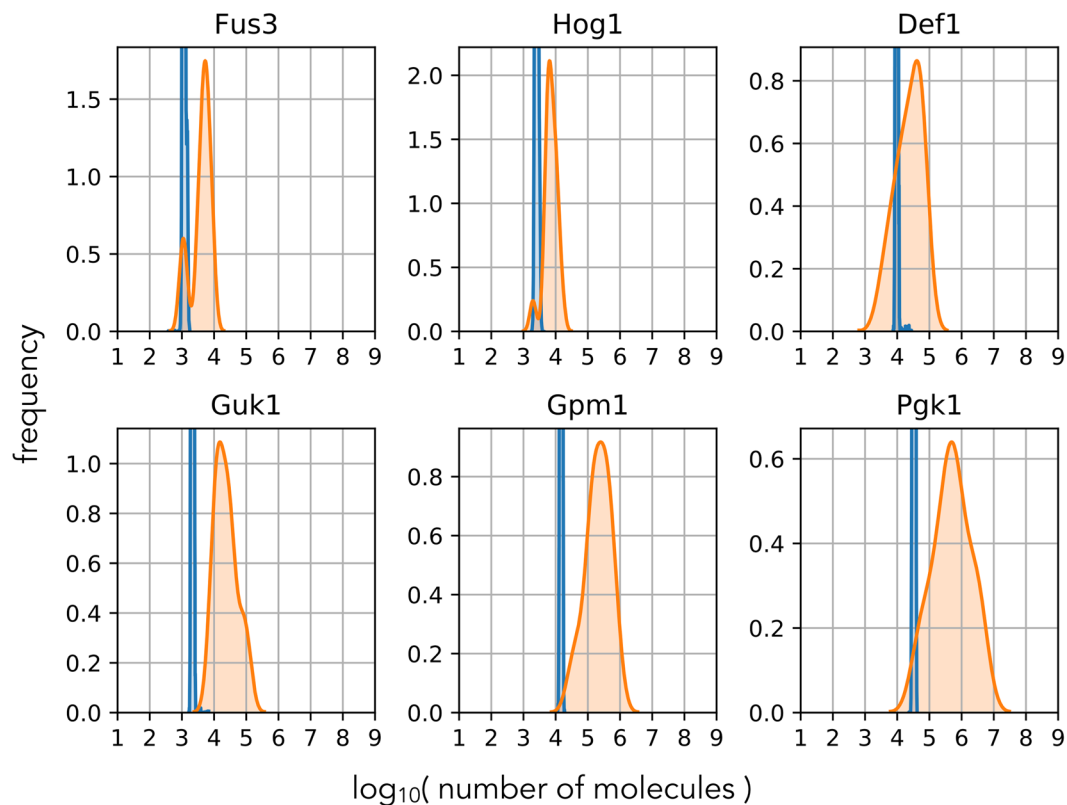
Figure 4. Inference with two heterogeneous decay processes and measurement noise systematically underestimates the numbers of molecules. We show the distribution of protein numbers estimated from the data and our inference of ν in blue and the distribution of results from 19 biochemical experiments in orange.

the heterogeneous parameters and to improve the mixing of the Markov chain, we propose the combination νx_0, referred to as y_0, rather than x_0.

To generate samples from the posterior, we use a Metropolis-within-Gibbs scheme[42] with the heterogeneous parameters updated separately from the homogeneous parameters. We employ adaptive parallel tempering to accelerate mixing in the Gibbs sampler[43], which performs well on benchmark biochemical models[44]. Specifically, we use 10 chains with their temperatures chosen adaptively. Parameters are proposed independently, with λ_1, λ_2, f and α proposed from normal distributions and y_0, the σ_{e_i}, and ν proposed from log-normal distributions. We adaptively select the scales for the proposal distributions for the heterogeneous and homogeneous parameters.

