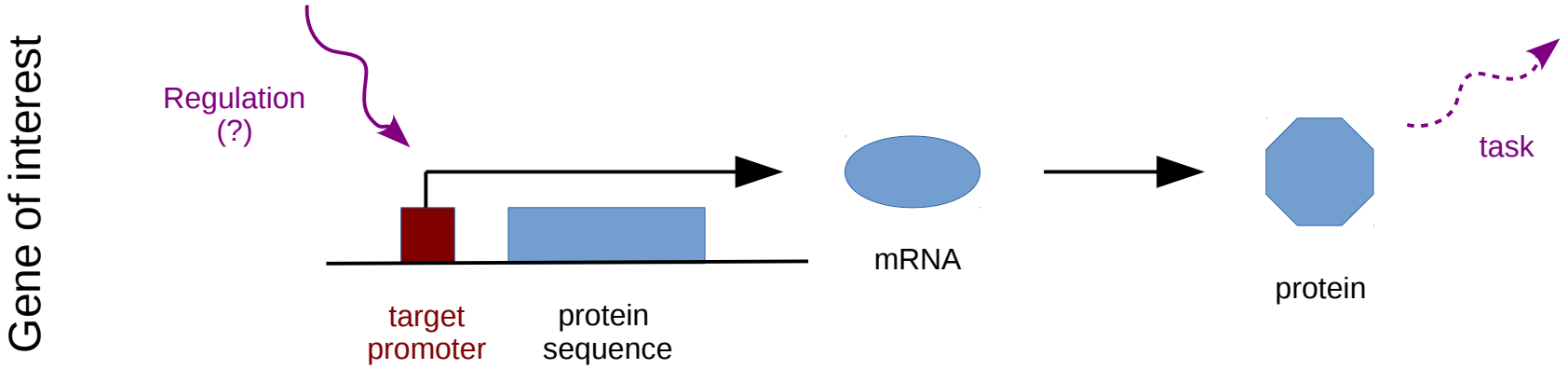
Extrinsic noise estimation in single-cell gene expression

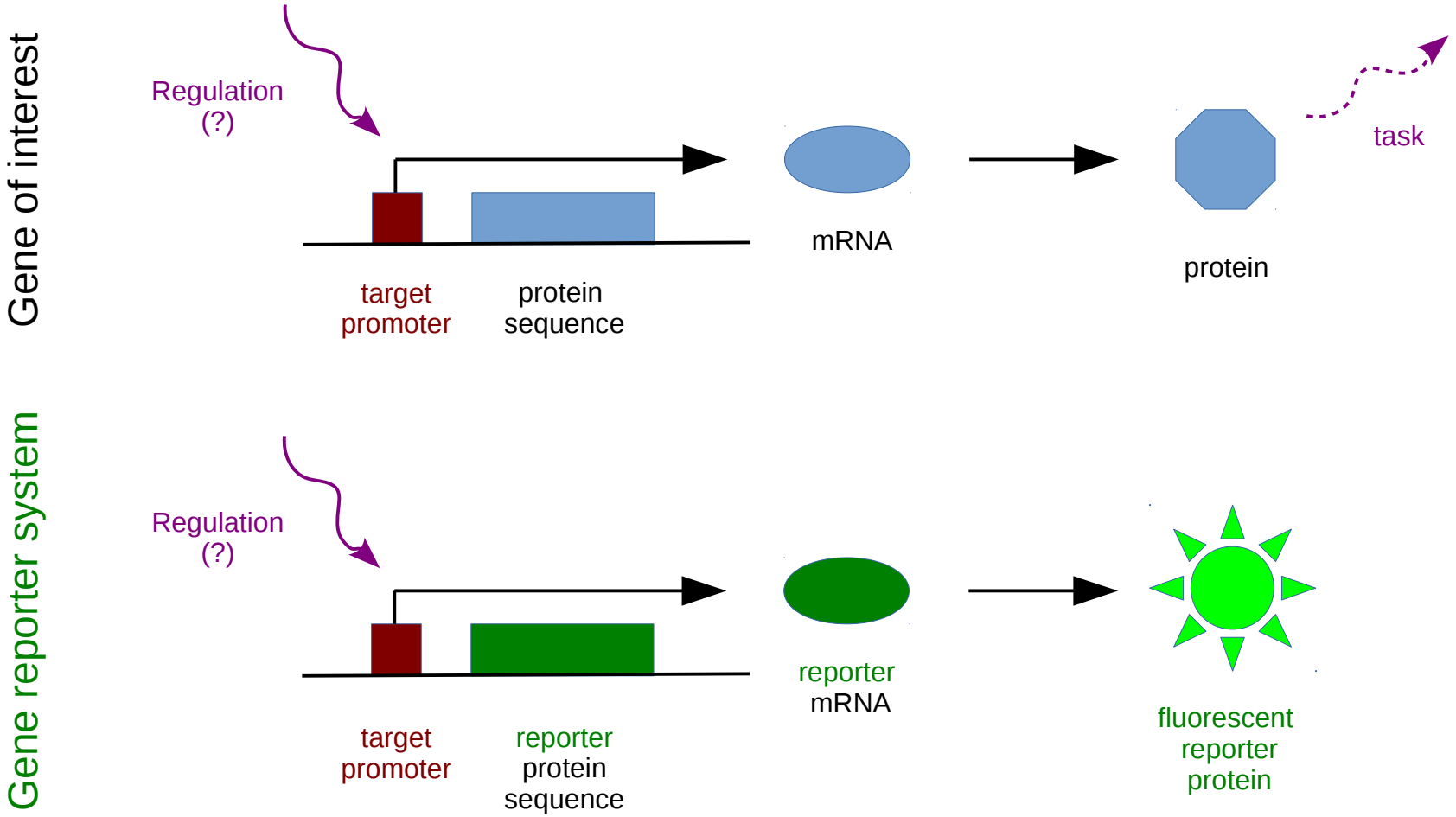
Eugenio Cinquemani
INRIA Grenoble – Rhône-Alpes

