

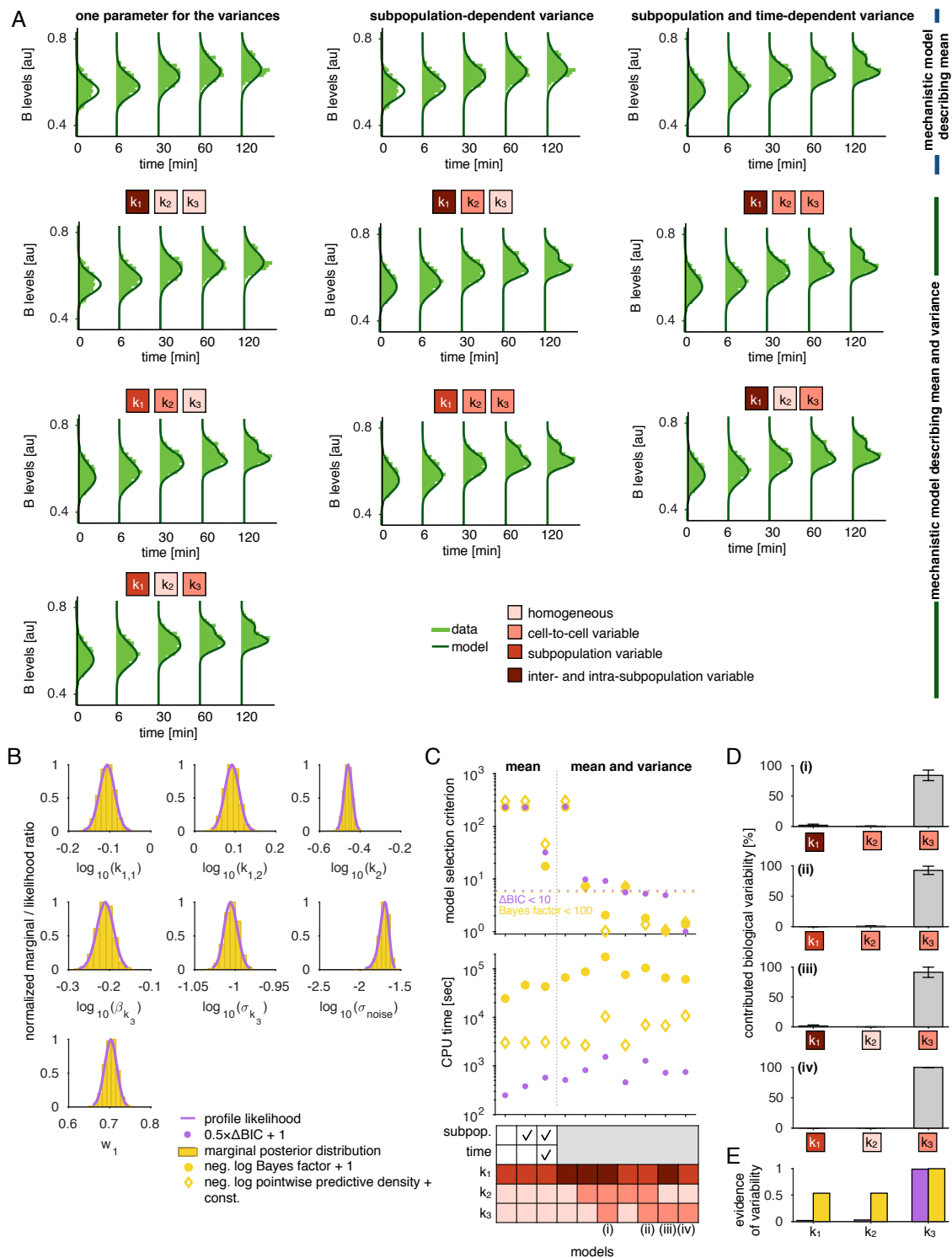
Cell Systems, Volume 6

Supplemental Information

A Hierarchical, Data-Driven Approach to Modeling Single-Cell Populations Predicts Latent Causes of Cell-To-Cell Variability

Carolin Loos, Katharina Moeller, Fabian Fröhlich, Tim Hucho, and Jan Hasenauer

Figure S1

**Figure S1, related to Figure 3. Analysis of the models for the conversion process.**

(A) Fitted models for the conversion process. Upper part: Models accounting only for the means using RREs according to Hasenauer et al. (2014). Lower part: Models accounting for the means and variances using sigma-point approximations.

