Estimating numbers of fluorescent molecules in single cells by analysing fluctuations in photobleaching

Elco Bakker*   Peter S. Swain *†

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 develop a general algorithm to infer numbers of molecules from fluctuations in the photobleaching of proteins tagged with Green Fluorescent Protein. To untangle measurement noise from stochastic fluctuations, we use the linear noise approximation and Kalman filtering within a framework of Bayesian inference. Not only do our results agree with biochemical measurements for multiple proteins in budding yeast, but we also provide a statistically verified model of measurement noise for fluorescence microscopes. 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.

Introduction

In fluorescence microscopy, converting measurements of fluorescence into numbers of molecules is a long-standing challenge. This deficit limits both the ability to combine fluorescence measurements from different experiments into one data set and the application of quantitative analyses of time-series that assumes the numbers of molecules are known [1, 2, 3, 4, 5].

Although methods exist to estimate the ratio between protein number and fluorescence, most require additional expertise and equipment beyond that typically needed for fluorescence microscopy [6]. Examples include fluorescence correlation spectroscopy [7, 8], image correlation spectroscopy [9], photon-counting histogram analysis [10], and fluorescence intensity distribution analysis [11], but all these approaches require at least a confocal microscope. Data is publicly available for alternative biochemical methods, such as quantitative Western blotting [12, 13] and mass spectrometry [14], but their accuracy is difficult to assess [15].

Ideally we would like to have a technique that requires only a wide-field fluorescence microscope and can be applied directly to the cells whose fluorescence is of interest. One promising set of approaches are fluctuation-based methods. These methods were first proposed in 1974 [7], but since then no standard approach has been developed for wide-field microscopy. Fluctuation analyses exploit that the magnitude of fluctuations in fluorescence are determined not by the concentration of the fluorescent molecules but by their numbers [16] and infer the fluorescence per molecule by analysing deviations from the mean (Fig. 1A).

Bespoke fluctuation techniques have been developed for wide-field microscopes, which do not focus on diffusion as fluorescence correlation spectroscopy does but instead use other stochastic events. There are two main approaches. One is to study fluctuations in the distribution of fluorescent proteins between daughter cells at cell division [17, 18, 19]. These techniques have been

---

*: SynthSys – Synthetic and Systems Biology, School of Biological Sciences, University of Edinburgh, Edinburgh EH9 3BF, U.K.; †: Corresponding author (peter.swain@ed.ac.uk)

mostly applied to bacteria, are unsuitable for non-dividing cells [20] and do not straightforwardly extend to species that exhibit differences in size between mothers and daughters. Further, the data must be obtained over several tens of divisions, which may be prohibitive for slowly dividing cells. The other approach is to study fluctuations in stochastic decay processes. Interventions to inhibit translation and transcription have allowed the fluorescence per molecule to be estimated in mammalian cells [4], but the stability of fluorescent proteins can make these experiments time consuming. A faster approach is to deliberately induce photobleaching [21]: the process whereby fluorophores cease to fluoresce when continuously excited [22]. This method has been applied in vitro [23, 24] and to bacteria [25], but the proposed analysis relies on photobleaching exhibiting an exponential decay [21], which is expected for single molecules but not for the fluorescence of cells [22].

Analysing such data is challenging. We must fit not the mean behaviour but fluctuations around the mean, and these fluctuations are typically confounded by measurement noise. To be successful, any inference procedure must be able to distinguish between measurement noise and stochastic effects (Fig. 1 inset).

Here we calibrate data from fluorescence microscopy through a Bayesian analysis of the fluctuations during photobleaching of eukaryotic cells. We model photobleaching as a multi-exponential process [22] and include measurement noise, unlike earlier analyses [21, 23, 25, 24], which neglect both. Using budding yeast, we show that photobleaching is indeed not described by a single exponential decay and introduce, fit, and statistically compare models that differ in their description of measurement noise. We therefore find both an empirically verified distribution for the measurement noise present in wide-field fluorescence microscopy, which can be used in the analysis of other data, and the fluorescence per molecule for proteins tagged with Green Fluorescent Protein (GFP). The corresponding predictions of protein numbers agree with estimates found by biochemical methods for the six proteins tested.
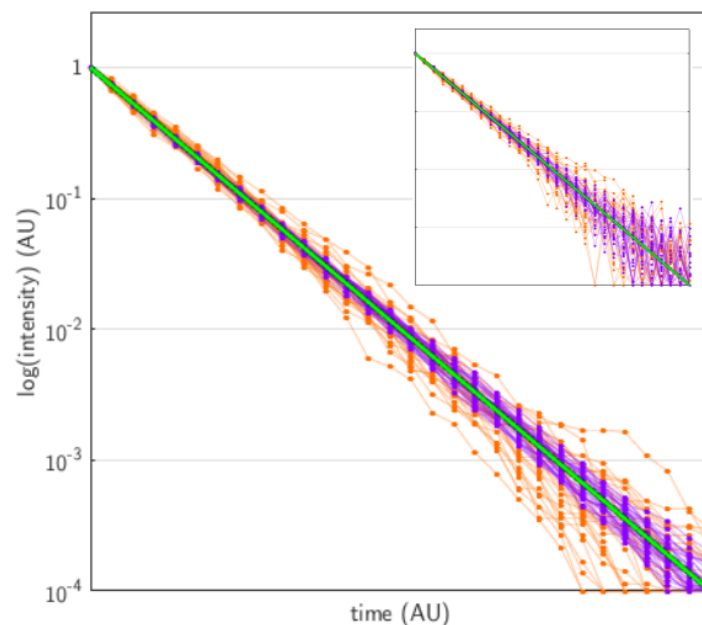

Figure 1. The magnitude of fluctuations is determined by numbers of molecules. Simulated time-series for stochastic exponential decay are normalized to the same starting value and plotted on a log scale for both low and high numbers of proteins. The mean behaviour is common (green line), but the data for low numbers of proteins (orange) shows larger deviations from the mean than the data for high numbers of proteins (purple). Inset: Stochastic fluctuations become masked with measurement noise (here an independent, additive Gaussian noise with zero mean and a standard deviation of 0.01): the orange and purple time-series now substantially overlap.

Results

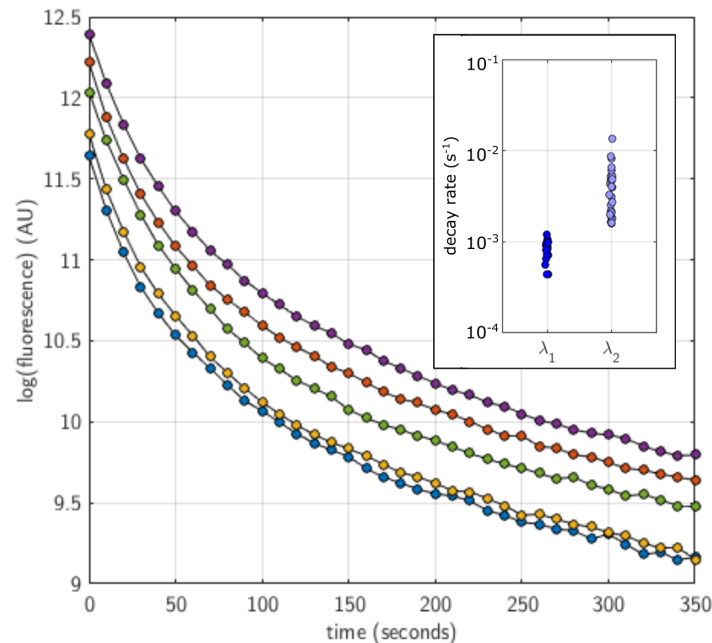## A biophysical model of photobleaching

Figure 2. Photobleaching is not described by a single exponential decay. The logarithm of the fluorescence of 5 cells of budding yeast expressing Pgk1-GFP and undergoing photobleaching does not fall on a straight line as a function of time (in contrast to the data of Fig. 1). The inset shows all the pairs of decay rates found from fitting a bi-exponential decay to the time-series of each cell (using maximum likelihood). There is substantial cell-to-cell variation (each circle represents a cell and notice the log scale on the y-axis).

We obtained time-series of photobleaching using budding yeast and GFP-tagged proteins [26]. Cells were fixed and photobleached in sustained illumination with fluorescence measurements taken every 10 seconds (Fig. 2 and Methods).

Our data is incompatible with earlier methodologies (Methods) because it is not well described by a single exponential decay (Fig. 2). First, the data for any individual cell 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 inset). Multi-exponential photobleaching is common [22, 27, 28, 29, 30] and can be caused by heterogeneous illumination [31], differing intracellular micro-environments [32], molecular rotation [33], and higher order interactions between excited fluorophores [34, 35]. These phenomena can also cause heterogeneity in parameters between cells if there is variation in either illumination or chemical composition across the population of cells.

We therefore consider each cell to have two populations of fluorescent proteins that photobleach with different rates. Writing the number of fluorescent proteins in cell j in the first population as $\mathbf{x}_{1,j}$ and the number of fluorescent proteins in the second population as $\mathbf{x}_{2,j}$, we model photobleaching as

$$x_{1,j} \xrightarrow{\lambda_{1,j}} \emptyset \qquad x_{2,j} \xrightarrow{\lambda_{2,j}} \emptyset \tag{1}$$

where $\lambda_{1,j}$ is the rate of photobleaching of the first population in cell j and $\lambda_{2,j}$ is the rate of photobleaching of the second population in that cell. We emphasise that these rates are specific to each cell.

