Supplementary material

Contents

1 Supplementary text for in vitro experiment

1.1 Supplementary Result

1.1.1 Re-establishment of parental distribution after isolation of cells with high and low CD10 expression in Rael cells.

CD 10 expression is characteristic for Rael cells while CBM1 are typically CD10 negative. Using this marker we next characterized the dynamics of repopulating the basin of attraction. The population fractions with the highest and the lowest CD10 edge cells from Rael cell population (1,3% of the total population) were sorted by FACS and recultured separately(Fig.S1). These two edge subpopulations exhibited almost complete reconstitution of the parental distribution within 5 days of culture (Fig.S1). Thus with respect to a phenotype consistent marker (CD10) isolated edge cells (<2% of the whole parental population) were capable of re-establishment of the parental distribution. This behavior of edge cells was robust since it is observed in the various cell populations and from edges at the different 'sides' of the parental population.

1.1.2 Irregularities in the relaxation process and reduced proliferation and viability of the 'edge cells'.

A second peak with the small, distinct population to the CD10 negative side in the 1st fraction on day 2 and day 3 attracted our attention (Fig.S3). It appears to be not the result of erratic behavior or experimental noise since it was observed in two consecutive days and also in repeated experiments, although the time point of appearance varied somewhat between 1 and 8 days. This transient appearance of a subpopulation in the opposite direction of the relaxation is also seen in the relaxation of CD54+ cells. One is tempted to interpret such behavior as manifestation of a small group of cells that 'spill over' to a neighboring attractor state representing a cell type not viable in the conditions of the main cell population.

1.1.3 A key regulatory gene can facilitate a switch between attractors.

Oct2 binds to the major enhancer (FR) for the viral C-promoter where transcription for the six viral nuclear EBNA proteins expressed in latency III (21). Oct2 shows higher expression in latency I (Rael) than in latency III cells(21). Based on these findings we employed a statistical mechanical model of the switch between latency I (type I) and III (type III), showing a possible key role of Oct 2 levels for this molecular switch(22). Transient transfection of shRNA to Oct 2 suppressed Oct 2 levels (Fig.S 4).

1.2 Supplementary Material and Methods

1.2.1 Cell lines and tissue culture.

Cell lines and tissue culture. Four human Epstein-Barr virus (EBV) carrying B-cell lines were used: Rael, Mutu I, CBM1-Ral-Sto and Mutu III. The cell lines were maintained as suspension cultures in RPMI 1640 medium (Sigma-Aldrich, St Louis, MO, USA) supplemented with 10% fetal bovine serum (FBS), L-glutamine, streptomycin and penicillin (Sigma-Aldrich, St Louis, MO, USA).

1.2.2 Antibodies and labeling of cells.

Antibody of anti-CD10 (ALB1) conjugated with PE, anti-CD19 (J3-119) conjugated with APC and (anti-CD23 (9P25) conjugated with FITC were from Beckman Coulter (Bromma, Sweden). Antibody anti-CD54 (HCD54) conjugated with Pacific-blue was from Biolegend (San Diego, CA, USA). Appropriate isotype control antibodies (Biolegend, San Diego,

CA) were used to establish background signal caused by non-specific antibody binding. BL (type I) cells are CD10 positive and CD54 negative, while the reverse is true for LCL (type III) cells. CD19 and CD23 are ubiquitous markers for most B-cells and are shared by all four cell lines, while CD10 and CD 54 are cell type specific. CD10 is a marker of germinal center B cells and some hematopoietic stem cells. LCLs express CD54, also known as ICAM-1 (Inter-Cellular Adhesion Molecule 1), a transmembrane protein belonging to the immunoglobulin superfamily, related to their property to form adhesion and cell conglomerates in cell. Cells for FCM (see below) were harvested 48h after feeding, and centrifuged. The pellets were washed twice with PBS and resuspended in PBS with 2% FBS. For sorting, 10 million cells/ml were labelled with antibodies and incubated at +4° C for 30 min, washed with 1ml PBS with 2 % FBS and centrifuged at 800 r/min for 5 min. The supernatant was removed and 200-400ul 2% FBS was added. Appropriate isotype control antibodies were used to establish background signal caused by non-specific antibody binding.

1.2.3 Flow cytometry (FCM), FACS including calibration.

Flow cytometry was performed with a Becton Dickinson FACSAria (FACS facility at the Swedish Center for Disease Control, KI), or with MoFlo XDP (Beckman Coulter, Bromma, Sweden) at the FACS facility at MTC, KI. Computational data analysis was done with FlowJo 2.2.2 (Ashland, OR, USA). For cell sorting, the input cell number ranged from 100×10^6 cells to 150×10^6 cells. Cells were sorted into ice-cold medium for a maximal duration of 3 h. Original gating was based on the isotype and unstained controls. Quantum PE MESF beads for CD10 and Quantum APC MESF beads for CD 19 (Bangs Laboratories, Fishers, IN, USA) were used to correct for the effect of day-to-day fluctuations of the flow cytometer following the manufacturer's instructions. The MESF kits were comprised of five populations of calibrated fluorescent standards: four populations of microspheres having different levels of PE or APC fluorescence intensity and one Certified Blank^{TM} population. After the cells were sorted out, they were cultured in proper microwell tissue culture plates (Costar, Corning B.V. Life Science, Amsterdam, The Netherlands) with the RPMI 1640 medium (Sigma-Aldrich, St Louis, MO, USA) and then transferred to tissue culture bottles when there were enough numbers of cells.