GdT MathBio
April 25 (Italy's liberation day), 2018

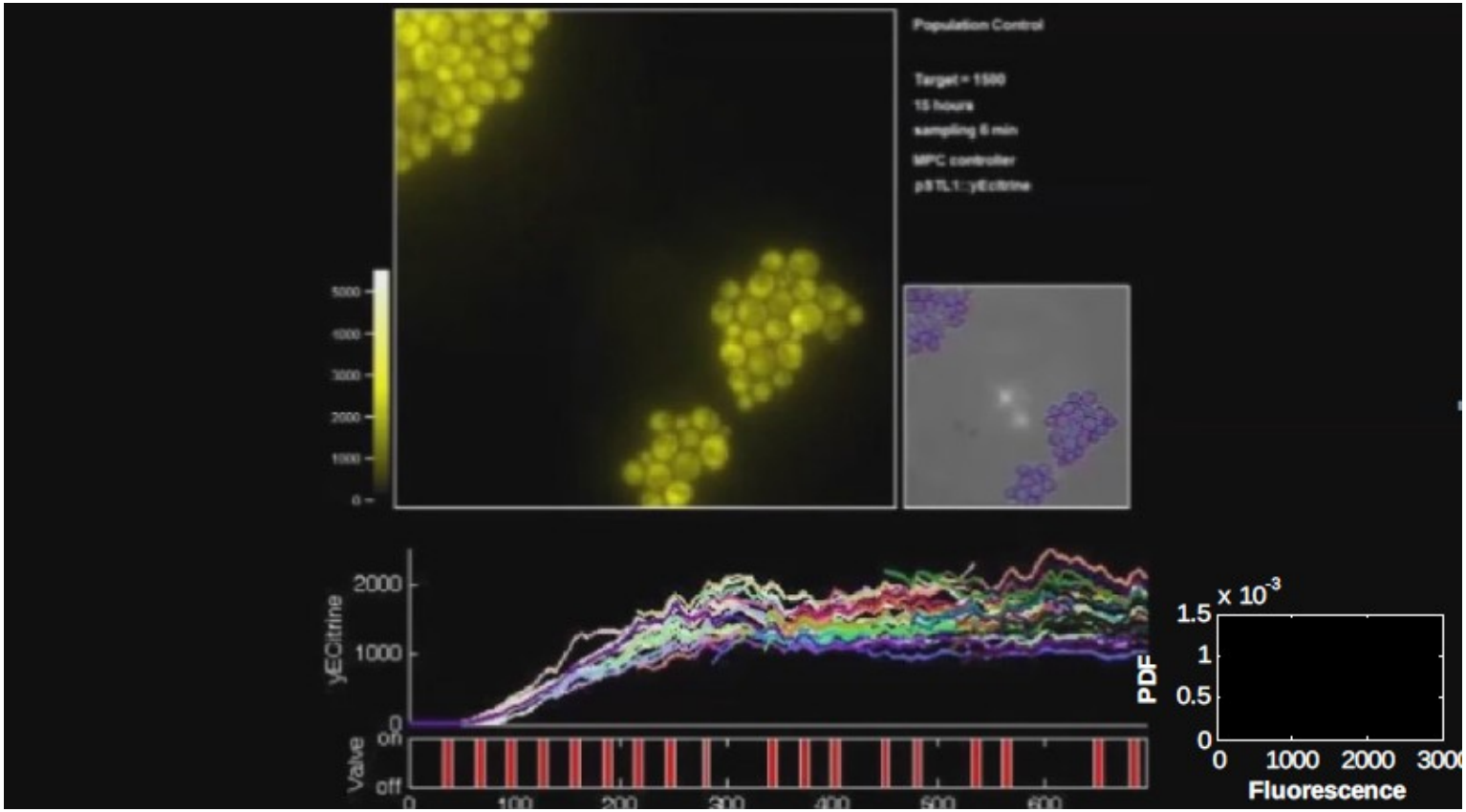
Gene expression



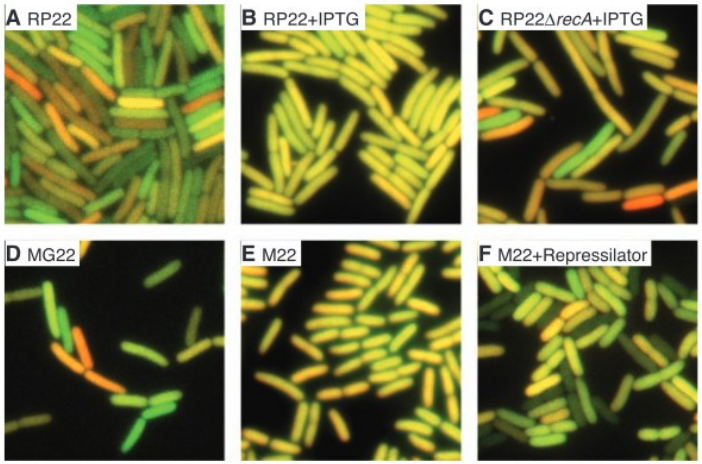
Gene expression and fluorescence reporting

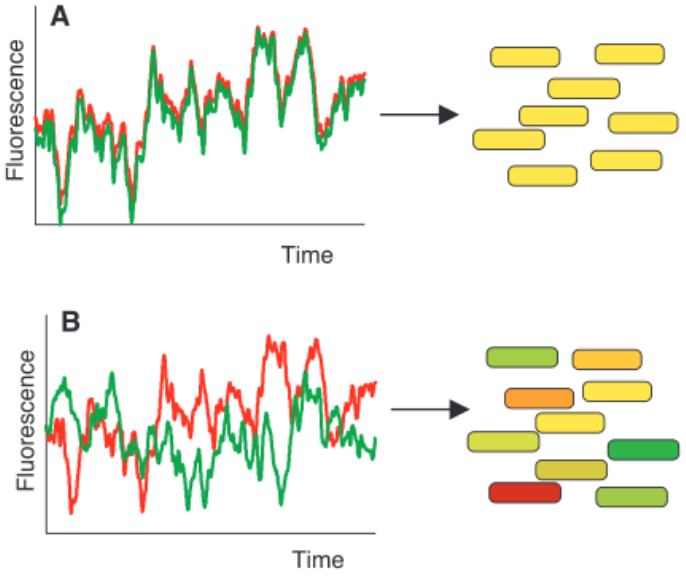


Gene expression variability



Intrinsic vs. extrinsic noise





(Elowitz et al,Science 2002)

- Intrinsic : Random transcription and translation events
- Extrinsic : Other sources of variability (parameters, promoter activity, ...)

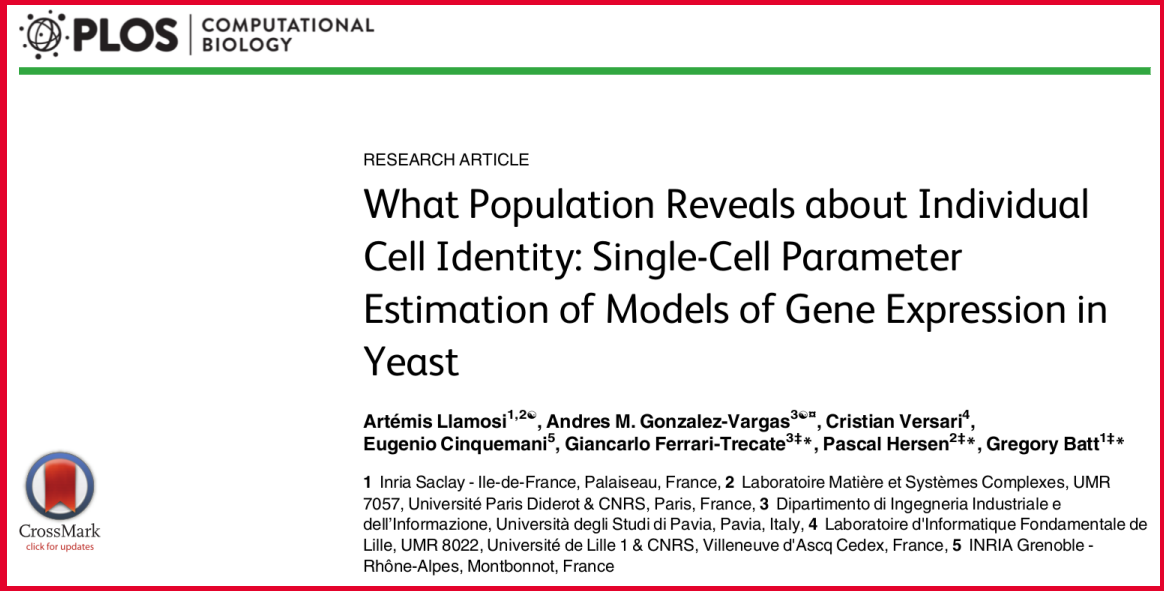
Outline

- Cell-to-cell parameter variability in osmotic shock response

  (Llamosi et al., PloS 2016)

- Stochastic variability in promoter activity

  (Cinquemani, arXiv 2017 and under review)