3

Modelling measurement noise

To infer numbers of molecules, we require a model for the measurement noise generated by fluorescence microscopes.

A common description of this error is additive Gaussian noise [1, 2, 4, 18, 36]. Writing $\mathcal{N}(y; \mu, \sigma^2)$ for a Gaussian distribution in y with mean μ and standard deviation σ, $y_j(t)$ for the measured fluorescence of cell j at time point t, and ν for the fluorescence per molecule – the parameter that we wish to infer, we have

$$y_j(t) \sim \mathcal{N}\Big(y_j(t); \nu[x_{1,j}(t) + x_{2,j}(t)] + f_j, \sigma_e^2\Big) \tag{2}$$

where j again indexes an individual cell. Here σ_e is a constant, common to all cells, that determines the magnitude of the measurement noise, and we include a heterogeneous autofluorescence for each cell as f_j (Methods).

The additive Gaussian model for noise has been criticised, and multiplicative Gaussian noise [37] and log-normal noise [3] proposed as alternatives. Unlike additive Gaussian noise, both these models have the intuitively attractive property of a standard deviation that scales with the mean: we do not expect to have the same accuracy when measuring, for example, 100 and 10,000 proteins. This scaling is physically justified if the major sources of measurement noise are inconsistent illumination, segmentation errors, and small movements of the stage. Poisson noise, which is often used to model errors generated by CCD cameras [38], has the similar property that the standard deviation scales with the square root of the mean.

We therefore considered a second model for measurement noise, which we call polynomial state-dependent normal noise, in which the measurement is distributed as:

$$y_j(t) \sim \mathcal{N}\Big(y_j(t); \nu[x_{1,j}(t) + x_{2,j}(t)] + f_j,$$
$$\sigma_{e,0}^2 + \nu[x_{1,j}(t) + x_{2,j}(t)]\sigma_{e,1}^2 + \nu^2[x_{1,j}(t) + x_{2,j}(t)]^2\sigma_{e,2}^2\Big) \tag{3}$$

for constant $\sigma_{e,i}$. This noise model is general: if $\sigma_{e,1} = \sigma_{e,2} = 0$, the model recovers Eq. 2; if $\sigma_{e,1} = 0$, the model's standard deviation scales with the mean. The model does not have the positive skewness of log-normal noise and has support for negative fluorescence values, which can be observed after correcting images for background.

A fluctuation analysis

Due to their complexity, these models are no longer amenable to previous approaches [18, 21], and we instead use Bayesian methods for the analyses of time-series.

Briefly (see Methods for details), we employ the linear noise approximation [16, 39] to describe fluctuations in the dynamics of bleaching and combine this approximation with a Kalman filter [36, 40] to estimate the likelihood of the parameters. These parameters include homogenous parameters shared between all cells (ν and σ_e) and heterogeneous parameters that vary from cell to cell ($\lambda_{1,j}$, $\lambda_{2,j}$, and f_j). To move from this likelihood to inference, we first use an optimisation routine to estimate the parameter values with maximum likelihood and then take these values as a starting point for an adaptive Markov chain Monte Carlo scheme to sample the parameters' posterior distribution [41].

For additive Gaussian noise (Eq. 2), a standard Kalman filter can be used [40], but for state-dependent noise (Eq. 3), we use the predicted values of the protein concentrations at each time point to find an estimated value for the variance in Eq. 3. As the inference proceeds sequentially, this approximation means that the variance is known at each time point allowing the standard Kalman filter to be employed again, speeding up the analysis.

We verified the algorithm for both types of measurement noise using simulated data and use the evidence for each model, $P(\text{data}|\text{model})$, to discriminate between models [42].

To compare with biochemical measurements of protein numbers, we use the fluorescence data at the initial time-point before we bleach the cells and the posterior distribution for the parameters of the model given the bleaching data to infer the expected number of proteins in log-phase cells.

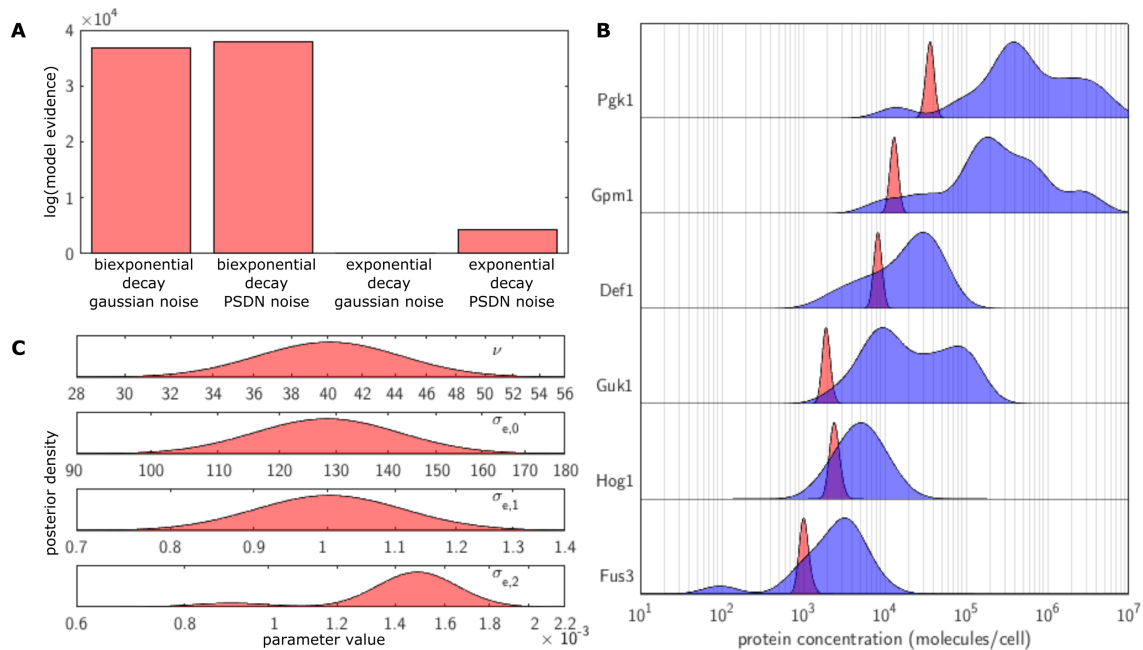## Comparison with biochemical estimates of protein numbers

Figure 3. For multiple proteins in budding yeast, estimates of numbers of proteins by our analysis of fluctuations in photobleaching agree with those obtained biochemically. (A) Considering the log evidence for the four models (either mono- or bi-exponential bleaching and either additive Gaussian or state-dependent measurement noise), the data strongly favour bleaching with a biexponential decay and a state-dependent measurement noise (the evidence for this model is at least 5000 dB higher than any other). We normalize so that the lowest log evidence is zero (for the model with mono-exponential bleaching and Gaussian measurement noise). (B) Our analyses, using the favoured model of Eqs. 1 and 3, predicts numbers of proteins in agreement with biochemical estimates. For each protein, the blue histogram is a kernel density estimate of the distribution of protein numbers measured by diverse biochemical methods [15]; the red histogram shows the distribution of inferred numbers for log-phase cells using our analysis (from multiple, replicate photobleaching experiments – Methods). (C) The posterior distributions for the homogenous parameters peak at physically reasonable values and provide, with Eq. 3, a model for the measurement noise of wide-field fluorescence microscopes. The plots show the marginal posteriors (from all replicates – Methods).

To test our method, we looked at biochemical measurements of the numbers of molecules for the proteome of budding yeast [15] and selected six proteins (Fus3, Hog1, Guk1, Def1, Gpm1 and Pgk1), which have a range of copy numbers. GFP fusions for these proteins [43] were subjected to our analysis.

The data favour both bi-exponential photobleaching and a state-dependent measurement noise. Running our inference procedure (Methods), we estimated the evidence, $P(\text{data}|\text{model})$ or the marginal likelihood, for four models: either mono- or bi-exponential bleaching and either additive Gaussian or state-dependent measurement noise. Following standard interpretations of the numerical values of evidence [44], bi-exponential photobleaching is highly favoured and a

5

state-dependent measurement noise preferred over Gaussian noise (Fig. 3A).

Using this model, we infer numbers of proteins in agreement with the biochemical estimates (Fig. 3B). There appears to be a general tendency to underestimate the number of molecules, especially for proteins with high-copy numbers, but there is considerable variation in the biochemical results, which may themselves be biased, and GFP-tagging could affect the stability of a protein and its mRNA.

Understanding the possible reasons the algorithm might give an underestimation requires developing an intuition for how protein number is inferred. 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 not informative about their individual values. The magnitude of the fluctuations in the underlying biochemical process changing x (here photobleaching) are expected to scale with the mean [16]: $\text{Var}[x] \sim E[x]$. Therefore the variance in the fluorescence, $y = \nu x$, scales as $\text{Var}[y] \sim \nu^2 E[x] = \nu E[y]$. For a given fluorescence (a given $E[y]$), large fluctuations therefore imply a high value of ν and correspondingly low numbers of proteins because $E[x] = E[y]/\nu$, and vice versa (Fig. 1). In our algorithm, we make the common assumption that the measurement noise on each measurement is independent. Any correlation in the measurement noise over time will therefore be interpreted as originating from bleaching. The magnitude of the fluctuations in bleaching will correspondingly be overestimated and the number of molecules underestimated. Physically, there are multiple possible sources of correlated error: for example, a gradual drift in the brightness of the illuminating LED or if the microscope slide or cells were to slowly move during the experiment. Furthermore, shortcomings in our biochemical model of bleaching will lead to a discrepancy between the model's expected behaviour and the cells' actual behaviour that would appear too as correlated errors.