1.2.4 Apoptosis analysis.

APC Annexin V/ Dead Cell Apoptosis Kit with APC annexin V and STYOX Green for Flow Cytometry (Invitrogen, Fisher Scientific, Gothenburg, Sweden) was used for analysis of apoptosis according to the protocol provided by the manufacturer. In additional experiments we evaluated frequency of apoptotic and dead cells by gating those in the Flow Cytometry analysis.

1.2.5 CFSE staining.

The CellTraceTM CFSE Cell Proliferation Kit (Invitrogen, Fisher Scientific, Gothenburg, Sweden) was used for intracellular labeling of cells according to the protocol provided by the manufacture(37).

1.2.6 Gene expression profiling.

Gene expression profiling was performed using HuGene-1_0-st-v1 chips at the Bioinformatics and Expression Analysis core facility (BEA) at the Huddinge campus, Karolinska Institutet.

1.2.7 Microarray and data analysis.

After gene expression profiling (see Supplementary text on Material and Methods 1.2.5),

the GEDI maps for visual representation of global gene expression based on self-organizing maps were generated using the program GEDI (http://www.childrenshospital.org/research/ingber/GEDI/gedihome.htm).

1.2.8 Relative quantitative RT-PCR.

Comparative CT method was used. The experiments were performed using the ABI 7500 Sequence Detection System (Applied Biosystems, Foster City, CA). Power SYBR Green master mixture was from Life Technologies (Applied Biosystem, Foster City, CA USA). All reactions were conducted in triplicates with reaction volume 25 ul/tube.

Primers: CD54: Sense: CAGAGGTTGAACCCCACAGT.

Antisence:CCTCTGGCTTCGTCAGAATC;

CD10: Sense: CCTTCTTTAGTGCCCAGCAG.

Antisense: GCAGACTGTTGAGTCCACCA.

GAPDH: Sense: GAGTCAACGGATTTGGTCGT.

Antisense: AATTTGCCATGGGTGGAA.

1.2.9 Knock down of Oct 2 by shRNA.

Knock down of transcription factor Oct 2 was done with shRNA (anti-oct2 shRNA TI341098, Origene Technologies Inc. Rockville, MD, USA)). 2×10^6 Rael cells/tube were centrifuged at 1000 rpm $\times$ 5 min, the supernatant was discarded and 100 ul Mirus Ingenio^{TM} Electroporation solution (Mirus Bio LLC, Madison, WI, USA) was added with 2 ug GFP and 10 ul shRNA vector solution (100ng/ul). In addition, vector control (TR20003) and negative GFP control (TR 30003 shRNA pRS non-effective GFP plasmid) were used. After mixing, the mixture was transferred to an electroporation cuvette. After mixing in the cuvette, it was placed into Nucleofector I (Amaxa Biosystems, Cologne, Germany) with program D-26. After electroporation 1 ml pre-warmed culture medium was added and the cells were transferred to a 6-well tissue microplate (Becton Dickinson, Franklin Lakes, NJ, USA) and incubated at 37 °C.

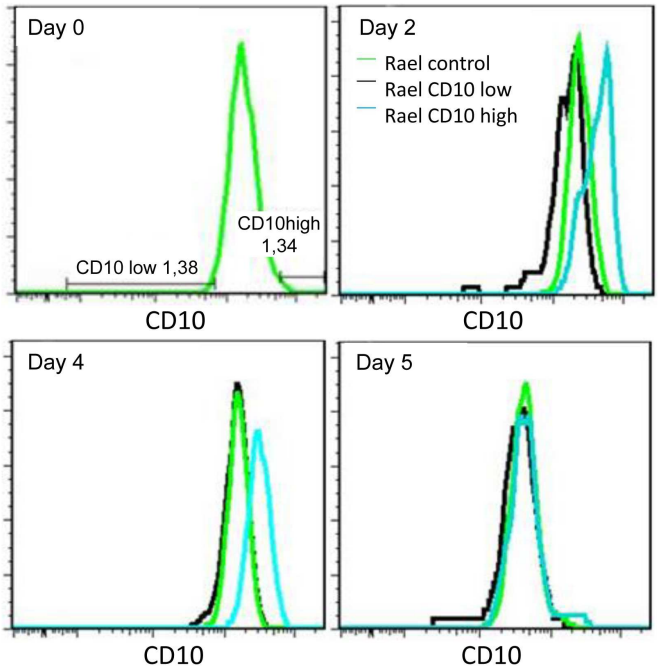

Figure S. 1: The re-establishment of the Rael CD10 parental distribution from the highest and lowest CD 10 expressing, selected Rael cells. The Rael cells with the highest and lowest CD10 expression were sorted out separately (day 0) and cultured. Their CD10 levels were re-analyzed on day 2,4 and 5 after isolation.