PLOS | COMPUTATIONAL BIOLOGY

RESEARCH ARTICLE

What Population Reveals about Individual Cell Identity: Single-Cell Parameter Estimation of Models of Gene Expression in Yeast

Artémis Llamosi[1,2☯], Andres M. Gonzalez-Vargas[3☯¤], Cristian Versari[4], Eugenio Cinquemani[5], Giancarlo Ferrari-Trecate[3‡*], Pascal Hersen[2‡*], Gregory Batt[1‡*]

1 Inria Saclay - Ile-de-France, Palaiseau, France, **2** Laboratoire Matière et Systèmes Complexes, UMR 7057, Université Paris Diderot & CNRS, Paris, France, **3** Dipartimento di Ingegneria Industriale e dell'Informazione, Università degli Studi di Pavia, Pavia, Italy, **4** Laboratoire d'Informatique Fondamentale de Lille, UMR 8022, Université de Lille 1 & CNRS, Villeneuve d'Ascq Cedex, France, **5** INRIA Grenoble - Rhône-Alpes, Montbonnot, France

CrossMark
click for updates

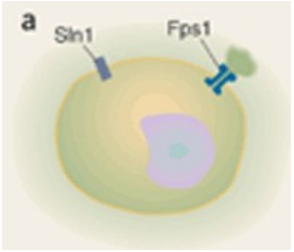
Intercellular variability in osmotic shock response

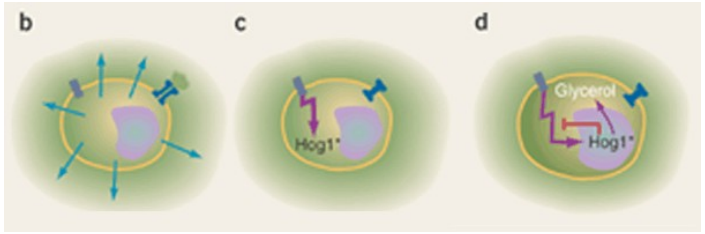
Target system :
Osmotic shock response in yeast

- Low-osmolarity environment :

  Cells are fine
- Increase of environmental osmolarity :

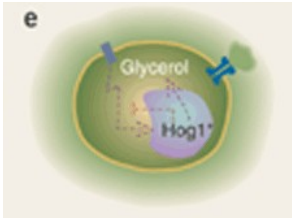
  Cells feel bad and react





(**b**) Loss of turgor, water rushes out of the cells.
(**c**) glycerol efflux channel closed, accumulation of glycerol, Hog1 phosphorylation.
(**d**) Hog1 migrates to nucleus, upregulates expression in several genes

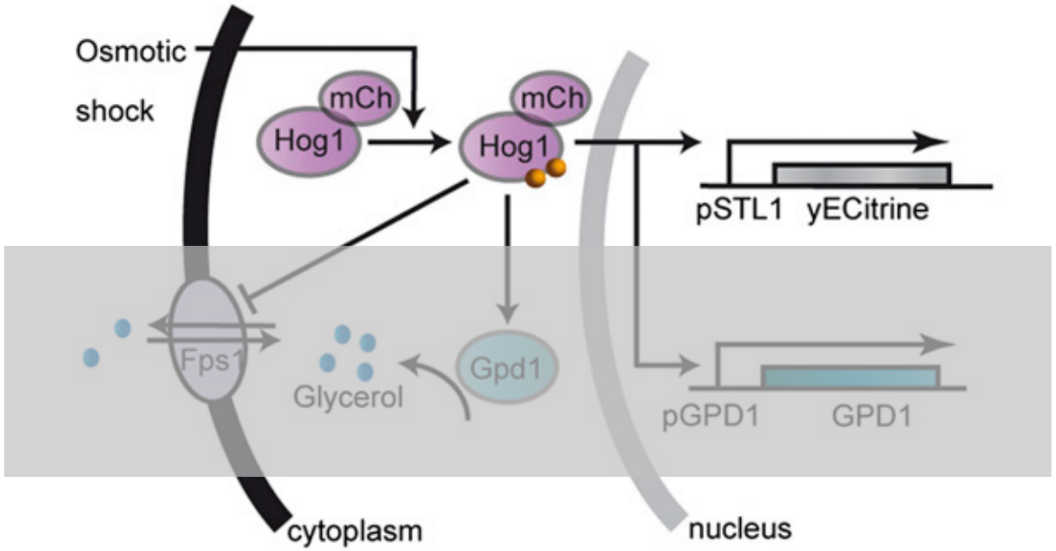
- Eventually, cell adapt to the new medium



(**e**) cell reinflation decreases HOG signaling. Phosphatases deactivate HOG pathway, turgor is restored and Fps1 reopens.

HOG1 pathway : Gene expression monitoring

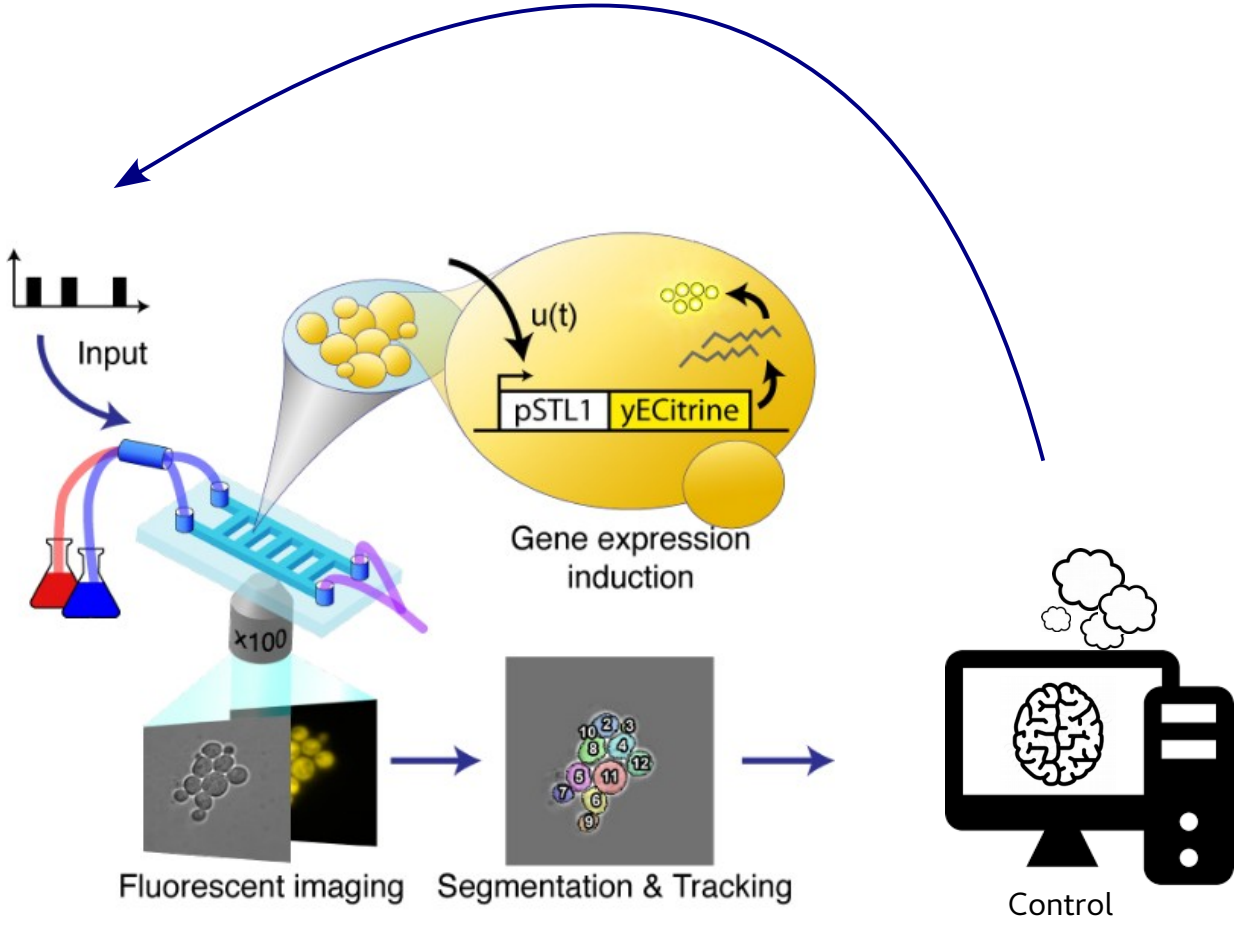
- Focus on expression dynamics of osmosensitive genes
- Fluorescent reporter monitoring of STL1 activity under short osmolarity pulses :