(B) Normalized marginal posterior distribution computed from samples of the posterior distribution and likelihood ratio obtained by profile likelihoods for all parameters.

(C) Model selection criteria and required computation times for all models. Lower values indicate a higher evidence for the corresponding model. The horizontal dotted lines indicate the cutoff corresponding to a BIC difference of 10 and a Bayes factor of 100.

(D) Contribution to overall cell-to-cell variability of the observable for the models with Bayes factor < 100. The errorbars indicate deviation over time points.

(E) Evidence for variabilities in parameters computed based on BIC weights (left, purple) and marginal likelihoods (right, yellow).

Figure S2

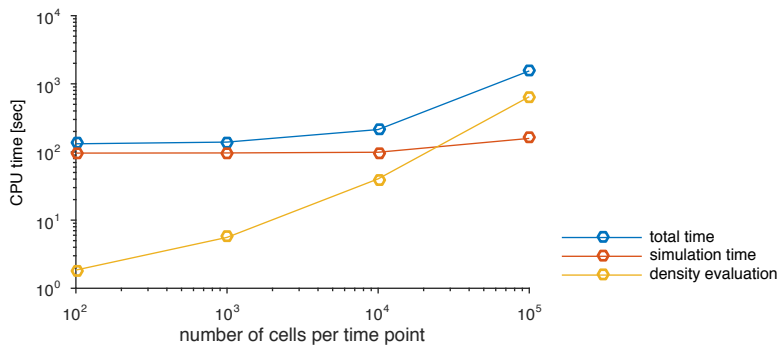


Figure S2, related to Figure 3. Computation time of the method for the conversion process. The overall computation time for 10 starts is depicted for varying number of measured cells per time point. The circles indicate the mean for three replicates. Different contributions to the overall computation time needed are shown: The time needed for the evaluation of the values for the cells under the density of the mixture distribution (yellow) and the time needed for simulation (orange).

Figure S3

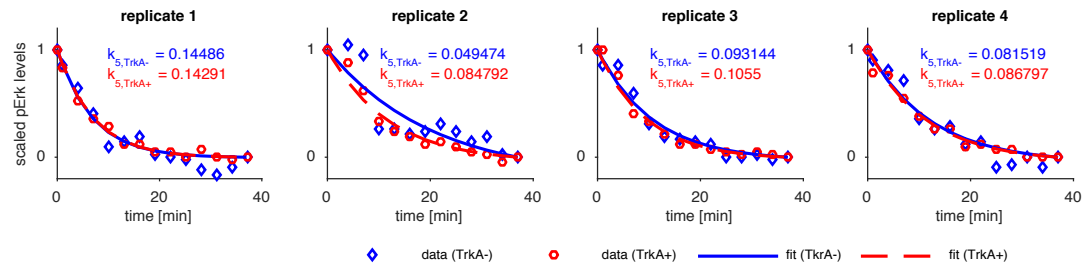


Figure S3, related to Figure 5. Erk1/2 dephosphorylation in TrkA- and TrkA+ subpopulations. Erk1/2 phosphorylation was induced in both neuronal subgroups TrkA- (blue) and TrkA+ (red) by an 1 h treatment with the combined stimuli NGF (acting on TrkA+ neurons) and GDNF (acting on TrkA- neurons expressing the GDNF receptor Ret). Subsequent inhibition of Mek by U0126 induced a pErk1/2 decline. Data of four individual experiments are shown with estimated exponential decay fit. The corresponding values k_5 for the dephosphorylation are noted for both subpopulations.