1.3 Supplementary Figure

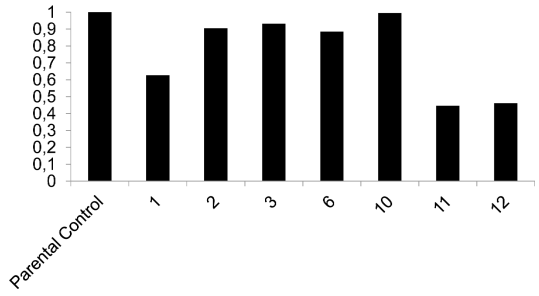

Figure S. 2: Analysis of viability in different subfraction of CD 10-selected Rael. The relative fold change of viability of the Rael cells sorted out by different CD10 levels. The viability of the Rael parental control, the three CD10highest fractions, the three CD10lowest fractions and one of the middle (the 6th fraction) according to figure 3A.

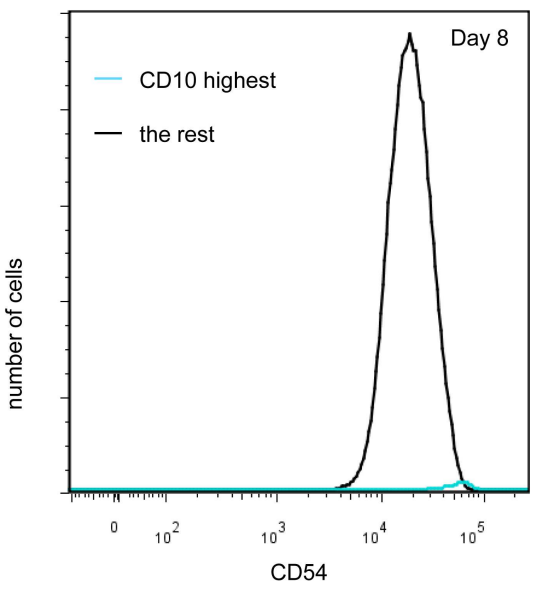

Figure S. 3: The original FCM data of figure 4C without adjustment of the Y-scale. CD10 level of the CD10highest Rael cells (CFSE+) and the rest of the Rael population (CFSE-, black) after CFSE staining and co-culturing for 8 days.

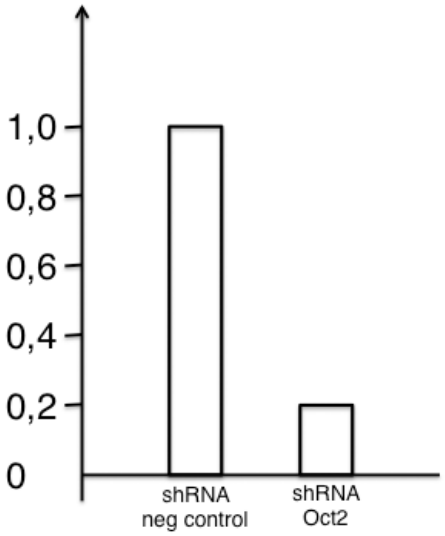

Figure S. 4: Transient transfection of shRNA to Oct 2 suppressed Oct 2 levels. Oct 2 mRNA expression 72 hours after transfection of Rael cells with control shRNA or anti-OCT2 shRNA by qPCR.

2 Cancer Cell Attractor Explored By In Silico Modelling With Fokker-Planck Equation

Abstract

This document contains supplementary material detailing how a postulated dynamics of the Fokker-Planck type can be reliably and robustly inferred from FACS data, even using only one marker.

2.1 Introduction

The purpose of this Supplementary Material is to discuss a mathematical approach to explore the cancer attractor hypothesis and its properties. The data, described in the main paper, consist of single-cell measurements of cell surface markers. Once measured, cells are removed from the system, and we cannot follow how individual cells change over time. We can however follow the distribution of a cell surface marker in the population, and how this distribution changes over time, which suggests a description in terms of Kolmogorov forward equations (Fokker-Planck equations) [1]. If we want to deduce properties of the dynamics of single cells moving within or between attractors we hence have to solve the inverse problem of deducing the dynamics (drift and diffusion terms in a Fokker-Planck equation) from snapshots of a distribution at discrete times, which is the topic addressed here.