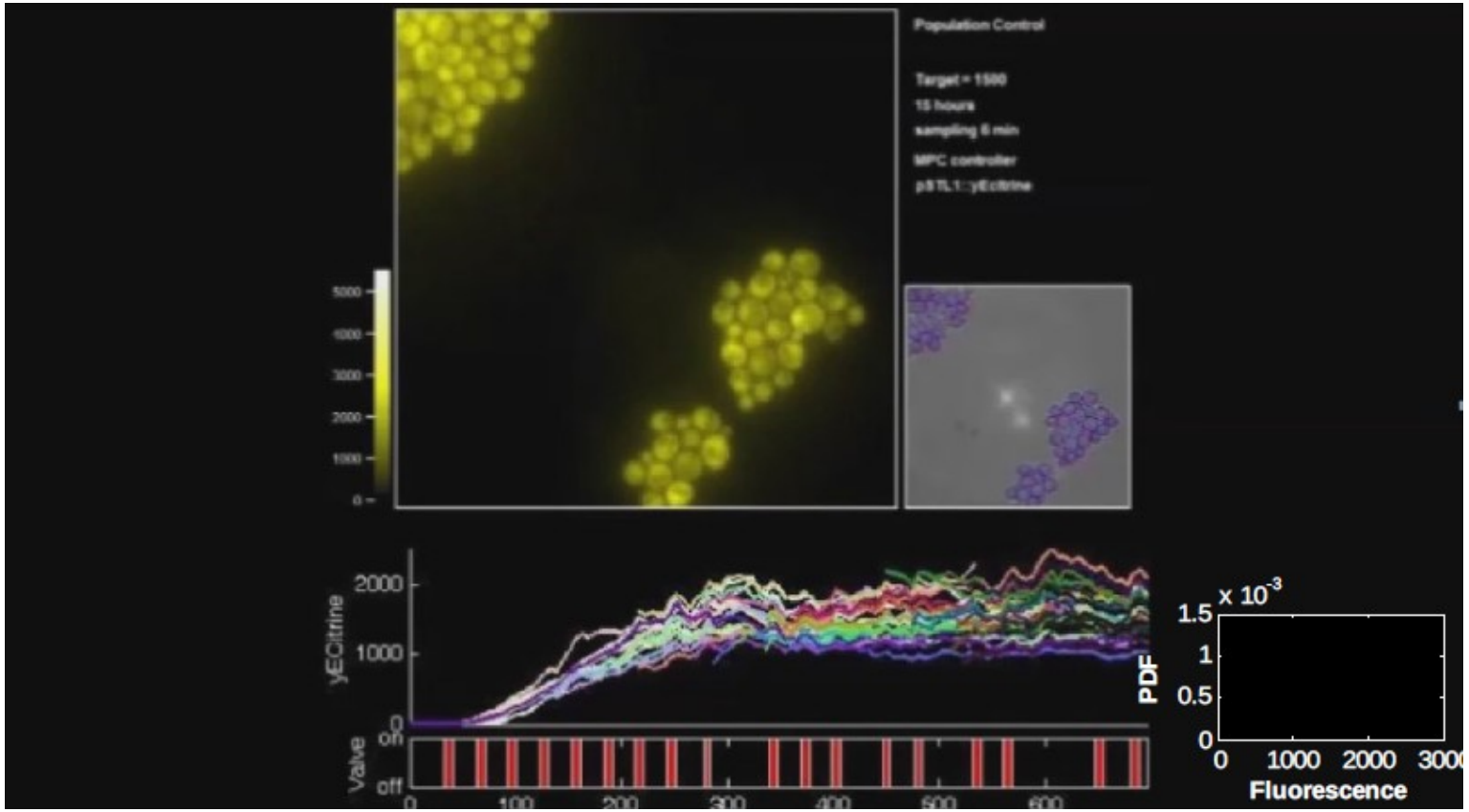
(Uhlendorf et al., *PNAS* 2012)

- Feedback regulation is cut or not triggered (adaptation prevented)

Experimental setup



Life according to single cells...



- Question #1 : How to better control this ?
- Question #0: How to make sense of (i.e. model) this ?

Single-cell kinetic model

Input: Osmotic shock signal *u(t)*

Gene expression:

$$\dot{m}(t) = k_m u(t) - g_m m(t),$$
$$\dot{p}(t) = k_p m(t) - g_p p(t),$$

Protein maturation:

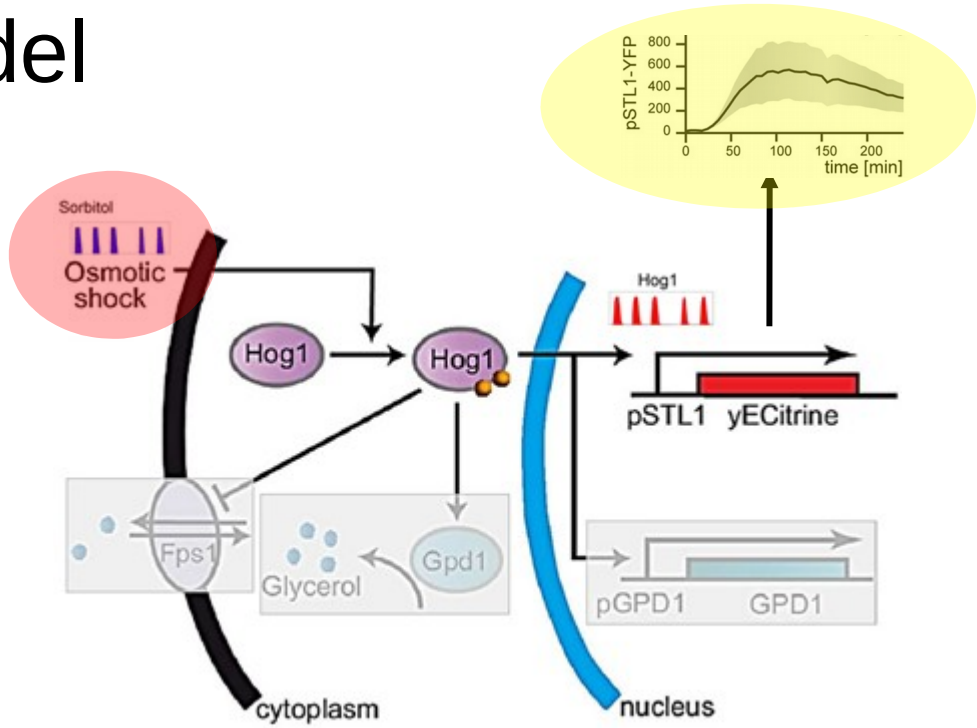
$$f(t) = e^{-g_p \tau} p(t - \tau).$$

Measurement noise intensity:

$$h(t) = (\varepsilon_a + f(t)\varepsilon_b)$$

Output:

$$y(t) = f(t) + h(t)\eta$$



Parameters:

$$\theta = (k_p, g_p, k_m, g_m),\ \tau,\ \text{initial conditions}$$

Our focus:

$$k_{mp} = k_m \cdot k_p,\ g_m,\ g_p$$

Mixed-Effects paradigm

- *N* cells are different individuals of the same population
- Different cell parameters following a common population distribution

$$\theta_1, \ldots, \theta_N \text{ i.i.d.}, \quad \log \theta_i \sim \mathcal{N}(\mu, \Sigma)$$