Our method not only returns an estimate for the number of molecules but also posterior probabilities for the parameters necessary to describe a wide-field fluorescence microscope's distribution of measurement noise (Fig. 3C). It is difficult to find published data with which to compare these parameters. Nevertheless, the range of supported $\sigma_{e,1}$ values is close to 1, which is the value expected for Poissonian noise. These estimates are robust: using an informative prior for the brightness parameter, ν, the modes of the posterior distributions maintain the same order of magnitude and are shifted at most by a factor of two (Methods).

Discussion

A long standing problem in quantitative fluorescence microscopy is converting measurements of fluorescence to absolute units. Absolute units enable the data from different laboratories to be combined and are needed both for models to accurately fit single-cell data [3] and to assess the validity of the magnitude of fitted parameters [45]. As such, absolute units are an important facilitator for the long-term success of systems and synthetic biology [46, 47].

Here we have presented a fluctuation analysis that estimates the numbers of fluorescent proteins by deliberately photobleaching cells on a wide-field fluorescence microscope. Fluctuation analyses have long been used to estimate numbers of molecules on confocal microscopes, but the challenges of untangling measurement noise from fluctuations in stochastic biochemistry has prohibited a similar approach becoming a standard for the more commonly used wide-field fluorescence microscope.

Building on advances in the study of stochastic gene expression [40], we have used the linear noise approximation to model stochastic biochemical events combined with a Kalman filter and a state-dependent measurement noise to infer numbers of proteins from fluctuations in photobleaching. For budding yeast, we have shown that photobleaching has two distinct timescales, corresponding to two decay processes, but our framework remains valid for more complex reactions. The approximations developed allow faster inference than alternative approaches, such as particle-based methods, and the data we require can be gathered on a standard wide-

field fluorescence microscope, requiring no specialist equipment.

Further, we have rigorously compared different models of measurement noise, and we anticipate that the state-dependent measurement noise, Eq. 3, which is most consistent with our data, together with its inferred parameters (Fig. 3C), will underpin future work both for fitting mathematical models and for quantitative microscopy [48].

To bring together discoveries from different laboratories and to build those discoveries into a larger predictive framework, measurements both in single cells and in absolute numbers are necessary. We believe that methods that extend the fluctuation analyses available for confocal microscopes to wide-field microscopes, such as the one we present here, are key for a broad transition to a quantitative single-cell biology.

Methods

Selecting proteins to study

When selecting proteins to test our method, we looked at three whole-proteome datasets of absolute numbers of protein obtained by quantitative Western blot [26], mass spectrometry [14], and fluorescence microscopy [49]. 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 [50] 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 [51] and Fus3 because of the availability of a measurement by fluorescence correlation spectroscopy (FCS) [52]. Taking a cytoplasmic concentration of 180 nM for vegetative *Saccharomyces cerevisiae* cells and a three times higher concentration for the nucleus [52] along with cellular and nuclear volumes of 42 μl and 3 μl [53, 54, 55], we estimate 5200 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 [15], 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, being 1.75 times the median.

Cell preparation

Following [26], cells from the ORF GFP collection [43] 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 ~ 0.5, the cells were fixed [56].

Microscopy and image analysis

All experiments were performed on a Nikon Eclipse Ti inverted microscope controlled using custom MATLAB scripts (Mathworks) written for Micromanager [57] and the Perfect Focus System. We used a 60X 1.2 NA water immersion objective (Nikon), and images were acquired using an Evolve EMCCD camera (Photometrics) with a 512×512 sensor in CCD mode. Cells were adhered to slides using concanavalin A.

To effect photobleaching, the GFP excitation LED was kept at full power for the duration of the experiment with cells imaged for GFP every 10 seconds. This procedure was repeated for multiple, isolated fields of view for each slide. Wild-type cells, which do not express GFP, were also imaged.

Fluorescence images were corrected for flat field and background. We obtained a flat field image by flowing 0.001% fluorescein (by mass) through a narrow microfluidic device [58] and
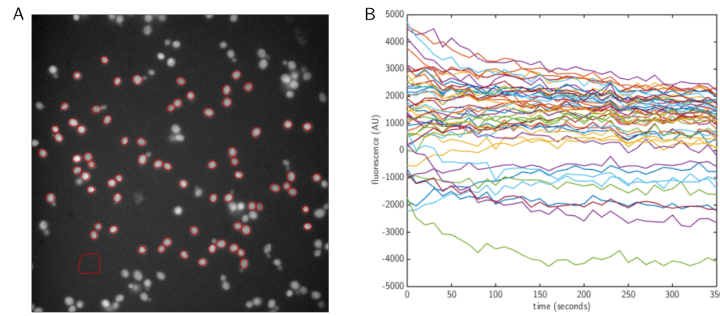
Figure 4. After correction for flat field and background, cells have a residual autofluorescence. (A) Examples of selected cells with their outlines in red. The area in the bottom left without cells was used to correct the fluorescence for background. (B) The residual fluorescence of each wild-type cell, after all corrections, has a non-zero mean, which we include in our models of measurement noise, Eqs. 2 and 3, as the heterogenous parameter f_j.

imaging multiple positions over an extended 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 the slide, we subtracted the average pixel fluorescence for a region of the image containing no cells from each pixel value (Fig. 4A). Fluorescence images were also registered to correct for drift in the field of view.

Date	Strain	n for each replicate
11 Oct 2016	Def1-GFP	30, 30, 30
	Fus3-GFP	30, 25, 30
	Gpm1-GFP	17, 30
	Guk1-GFP	30, 30, 30
	Hog1-GFP	8, 17, 30
	Pgk1-GFP	9★, 21★, 12★
23 Mar 2017	Def1-GFP	30, 30
	Fus3-GFP	30, 30, 30, 30, 30
	Gpm1-GFP	30, 30, 30, 30
	Guk1-GFP	30, 30
	Hog1-GFP	24, 24, 30, 30
	Pgk1-GFP	30, 30, 30, 30

Table 1. Numbers of cells for each experimental replicate for each protein. On two occasions, cells were brought up from stocks, cultured, fixed, and photobleached. Cells were imaged for 400 ms except for asterisked replicates, which were imaged for 200 ms to avoid saturating the camera. We compensate for this difference when combining estimates (Methods).

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 time points. 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.

Cells were also corrected for autofluorescence. We corrected wild-type cells for flat field and background, and their average value was subtracted from all fluorescent cells. When this procedure is performed on wild-type cells, the residual fluorescence time-series display a systematic

error that varies from cell to cell (Fig. 4B). We therefore assume a similar systematic error in the time-series of our fluorescent cells and include a heterogeneous basal fluorescence in our model of measurement noise (f_j in Eqs. 2 and 3).

Finally, cell fluorescence was calculated as the sum of the values of brightest 80% of pixels within the cell boundary. This measure reduced the effect of movements of the stage. We used the 35 time points after the 5th because the first 5 time points were often erratic. Cells or replicates that displayed large systematic deviations from bi-exponential decay (such as a sudden drop in fluorescence) were discarded.

We twice repeated the entire experimental procedure (bringing colonies up from glycerol stocks, cell fixation and imaging). The list of all replicates used, with cell numbers per replicate, are shown in Table 1.

Inference fails assuming no measurement noise and mono-exponential decay

Analysing the data using the previously proposed estimator [21], which assumes no measurement noise and that photobleaching decays exponentially, the number of proteins inferred are inconsistent with the biochemical data. To correct for residual autofluorescence, we first fitted an exponential decay with an offset to each cell and then subtracting the fitted offset from the data before processing (a procedure that worked for data simulated from a single exponential decay with parameters shared by all cells and no measurement noise). Other than this correction, we followed the procedure of Kim *et al.* [25].

This naive inference underestimated numbers of proteins by several orders of magnitude. For Fus3, 174 and 1 proteins are estimated for the two replicates (the biochemical estimate is 3835 [15] – an underestimate by at least a factor of 20); for Hog1, the estimates are 69 and 77 (an underestimate by a factor of 90); for Guk1, the estimates are 153 and 1 (an underestimate by a factor of over 420); for Def1, the estimates are 7 and 47 (an underestimate by a factor of over 750); for Gpm1, the estimates are 134 and 11 (an underestimate by a factor of over 4500); for Pgk1, the estimates are 34 and 9 (an underestimate of over 45000).

Inference with measurement noise and complex decay

Given data $\mathbf{Y}_t = \{\mathbf{y}_1, \cdots, \mathbf{y}_t\}$, we would like to infer $\boldsymbol{\theta} = \{\boldsymbol{\theta}_m, \boldsymbol{\theta}_e\}$, where $\boldsymbol{\theta}_m$ are the parameters for the biophysical model and describe the dynamics of the underlying $\mathbf{x}$ variables and $\boldsymbol{\theta}_e$ are the parameters describing the distribution of the measurement noise (including the autofluorescence f_j). The $\mathbf{y}$ variables are related to the $\mathbf{x}$ variables through this measurement noise. We note that an analytical, maximum likelihood solution is possible for the ideal case without measurement noise [56].

Modelling protein concentration using the linear noise approximation

The dynamics of the protein concentrations, $\mathbf{x}$, are determined by chemical reactions and can be described by a chemical master equation for $P(\mathbf{x}, t)$, the probability distribution of the state $\mathbf{x}$ at time t [16]. We use the linear noise approximation (first-order terms in an expansion of the master equation in the size of the system, which for us is the volume of the cell [39]). This approximation makes $P(\mathbf{x}, t)$ a normal distribution if the initial distribution is either a normal or a delta distribution [16].