To start the Monte Carlo method, we try to find parameters that maximize the likelihood $P(\mathbf{Y}_i|\boldsymbol{\theta})$. We use a nested optimisation scheme[45]:

1. We find starting values for the heterogeneous parameters by fitting a bi-exponential decay to each cell's time series.
2. All parameters, including the heterogenous parameters, are then fitted for each cell, independently of all other cells, using a particle swarm.
3. We create an initial parameter set for the homogenous parameters by taking the median of the homogeneous parameters from the individual fits of Step 2.
4. We perform iterative optimisation: first optimising the homogeneous parameters with the heterogeneous parameters fixed then vice versa until a maximum is reached. Each individual optimisation uses Matlab's fmincon.
5. To provide diverse starting points for the chains, half of our chains are initialised at the parameter values found in Step 4 and the other half are initialised from optima found by performing the iterative optimisation of Step 4 from random parameters rather than from those found in Step 3.

Estimating the expected numbers of proteins.   Given a sample of parameters from the posterior distribution $P(\boldsymbol{\theta}|\mathbf{Y})$, and a fluorescence measurement y, we would like to find the posterior distribution of the number of proteins, x:

$$P(x|y, \mathbf{Y}) = \int d\boldsymbol{\theta} P(x|y, \boldsymbol{\theta}) P(\boldsymbol{\theta}|y, \mathbf{Y}) P(y). \tag{41}$$

We assume that the measurement y does not change the posterior probability of $\boldsymbol{\theta}$ (because there is substantially more data in $\mathbf{Y}$): $P(\boldsymbol{\theta}|y, \mathbf{Y}) \simeq P(\boldsymbol{\theta}|\mathbf{Y})$. Ignoring $P(y)$, which is independent of x, we have

$$P(x|y, \mathbf{Y}) \propto \int d\boldsymbol{\theta} P(x|y, \boldsymbol{\theta})P(\boldsymbol{\theta}|\mathbf{Y}). \tag{42}$$

By Bayes's rule

$$\begin{aligned} P(x|y, \boldsymbol{\theta}) &\propto P(y|x, \boldsymbol{\theta})P(x) \\ &\propto P(y|x, \boldsymbol{\theta}) \end{aligned} \tag{43}$$

assuming a constant prior, $P(x)$. We approximate νx by y in Eq. 29 so that

$$\begin{aligned} P(x|y, \boldsymbol{\theta}) &\propto \mathcal{N}(y; y + f, \sigma_{e,0}^2 + y\sigma_{e,1}^2 + (y\sigma_{e,2})^2) \\ &= \mathcal{N}\left(x; \frac{y - f}{\nu}, \frac{\sigma_{e,0}^2 + y\sigma_{e,1}^2 + (y\sigma_{e,2})^2}{\nu^2}\right) \end{aligned} \tag{44}$$

normalizing over x and using the properties of normal distributions. We use Eq. 44 to evaluate Eq. 42 as an average over the Monte Carlo samples generated from $P(\boldsymbol{\theta}|\mathbf{Y})$ on a grid of x values.

To extent these results to measurements of fluorescence over a population of cells where we are interested in the distribution of $\bar{x} = \sum x_i/N$, we can use that the distribution of a sum of independent normal variables is also normal with a mean equal to the sum of the means of the variables in the sum and with a variance equal to the sum of the variances[19]. Equation 44 then becomes

$$\begin{aligned} P(x|\mathbf{y}, \boldsymbol{\theta}) &= \mathcal{N}(\bar{x}; \frac{\bar{y} - \bar{f}}{\nu}, \frac{\sigma_{e,0}^2 + \bar{y}\sigma_{e,1}^2 + \overline{y^2}\sigma_{e,2}^2}{N\nu^2}) \\ &\simeq \delta(\bar{x} - \frac{\bar{y}}{\nu}) \end{aligned} \tag{45}$$

if $\sigma_{e,i}^2/N$ and $\bar{f}/\bar{y}$ are sufficiently small. Using a kernel density representation for $P(\bar{x}|\mathbf{y}, \mathbf{Y})$, with kernel $K(x, x') = \log \mathcal{N}(x; x', \sigma_K)$ and $\sigma_K^2 = 10^{-2}$, we can write that

