Cell Systems, Volume *9*

Supplemental Information

Of Gene Expression and Cell Division Time: A Mathematical Framework for Advanced Differential Gene Expression and Data Analysis

Katharina Baum, Johannes Schuchhardt, Jana Wolf, and Dorothea Busse

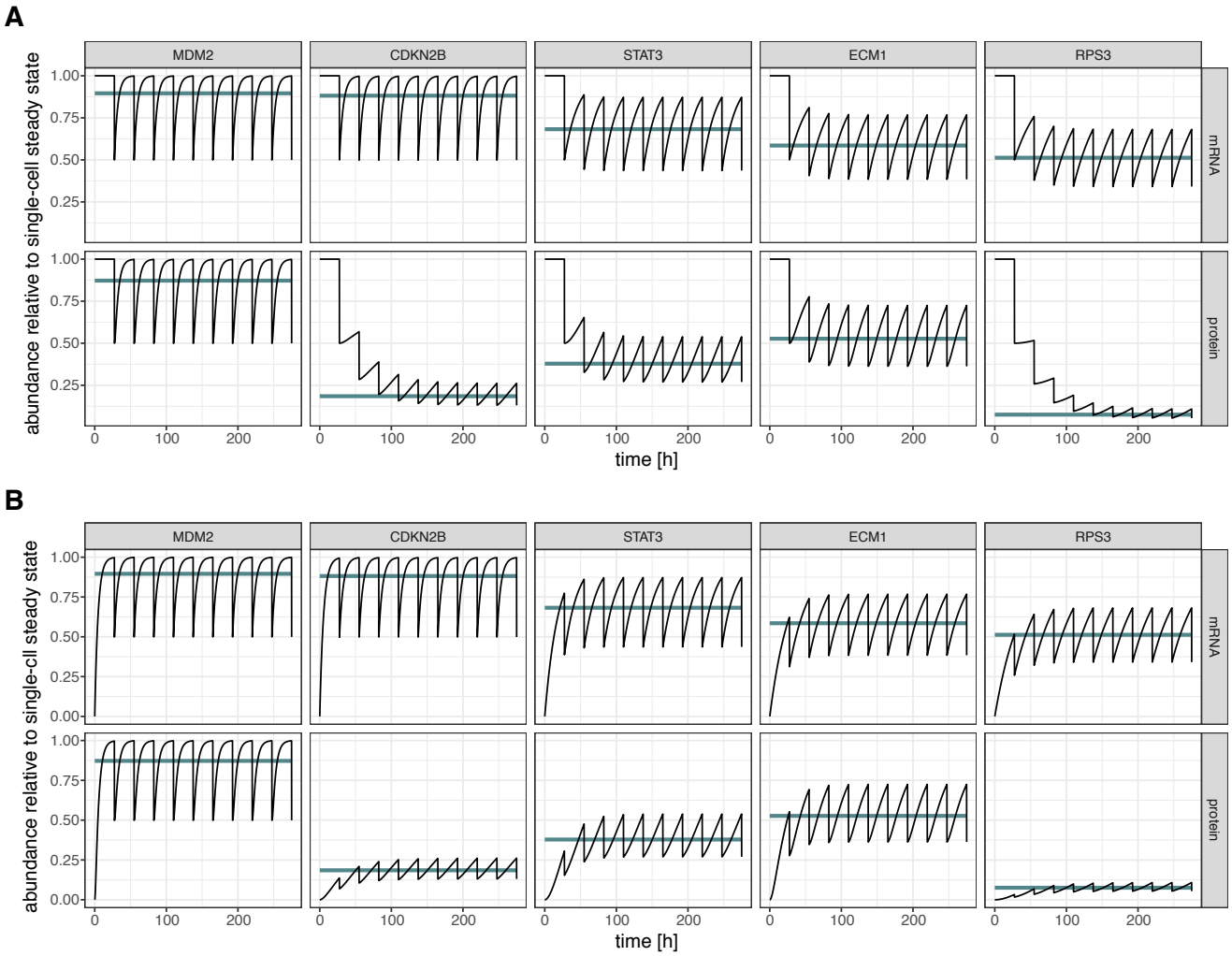

Figure S1. Single cell kinetics of mRNA and protein abundance. Related to Figure 2, Figure 3.
Depicted are relative mRNA and protein abundances of MDM2, CDKN2B, STAT3, ECM1 and RPS3 (black lines) over time for a cell which divides into two cells at $\tau = 27.5$ h; only one descendant cell is tracked (values of kinetic parameters given in Table S1). The blue lines give the population average abundances (relative to single cell steady state levels). Starting abundances are the steady state levels (**A**), or zero abundance (**B**). Two observations are made: First, the different mRNA and protein half-lives influence the relationship between the population average mRNA and protein abundances R and P (blue lines) and the corresponding single cell steady states (values are normalized to steady state, therefore the steady state values equal one). The longer the half-lives the more distant the population averages are from the steady states (compare also Figures 2B and 3). Second, the mRNA abundances and protein abundances within the single cells converge very fast: After 1-7 divisions, the abundance at cell birth of a daughter cell is similar to the abundance at cell birth of its mother cell, for both mRNA and protein ($r(\tau) = 2 \cdot r(0)$ and $p(\tau) = 2 \cdot p(0)$). mRNAs and proteins with long half-lives tend to take longer until the transient phase for reaching this state is completed.