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

$$P(\mathbf{x}, t) = \mathcal{N}\big(\mathbf{x}; \boldsymbol{\mu}(t), \boldsymbol{\Sigma}(t)\big) \tag{4}$$

where $\boldsymbol{\mu}(t)$, the mean, and $\boldsymbol{\Sigma}(t)$, the covariance matrix, obey [40]

$$\frac{d}{dt}\boldsymbol{\mu} = \mathbf{S}^T \mathbf{h}(\boldsymbol{\mu}, t) \tag{5}$$

9

and

$$\frac{d}{dt}\mathbf{\Sigma} = \mathbf{J}\mathbf{\Sigma} + \mathbf{\Sigma}\mathbf{J}^T + \mathbf{S}^T\mathbf{H}\mathbf{S}. \tag{6}$$

Here $\mathbf{J}$ is the Jacobian:

$$J_{ij} = \frac{\partial}{\partial x_j}\sum_k S_{ik}h_k(\boldsymbol{\mu}, t) \tag{7}$$

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

Linear noise and sequential data − a Kalman filter

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

$$P(\boldsymbol{\theta}|\mathbf{Y}_t) \sim P(\mathbf{Y}_t|\boldsymbol{\theta})P(\boldsymbol{\theta}) \tag{8}$$

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

$$P(\boldsymbol{\theta}|\mathbf{Y}_t) \sim P(\boldsymbol{\theta})\prod_{i=1}^t P(\mathbf{y}_i|\mathbf{Y}_{i-1}, \boldsymbol{\theta}). \tag{9}$$

We find each term in Eq. 9 sequentially by considering the dynamics of $\mathbf{x}$ from one time point to the next and then correcting that dynamics given the observed data [40]. Assume that at time point $i-1$ the distribution $P(\mathbf{x}_{i-1}|\mathbf{Y}_{i-1}, \boldsymbol{\theta})$ is Gaussian with a known mean $\boldsymbol{\mu}_{i-1}^*$ andcovariance matrix $\mathbf{\Sigma}_{i-1}^*$:

$$P(\mathbf{x}_{i-1}|\mathbf{Y}_{i-1}, \boldsymbol{\theta}) = \mathcal{N}(\mathbf{x}_{i-1}; \boldsymbol{\mu}_{i-1}^*, \mathbf{\Sigma}_{i-1}^*). \tag{10}$$

Using the linear noise approximation to describe the dynamics of $\mathbf{x}$, we can, with Eq. 10 providing the initial condition, integrate Eqs. 5 and 6 over one time interval to time point i to find $\boldsymbol{\mu}_i$ and $\mathbf{\Sigma}_i$ and that

$$P(\mathbf{x}_i|\mathbf{Y}_{i-1}, \boldsymbol{\theta}) = \mathcal{N}(\mathbf{x}_i; \boldsymbol{\mu}_i, \mathbf{\Sigma}_i). \tag{11}$$

We next wish to extend the conditioning in Eq. 11 to include the data point, $\mathbf{y}_i$, at time point i. To do so, note that

$$\begin{aligned} P(\mathbf{x}_i|\mathbf{Y}_i, \boldsymbol{\theta}) &= P(\mathbf{x}_i|\mathbf{y}_i, \mathbf{Y}_{i-1}, \boldsymbol{\theta}) \\ &= \frac{P(\mathbf{y}_i|\mathbf{x}_i, \mathbf{Y}_{i-1}, \boldsymbol{\theta})P(\mathbf{x}_i|\mathbf{Y}_{i-1}, \boldsymbol{\theta})}{P(\mathbf{y}_i|\mathbf{Y}_{i-1}, \boldsymbol{\theta})} \\ &\sim P(\mathbf{y}_i|\mathbf{x}_i, \boldsymbol{\theta})P(\mathbf{x}_i|\mathbf{Y}_{i-1}, \boldsymbol{\theta}) \end{aligned} \tag{12}$$

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

$$P(\mathbf{y}|\mathbf{x}, \boldsymbol{\theta}) = \mathcal{N}(\mathbf{y}; \mathbf{U}\mathbf{x}, \mathbf{V}) \tag{13}$$

with $\mathbf{U}$ being a constant projection matrix and $\mathbf{V}$ being a covariance matrix, then we can simplify Eq. 12 using the properties of normal distributions [59]. We find that $P(\mathbf{x}_i|\mathbf{Y}_i, \boldsymbol{\theta})$ is also normal with a mean $\boldsymbol{\mu}_i^*$ and a covariance $\mathbf{\Sigma}_i^*$ that satisfy [40]

$$\begin{aligned} \boldsymbol{\mu}_i^* &= \boldsymbol{\mu}_i + \mathbf{\Sigma}_i\mathbf{U}^T(\mathbf{U}\mathbf{\Sigma}_i\mathbf{U}^T + \mathbf{V})^{-1}(\mathbf{y}_i - \mathbf{U}\boldsymbol{\mu}_i) \\ \mathbf{\Sigma}_i^* &= \mathbf{\Sigma}_i + \mathbf{\Sigma}_i\mathbf{U}^T(\mathbf{U}\mathbf{\Sigma}_i\mathbf{U}^T + \mathbf{V})^{-1}(-\mathbf{U}\mathbf{\Sigma}_i). \end{aligned} \tag{14}$$

We see that the prediction of $\boldsymbol{\mu}_i$ and $\mathbf{\Sigma}_i$ found from $\boldsymbol{\mu}_{i-1}^*$ and $\mathbf{\Sigma}_{i-1}^*$ using the linear noise approximation are corrected because of the new data point and the measurement noise.

10

The factors in Eq. 9, $P(\mathbf{y}_i|\mathbf{Y}_{i-1}, \boldsymbol{\theta})$ obey

$$P(\mathbf{y}_i|\mathbf{Y}_{i-1}, \boldsymbol{\theta}) = \int d\mathbf{x}_i P(\mathbf{y}_i|\mathbf{x}_i, \boldsymbol{\theta}) P(\mathbf{x}_i|\mathbf{Y}_{i-1}, \boldsymbol{\theta}) \tag{15}$$

and are therefore the normalizing factors for Eq. 12, satisfying

$$P(\mathbf{y}_i|\mathbf{Y}_{i-1}, \boldsymbol{\theta}) = \mathcal{N}(\mathbf{y}_i; \mathbf{U}\boldsymbol{\mu}_i, \mathbf{U}\boldsymbol{\Sigma}_i\mathbf{U}^T + \mathbf{V}). \tag{16}$$

Hence from a normal prior distribution for $\mathbf{x}_1$, $P(\mathbf{x}_1|\boldsymbol{\theta})$, we use Eq. 15 to find $P(\mathbf{y}_1|\boldsymbol{\theta})$, the first term in the factorization of the likelihood in Eq. 9, and Bayes's rule to find $P(\mathbf{x}_1|\mathbf{y}_1, \boldsymbol{\theta}) \sim P(\mathbf{y}_1|\mathbf{x}_1, \boldsymbol{\theta})P(\mathbf{x}_1|\boldsymbol{\theta})$, the starting normal distribution in Eq. 10 for the sequential inference.

To specialize the algorithm to photobleaching, we subtract the autofluorescence, f_j, from each data point before applying the Kalman filter. The Kalman update step, Eq. 14, can result in unphysical, negative components of $\boldsymbol{\mu}$, which we set to zero and then continue the Kalman filter.

Application to biophysical models of photobleaching

To explain the observed biexponential decay during bleaching, we consider two populations of molecules that bleach at different rates

$$x_1 \xrightarrow{\lambda_1} \emptyset \qquad x_2 \xrightarrow{\lambda_2} \emptyset$$

for which we can analytically solve Eqs. 5 and 6.

For this model

$$\mathbf{S} = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} \qquad \text{and} \qquad \mathbf{h} = \begin{pmatrix} \lambda_1 x_1 \\ \lambda_2 x_2 \end{pmatrix}$$

and

$$\mathbf{U} = [\nu \quad \nu] \quad \text{and} \quad \mathbf{V} = \sigma_e^2$$

in Eq. 13.