2.2 Mathematical Background

2.2.1 The Cell-State Space

Phenotypic cell-to-cell variability within a population is a common feature of biological systems. In order to characterize the property of cell quantitatively, we can enumerate a list of variables which can be used to identify different cell-types under a desired resolution for each cell, such as the expression of N different proteins $X_1, X_2, ...X_N$ (and/or other quantities). A set of X_i represents a specific cell-state, which can be written as a vector $\vec{X} = (X_1, X_2, ...X_N)$, and each cell can then be mapped onto this abstract high-dimensional space to represent its phenotype. Technologies have successfully been developed to measure the average of variables over a population (e.g. proteomics, transcriptomics, other -omics technologies) but one is still far from being able to measure anything approaching a complete set of variables in single cells. The single cell phenotypic description as $\vec{X}$ is therefore an abstract quantity, which is not directly observable. Under the control of the gene-regulatory network, other biological mechanisms or external stimuli, some property of a cell may change over time, that is to say, each $X_i = X_i(t)$ is a function of time. The changes in cell state are then represented by $\vec{X}(t)$ as a point moving along some trajectory in state-space. In dynamical systems theory one says that if different trajectories converge to the same set of points that set is called an attractor, and the domain of initial conditions that eventually falls into the same attractor is called its basin of attraction [2, 3]. In the problem at hand we have two complications; that the dimension of the system is very large, and that due to the intrinsic stochasticity of biological systems, the dynamics of a single cell state $\vec{X}$ is not purely deterministic. We refer to the monograph [4] for general aspects of dynamical systems perturbed by noise and to [5] for considerations pertaining in very high dimension, and assume that the stochasticity can be described as Gaussian noise. The cell state is thus under the influence of deterministic changes imposed by the underlying regulatory network (drift towards the attractor) as well as random noise (diffusion, corresponding to gene expression noise). This type of system can be modeled by assigning functions to explicitly describe these two forces, as is done e.g.

with physical systems subject to thermal fluctuations. The change of state will therefore be described by an Itô stochastic differential equation (SDE):

$$d\vec{X}_t = \boldsymbol{\mu}(\vec{X}_t, t)dt + \boldsymbol{\sigma}(\vec{X}_t, t)d\boldsymbol{W}_t \tag{1}$$

In (1) $\boldsymbol{\mu}(\vec{X}_t, t)$ is a N-dimensional function (deterministic part), $\boldsymbol{W}_t$ is a standard M-dimensional Wiener process (random part), and $\boldsymbol{\sigma}(\vec{X}_t, t)$ is a N-by-M dimensional matrix, the elements of which are also functions of the state $\vec{X}$. The $\boldsymbol{W}_t$ describe noise sources and the number of independent noise sources (M) does not have to be the same as the number of observed variables (N). However, in the Itô stochastic differential equation we may without restriction take $M \leq N$, and for simplicity of the argument we will here assume that M equals N such that $\boldsymbol{\sigma}(\vec{X}_t, t)$ is a square matrix. The Itô prescription means that the matrix $\boldsymbol{\sigma}(\vec{X}_t, t)$ is evaluated *before* the increment is taken and here means that the production rates of any protein (and other state variable) are determined by the amounts of this and other proteins (or other state variables) present immediately before the actual production process begins.

In general one cannot say much about the solutions to an equation such as (1). However, one can argue on qualitative grounds that it may often be of interest to look at analogous equations of much lower dimensionality. The argument goes as follows: we assume that the deterministic part of (1), $\frac{d\vec{X}_t}{dt} = \boldsymbol{\mu}(\vec{X}_t, t)$, has solutions characterized by low dimensionality. The main example would be (attractive) fixed points $\vec{X}^*$ (which have dimension zero). We assume this is the case and write $\vec{X}_t = \vec{X}^* + \vec{Y}_t + \vec{Z}_t$ where $\vec{Y}_t$ are *slow modes* and $\vec{Z}_t$ are *fast modes*. To linear order we hence have $\frac{d\vec{Y}_t}{dt} = A_{YY}\vec{Y}_t + A_{YZ}\vec{Z}_t$ and $\frac{d\vec{Z}_t}{dt} = A_{ZY}\vec{Y}_t + A_{ZZ}\vec{Z}_t$ where A_{YY}, A_{YZ}, A_{ZY} and A_{ZZ} are suitable matrices (linear operators) which depend on $\vec{X}^*$. The separation into slow and fast modes means that the matrix A_{ZZ} is strongly contracting such that for every $\vec{Y}_t$ and every time t we will, to good accuracy, have $\vec{Z}_t \approx -A_{ZZ}^{-1}A_{ZY}\vec{Y}_t$. The Y directions are then *marginal* and the Z are *stable*. By techniques that are well developed in dynamical systems theory it is possible to extend such reasoning to *center manifolds* and *stable manifolds* where the slow motion takes place on a hyper-surface given by $\vec{Z}_t = g(\vec{Y}, \vec{X}^*)$, where g is some function. By a change of variables $\vec{Y} = \vec{Y'} + h_Y(\vec{Y'})$ and $\vec{Z} = \vec{Z'} + h_Z(\vec{Y'}, \vec{Z'})$ and using properly chosen transformations h_Y and h_Z the deterministic part of (1) can then be written as $\frac{d\vec{Y'}_t}{dt} = F(\vec{Y'}_t, t)$ and $\frac{d\vec{Z'}_t}{dt} = A_{ZZ}\vec{Z'}_t$. where A_{ZZ} is the same linear operator as above. The new function $F(Y')$ is called the *normal form* and is equal to $A_{YY}(\vec{Y'}) + N(\vec{Y'})$ where A_{YY} is the same as above and N contains nonlinear (resonant) terms determined by A_{YY}. The actual values of these resonant terms can be computed from $\boldsymbol{\mu}$ by an established procedure [2, 6].