Figure S4

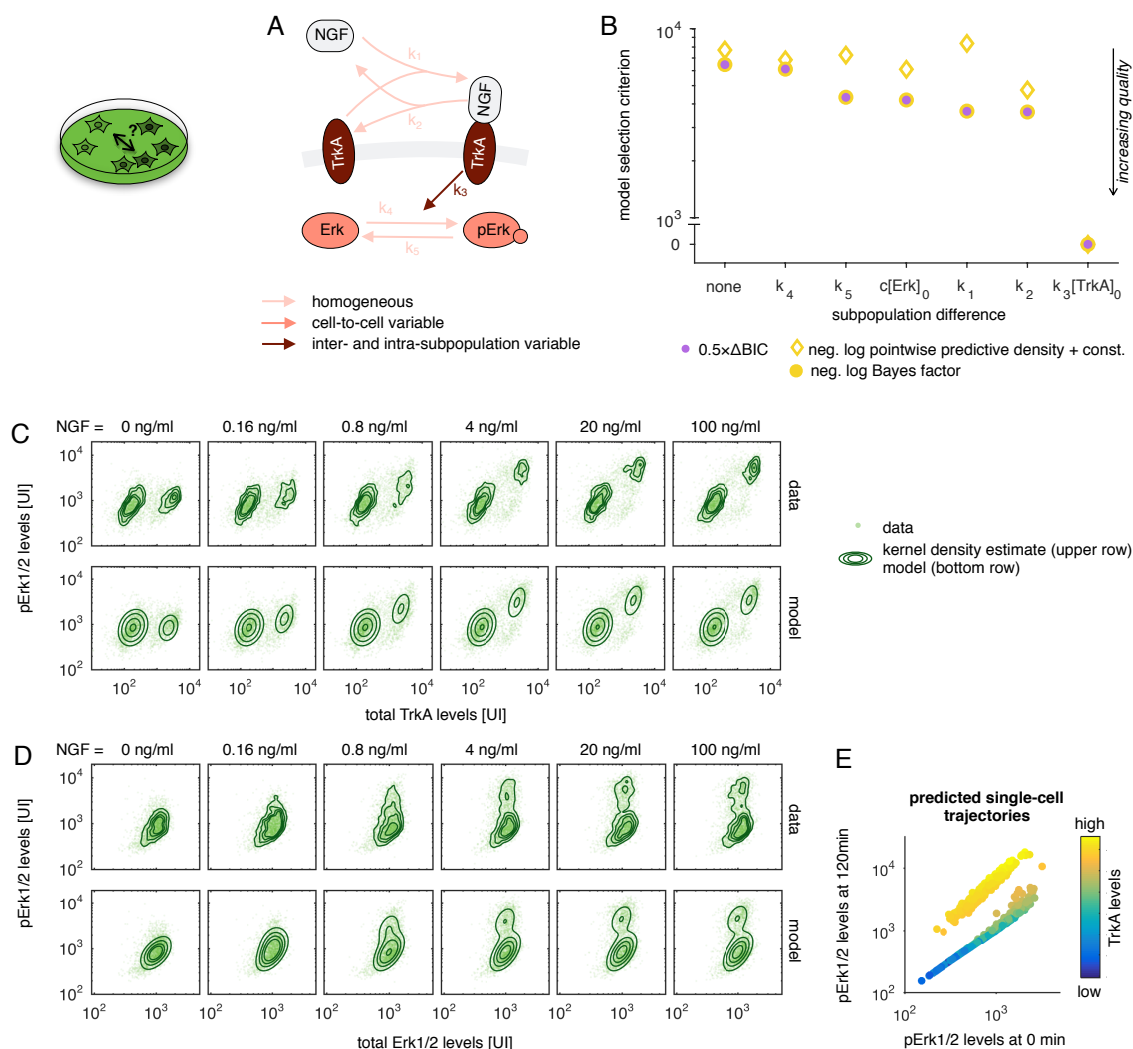


Figure S4, related to Figure 5. Data and fit for NGF-induced Erk1/2 signaling on PDL.

(A) Pathway model with color-coding of the variability of cellular properties.

(B) Comparison of models, which account for one or no difference between subpopulations, using BIC, log pointwise predictive densities, and Bayes factors.

(C,D) Data and fit for combined measurements of (C) TrkA and pErk1/2 levels and (D) Erk1/2 and pErk1/2 levels. The upper rows illustrate the data together with a kernel density estimate. The bottom rows visualize the data together with the contour lines of the hierarchical model.

(E) Predicted single-cell trajectories for the optimal parameter values, showing the relation between pErk1/2 levels in steady state (0 min) and after stimulation with NGF (120 min). The color of the cells indicates the TrkA level, which is assumed to be constant over time.