Eq. 5 becomes

$$\frac{d}{dt}\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} = \begin{pmatrix} -\lambda_1 \mu_1 \\ -\lambda_2 \mu_2 \end{pmatrix} \tag{17}$$

so that

$$\boldsymbol{\mu} = \begin{pmatrix} x_{1,0}\, e^{-\lambda_1 t} \\ x_{2,0}\, e^{-\lambda_2 t} \end{pmatrix} \tag{18}$$

for initial condition $\mathbf{x}_0$. In Eq. 6,

$$\mathbf{J} = \begin{pmatrix} -\lambda_1 & 0 \\ 0 & -\lambda_2 \end{pmatrix}$$

and

$$\mathbf{S}^{\mathrm{T}}\mathbf{H}\mathbf{S} = \begin{pmatrix} \lambda_1\, x_{1,0}\, e^{-\lambda_1 t} & 0 \\ 0 & \lambda_2\, x_{2,0}\, e^{-\lambda_2 t} \end{pmatrix}$$

which upon integration implies

$$\boldsymbol{\Sigma}(t) = \begin{pmatrix} e^{-\lambda_1 t} & 0 \\ 0 & e^{-\lambda_2 t} \end{pmatrix} \left[\begin{pmatrix} x_{1,0}\,(e^{\lambda_1 t} - 1) & 0 \\ 0 & x_{2,0}\,(e^{\lambda_2 t} - 1) \end{pmatrix} + \boldsymbol{\Sigma}(0) \right] \begin{pmatrix} e^{-\lambda_1 t} & 0 \\ 0 & e^{-\lambda_2 t} \end{pmatrix}. \tag{19}$$

11

Starting the inference

To begin the inference scheme, we must first specify a prior $P(\mathbf{x}_1|\boldsymbol{\theta}_m)$. To do so, we introduce two new parameters: x_0, which is the total amount of fluorescent protein at time point 0, and α, which is the partitioning of this fluorescent protein between the slow- and fast-bleaching 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. We therefore assume that $P(\mathbf{x}_1|\boldsymbol{\theta}_m) = \mathcal{N}(\mathbf{x}_1; \boldsymbol{\mu}_1^*, \boldsymbol{\Sigma}_1^*)$, with

$$\boldsymbol{\mu}_1^* = x_0 \begin{pmatrix} 1-\alpha \\ \alpha \end{pmatrix} \qquad \text{and} \qquad \boldsymbol{\Sigma}_1^* = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}. \tag{20}$$

This distribution is a delta function because $\boldsymbol{\Sigma}_1^*$ is a matrix of zeros and is not strictly normal. Nevertheless, Eq. 20 imposes the known initial value $\boldsymbol{\mu}_1^*$.

Sampling the posterior probability

To sample from the posterior probability, $P(\boldsymbol{\theta}|\mathbf{Y}_t)$, in Eq. 8 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 (σ_e 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 reasonable physical values. To ensure the expected behaviour is dependent only on the heterogeneous parameters and to improve the mixing of the Markov chain, we propose the combination νx_0, referred to as b_0, rather than x_0.

To generate samples from the posterior, we use a Metropolis-within-Gibbs scheme [60] with the heterogeneous parameters updated separately from the homogeneous parameters. We employ adaptive parallel tempering to accelerate mixing in the Gibbs sampler [61], which has been found to perform well on benchmark biochemical models [62]. Specifically, we use 10 chains with their temperatures chosen adaptively. Parameters are proposed independently, with λ_1, λ_2, f and α proposed from a normal distribution and b_0, σ_e and ν proposed from a log-normal distribution. The scales for the proposal distributions for the heterogeneous and homogeneous parameters are adaptively selected.

To start the Monte Carlo method, we try to find values of the parameters that maximize the likelihood $P(\mathbf{Y}_t|\boldsymbol{\theta})$. We use a nested optimisation scheme [37]:

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 used Matlab's `fmincon` with the default algorithm.

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.

Extending to state-dependent measurement noise

The variance of the measurement noise of Eq. 3 depends on the state of the system — the $\mathbf{x}_{1,j}$ and $\mathbf{x}_{2,j}$. Our inference scheme, however, assumes a constant variance and so is no longer valid.

We therefore approximate Eq. 3 by replacing the explicit state dependence in the variance by its expected value given $\mathbf{Y}_{t-1}$:

$$
\begin{aligned}
y_j | \mathbf{x}_j &\sim \mathcal{N}\Big(y_j \,;\, \mathbf{U}\mathbf{x}_j + f_j \,,\, \sigma_{e,0}^2 + \mathbf{U}\mathbf{x}_j \sigma_{e,1}^2 + (\mathbf{U}\mathbf{x}_j \sigma_{e,2})^2 \Big) \\
&\simeq \mathcal{N}\Big(y_j \,;\, \mathbf{U}\mathbf{x}_j + f_j \,,\, \mathbb{E}_{\mathbf{x}_j | \mathbf{Y}_{j-1}}\big[\sigma_{e,0}^2 + \mathbf{U}\mathbf{x}_j \sigma_{e,1}^2 + (\mathbf{U}\mathbf{x}_j \sigma_{e,2})^2 \big] \Big) \\
&= \mathcal{N}\Big(y_j \,;\, \mathbf{U}\mathbf{x}_j + f_j \,,\, \sigma_{e,0}^2 + \mathbf{U}\boldsymbol{\mu}_j \sigma_{e,1}^2 + \mathbf{U}(\boldsymbol{\mu}_j \boldsymbol{\mu}_j^{\mathrm{T}} + \boldsymbol{\Sigma}_j)\mathbf{U}^{\mathrm{T}} \sigma_{e,2}^2 \Big)
\end{aligned}
\tag{21}
$$

The usual update step in the Kalman filter, Eq. 14, can then be used.

This estimate of the likelihood performs comparably with a particle filter and is more robust [56]. We use the Metropolis-within-Gibbs scheme to sample the posterior probability with log-normal proposal distributions and scale free priors for the additional homogeneous parameters: $\sigma_{e,1}$ and $\sigma_{e,2}$.
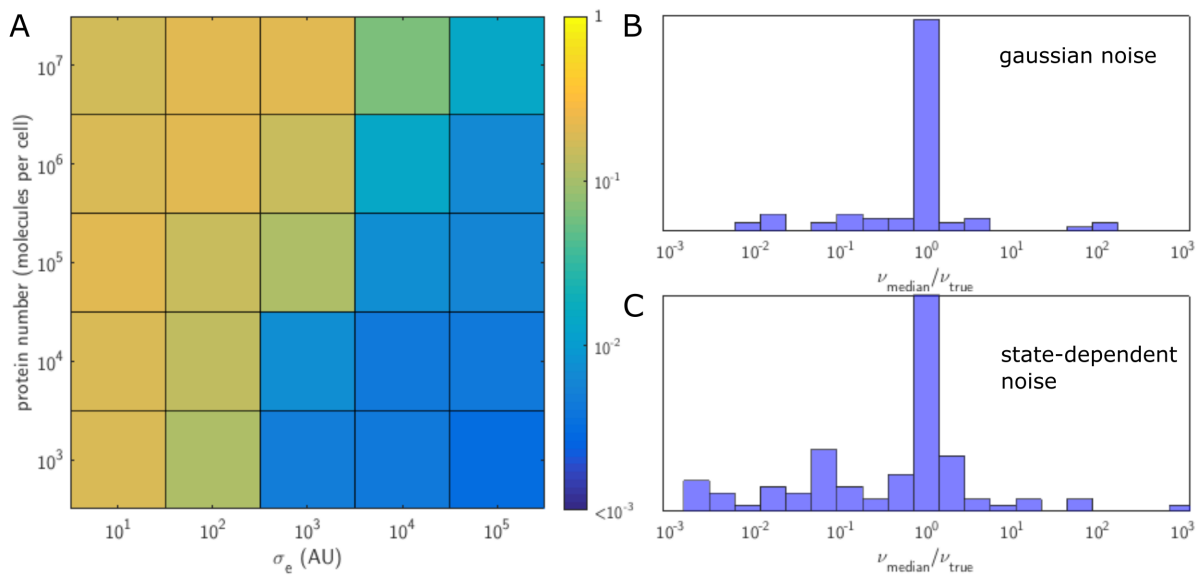
In silico **verification**

Figure 5. Parameters are well identified for simulated data. We simulated time-series for the model of Eq. 1 with a distribution of decay rates and initial numbers of proteins similar to those observed experimentally. (A) For Gaussian measurement noise, we show the probability of the true value of ν averaged over 5 simulated datasets (and normalized to correct for the kernel density estimates we use – Methods). A value of 1 would indicate total confidence in the true value. When the initial number of proteins (y-axis) is large compared to the measurement noise (x-axis), ν can be accurately inferred. Here $\nu = 10$. (B) Letting $\nu = 1$ and $\nu = 100$ too, we show the results for all three values of ν as a histogram of $\log_{10}(\nu_{\mathrm{median}}/\nu_{\mathrm{true}})$. The spike at 1 indicates ν can usually be correctly inferred. (C) A similar histogram for state-dependent measurement noise also shows that inference of ν typically remains accurate if more challenging.

We verified our methodology using simulated data (Fig. 5). For additive Gaussian measurement noise, the inference is able to accurately infer ν if the measurement noise is not so high

13

as to mask the stochastic fluctuations. This threshold depends on the initial number of proteins. For any magnitude of measurement noise, there will be a sufficiently high initial number that stochastic fluctuations are detectable. For state-dependent noise, however, the inference no longer necessarily improves with increasing numbers of protein when the state-dependent noise is sufficiently large (data not shown).

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:

$$P(x|y, \mathbf{Y}) = \int d\boldsymbol{\theta} P(x|y, \boldsymbol{\theta}) P(\boldsymbol{\theta}|y, \mathbf{Y}) P(y). \tag{22}$$

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

$$P(x|y, \mathbf{Y}) \sim \int d\boldsymbol{\theta} P(x|y, \boldsymbol{\theta}) P(\boldsymbol{\theta}|\mathbf{Y}). \tag{23}$$

By Bayes's rule

$$\begin{aligned} P(x|y, \boldsymbol{\theta}) &\sim& P(y|x, \boldsymbol{\theta}) P(x) \\ &\sim& P(y|x, \boldsymbol{\theta}) \end{aligned} \tag{24}$$

assuming a constant prior, $P(x)$.

For Gaussian measurement noise, Eq. 2, we use that

$$\begin{aligned} P(y|x, \boldsymbol{\theta}) &=& \mathcal{N}(y; \nu x + f, \sigma_e^2) \\ &=& \frac{1}{\nu} \mathcal{N}\left(x; \frac{y-f}{\nu}, \frac{\sigma_e^2}{\nu^2}\right) \end{aligned} \tag{25}$$

from the properties of Gaussian distributions. From Eq. 24 then

$$P(x|y, \boldsymbol{\theta}) = \mathcal{N}\left(x; \frac{y-f}{\nu}, \frac{\sigma_e^2}{\nu^2}\right) \tag{26}$$

normalizing over x.

