A Mathematical Insight Into Cell Labelling Experiments for Clonal Analysis

Noemi Picco, noemi.picco@swansea.ac.uk

Supplementary File

`EmbryonicEmx1CreERControlClones(HippenmeyerLab)-2018.xlsx`

Supplementary Figures

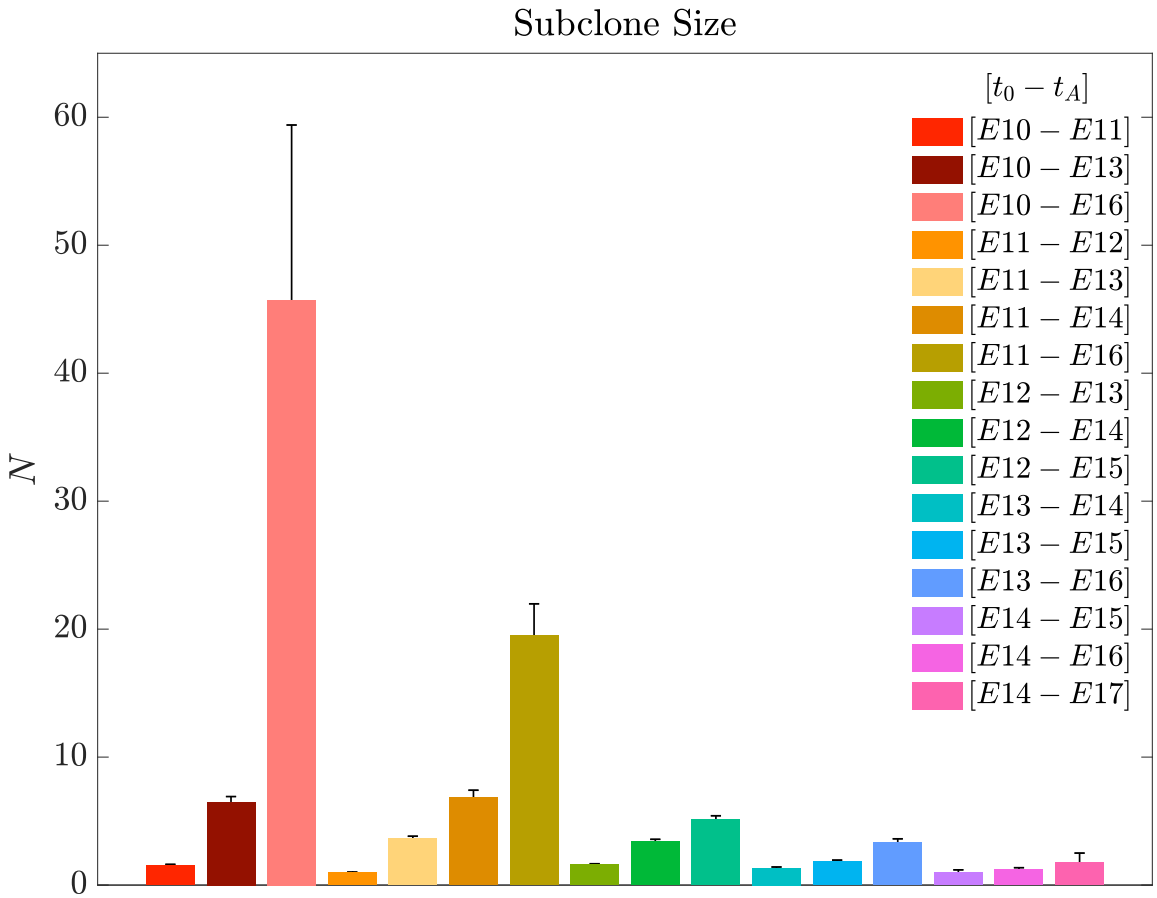


Figure S1: Subclonal distributions for all experimental windows in the MADM dataset. Bars show mean, errorbars show standard error of mean. Full dataset in `EmbryonicEmx1CreERControlClones(HippenmeyerLab)-2018.xlsx`