We now consider the new center manifold coordinate $\vec{Y'}$ a function of the full state $\vec{X} = (\vec{Y}, \vec{Z})$ with its dynamics given by (1). By the Itô lemma we generally have

$$\begin{aligned} d\vec{Y'}_t &= \frac{\partial Y'}{\partial Y}dY + \frac{\partial Y'}{\partial Z}dZ + \frac{1}{2}\partial_Y\left(\frac{\partial Y'}{\partial Y}\right)(dY)^2 \\ &+\frac{1}{2}\partial_Z\left(\frac{\partial Y'}{\partial Y}\right)(dYdZ) + \frac{1}{2}\partial_Y\left(\frac{\partial Y'}{\partial Z}\right)(dZdY) \\ &+\frac{1}{2}\partial_Z\left(\frac{\partial Y'}{\partial Z}\right)(dZ)^2 \end{aligned} \tag{2}$$

It may be permissible to ignore the cross-term $(dZdY)$ as well as the squared fluctuations in the stable directions ($(dZ)^2$). The dynamics for Y' would then look like

$$d\vec{Y'}_t = \tilde{F}(Y', t)dt + \boldsymbol{\sigma}_Y(\vec{Y'_t}, t)d\boldsymbol{W}_t^Y \tag{3}$$

where $\tilde{F} = F + \frac{1}{2}\partial_Y\left(\frac{\partial Y'}{\partial Y}\right) < (dY)^2 >$ and $\boldsymbol{\sigma}_Y d\boldsymbol{W}_t^Y$ describes the net noise acting tangentially to the center manifold.

Equation (3) is of the same kind as (1), only of much lower dimensionality. The approach of this paper is not to try to derive (3) from (1), but to assume

that a form like (3) is valid for a low-dimensional (indeed one-dimensional) slow variable, and then see if we can reconstruct the dynamics, in this variable, from the data.

2.2.2 Fokker-Planck Equation

We now revert to the earlier notation and call the state variable $\vec{X}_t$. The Fokker-Planck Equation (FPE) describes the time evolution of probability density function f($\vec{X}_t$,t) for a system when individual trajectories follow (1) and is

$$\frac{\partial f(\vec{x},t)}{\partial t} = -\sum_{i=1}^{n} \frac{\partial}{\partial x_i}(\mu_i f) + \sum_{i=1}^{n}\sum_{j=1}^{n} \frac{\partial^2}{\partial x_i \partial x_j}(B_{ij} f) \tag{4}$$

where $B_{ij} = \frac{1}{2}(\boldsymbol{\sigma} \cdot \boldsymbol{\sigma}^{\boldsymbol{T}})_{ij}$ [1].

We will below consider the case where the variable $\vec{X}$ is one-dimensional and where $B_{ij}(\vec{x},t) = \delta_{ij} B_0$ is a space-time constant. The field $A_i(\vec{x},t)$ is then a gradient field *i.e* $A_i(\vec{x},t) = -\frac{\partial}{\partial x_i} V(\vec{x},t)$ and the FPE can be written in a simpler form:

$$\begin{aligned} \frac{\partial f(\vec{x},t)}{\partial t} &= \nabla \cdot (f(\vec{x},t)\nabla V(\vec{x},t)) + B_0 \nabla^2 f(\vec{x},t) \\ &= \nabla \cdot (f \nabla V + B_0 \nabla f) \end{aligned} \tag{5}$$

We note that the stationary states ($\frac{\partial f(\vec{x},t)}{\partial t} = 0$) of (5) are those distributions that satisfy $f\nabla V + B_0 \nabla f =$ Constant. If we consider flux-less solutions (no sources or sinks) the constant is zero which gives the familiar result $f \propto \exp(-\frac{V}{B_0})$.

2.3 Reconstruction tasks

In the experiments described in the main text we register/observe FCM-data at different time points, which can be normalized to give probability distributions $f(x,t)$. Data is now one-dimensional (one cell surface marker per cell) and the variable (normalized signal of the cell surface marker) is written x. The mathematical task is then to determine the functions $A(x)$ and $B(x)$ in a one-dimensional Fokker-Planck equation

$$\frac{\partial f(x,t)}{\partial t} = -\frac{\partial}{\partial x}(A(x) f(x,t)) + \frac{\partial^2}{\partial x^2}(B(x) f(x,t)) \tag{6}$$

from observations of $f(x,t)$. To illustrate possible extensions and limitations we will also be interested in models of more than one underlying state where the change in each state is described by (6) and we may also include different growth rates in different states and/or transitions between the states.

The number of different model classes one can choose is unlimited, but given finite data there is a trade-off between goodness of fit and the complexity of a model class. Different procedures have been developed to strike this balance such as *e.g.* the Akaike information criterion [7] and the Bayesian information criterion [8]. Our objective here is to illustrate that with models of different complexity different aspects of behavior can be captured by including more details in the description, and we will therefore describe results obtained with the following simple model classes:

1. A simple base-line one-attractor model "Model 1" given by

$$A(x) = -k(x - x^*) \quad B(x) = B \tag{7}$$

   This model has three parameters, k, x^* and B.