- *Inference from data :* Learn population statistics from the ensemble of the individual-cell data

$$Y = \{y_i(t) : \ \forall t, \ i = 1, \ldots, N\}$$

- *Key point :* Every individual contibutes to learning population features, and so it helps learning features of all other individuals

Mixed-Effects model inference

- Estimate population statistics by marginal likelihood maximization

$$(\hat{\mu}, \hat{\Sigma}) \leftarrow \max_{\mu,\Sigma} p(y|\mu,\Sigma), \quad p(y|\mu,\Sigma) = \prod_i \int p(y_i|\theta_i)\, p(\theta_i|\mu,\Sigma)\, d\theta_i$$

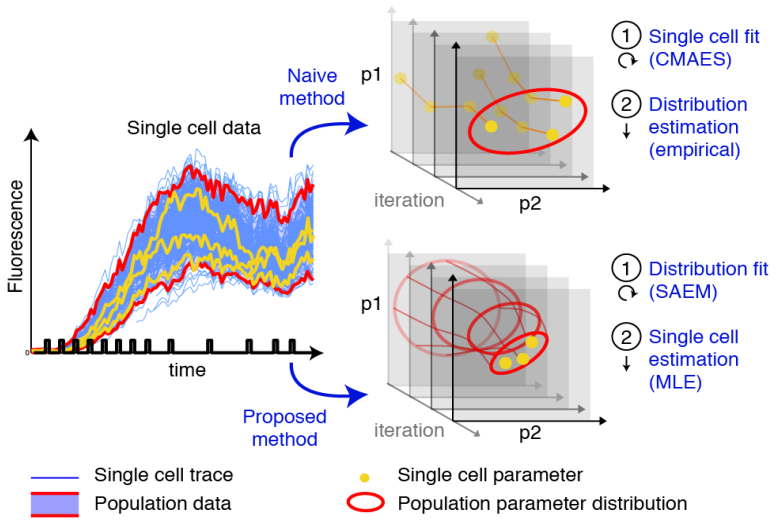
- Estimate single-cell parameters via Maximum-A-Posteriori

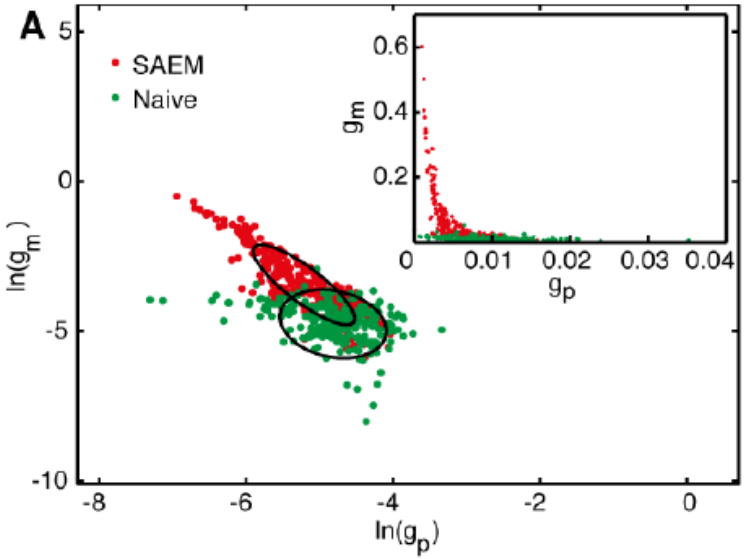
$$\hat{\theta}_i \leftarrow \max_{\theta_i} p(\theta_i|y_i,\hat{\mu},\hat{\Sigma}), \qquad p(\theta_i|y_i,\hat{\mu},\hat{\Sigma}) \propto p(y_i|\theta_i) \cdot p(\theta_i|\hat{\mu},\hat{\Sigma})$$

- Effective implementation (SAEM):
  Stochastic Approximation of Expectation-Maximization

Mixed-Effects vs. naive inference

- Naive approach: Empircal parameter statistics from individual cell fits





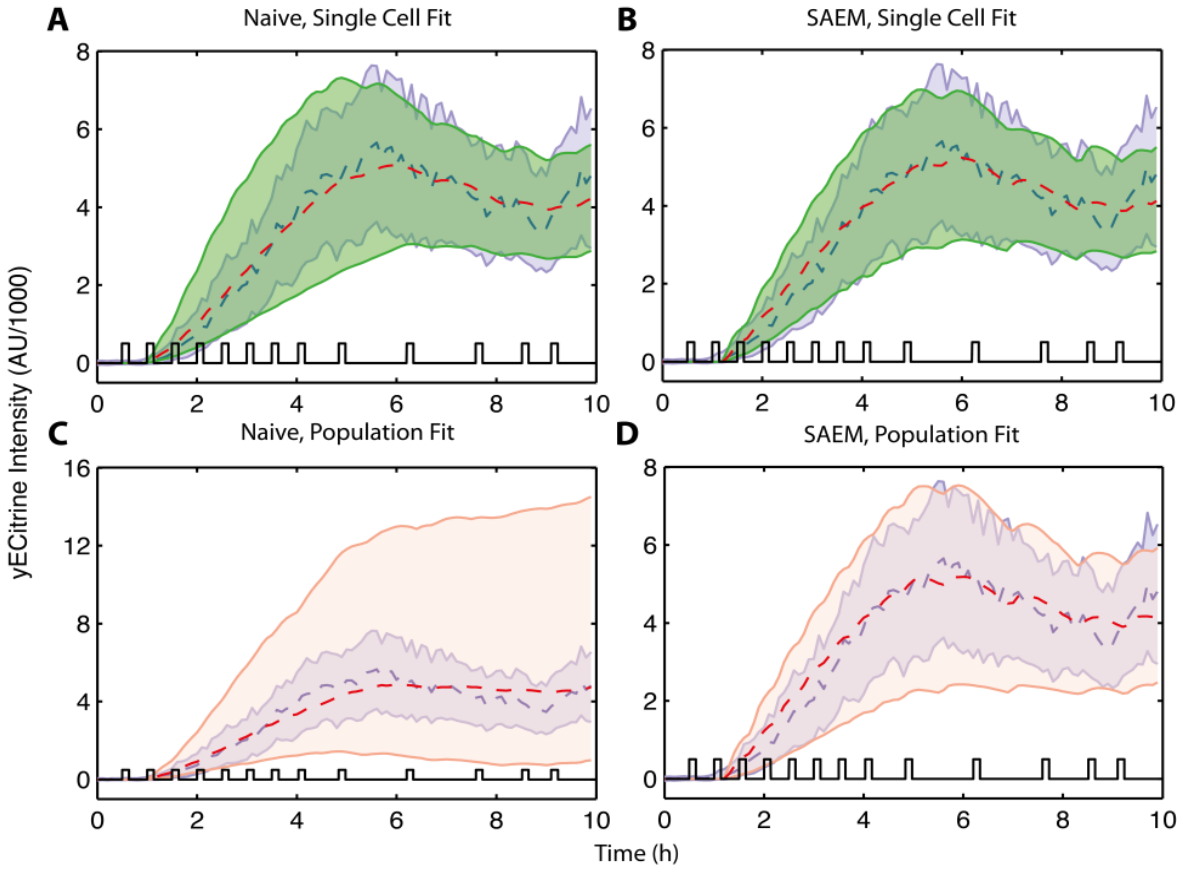
	Parameter distribution structure	Parameter distribution spread
SAEM	0.62	6.7
Naive	0.19	37