$$\begin{aligned} P(x|\mathbf{y}, \mathbf{Y}) &= \int dx' K(\bar{x}, x')P(x'|\mathbf{y}, \mathbf{Y}) \\ &\simeq \int d\nu dx' K(\bar{x}, x')\delta\left(x' - \frac{\bar{y}}{\nu}\right)P(\nu|\mathbf{Y}) \\ &= \int d\nu K\left(\bar{x}, \frac{\bar{y}}{\nu}\right)P(\nu|\mathbf{Y}) \end{aligned} \tag{46}$$

using Eqs 42 and 45. We evaluate the integral in Eq. 46 using the Monte Carlo samples of $P(\nu|\mathbf{Y})$.

Combining estimates from different experimental replicates.

Writing $\{\mathcal{D}\} = \{\mathcal{D}_1, \mathcal{D}_2 \ldots \mathcal{D}_N\}$ as the set of data from all replicates, we wish to sample from $P(\boldsymbol{\theta}|\{\mathcal{D}\})$. We note that the datasets are conditionally independent given $\boldsymbol{\theta}$ so that

$$\begin{aligned} P(\boldsymbol{\theta}|\{\mathcal{D}\}) &\propto P(\boldsymbol{\theta}) \prod_i^N P(\mathcal{D}_i|\boldsymbol{\theta}) \\ &\propto P(\boldsymbol{\theta}) \prod^N \frac{P(\boldsymbol{\theta}|\mathcal{D}_i)}{P(\boldsymbol{\theta})} \\ &= \frac{\prod_i P(\boldsymbol{\theta}|\mathcal{D}_i)}{P(\boldsymbol{\theta})^{N-1}}. \end{aligned} \tag{47}$$

We approximate the $P(\boldsymbol{\theta}|\mathcal{D}_i)$ in Eq. 47 with kernel density estimates. If $\{\boldsymbol{\theta}_{i,1}, \ldots \boldsymbol{\theta}_{i,M}\}$ is the set of parameter samples from the posterior $P(\boldsymbol{\theta}|\mathcal{D}_i)$ obtained from our Monte Carlo method, then

$$P(\boldsymbol{\theta}|\mathcal{D}_i) \simeq \frac{1}{M} \sum_{j=1}^M K(\boldsymbol{\theta}, \boldsymbol{\theta}_{i,j}) \tag{48}$$

where the kernel functions are isotropic log normal distributions with variance σ_K^2 as before. In principle, these kernel density estimates allow Eq. 47 to be evaluated at any $\boldsymbol{\theta}$; in practice, we use a Metropolis algorithm to sample $\boldsymbol{\theta}$ to overcome the dimensionality of the parameter space.

Results

Our results give a lower bound on the number of molecules (Fig. 4). Note that here we ignore the first five time points of each time series because these points can have unexpectedly large fluctuations.

Modelling measurement noise. To determine the importance of the different contributions to Eq. 29, we used an informative log-normal prior for the parameter ν with its mode equal to the median of the value of ν estimated from the biochemical data on numbers and its empirical standard deviation equal to half of their inter-quartile range. Repeating the inference with this prior, we find that $\nu = 19.2$ (interquartile range: 19.0 to 19.5), $\sigma_{e,0} = 136.6$ (interquartile range: 135.1 to 138.5), $\sigma_{e,1} = 1.31$ (interquartile range: 1.29 to 1.32), and $\sigma_{e,2} = 0.0027$ (interquartile range: 0.0027 to 0.0028).

These results imply that the term proportional to the number of molecules in the variance of the measurement noise dominates the constant term. For all the proteins studied, the linear term (proportional to $\sigma_{e,1}$ in Eq. 29) is at least an order of magnitude larger that the constant term ($\sigma_{e,0}$), and the quadratic term (proportional to $\sigma_{e,2}$) dominates for proteins with high numbers (Gpm1 and Pgk1).

Data availability

Data generated in this work is available at https://doi.org/10.7488/ds/2594.