2. A two-attractor model "Model 2" given by two equations of the same type as (6) for two unobservable separate densities f_1 and f_2, and an output equation

$$\begin{aligned} f(x,t) \quad = \quad & (1 - \frac{a_1}{1 + t/a_2}) f_1(x,t) \\ & + \frac{a_1}{1 + t/a_2} f_2(x,t) \end{aligned} \tag{8}$$

   This model has eight parameters: k_1, x_1^* and B_1, k_2, x_2^* and B_2, and a_1 and a_2. This model describes the biologically plausible scenario of a

mixture of two underlying populations (which cannot be distinguished experimentally), with no transitions between them, and where the fraction of cells in the respective population change linearly over time. As any small change is approximately linear this is a proxy for more realistic but also more involved changes such as different growth rates.

3. A three-attractor model "Model 3" given by three equations of the same type as (6) with transitions between the states. This model is described in more detail below. In its full version it has 13 parameters, but we find it convenient to primarily consider special cases that have from three to eight parameters.

2.3.1 Numerical procedure

For each of the models the basic steps when searching for the appropriate parameters are

1. Initialize by choosing a set of reasonable parameters $\vec{p_0}$ as initial value
2. In step k with parameter value $\vec{p_k}$, numerically simulate $f_{p_k}(x,t)$ from original distribution: use the first distribution at t_0 as the true initial f, and use FPE with parameter $\vec{p_k}$ to calculate distributions at later time points
3. Compare $f_{p_k}(x,t)$ with experimental distribution $f_E(x,t)$, and use a functional $Q(f_{p_k}(x,t), f_E(x,t))$ to score their difference
4. Minimize Q($f_{p_k}(x,t)$,$f_E(x,t)$), use some strategy to find better parameters $\vec{p_{k+1}}$, which brings lower Q-value.
5. Repeat step 2 to step 4 until convergence
6. When a minimum Q is reached, obtain the best parameters p_*, such that $f_{p_*}(x,t)$ closest to $f_E(x,t)$

We solved the time development of $f_{p_k}(x,t)$ by a standard first-order discretization in time and second-order discretization in space, accuracy and stability checked by varying the resolution. As a reconstruction quality criterion Q we used mean square error. The minimum was found using the Numerical Recipies package 'amoeba.c', which performs a descent in the parameter landscape, without explicitly computing gradients [9].

2.4 Sample results

In this section we show experimental distributions (full lines) compared to numerical simulations (dotted lines), for each of the three models presented. The same color is used to plot the experimental and theoretical values for the same time point (day), in each figure. The full data, including Day 0 (which is partly off-scale) is shown in Fig.S5 for fraction of highest value of CD10 (CD10-high) and in Fig.S6 for the fraction of lowest value of CD10 (CD10-low). Day 0 is not only off-scale, the simplest "Model 1" also cannot fully reproduce the changes from Day 0 to Day 1. In the following data for Day 0 will mostly not be displayed, and for part of the modeling the initial data will be taken to be Day 1.

2.4.1 Model 1

In this model there are three parameters to be optimized, denoted $[B, k, x^*]$. For this example the stationary state *i.e.* the parental distribution is $f(x,\infty) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp(-\frac{(x-x^*)^2}{2\sigma^2})$ where $\sigma^2 = \frac{B}{k}$, which is a normal distribution with center x^* and standard deviation σ. We can use the ratio $\sigma = \sqrt{\frac{B}{k}}$ to check the width of curves, and as a benchmark to judge the quality of fitting. We use Day 1 data as initial condition, and fit the parameters to data in Day 2, Day 7, Day 16 and Day 28, respectively. We start from either the fraction of highest value of CD10 (CD10-high) or the fraction of lowest value of

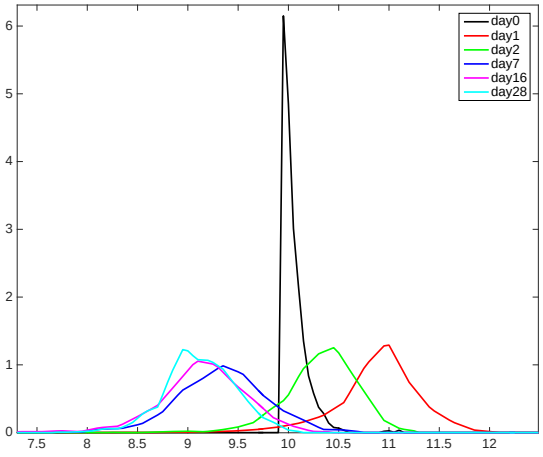

Figure S. 5: Development of the distribution of the fraction of highest value of CD10 (DC10-high). Black line (sharply peaked) represent the distribution on Day 0 where only cells with a FACS signal greater than about ten have been retained. This threshold is far out in the tail of the distribution (see main text), and the sharp peak in the distribution on Day 0 is a reflection of this tail. Observe that the distribution in the main moves out (moves to higher values) between Day 0 and Day 1 (red), and thereafter relaxes back to the parental distribution through Day 2 (green), Day 7 (navy blue), Day 16 (magenta) and Day 28 (turquoise).