- Parameter distribution from ME inference more structured and compact

ME inference *does* retrieve population statistics

Single-cell trajectories from parameter estimates (confidence bounds)

Resimulation of single-cell trajectories via parameter resampling (confidence bounds)



Naive approach

ME approach

Mother-daughter parameter inheritance

- Daughter cell parameters statistically closer to those of their mothers
- A formal yet 'irrefutable' observation to be investigated further

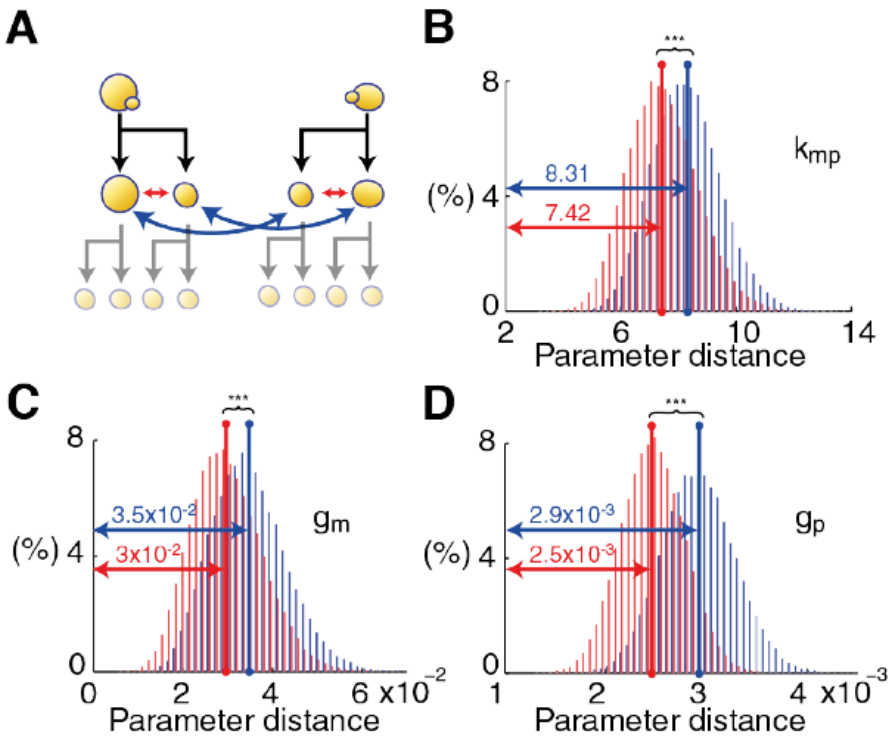
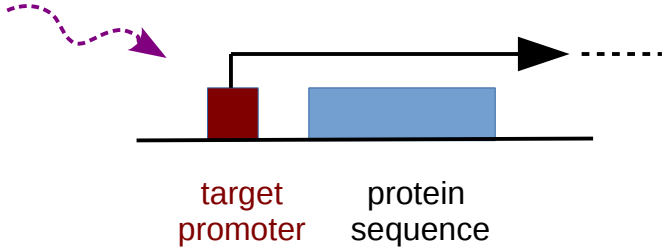


Fig. 7: Parameter values of individual cells are statistically closer to the parameters of their own mother than to the parameters of another mother cell. (A) The distance between parameters of related mother and daughter cells (MD) and non-related mother and daughter cells (nMD) were compared. (B-D) Distribution for each parameter of the average distance between 40 pairs of MD (red) and nMD (blue) for 50000 combinations obtained by bootstrapping (Text S1). All parameters are closer between mothers and daughters than on average (*** *p*-values < 10^{-15}).

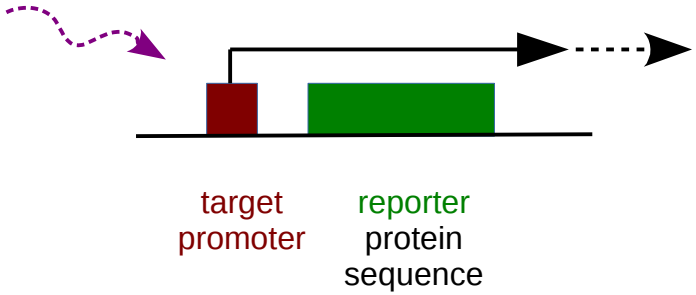
Stochastic variability in promoter activity