For state-dependent measurement noise, Eq. 3, we have

$$P(y|x, \boldsymbol{\theta}) = \mathcal{N}\left(y; \nu x + f, \sigma_{e,0}^2 + \nu x \sigma_{e,1}^2 + (\nu x \sigma_{e,2})^2\right) \tag{27}$$

which cannot be straightforwardly inverted to give a normalized distribution in x. We therefore approximate νx by y so that the variance in Eq. 27 becomes constant. Then

$$\begin{aligned} P(x|y, \boldsymbol{\theta}) &\propto& \mathcal{N}\left(y; \nu x + f, \sigma_{e,0}^2 + y\sigma_{e,1}^2 + (y\sigma_{e,2})^2\right) \\ &=& \mathcal{N}\left(x; \frac{y-f}{\nu}, \frac{\sigma_{e,0}^2 + y\sigma_{e,1}^2 + (y\sigma_{e,2})^2}{\nu^2}\right) \end{aligned} \tag{28}$$

similarly to Eq. 26.

Continuing with σ_e^2 as the variance for simplicity (although the results will hold too for the constant variance in Eq. 28), Eq. 23 becomes

$$P(x|y, \mathbf{Y}) = \int d\nu d\sigma_e df \, \mathcal{N}\left(x; \frac{y-f}{\nu}, \frac{\sigma_e^2}{\nu^2}\right) P(\nu, \sigma_e, f|\mathbf{Y}) \tag{29}$$

which we estimate on a grid of x values as an average over the Monte Carlo samples generated from $P(\nu, \sigma_e, f|\mathbf{Y})$.

Multiple measurements: If we have multiple, replicate measurements of fluorescence (each, for example, from a different cell) and are interested in the distribution of $\bar{x} = \sum x_i/N$, we can use that the distribution of a sum of Gaussian variables is also Gaussian with a mean and a variance equal to the sum of the mean and variances of all the Gaussian variables in the sum [16]:

$$P\left(\bar{x} \mid \mathbf{y}, \mathbf{Y}\right) = \int d\nu \, d\sigma_e \, df \, \mathcal{N}\left(\bar{x}; \frac{\bar{y} - \bar{f}}{\nu}, \frac{\sigma_e^2}{N\nu^2}\right) P\left(\nu, \sigma_e, f \mid \mathbf{Y}\right) \tag{30}$$

following Eq. 29. Practically, we assume that σ_e^2/N and $\bar{f}/\bar{y}$ sufficiently small that we can take the limiting case:

$$P\left(\bar{x} \mid \mathbf{y}, \mathbf{Y}\right) = \int d\nu \, \delta\left(\bar{x} - \frac{\bar{y}}{\nu}\right) P\left(\nu \mid \mathbf{Y}\right). \tag{31}$$

We use a kernel density representation for $P\left(\bar{x} \mid \mathbf{y}, \mathbf{Y}\right)$ with kernels $K(x, x') = \log \mathcal{N}\left(x; x', \sigma_K\right)$ and $\sigma_K^2 = 10^{-2}$:

$$\begin{aligned} P\left(\bar{x} \mid \mathbf{y}, \mathbf{Y}\right) &= \int dx' \, K(\bar{x}, x') P(x' \mid \mathbf{y}, \mathbf{Y}) \\ &\simeq \int d\nu \, dx' \, K(\bar{x}, x') \delta\left(x' - \frac{\bar{y}}{\nu}\right) P\left(\nu \mid \mathbf{Y}\right) \\ &= \int d\nu \, K\left(\bar{x}, \frac{\bar{y}}{\nu}\right) P\left(\nu \mid \mathbf{Y}\right) \end{aligned} \tag{32}$$

using Eq. 31. We evaluate the integral in Eq. 32 using the Monte Carlo samples of $P\left(\nu \mid \mathbf{Y}\right)$.

Combining estimates from all datasets

To have a final estimate of the homogeneous parameters for analysing other microscopy experiments, we combined posteriors from all replicates for all genes (the posteriors in Fig. 6). 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 $P\left(\boldsymbol{\theta} \mid \{\mathcal{D}\}\right)$. We note that the datasets are conditionally independent given $\boldsymbol{\theta}$ so that

$$\begin{aligned} P\left(\boldsymbol{\theta} \mid \{\mathcal{D}\}\right) &\sim P(\boldsymbol{\theta}) \prod_i^N P\left(\mathcal{D}_i \mid \boldsymbol{\theta}\right) \\ &\sim P(\boldsymbol{\theta}) \prod^N \frac{P(\boldsymbol{\theta} \mid \mathcal{D}_i)}{P(\boldsymbol{\theta})} \\ &= \frac{\prod_i P(\boldsymbol{\theta} \mid \mathcal{D}_i)}{P(\boldsymbol{\theta})^{N-1}}. \end{aligned} \tag{33}$$

We approximate the $P(\boldsymbol{\theta} \mid \mathcal{D}_i)$ in Eq. 33 with kernel density estimates. If $\{\boldsymbol{\theta}_{i,1}, \ldots \boldsymbol{\theta}_{i,M}\}$ is the set of parameter samples from the posterior $P\left(\boldsymbol{\theta} \mid \mathcal{D}_i\right)$ obtained from our Monte Carlo method, then

$$P\left(\boldsymbol{\theta} \mid \mathcal{D}_i\right) \simeq \frac{1}{M} \sum_{j=1}^M K(\boldsymbol{\theta}, \boldsymbol{\theta}_{i,j}) \tag{34}$$

where the kernel functions are isotropic log normal distributions with variance σ_K^2 as before. In principle, these kernel density estimates allow Eq. 33 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.

This procedure was used to both combine the posteriors for replicate datasets (Fig. 6) for a given protein (Fig. 3B) and combine the posteriors for all replicates and all proteins (Figs. 3C & 7).

For Pgk1-GFP, we accounted for the difference of a factor of two in the exposure time used in the two biological replicates (Table 1). Before combining samples, the samples of ν were doubled
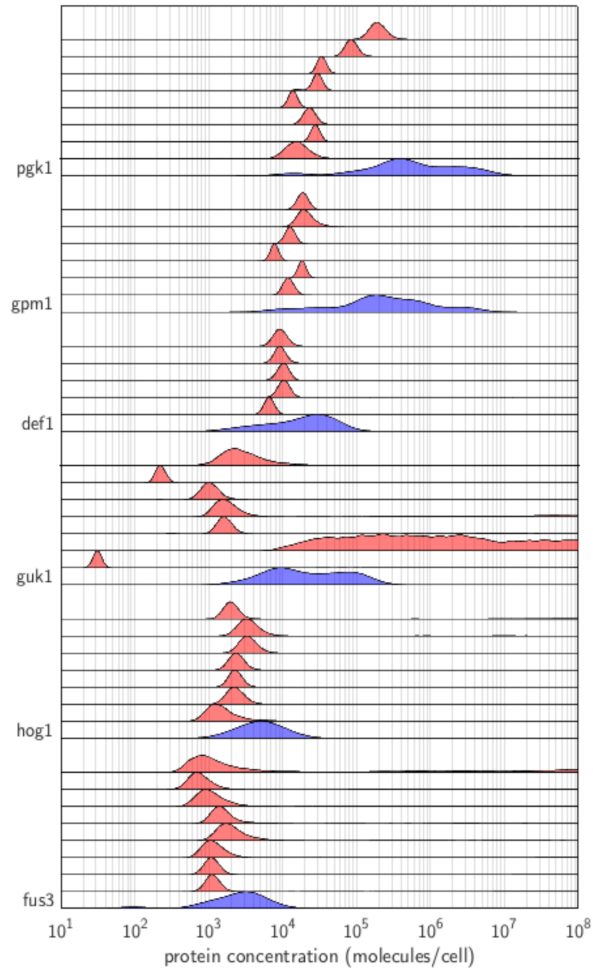
Figure 6. Inference from individual replicates is typically consistent. For each protein, the blue histogram is a kernel density estimate of the distribution of protein numbers using different biochemical measurements [15]; the red histograms show the distribution of inferred numbers using our analysis for each replicate.

for the three replicates obtained with the low exposure time. Similarly, we doubled the cellular fluorescence for these three replicates before using the full posterior for ν to estimate protein numbers.

Estimating the evidence for each model

To compare models of the measurement noise, we calculate Bayes factors, i.e. a ratio of the log evidence, $\log P(\text{data}|\text{model})$, for four models — either mono- or biexponential bleaching and either Gaussian or state-dependent measurement noise. For a given model, $\mathcal{M}$, we estimate the

16

evidence, $E(\mathcal{M})$, as follows:

$$
\begin{aligned}
E(\mathcal{M}) &= P\left(\{\mathcal{D}\}|\,\mathcal{M}\right) \\
&= \int d\boldsymbol{\theta}\, P\left(\{\mathcal{D}\}, \boldsymbol{\theta}|\,\mathcal{M}\right) \\
&= \int d\boldsymbol{\theta}\, P(\boldsymbol{\theta}) \prod_{i=1}^{N} P\left(\mathcal{D}_i|\,\boldsymbol{\theta}, \mathcal{M}\right) \\
&= \left[\prod_{i=1}^{N} E_i(\mathcal{M})\right] \int d\boldsymbol{\theta}\, \frac{\prod_{i=1}^{N} P\left(\boldsymbol{\theta}|\,\mathcal{D}_i, \mathcal{M}\right)}{P(\boldsymbol{\theta})^{N-1}}
\end{aligned}
\tag{35}
$$

where $E_i(\mathcal{M}) = P(\mathcal{D}_i|\mathcal{M})$ is the model evidence for the data set $\mathcal{D}_i$. We approximate the individual evidences, $E_i(\mathcal{M})$, using the Monte Carlo samples for the individual datasets. The second term we approximate using our samples for combining posteriors and the kernel density estimate (Eqs. 33 and 34). In both cases, we use the Laplace approximation to find the evidence and the covariance of the samples to estimate Hessians [42].