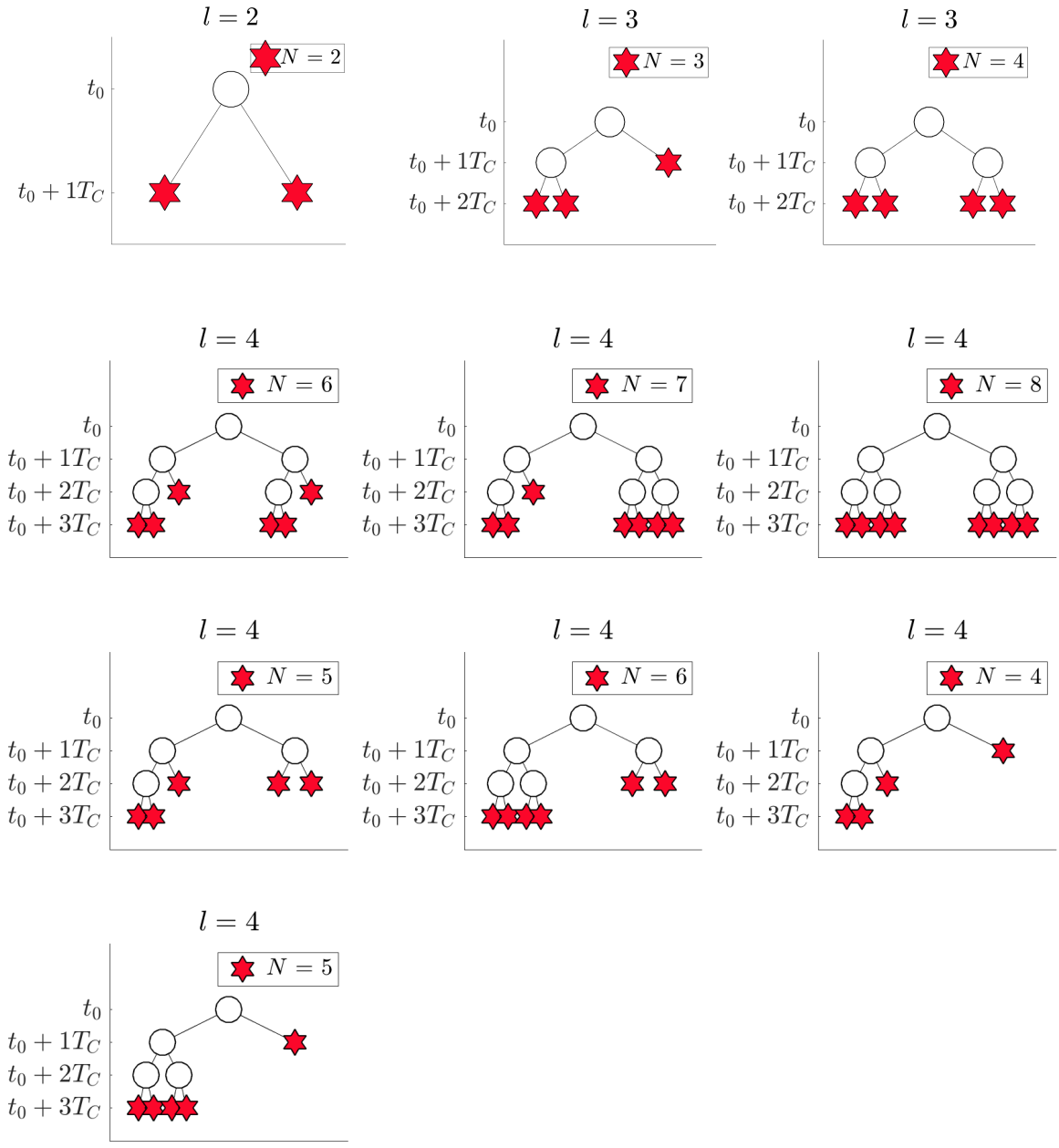


Figure S2: Full set of binary trees of depths $l = 2, 3, 4$ and corresponding clonal size N. A fixed cell cycling time T_C is assumed, as well as no cell death.