Population snapshot data

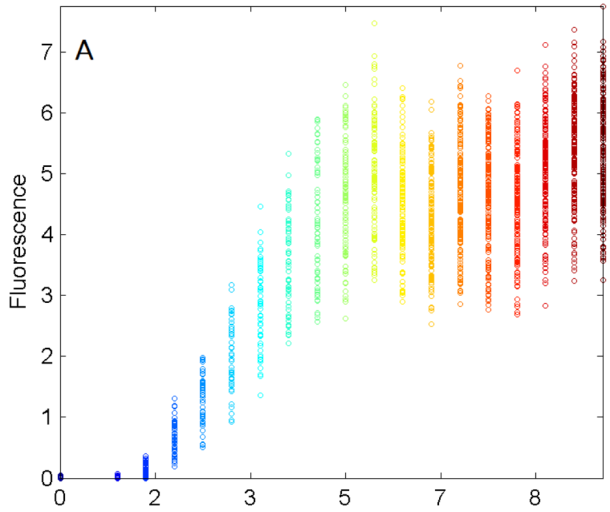
Gene of interest



Gene reporter system



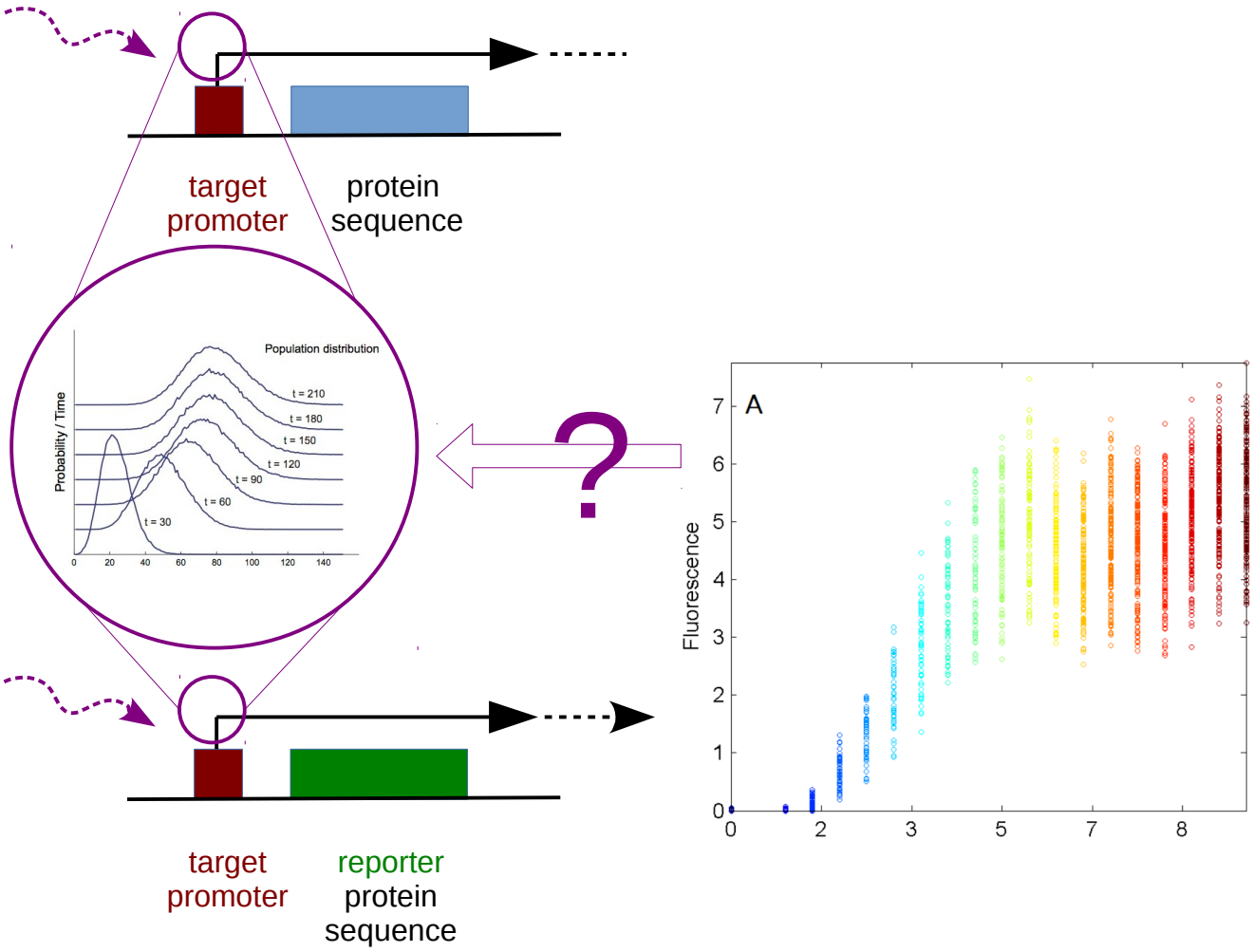
Fluorescence distribution in samples from a population of cells:



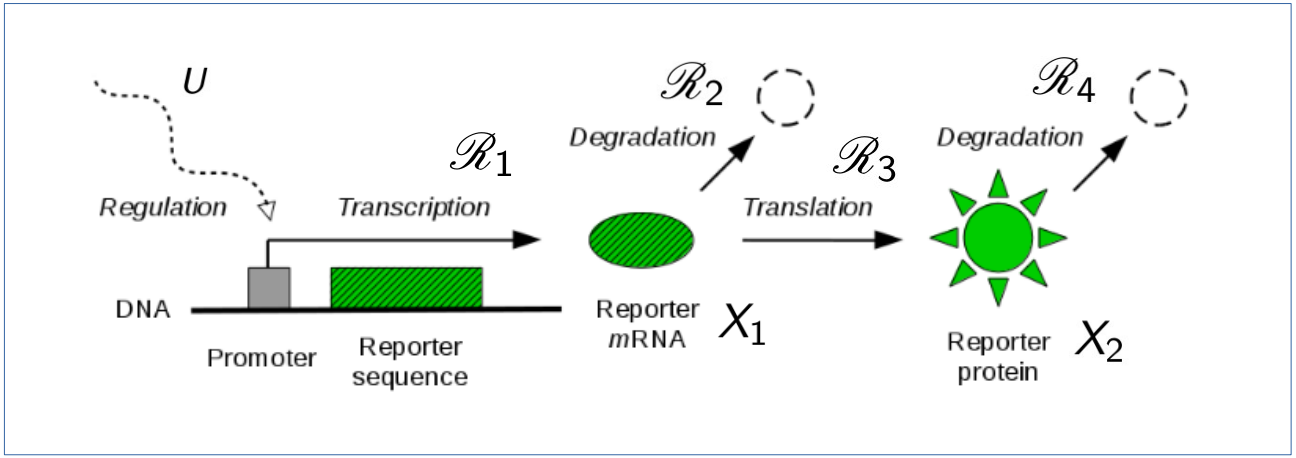
Goal : Inference of promoter activity statistics

Gene of interest

Gene reporter system



Random telegraph model



- Reactions :

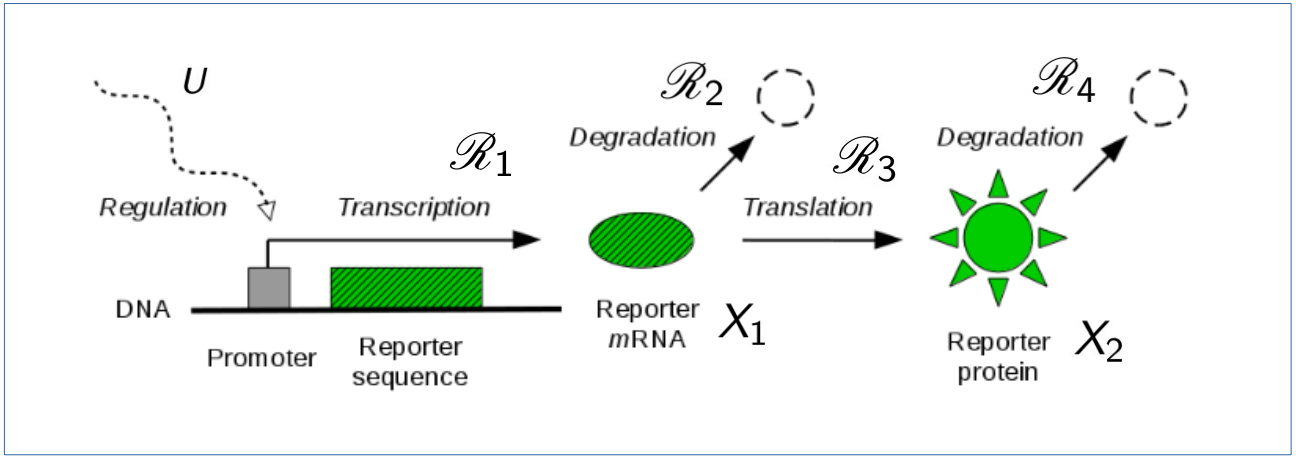
$$\mathscr{R}_1 : \emptyset \xrightarrow{k_M \cdot U} M \qquad \mathscr{R}_2 : M \xrightarrow{d_M} \emptyset$$

$$\mathscr{R}_3 : M \xrightarrow{k_P} M + P \qquad \mathscr{R}_4 : P \xrightarrow{d_P} \emptyset$$

- Stoichiometry matrix and reaction rates :

$$S = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix} \qquad w(t) = \begin{bmatrix} k_M U(t) \\ d_M X_1(t) \\ k_P X_1(t) \\ d_P X_2(t) \end{bmatrix}$$

Random telegraph model



- Reaction rates are affine in the state :

$$w(t) = WX(t) + F(t)$$

$$W = \begin{bmatrix} 0 & 0 \\ d_M & 0 \\ k_P & 0 \\ 0 & d_P \end{bmatrix} \qquad F(t) = \begin{bmatrix} k_M U(t) \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