In addition to being favoured by the total evidence for all datasets, the bi-exponential bleaching model with state-dependent noise has the highest evidence in the majority of data sets: 25 compared to 8 for bi-exponential decay and Gaussian measurement noise, 6 for mono-exponential decay and Gaussian noise, and 2 for mono-exponential decay and state-dependent noise.

An informative prior results in only a minor change to the inference

To further test the validity of our model, we repeated the analysis, imposing an informative log-normal prior for the parameter ν. In so doing, we integrate all available data to obtain a best estimate of all the parameters in our model. These estimates can be used in analysing future experiments, for simulations of microscopy data, and can be compared with the estimates obtained with uninformative priors to assess the reliability of the model. Further, we can again calculate the evidence for each model to identify, given all the information available, the best model for our data.

We used an informative log-normal prior for the parameter ν, which is the fluorescence per molecule, with parameters fitted to the published data of Ho *et al.* [15]. For each data set, this parameter ν was confined to be within 1 interquartile range of the median published value: the mode of the prior was set to the median of the published results and the scale factor set to half of their interquartile range.

With the informative prior, the bi-exponential model with state-dependent noise is still selected with a significant Bayes factor ($> 10^4$ dB compared to the next best model). The modes of the posterior distributions for the parameters are shifted at most by a factor of 2 (Fig. 7), which is negligible given that the correct order of magnitude is typically all that is required. As expected, the parameter ν is shifted down to amend the underestimation of protein numbers, and the measurement noise parameters ($\sigma_{e,2}$ and to a lesser extent $\sigma_{e,1}$) are increased to accommodate the fluctuations in fluorescence that are no longer explained by the now reduced stochastic fluctuations.

Data availability

Data generated in this work is available at `http://dx.doi.org/`.
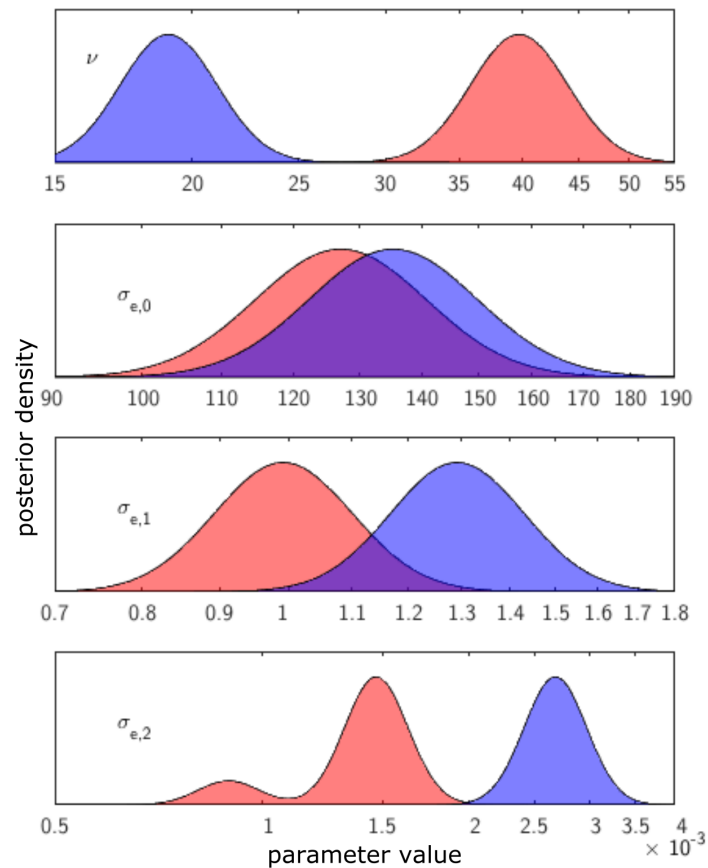
Acknowledgements

Figure 7. Tightly restricting the *a priori* range of ν does not substantially change the inference. The locations of the modes of the posterior distributions for the data of Fig. 3 maintain the same order of magnitude even when this data is used to impose an informative prior for the fluorescence per molecule (distributions with the informative prior in blue; distributions from Fig. 3C in red).

Author contributions

EB performed the experiments and developed the analysis; EB & PSS wrote the paper.

References

[1] Suter, D. D. M., Molina, N., Gatfield, D., Schneider, K., Schibler, U., and Naef, F. Mammalian Genes Are Transcribed with Widely Different Bursting Kinetics. *science* **332**(6028), 472–4, apr (2011).

[2] Harper, C. V., Finkenstädt, B., Woodcock, D. J., Friedrichsen, S., Semprini, S., Ashall, L., Spiller, D. G., Mullins, J. J., Rand, D. a., Davis, J. R. E., and White, M. R. H. Dynamic analysis of stochastic transcription cycles. *PLoS biology* **9**(4), e1000607, apr (2011).

[3] Zechner, C., Unger, M., Pelet, S., Peter, M., and Koeppl, H. Scalable inference of heterogeneous reaction kinetics from pooled single-cell recordings. *Nature Methods* **11**(2), 197–202, jan (2014).

[4] Finkenstädt, B., Woodcock, D. J., Komorowski, M., Harper, C. V., Davis, J. R. E., White, M. R. H., and Rand, D. a. Quantifying intrinsic and extrinsic noise in gene transcription using the linear noise approximation: An application to single cell data. *Annals of Applied Statistics* **7**(4), 1960–1982 (2013).

[5] Featherstone, K., Hey, K., Momiji, H., McNamara, A. V., Patist, A. L., Woodburn, J., Spiller, D. G., Christian, H. C., McNeilly, A. S., Mullins, J. J., Finkenstädt, B. F., Rand, D. A., White, M. R., and Davis, J. R. Spatially coordinated dynamic gene transcription in living pituitary tissue. *eLife* **5**(FEBRUARY2016), 1–25, feb (2016).

[6] Verdaasdonk, J. S., Lawrimore, J., and Bloom, K. Determining absolute protein numbers by quantitative fluorescence microscopy. *Methods in Cell Biology* **123**, 347–365, jan (2014).

[7] Elson, E. L. and Magde, D. Fluorescence correlation spectroscopy I. Conceptual basis and theory. *Biopolymers* **13**, 1–27 (1974).

[8] Magde, D., Elson, E. L., and Webb, W. W. Fluorescence correlation spectroscopy II. An experimental realization. *Biopolymers* **13**, 29–61 (1974).

[9] Wiseman, P. W., Squier, J. A., Ellisman, M. H., and Wilson, K. R. Two-photon image correlation spectroscopy and image cross-correlation spectroscopy. *J. Microsc* **200**, 14–25 (2000).

[10] Chen, Y., Muller, J. D., So, P. T., and Gratton, E. The photon counting histogram in fluorescence fluctuation spectroscopy. *Biophys. J* **77**, 553–567 (1999).

[11] Kask, P., Palo, K., Ullmann, D., and Gall, K. Fluorescence-intensity distribution analysis and its application in biomolecular detection technology. *Proc. Natl. Acad. Sci. U. S. A* **96**, 13756–13761 (1999).

[12] Ghaemmaghami, S., Huh, W.-K., Bower, K., Howson, R. W., Belle, A., Dephoure, N., O'Shea, E. K., and Weissman, J. S. Global analysis of protein expression in yeast. *Nature* **425**(6959), 737–41, oct (2003).

[13] Wu, J.-Q., McCormick, C. D., and Pollard, T. D. Counting proteins in living cells by quantitative fluorescence microscopy with internal standards. *Methods in cell biology* **89**, 253–273 (2008).

[14] Lu, P., Vogel, C., Wang, R., Yao, X., and Marcotte, E. M. Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation. *Nature biotechnology* **25**(1), 117–24, jan (2007).

[15] Ho, B., Baryshnikova, A., and Brown, G. W. Unification of Protein Abundance Datasets Yields a Quantitative Saccharomyces cerevisiae Proteome. *Cell systems* **0**(0), jan (2018).

[16] Van Kampen, N. *Stochastic Processes in Physics and Chemistry*. Elsevier Inc., 3rd edition, (2007).

[17] Rosenfeld, N., Young, J. W., Alon, U., Swain, P. S., and Elowitz, M. B. Gene regulation at the single-cell level. *Science* **307**(5717), 1962–1965, March (2005).

[18] Rosenfeld, N., Perkins, T. J., Alon, U., Elowitz, M. B., and Swain, P. S. A fluctuation method to quantify in vivo fluorescence data. *Biophysical journal* **91**(2), 759–66, jul (2006).

[19] Teng, S.-W., Wang, Y., Tu, K. C., Long, T., Mehta, P., Wingreen, N. S., Bassler, B. L., and Ong, N. P. Measurement of the copy number of the master quorum-sensing regulator of a bacterial cell. *Biophysical journal* **98**(9), 2024–31, may (2010).