References

1. Verdaasdonk, J. S., Lawrimore, J. & Bloom, K. Determining absolute protein numbers by quantitative fluorescence microscopy. *Methods Cell Biol* **123**, 347–365 (2014).
2. Suter, D. D. M. *et al.* Mammalian genes are transcribed with widely different bursting kinetics. *Science* **332**, 472–4 (2011).
3. Harper, C. V. *et al.* Dynamic analysis of stochastic transcription cycles. *PLoS Biol* **9**, e1000607 (2011).
4. Zechner, C., Unger, M., Pelet, S., Peter, M. & Koeppl, H. Scalable inference of heterogeneous reaction kinetics from pooled single-cell recordings. *Nat Methods* **11**, 197–202 (2014).
5. Elson, E. L. & Magde, D. Fluorescence correlation spectroscopy I. Conceptual basis and theory. *Biopolymers* **13**, 1–27 (1974).
6. Magde, D., Elson, E. L. & Webb, W. W. Fluorescence correlation spectroscopy II. An experimental realization. *Biopolymers* **13**, 29–61 (1974).
7. Chen, Y., Muller, J. D., So, P. T. & Gratton, E. The photon counting histogram in fluorescence fluctuation spectroscopy. *Biophys J* **77**, 553–567 (1999).
8. Locke, J. C. & Elowitz, M. B. Using movies to analyse gene circuit dynamics in single cells. *Nat Rev Microbiol* **7**, 383–392 (2009).
9. Rosenfeld, N., Young, J. W., Alon, U., Swain, P. S. & Elowitz, M. B. Gene regulation at the single-cell level. *Science* **307**, 1962–1965 (2005).
10. Rosenfeld, N., Perkins, T. J., Alon, U., Elowitz, M. B. & Swain, P. S. A fluctuation method to quantify *in vivo* fluorescence data. *Biophys J* **91**, 759–766 (2006).
11. Teng, S.-W. *et al.* Measurement of the copy number of the master quorum-sensing regulator of a bacterial cell. *Biophys J* **98**, 2024–2031 (2010).
12. Zamparo, L. & Perkins, T. J. Statistical lower bounds on protein copy number from fluorescence expression images. *Bioinformatics* **25**, 2670–2676 (2009).
13. Finkenstädt, B. *et al.* Quantifying intrinsic and extrinsic noise in gene transcription using the linear noise approximation: An application to single cell data. *Ann Appl Stat* **7**, 1960–1982 (2013).
14. Nayak, C. R. & Rutenberg, A. D. Quantification of fluorophore copy number from intrinsic fluctuations during fluorescence photobleaching. *Biophys J* **101**, 2284–2293 (2011).
15. Nelson, S. R., Trybus, K. M. & Warshaw, D. M. Motor coupling through lipid membranes enhances transport velocities for ensembles of myosin Va. *Proc Nat Acad Sci USA* **111**, E3986–E3995 (2014).
16. Lombardo, A. T. *et al.* Myosin Va molecular motors manoeuvre liposome cargo through suspended actin filament intersections *in vitro*. *Nat Commun* **8**, 1–9 (2017).
17. Kim, N. H. *et al.* Real-time transposable element activity in individual live cells. *Proc Nat Acad Sci USA* **113**, 7278–7283 (2016).
18. Diaspro, A., Chirico, G., Usai, C., Ramoino, P. & Dobrucki, J. Photobleaching. In *Handbook of Biological Confocal Microscopy*, Springer, 3rd edition (2006).
19. Van Kampen, N. *Stochastic Processes in Physics and Chemistry*. Elsevier Inc., 3rd edition (2007).
20. Huh, W.-K. *et al.* Global analysis of protein localization in budding yeast. *Nature* **425**, 686–691 (2003).
21. Song, L., van Gijlswijk, R. P., Young, I. T. & Tanke, H. J. Influence of fluorochrome labeling density on the photobleaching kinetics of fluorescein in microscopy. *Cytometry* **27**, 213–23 (1997).
22. Taneja, S. & Rutenberg, A. D. Photobleaching of randomly rotating fluorescently decorated particles. *J Chem Phys* **147**, 104105 (2017).
23. Song, L., Hennink, E. J., Young, T. & Tanke, H. J. Photobleaching kinetics of fluorescein in quantitative fluorescence microscopy. *Biophys J* **68**, 2588–2600 (1995).
24. Fearnhead, P., Giagos, V. & Sherlock, C. Inference for reaction networks using the linear noise approximation. *Biometrics* **70**, 457–466 (2014).
25. Rasmussen, C. E. & Williams, C. K. I. *Gaussian Processes for Machine Learning*. (MIT Press, Cambridge, Massachusetts, 2006).
26. Ho, B., Baryshnikova, A. & Brown, G. W. Unification of protein abundance datasets yields a quantitative *Saccharomyces cerevisiae* proteome. *Cell Syst* **6**, 192–205 (2018).
27. Urquiza Garcia, U. *A mathematical model in absolute units for the Arabidopsis circadian oscillator*. PhD thesis (University of Edinburgh, 2018).
28. Kitano, H. Systems biology: a brief overview. *Science* **295**, 1662–1664 (2002).
29. Endy, D. Foundations for engineering biology. *Nature* **438**, 449–453 (2005).
30. Milo, R., Jorgensen, P., Moran, U., Weber, G. & Springer, M. BioNumbers–the database of key numbers in molecular and cell biology. *Nucl Acids Res* **38**, D750–753 (2010).
31. Newberry, M. V. Signal-to-noise considerations for sky-subtracted CCD data. *PASP* **103**, 122 (1991).
32. Ghaemmaghami, S. *et al.* Global analysis of protein expression in yeast. *Nature* **425**, 737–41 (2003).
33. Lu, P., Vogel, C., Wang, R., Yao, X. & Marcotte, E. M. Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation. *Nat Biotech* **25**, 117–124 (2007).
34. Chong, Y. *et al.* Yeast proteome dynamics from single cell imaging and automated analysis. *Cell* **161**, 1413–1424 (2015).
35. Breker, M., Gymrek, M. & Schuldiner, M. A novel single-cell screening platform reveals proteome plasticity during yeast stress responses. *J Cell Biol* **200**, 839–850 (2013).
36. Granados, A. A. *et al.* Distributing tasks via multiple input pathways increases cellular survival in stress. *Elife* **6**, 3649 (2017).