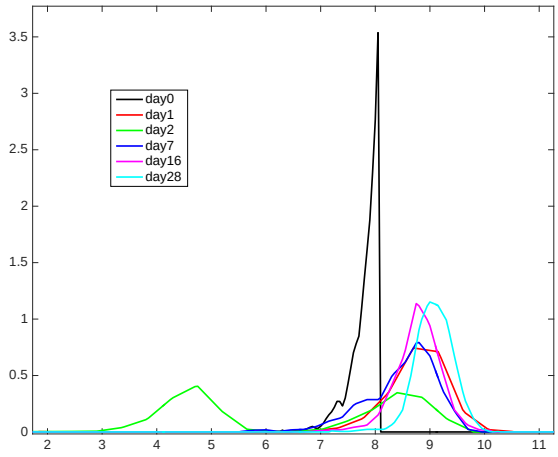

Figure S. 6: Development of the distribution of the fraction of lowest value of CD10 (DC10-low). Black line (sharply peaked) represent the distribution on Day 0 where only cells with a FACS signal less than about eight have been retained. Observe the double-peaked structure on Day 2 (green).

CD10 (CD10-low).

CD10-high Starting from Day 1 the distribution moves towards the center gradually, in qualitative agreement with the behavior predicted by the FPE. The optimized parameters were found to be $[B, k, x^*] = [0.0512, 0.367, 9.15]$ giving $\sigma_{high} = \sqrt{\frac{B}{k}} = \sqrt{\frac{0.0512}{0.3668}} = 0.373612$ which compares well with the parental distribution. Fig.S7 shows the theoretical curves in comparison with the experimental data and it is clear that the FPE also quantitatively captures the relaxation of the CD10-high fraction to the parental distribution starting from Day 1.

Between Day 0 and Day 1 (see main text and Fig.S5 above) other effects take place where the distribution temporarily moves away from the parental distribution (Day 0 data not shown in Fig.S7).

CD10-low This data behaves qualitatively quite differently from what one would expect from FPE. First the distribution quickly moves back close to the

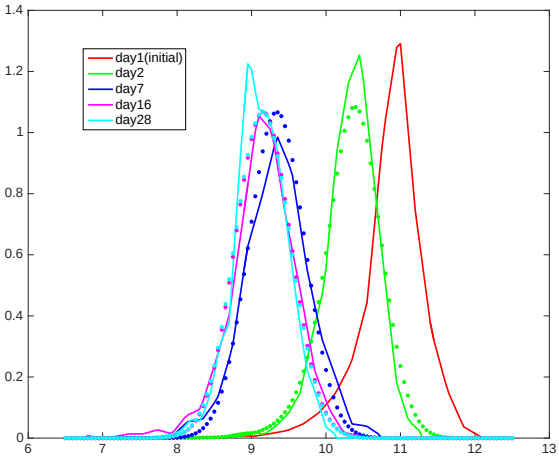

Figure S. 7: Results obtained by simulating the Fokker-Planck equation describing Model 1 and optimal parameters as explained in text. High CD10 fraction (CD10-high) on Day 1. used as initial conditions for the FPE. Note that CD10 value increases from Day 0 to Day 1 (Day 0 data is shown in Figure S. 5), a behavior which cannot be captured by Model 1, see main text and Fig.S5.

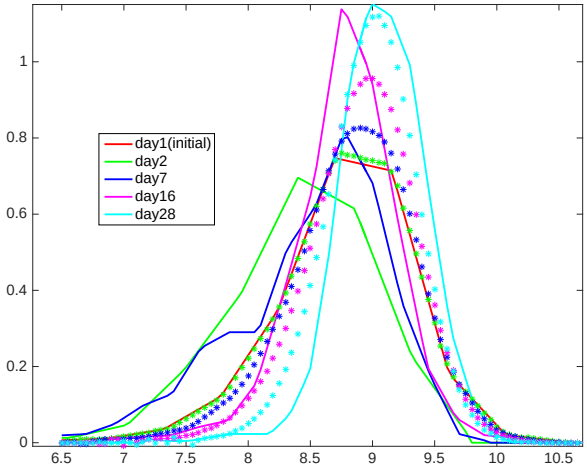

Figure S. 8: Results obtained by simulating the Fokker-Planck equation describing Model 1 and optimal parameters as explained in text. Low CD10 fraction (CD10-low) used as initial conditions for the FPE.

parental distribution (Day 1), then it moves in the opposite direction (Day 2), and then slowly returns again to the parental distribution (Days 7, 16 and 28). In Fig.S6 above we show that the distribution also develops a second peak on Day 2 (to the left of data displayed in Fig. 8) a behavior which cannot be captured by Model 1. Assuming nevertheless that Model 1 can be used on this data and proceeding as above, the optimized parameters are found to be $[B, k, x^*] = [0.0010, 0.0202, 9.25]$ indicating weak noise (low value of B) and a weak restoring force (low value of k). The asymptotic width would be $\sigma_{low} = \sqrt{\frac{B_0}{k}} = \sqrt{\frac{0.0010}{0.0202}} = 0.222497$ which is not a good fit to the parental distribution. In short, Model 1 does not describe this data well, since it cannot explain the 'outward motion' of the distribution from Day 0 to Day 1 seen in CD10-high nor the second peak on Day 2 seen in CD10-low.