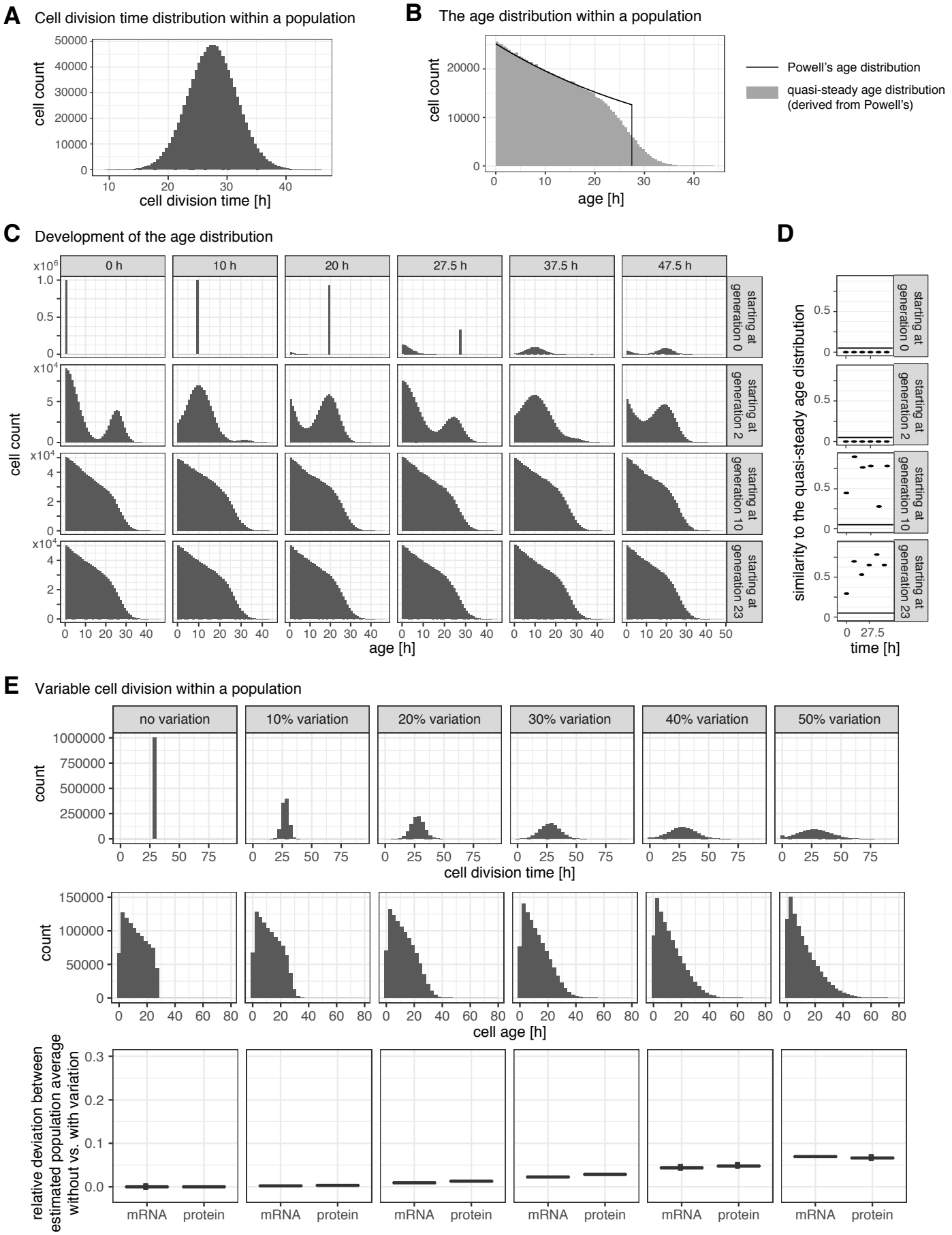

Figure S2. The age distribution within a population. Related to Figure 2.

A: Normal distribution of cell division times τ with a mean of 27.5 h and a standard deviation of 15%. Histogram for 10^6 cells.

B: Histogram of a quasi-steady age distribution for a population with cell division times as in A as derived by (Powell, 1956). Solid line: age distribution for cell populations with a cell division time of exactly 27.5 h.

C: Simulation of the development of an age distribution of 10^6 initially synchronized cells over time. The cell division time was updated three times per generation (0h, 10h, 20h) for 25 generations, cell division times of new-born cells were randomly assigned from a distribution as described in A. Distributions are shown only for generations 0-1 (first row), 2-3 (second row), 10-11 (third row), and 23-24 (fourth row). Finally, a stable age distribution evolves which is similar to that from B. Please note that the binwidths in the histograms in B and C are different and therefore the values on the y-axis differ.
D: Similarity of the simulated age distributions from C to the quasi-steady age distribution in B (as measured by the Benjamini-Hochberg corrected p-values of the Kolmogorov-Smirnov test for 10^4 cells at each update of the population). In each row, the similarities obtained for the six corresponding histograms from B are given as circles. Similarities above the black horizontal line mean that the distributions are statistically identical (corrected p-value < 0.05). Over time, the age distribution develops towards the quasi-steady distribution corresponding to Powell's.
E: Histograms of the cell division times τ within a population of 10^6 cells for different degrees of variation around 27.5 h (top) and the corresponding quasi-steady age distributions within the populations (middle). Bottom: Boxplots of relative deviation of the simulated STAT3 mRNA and protein population average abundances with variation in cell division times (and otherwise identical kinetic parameters) from the respective abundances without variation, for 100 populations each. Only slight differences up to 7% are observed even for large variation in the cell division times and consequently age distributions.