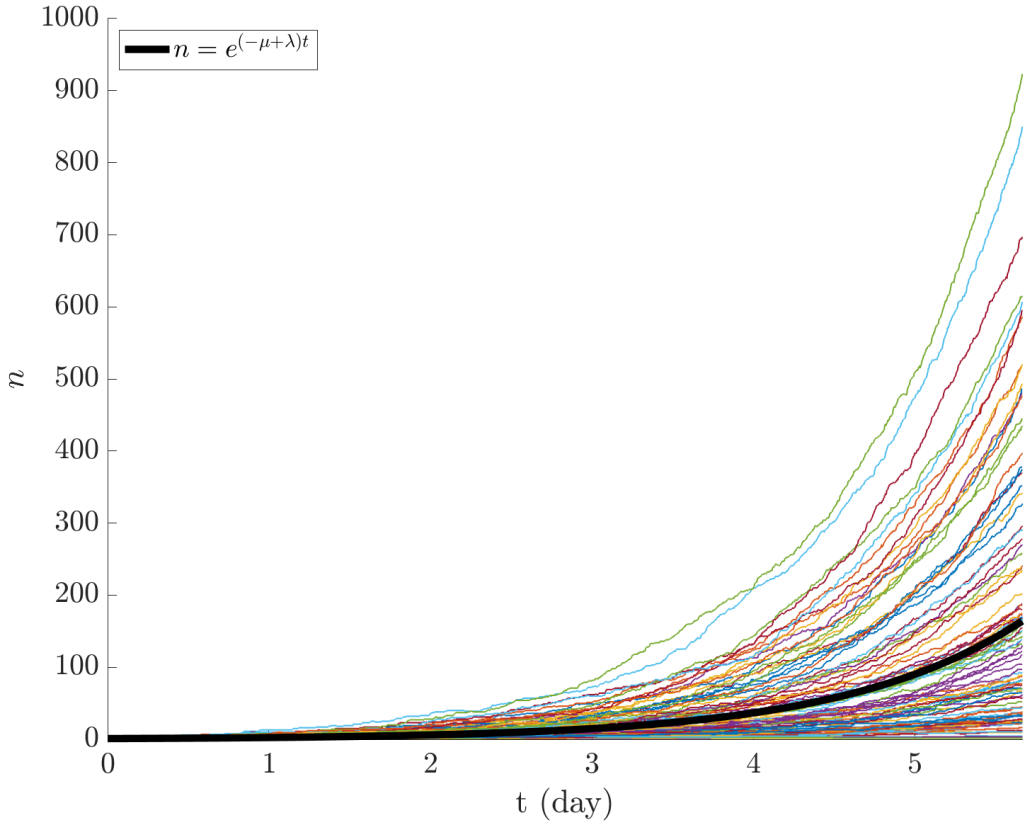


Figure S3: 100 realisations of the Gillespie algorithm for the stochastic one species birth-death process (P1) with $\lambda = 1$ day^{-1} and $\mu = 0.1$ day^{-1}. The vertical axis shows the size of the clone at time t. The deterministic solution is shown in black.