Figure S5

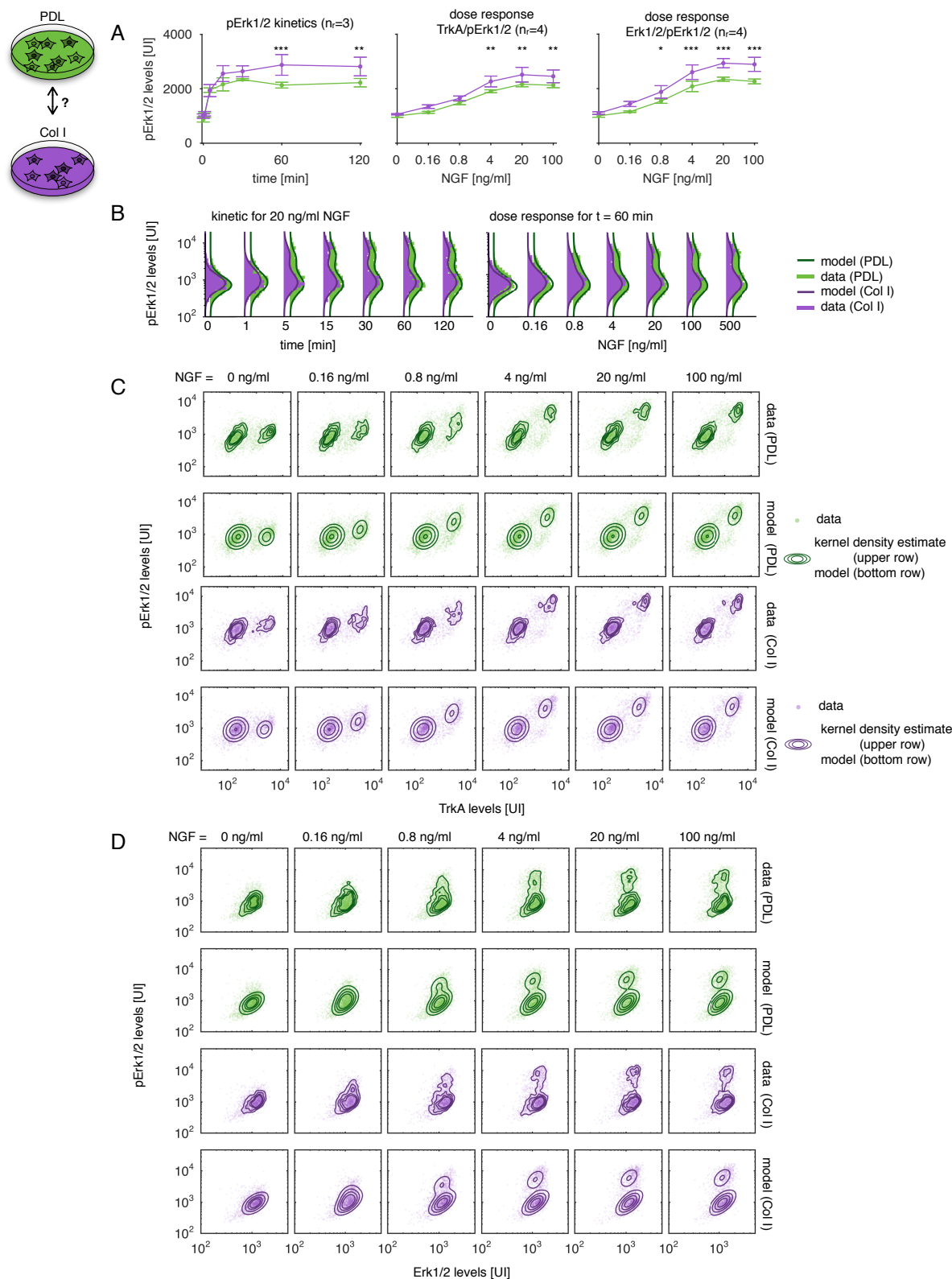


Figure S5, related to Figure 6. NGF-induced Erk1/2 signaling on different extracellular scaffolds.

(A) Mean response to NGF stimulation on Col I compared to PDL. The error bars indicate the standard deviation over replicates. A two-way ANOVA showed significant differences ($p < 0.01$) between the extracellular scaffolds for each experiment. Significances for individual time points/doses obtained by Sidak's multiple comparisons test are indicated by * ($p < 0.05$), ** ($p < 0.01$), and *** ($p < 0.001$).

(B-D) Data and fit for NGF-induced Erk1/2 signaling on different extracellular scaffolds.

(B) pErk1/2 kinetics and dose responses for PDL (green) and Col I (purple).

(C,D) Multivariate measurements of (C) pErk/TrkA levels and (D) pErk/Erk levels. The upper rows illustrate the data together with a kernel density estimate. The bottom rows visualize the data together with the contour lines of the hierarchical model, accounting for differences in Erk1/2 levels, Erk1/2 dephosphorylation, and cellular TrkA activity.