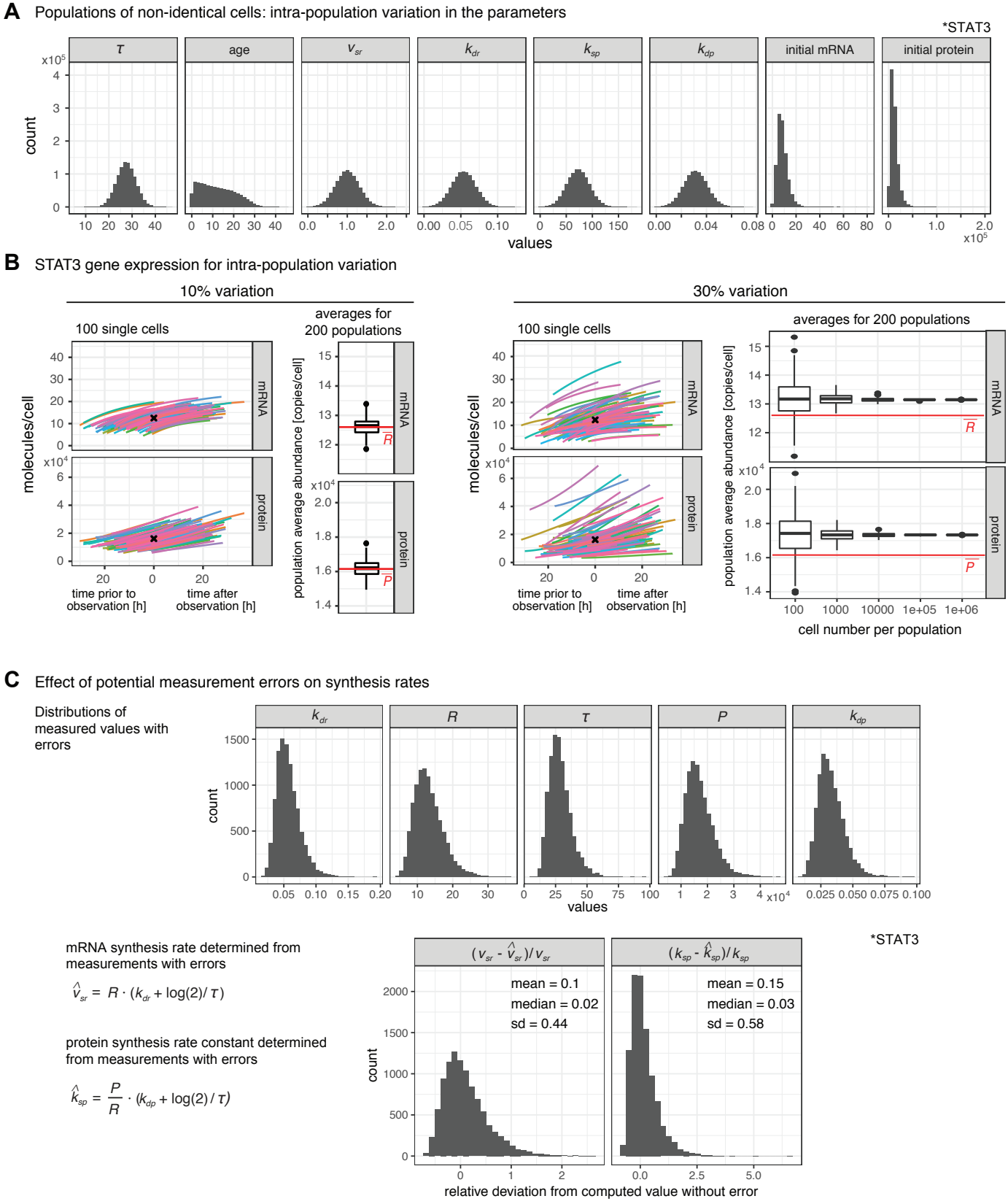

Figure S3. Populations of non-identical cells, and effect of potential measurement errors on synthesis rates. Related to Figure 2.

A: Distributions assumed for the cell division time, cell age, kinetic parameters and initial conditions for representing variation between single cells within a population (histograms of 10^6 sampled values for the example of STAT3, parameter values given in Table S1, τ = 27.5 h).

B: Simulations of STAT3 gene expression dynamics (mRNA top and protein bottom) in 100 single cells are shown allowing for 10% (left) and 30% (right) variation in v_{sr}, k_{dr}, k_{sp}, k_{dp} and the initial abundances at cell birth, and 15% for the cell division time. The population averages (arithmetic mean) of 200 such simulations of a population of ≥ 100 cells (100, 1000, 10^4, 10^5 or 10^6 cells, right) are presented in boxplots. The red lines show the calculated average population mRNA and protein abundances using the equations in Figure 2D.

With increasing variability, the small positive shift between the simulated average mRNA and protein abundances and the analytically derived abundances assuming identical cells increases slightly. With increasing cell number per population, the variation between populations is reduced and the shift turns more stable.

C: Effect of possible measurement errors on the calculated mRNA and protein synthesis rate constants. For the measured quantities: degradation rate constant k_{dr}, population average mRNA abundance R, cell division time τ, population average protein abundance P and protein degradation rate constant k_{dp} we assumed a log-normal distribution with standard deviation of 30% (top, shown for quantities of STAT3 for 10^4 sampled values, see Table S1). The relative deviations of the calculated synthesis rate constants with and without measurement error are characterized by the resulting distributions. These have a larger width, standard deviations are 44% and 58% for mRNA (bottom left) and protein (bottom right), respectively. Similar dispersions between 41-46% for the transcription rate and 57-64% for the translation rate constant are obtained for other mRNA-protein pairs (Table S1) for cell division times sampled around 27.5 h.