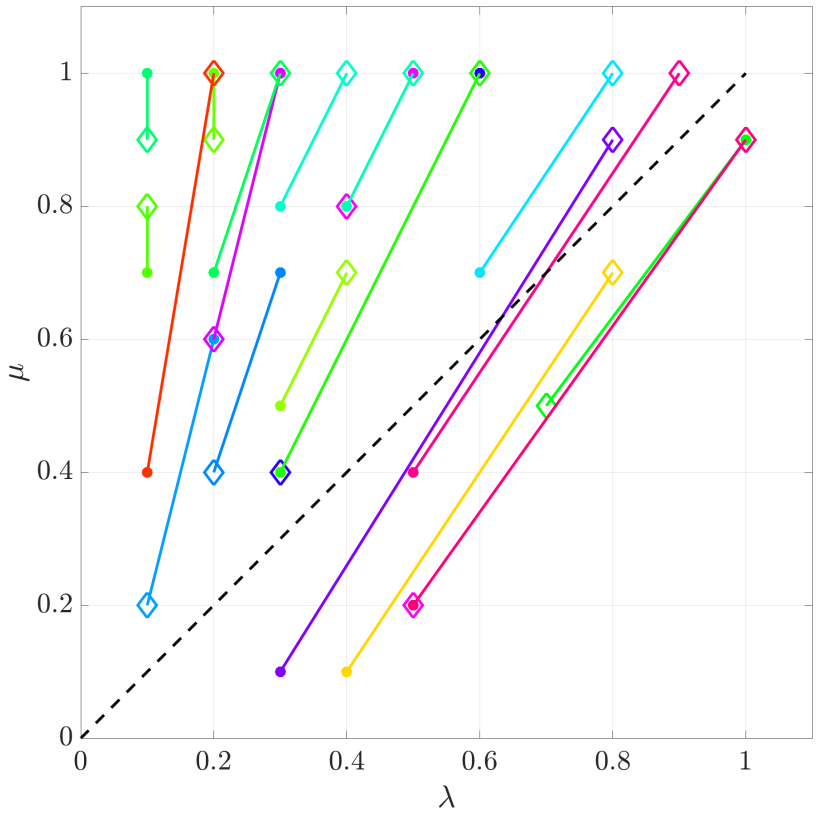


Figure S4: Testing the "recovery ability" of the least squares search algorithm. The recovery of known birth and death parameters is attempted for several points in the (λ, μ) space. A circle indicates the ground truth, and the diamond it links to indicates the recovered value $\left(\bar{\lambda}, \bar{\mu}\right)$. Only missed guesses are shown. The search algorithm is able to recover the ground truth in the region of biological interest (large birth rate λ, small death rate μ). The dashed black line indicates the $\lambda = \mu$ space where the model is not meaningful.

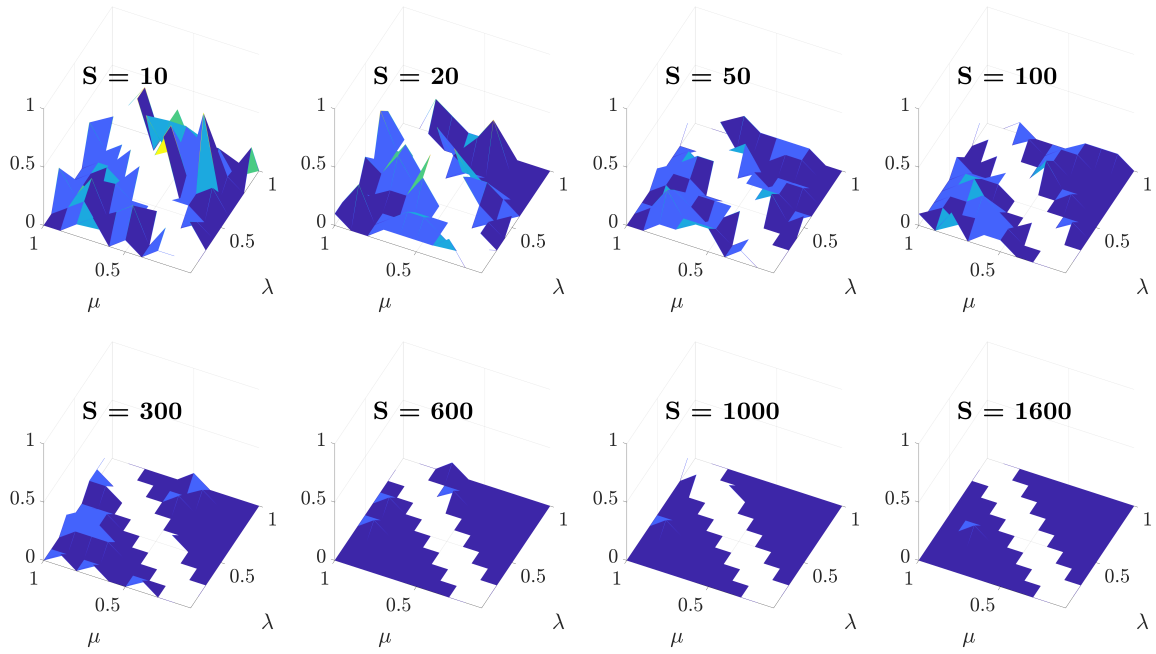


Figure S5: Testing the convergence of the estimate method for increased sample size S. The recovery of known birth and death parameters is attempted for several points in the (λ, μ) space, using an increasingly large dataset (S clones) extracted from the set of 2000 realisations of virtual clonal analysis. For each point (λ, μ) the error (plotted on the vertical axis) is calculated as the distance between the probability distribution functions evaluated at the ground truth point (λ, μ) and at the estimated point $\left(\bar{\lambda}, \bar{\mu}\right)$, across analysis times, according to:

$$error = \sqrt{\sum_{t_A} \sum_{n} \left(\mathbb{P}(N = n)|_{(\lambda,\mu)} - \mathbb{P}(N = n)|_{\left(\bar{\lambda},\bar{\mu}\right)} \right)^2}, \tag{S1}$$

where $\mathbb{P}(N = n)$ is defined in Eq. (1) of the main text.

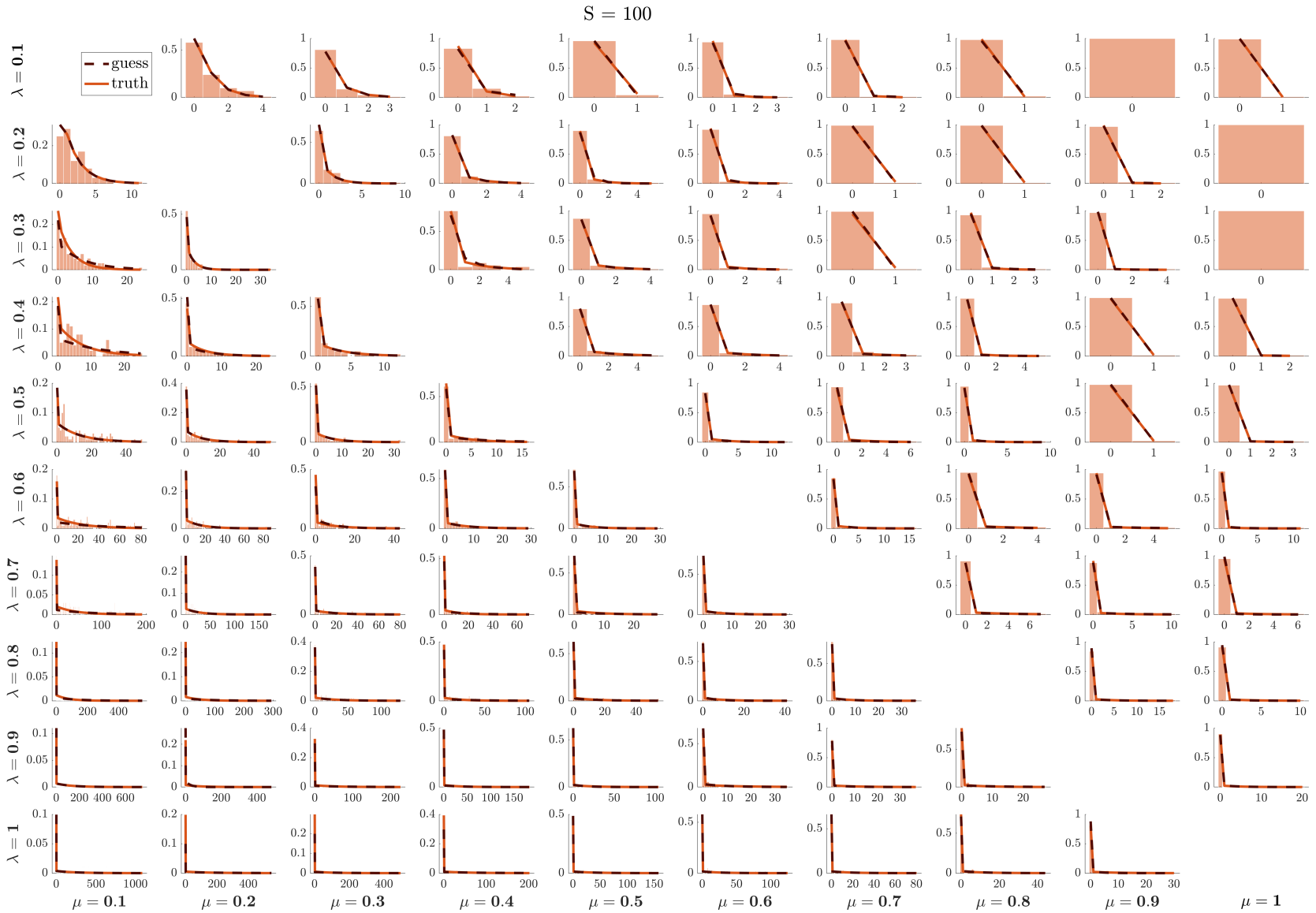
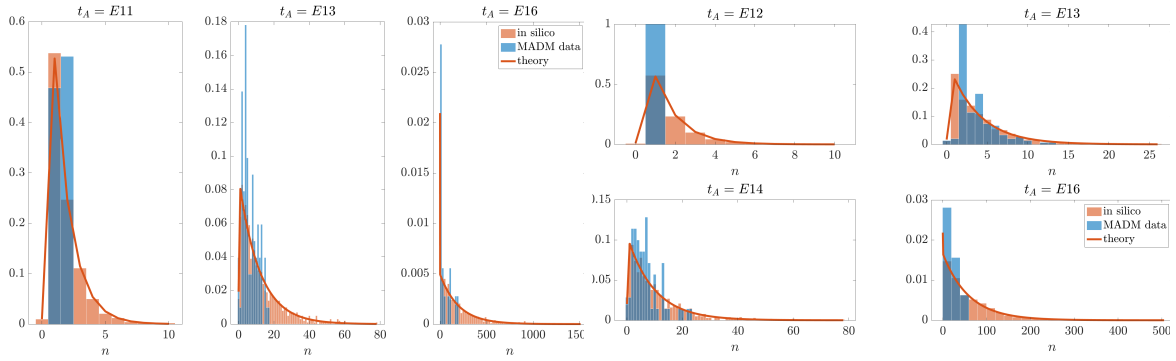
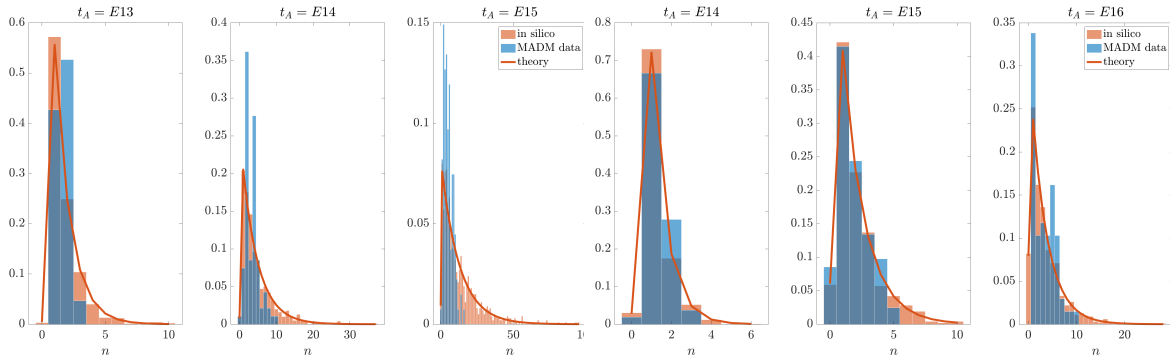