37. Maeder, C. I. *et al*. Spatial regulation of Fus3 MAP kinase activity through a reaction-diffusion mechanism in yeast pheromone signalling. *Nat Cell Biol* **9**, 1319–1326 (2007).
38. Bakker, E. *Quantitative fluorescence microscopy methods for studying transcription with application to the yeast GAL1 promoter*. PhD thesis (University of Edinburgh, 2017).
39. Edelstein, A., Amodaj, N., Hoover, K., Vale, R. & Stuurman, N. Computer control of microscopes using μmanager. *Curr Protoc Mol Biol* **92**, 14–20 (2010).
40. Crane, M. M., Clark, I. B. N., Bakker, E., Smith, S. & Swain, P. S. A microfluidic system for studying ageing and dynamic single-cell responses in budding yeast. *PLoS One* **9**, e100042 (2014).
41. Swain, P. S. *et al*. Inferring time derivatives including cell growth rates using Gaussian processes. *Nat Commun* **7**, 13766 (2016).
42. Neal, R. M. *Probabilistic inference using Markov chain Monte Carlo methods*. Technical report, University of Toronto (1993).
43. Miasojedow, B., Moulines, E. & Vihola, M. An adaptive parallel tempering algorithm. *J Comput Graph Stat* **22**, 649–664 (2013).
44. Ballnus, B. *et al*. Comprehensive benchmarking of Markov chain Monte Carlo methods for dynamical systems. *BMC Syst Biol* **11**, 63 (2017).
45. Llamosi, A. *et al*. What population reveals about individual cell identity: Single-cell parameter estimation of models of gene expression in yeast. *PLoS Comput Biol* **12**, e1004706 (2016).

Acknowledgements

Author contributions

E.B. performed the experiments; E.B. & P.S.S. developed the analysis and wrote the paper.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to P.S.S.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.