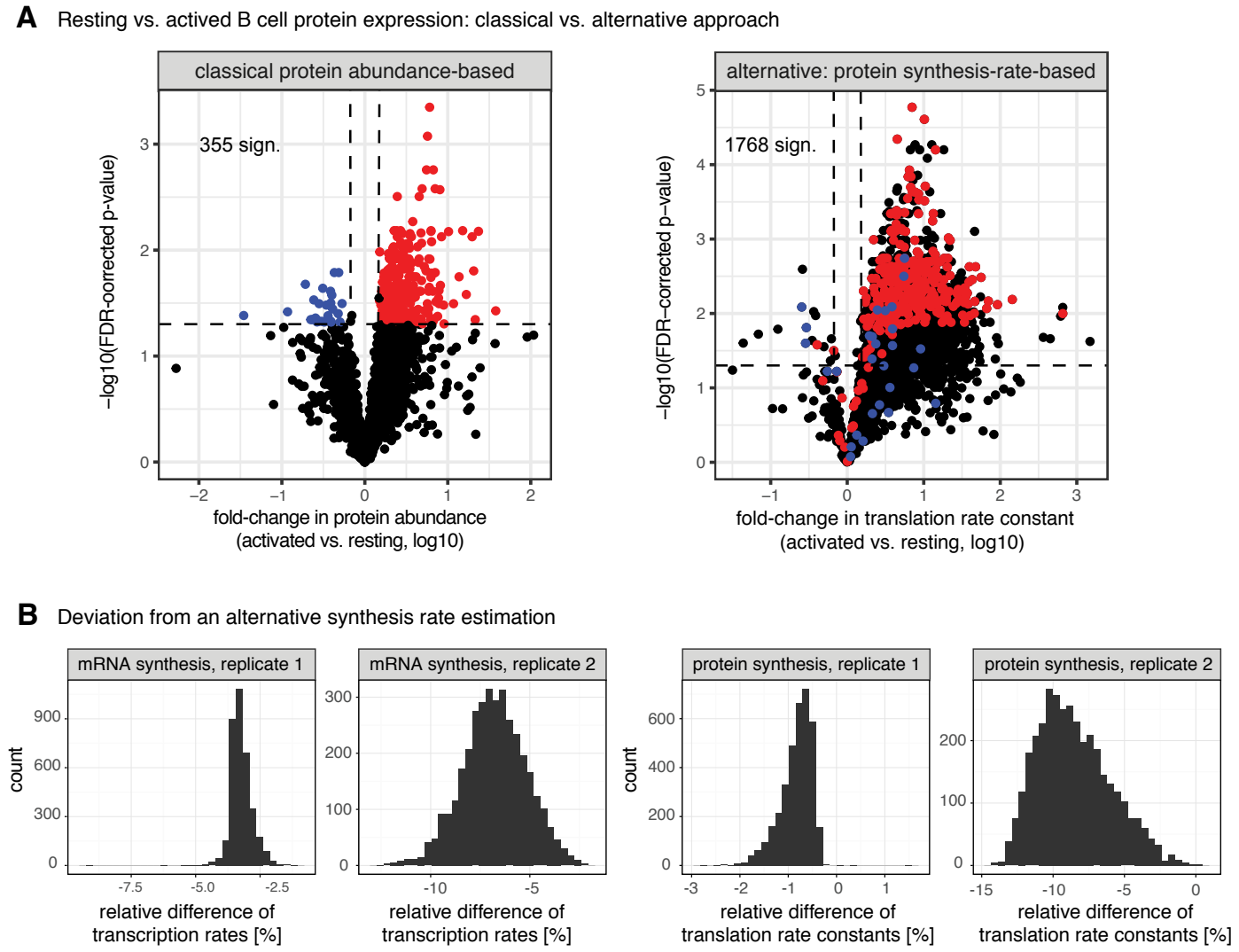

Figure S4. Application of the derived formulas. Related to Figure 4.

A: Differential protein expression in resting vs. activated B cells. Benjamini-Hochberg-corrected Welch's test p-values vs. fold changes of protein abundances as measured by IBAQ (Rieckmann et al., 2017) (classical approach, left) or of protein synthesis rate constants as computed from Eq. Figure 2D (STAR methods Eq. 17, alternative approach, right). The 327 proteins with significantly increased abundance using the classical approach are marked in red, the 28 proteins with significantly decreased abundance using the classical approach in blue (both left and right). 1768 proteins were detected as significantly different, 1442 of which up-regulated, using the alternative approach.