[20] Zamparo, L. and Perkins, T. J. Statistical lower bounds on protein copy number from fluorescence expression images. *Bioinformatics* **25**(20), 2670–2676 (2009).

[21] Nayak, C. R. and Rutenberg, A. D. Quantification of fluorophore copy number from intrinsic fluctuations during fluorescence photobleaching. *Biophysical journal* **101**(9), 2284–93, nov (2011).

[22] Diaspro, A., Chirico, G., Usai, C., Ramoino, P., and Dobrucki, J. Photobleaching. In *Handbook of Biological Confocal Microscopy, third edition*, chapter 39, 690–702. (2006).

[23] Nelson, S. R., Trybus, K. M., and Warshaw, D. M. Motor coupling through lipid membranes enhances transport velocities for ensembles of myosin Va. *Proceedings of the National Academy of Sciences* **111**(38), E3986–E3995 (2014).

[24] Lombardo, A. T., Nelson, S. R., Ali, M. Y., Kennedy, G. G., Trybus, K. M., Walcott, S., and Warshaw, D. M. Myosin Va molecular motors manoeuvre liposome cargo through suspended actin filament intersections in vitro. *Nature Communications* **8**, 1–9 (2017).

[25] Kim, N. H., Lee, G., Sherer, N. A., Martini, K. M., Goldenfeld, N., and Kuhlman, T. E. Real-time transposable element activity in individual live cells. *Proceedings of the National Academy of Sciences* **113**(26), 7278–7283 (2016).

[26] Ghaemmaghami, S., Huh, W.-K., Bower, K., Howson, R. W., Belle, A., Dephoure, N., O'Shea, E. K., and Weissman, J. S. Global analysis of protein expression in yeast. *Nature* **425**(6959), 737–41, oct (2003).

[27] Patterson, G. H., Knobel, S. M., Sharif, W. D., Kain, S. R., and Piston, D. W. Use of the green fluorescent protein and its mutants in quantitative fluorescence microscopy. *Biophysical journal* **73**(November), 2782–2790 (1997).

[28] Ghauharali, R. and Brakenhoff, G. Fluorescence photobleaching-based image standardization for fluorescence microscopy. *Journal of Microscopy* **198**(April 1999), 88–100 (2000).

[29] Kalies, S., Kuetemeyer, K., and Heisterkamp, a. Mechanisms of high-order photobleaching and its relationship to intracellular ablation. *Biomedical optics express* **2**(4), 805–16, jan (2011).

[30] Füreder-Kitzmüller, E., Hesse, J., Ebner, A., Gruber, H. J., and Schütz, G. J. Non-exponential bleaching of single bioconjugated Cy5 molecules. *Chemical Physics Letters* **404**(1-3), 13–18, mar (2005).

[31] Berglund, A. J. Nonexponential statistics of fluorescence photobleaching. *The Journal of chemical physics* **121**(7), 2899–903, aug (2004).

[32] Song, L., van Gijlswijk, R. P., Young, I. T., and Tanke, H. J. Influence of fluorochrome labeling density on the photobleaching kinetics of fluorescein in microscopy. *Cytometry* **27**(3), 213–23, mar (1997).

[33] Taneja, S. and Rutenberg, A. D. Photobleaching of randomly rotating fluorescently decorated particles. *The Journal of Chemical Physics* **147**(10), 104105, sep (2017).

[34] Rettig, W., Strehmel, B., Schrader, S., and Seifert, H. *Applied Fluorescence in Chemistry, Biology and Medicine*. Springer Berlin Heidelberg, Berlin, Heidelberg, (1999).

[35] Song, L., Hennink, E. J., Young, T., and Tanke, H. J. Photobleaching kinetics of fluorescein in quantitative fluorescence microscopy. *Biophysical Journal* **68**(June 1995), 2588–2600 (1995).

[36] Hey, K. L., Momiji, H., Featherstone, K., Davis, J. R. E., White, M. R. H., Rand, D. a., and Finkenstadt, B. A stochastic transcriptional switch model for single cell imaging data. *Biostatistics* **16**, 655–669 (2015).

[37] Llamosi, A., Gonzalez-Vargas, A. M., Versari, C., Cinquemani, E., Ferrari-Trecate, G., Hersen, P., and Batt, G. What Population Reveals about Individual Cell Identity: Single-Cell Parameter Estimation of Models of Gene Expression in Yeast. *PLOS Computational Biology* **12**(2), e1004706, feb (2016).

[38] Waters, J. C. Accuracy and precision in quantitative fluorescence microscopy. *The Journal of cell biology* **185**(7), 1135–48, jun (2009).

[39] Schnoerr, D., Sanguinetti, G., and Grima, R. Approximation and inference methods for stochastic biochemical kinetics -a tutorial review. *J. Phys. A: Math. Theor.* **50**(093001) (2017).

[40] Fearnhead, P., Giagos, V., and Sherlock, C. Inference for reaction networks using the linear noise approximation. *Biometrics* **70**(2), 457–466, jun (2014).

[41] Haario, H., Laine, M., Mira, A., and Saksman, E. DRAM: Efficient adaptive MCMC. *Statistics and Computing* **16**(4), 339–354, dec (2006).

[42] Mackay, D. J. C. *Information Theory , Inference , and Learning Algorithms*. Cambridge University Press, (2003).

[43] Huh, W.-K., Falvo, J. V., Gerke, L. C., Carroll, A. S., Howson, R. W., Weissman, J. S., and O'Shea, E. K. Global analysis of protein localization in budding yeast. *Nature* **425**(6959), 686–691, oct (2003).

[44] Kass, R. E., Carlin, B. P., Gelman, A., and Neal, R. M. Markov Chain Monte Carlo in Practice: A Roundtable Discussion. *The American Statistician* **52**(2), 93–100, may (1998).

[45] Urquiza Garcia, U. *A mathematical model in absolute units for the Arabidopsis circadian oscillator*. PhD thesis, University of Edinburgh, (2018).

[46] Kitano, H. Systems biology: a brief overview. *Science* **295**(5560), 1662–1664 (2002).

[47] Endy, D. Foundations for engineering biology. *Nature* **438**(7067), 449–453 (2005).

[48] Gordon, A., Colman-Lerner, A., Chin, T. E., Benjamin, K. R., Yu, R. C., and Brent, R. Single-cell quantification of molecules and rates using open-source microscope-based cytometry. *Nature Methods* **4**(2), 175–181, jan (2007).

[49] Chong, Y., Koh, J., Friesen, H., Duffy, K., Cox, M., Moses, A., Moffat, J., Boone, C., and Andrews, B. Yeast Proteome Dynamics from Single Cell Imaging and Automated Analysis. *Cell* **161**(6), 1413–1424 (2015).

[50] Breker, M., Gymrek, M., and Schuldiner, M. A novel single-cell screening platform reveals proteome plasticity during yeast stress responses. *Journal of Cell Biology* **200**(6), 839–850 (2013).

[51] Granados, A. A., Crane, M. M., Montano-Gutierrez, L. F., Tanaka, R. J., Voliotis, M., and Swain, P. S. Distributing tasks via multiple input pathways increases cellular survival in stress. *Elife* **6**, 3649, May (2017).

[52] Maeder, C. I., Hink, M. a., Kinkhabwala, A., Mayr, R., Bastiaens, P. I. H., and Knop, M. Spatial regulation of Fus3 MAP kinase activity through a reaction-diffusion mechanism in yeast pheromone signalling. *Nature cell biology* **9**(11), 1319–26, nov (2007).

[53] Milo, R., Jorgensen, P., Moran, U., Weber, G., and Springer, M. BioNumbers–the database of key numbers in molecular and cell biology. *Nucleic acids research* **38**(Database issue), D750–3, jan (2010).

[54] Jorgensen, P., Edgington, N. P., Schneider, B. L., Rupes, I., Tyers, M., and Futcher, B. The Size of the Nucleus Increases as Yeast Cells Grow. *Molecular Biology of the Cell* **18**(9), 3523–3532, jun (2007).

[55] Jorgensen, P., Nishikawa, J. L., Breitkreutz, B.-J., and Tyers, M. Systematic Identification of Pathways That Couple Cell Growth and Division in Yeast. *Science* **297**(5580), 395–400, jul (2002).

[56] Bakker, E. *Quantitative fluorescence microscopy methods for studying transcription with application to the yeast GAL1 promoter.* PhD thesis, University of Edinburgh, (2017).

[57] Edelstein, A., Amodaj, N., Hoover, K., Vale, R., and Stuurman, N. Computer control of microscopes using manager, , oct (2010).

[58] Crane, M. M., Clark, I. B. N., Bakker, E., Smith, S., and Swain, P. S. A Microfluidic System for Studying Ageing and Dynamic Single-Cell Responses in Budding Yeast. *PLoS ONE* **9**(6), e100042, jun (2014).

[59] Bishop, C. M. *Pattern recognition and machine learning.* Springer, (2006).

[60] Neal, R. M. *Probabilistic Inference Using Markov Chain Monte Carlo Methods.* (1993).

[61] Miasojedow, B., Moulines, E., and Vihola, M. An Adaptive Parallel Tempering Algorithm. *Journal of Computational and Graphical Statistics* **22**(3), 649–664, jul (2013).

[62] Ballnus, B., Hug, S., Hatz, K., Görlitz, L., Hasenauer, J., and Theis, F. J. Comprehensive benchmarking of Markov chain Monte Carlo methods for dynamical systems. *BMC Systems Biology* **11**(1), 63 (2017).