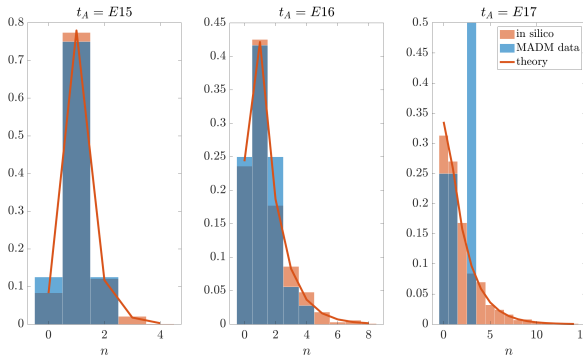
Figure S6: Recovering known values of λ and μ from $S = 100$ experiments of virtual clonal analysis with analysis times 1 day apart ($t_A = \{1, 2, \ldots 6\}$ days). For each ground truth point (λ, μ) only plots for $t_A = 6$ are shown. Histograms show the distribution of virtual clonal analysis data (on the horizontal axis is clone size n). The solid orange and dashed brown lines show the theoretical distributions $\mathbb{P}(N = n)$ parameterised on the ground truth (λ, μ) and the guess $\left(\bar{\lambda}, \bar{\mu}\right)$, respectively.



(a) $t_0 = E10 + 8$ hrs. (b) $t_0 = E11 + 9$ hrs.



(c) $t_0 = E12 + 10$ hrs. (d) $t_0 = E13 + 11$ hrs.



(e) $t_0 = E14 + 16$ hrs.

Figure S7: Comparing parameterised stochastic birth-death process (orange histograms) with the MADM dataset (blue histograms). The *in silico* data consist of 1000 realisations. The solid line is the theoretical distribution parameterised on the guessed values (λ, μ) for each injection time.

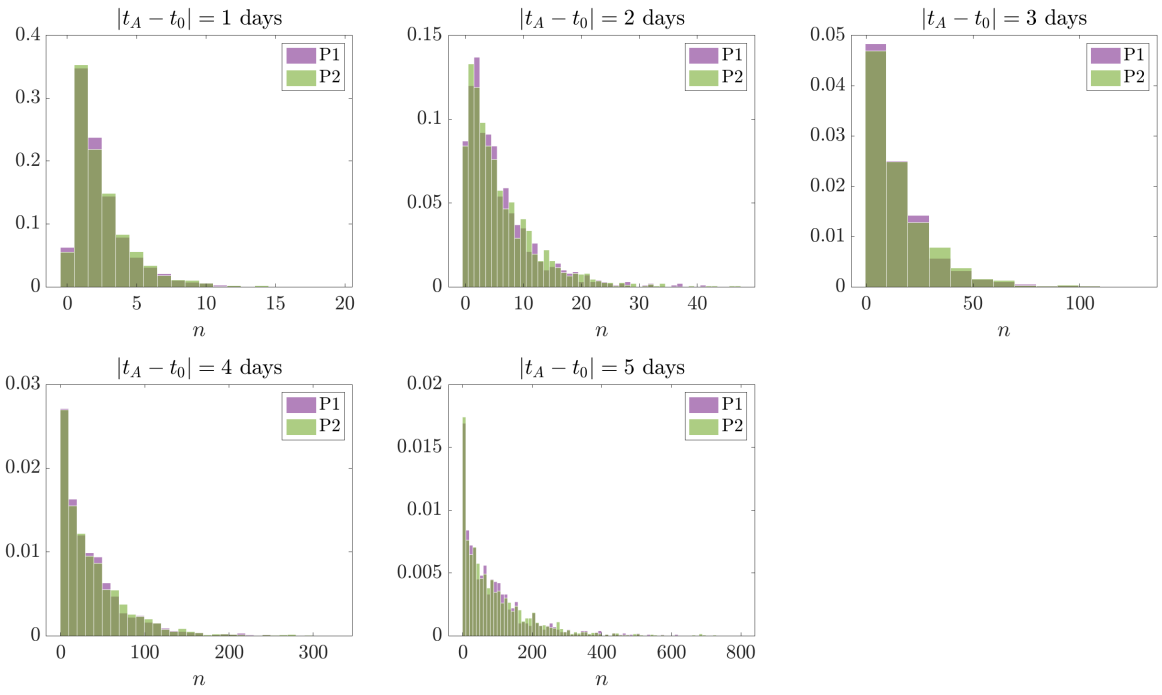


Figure S8: Equivalence of stochastic processes $P1$ and $P2$. One species process ($P1$) and two species process ($P2$) are identical when $\lambda_1 = \lambda$, and $\lambda_2 = 0$. Histograms show distributions of A cells, and $B + C$ cells, respectively.