Moment Equations

- First- and second-order statistics of X :

$$\mu(t) = \mathbb{E}[X(t)], \quad \Sigma(t) = \mathrm{Var}\big(X(t)\big), \quad \rho(z,t) = \mathrm{Cov}\big(X(z), X(t)\big)$$

- If U (that is, F) is a deterministic function :

$$d\mu(t)/dt = SW\mu(t) + SF(t)$$
$$d\Sigma(t)/dt = SW\Sigma(t) + \Sigma(t)W^T S^T + S\mathrm{diag}\big(W\mu(t) + F(t)\big)S^T$$

$$\partial\rho(z,t)/\partial z = SW\rho(z,t)$$

Generalized Moment Equations

- Now assume F is a(ny) stochastic process :

$$\mu_F(t) = \mathbb{E}[F(t)], \quad \rho_F(z,t) = \mathrm{Cov}\big(F(z), F(t)\big), \quad \xi_F(t) = \mathrm{Cov}\big(X(0), F(t)\big)$$

- Assuming *absence of feedback* from X to F :

$$\begin{aligned} d\mu(t)/dt &= SW\mu(t) + S\mu_F(t) \\ d\Sigma(t)/dt &= SW\Sigma(t) + \Sigma(t)W^T S^T + S\mathrm{diag}\big(W\mu(t) + \mu_F(t)\big)S^T \\ &\quad + V_{\xi_F}(t,t) + V_{\xi_F}^T(t,t) + V_{\rho_F}(t,t) + V_{\rho_F}(t,t)^T \\ \partial\rho(z,t)/\partial z &= SW\rho(z,t) + V_{\xi_F}(z,t) + V_{\rho_F}(z,t) \end{aligned}$$

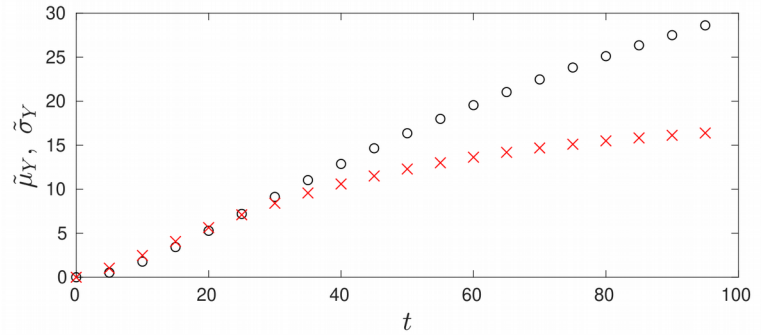
V_{ξ_F} linear (integral) functional of ξ_F

V_{ρ_F} linear (integral) functional of ρ_F

Inference of promoter activity statistics

- Given empirical statistics (population snapshot data)

$$\tilde{\mu}_2(t_k) = \mu_2^*(t_k) + e_k^{\mu}$$
$$\tilde{\sigma}_2^2(t_k) = \Sigma_{2,2}^*(t_k) + e_k^{\sigma}$$



with *k=1,...,K* and i.i.d. approx Gaussian noise

- Estimate unknown mean and autocovariance function of *U*

- Ill-posed linear inversion : Solve with (Tikhonov) regularization

Special case : *U* stationary, $X_0=0$

$$F(t) = \begin{bmatrix} k_M U(t) \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

- Unknown stationary statistics : $\bar{\mu}_U = \mu_U(\cdot), \quad \bar{\rho}_U(\delta) = \rho_U(\cdot + \delta, \cdot)$
- Estimate constant mean by fitting mean data with the solution of

$$\frac{d}{dt}\mu(t) = SW\mu(t) + \begin{bmatrix} k_M \bar{\mu}_U \\ 0 \end{bmatrix}$$

- Estimate autocovariance as the solution of the convex optimization

$$\min_{\bar{\rho}_U \in \mathscr{C}} \sum_{k=1}^{K} \alpha_k^2 \left(\tilde{\sigma}_2^2(t_k) - \mathscr{L}(t_k | \bar{\rho}_U) \right)^2 + \gamma \mathscr{Q}(\bar{\rho}_U)$$

PSD functions (convex cone)

Inverse of meas. error variance

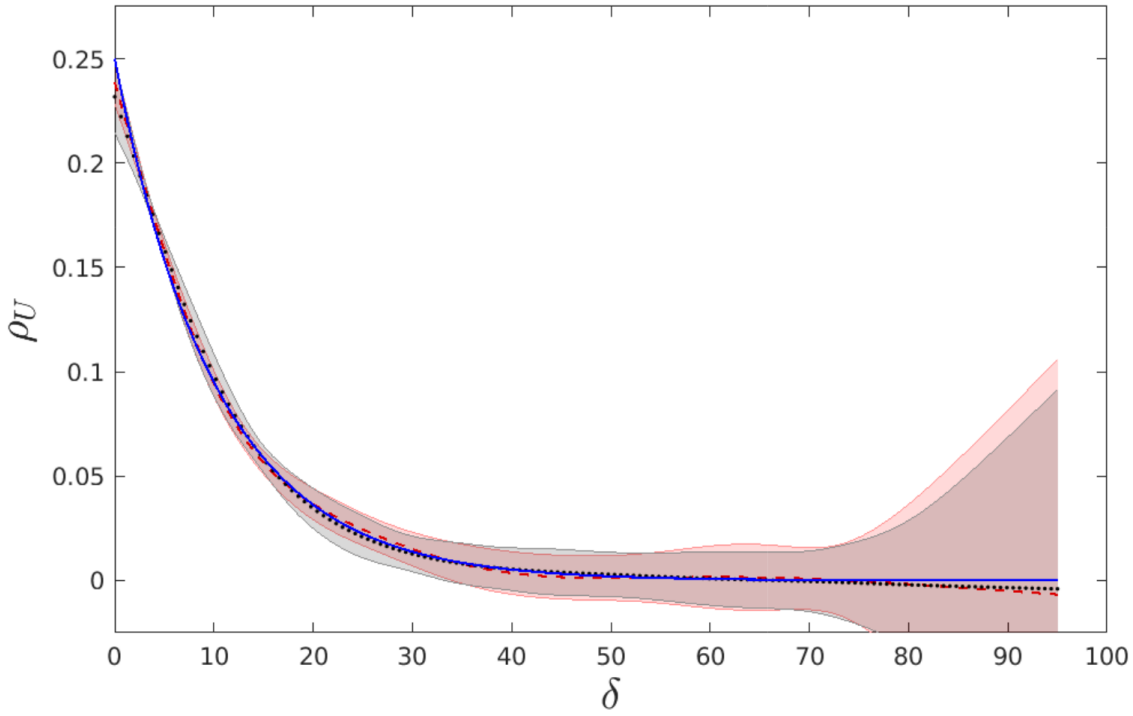
Solution at t_k of $d\Sigma(t)/dt = SW\Sigma(t) + \Sigma(t)W^T S^T + Q(t) + \Lambda(t|\bar{\rho}_U)$ with Q known function of the mean and Λ known functional

Convex penalization of irregular solutions

- Implementation : Finite-dimensional LQP

Results from numerical simulations

- Binary process *U* with stochastic switching rates
- Estimator mean ± 2 std for samples of 10^5 cells
- Automated vs. fixed choice of γ (vs. true autocovariance)



… Thanks !

eugenio.cinquemani@inria.fr
team.inria.fr/ibis