Figure S6

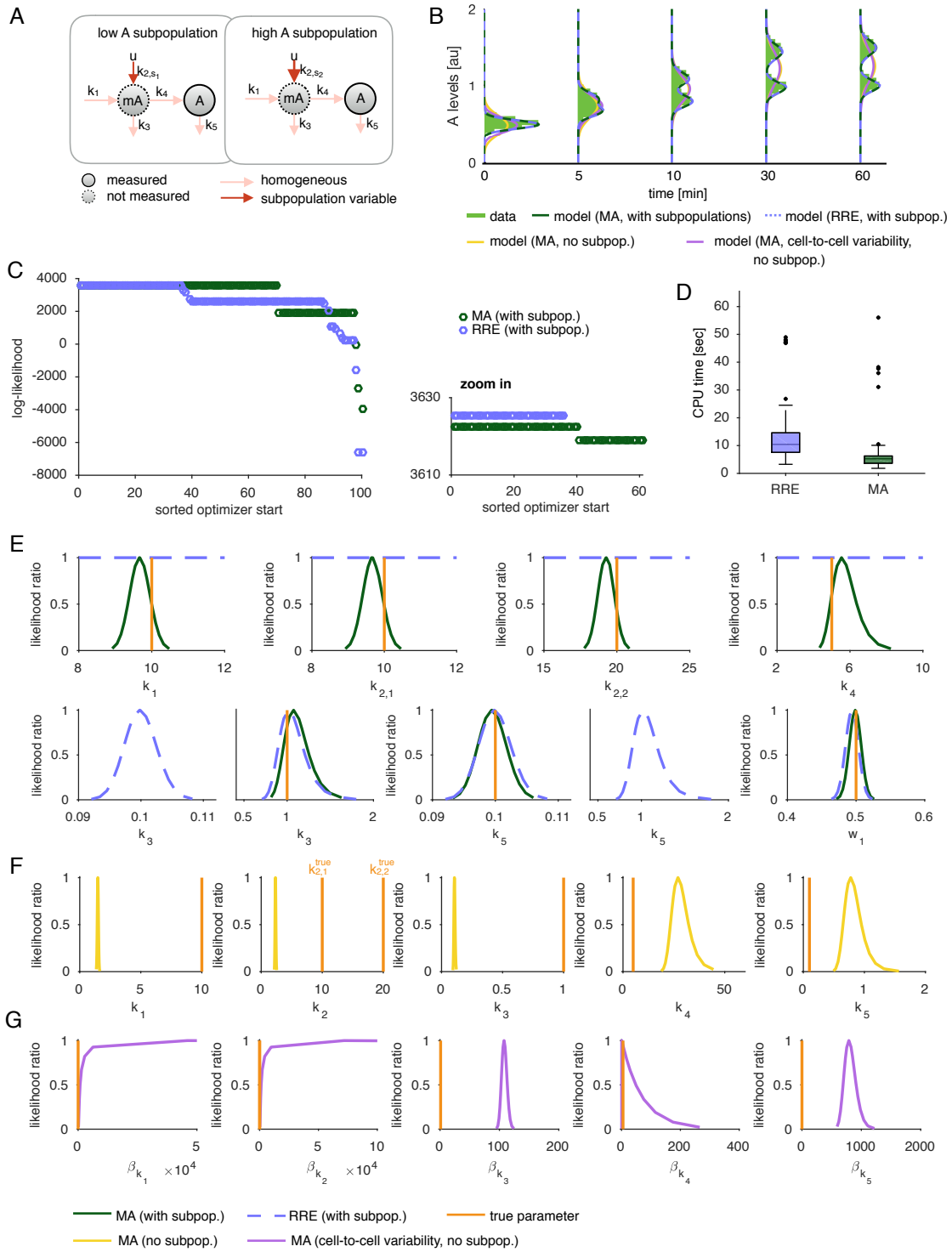


Figure S6, related to STAR methods, Accounting for intrinsic noise. Analysis of a stochastic two stage gene expression model.

(A) Illustration of the system.

(B) Data and fitted models for the moment-closure approximation (MA), for the case of accounting for subpopulation structures and disregarding subpopulation structures, and reaction rate equations (RRE).

(C) Log-likelihood values for 100 optimization starts sorted decreasingly for MA (green) and RRE (blue). The zoom in shows the 60 best optimization runs.

(D) Boxplot for the CPU time needed for one optimizer start.

(E) Profile likelihoods of the parameters for the models capturing the subpopulation structure.

(F) Profile likelihoods of the parameters for the model using the MA without accounting for subpopulations.

(G) Profile likelihoods of the means rates for the model using the MA, accounting for cell-to-cell variability of all parameters but not for subpopulations. This corresponds to the method proposed by Zechner et al. (2012). Note that the range in x-direction differs for subplots (E)-(G).