B: Deviation from an alternative synthesis rate estimation. Transcription and translation rates calculated according to the transformed equations in Figure 2D (STAR methods Eqs. 15 and 17) are compared to those calculated in (Schwanhäusser et al., 2013) (the latter indexed by 'Schw'). Shown are the differences of the values calculated for the 3569 mRNA-protein pairs with complete data for the two replicate data sets reported in (Schwanhäusser et al., 2013). Relative differences of the transcription rates, $(v_{sr}^{Schw} - v_{sr})/v_{sr}$, (left panels) and of the translation rate constants, $(k_{sp}^{Schw} - k_{sp})/k_{sp}$, (right panels) are given in percent for each replicate. In contrast to the consideration of a heterogeneous age-distribution in a population of growing cells, our earlier approach (Schwanhäusser et al., 2011) considers the time-average over a cell cycle. Overall, we found only small relative differences between the two approaches due to the near homogeneous age distribution of NIH3T3 cells.

Table S1: Parameter values for gene expression of STAT3, MDM2, CDKN2B, ECM1, RPS3. Related to Figure 2, Figure 3, Figure S1, Figure S2, Figure S3 and Table S2.
Kinetic parameter values derived from (Schwanhäusser et al., 2013) in NIH3T3 cells.

species	v_{sr} [no./h]	k_{sp} [1/h]	k_{dr} [1/h]	k_{dp} [1/h]	species long name
STAT3	1	72.5	log(2)/12.8	log(2)/22.1	Signal transducer and activator of transcription 3
MDM2	610.5	11.5	log(2)/3.2	log(2)/0.74	E3 ubiquitin-protein ligase
CDKN2B	2.98	386.62	log(2)/3.69	log(2)/103.51	Cyclin-dependent kinase inhibitor 2B
ECM1	3.72	15.92	log(2)/19.48	log(2)/3.06	Extracellular matrix protein 1
RPS3	15.25	913.95	log(2)/26.14	log(2)/159.34	40S ribosomal protein S3

Table S2: Sensitivity to variability between cells of the populations. Related to Figure 2, Figure S3.
We considered populations of non-identical cells with respect to the kinetic parameters of gene expression and cell division, and non-exact doubling of the abundances from cell birth to division. The sensitivity of the derived formulas (Eqs. Figure 2D) towards this intra-population variation is quantified by the relative difference, $(\widehat{R}\text{-}\overline{R})/\overline{R}$ or $(\widehat{P}\text{-}\overline{P})/\overline{P}$, between the population averages obtained when sampling 200 populations of 10^6 variable cells, $\widehat{R}$ or $\widehat{P}$, and the population average obtained for a population of identical cells, $\overline{R}$ or $\overline{P}$ (Eqs. Figure 2D). This relative shift is reported for five mRNA-protein pairs (Table S1) and a standard deviation of 30% in the kinetic parameters of gene expression and the initial abundances, and a standard deviation of 15% for different cell division times τ of 16 h, 27.5 h or 65.5 h. The sensitivity towards intra-population variability for the presented combinations of half-lives and cell division times is on average 5.8% for mRNA and 12.8% for protein. Even for the special case of the very unstable MDM2 mRNA and protein, the shift is only at most 32%.

species	mRNA half-life	protein half-life	τ	sensitivity mRNA	sensitivity protein
MDM2	short	short	16 h	7.6%	21.3%
	short	short	27.5 h	9.7%	25.3%
	short	short	65.5 h	12.4%	31.9%
CDKN2B	short	long	16 h	1.4%	9.6%
	short	long	27.5 h	2.9%	13.1%
	short	long	65.5 h	6.1%	19.5%
STAT3	intermediate	intermediate	16 h	2.5%	4.3%
	intermediate	intermediate	27.5 h	4.3%	7.3%
	intermediate	intermediate	65.5 h	7.7%	13.9%
ECM1	long	short	16 h	7.1%	7.5%
	long	short	27.5 h	9.2%	9.5%
	long	short	65.5 h	12.0%	13.2%
RPS3	long	long	16 h	0.8%	2.1%
	long	long	27.5 h	2.0%	2.9%
	long	long	65.5 h	4.9%	5.8%