2.4.2 Model 2

To explore if these effects could be explained in a simple way we hypothesized that what we observe is a mixture with qualitatively different behaviours, particularly at early times. We then assume that the populations we observe are composed of two underlying sub-populations, each one obeying independent and different FPE. We assume that we observe a mixture of the sub-populations in proportions which vary over time, which is a simple way to introduce growth rates of the two sub-populations into the model in a qualitative manner. Mathematically Model 2 is then defined by the system of equations

$$\begin{aligned}
&\frac{\partial f_1(x,t)}{\partial t} = -\frac{\partial}{\partial x}(A_1(x)f_1(x,t)) + \frac{\partial^2}{\partial x^2}(B_1(x)f_1(x,t)) \\
&A_1(x) = -k_1(x - x_1^*), \quad B_1(x) = B_1 \\
&\frac{\partial f_2(x,t)}{\partial t} = -\frac{\partial}{\partial x}(A_2(x)f_2(x,t)) + \frac{\partial^2}{\partial x^2}(B_2(x)f_2(x,t)) \\
&A_2(x) = -k_2(x - x_2^*), \quad B_2(x) = B_2 \\
&f(x,t) = (1 - \frac{a_1}{1 + t/a_2})f_1(x,t) + \frac{a_1}{1 + t/a_2}f_2(x,t)
\end{aligned} \tag{9}$$

where the observed population is given by $f(x,t)$. Model 2 has eight parameters: $[B_1, k_1, x_1^*, B_2, k_2, x_2^*, a_1, a_2]$.

CD10-high Starting from Day 0 we can now get a fit to the data with parameters $[0.0404, 0.212, 9.16, 0.0968, 0.0576, 25.0, 1.00, 2.94]$ representing one sub-population behaving similarly to the solution obtained in Model 1 ($B_1 \approx 0.04, k_1 \approx 0.4, x_1^* \approx 9$) and one sub-population centered at a larger value ($x_2^* \approx 25$). On Day 0 (data not displayed) Model 2 would give all the weight to the second sub-population. This does not mean that the optimal fit has a peak at $x_2^* \approx 25$ on Day 0, obviously not a good fit to the data in Fig.S5, but that all of the cells initially start out in the sub-population attracted to and moving towards $x_2^* \approx 25$. On Day 1 the second sub-population dominates approximately in ratio 3 : 1, and we can

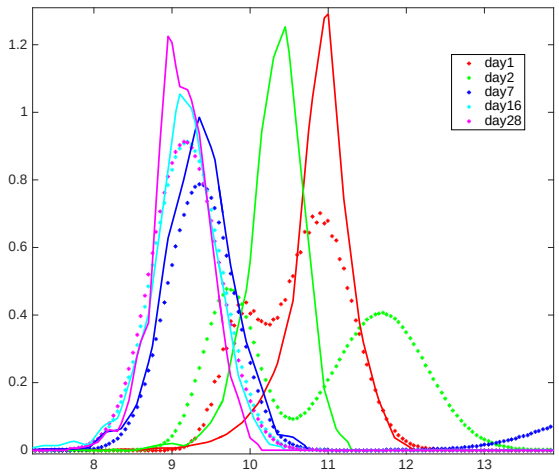

Figure S. 9: Results obtained by simulating the two Fokker-Planck equations describing Model 2 and the mixture using optimal parameters as explained in text. The high CD10 fraction (CD10-high) is used as initial data such that the two sub-populations are initially in proportion $a_1 : (1 - a_1)$.

therefore observe that the bulk of the population moves out from Day 0 to Day 1, but beyond Day 3 ($a_2 \approx 3$) the first sub-population takes over. The width of stationary distribution of center attractor is $\sigma_{c,high} = \sqrt{\frac{B_1}{k_1}} = 0.436848$ which again compares well with the parental distribution.

On the negative side, the optimal fit leads to a double-peak distribution on both Day 1 and Day 2, which obviously does not fit the data well. We therefore conclude that Model 2 is not an appropriate model for the CD10-high fraction.

CD10-low Starting from Day 1 we can get a fit to the data with parameters $[0.0595, 0.332, 8.95, 0.0596, 0.100, -9.52, 0.413, 28.0]$. which again represents one sub-population behaving similarly to the solution obtained in Model 1 ($B_1 \approx 0.06, k_1 \approx 0.3, x_1^* \approx 9$) and one sub-population centered at value away from the parental distribution

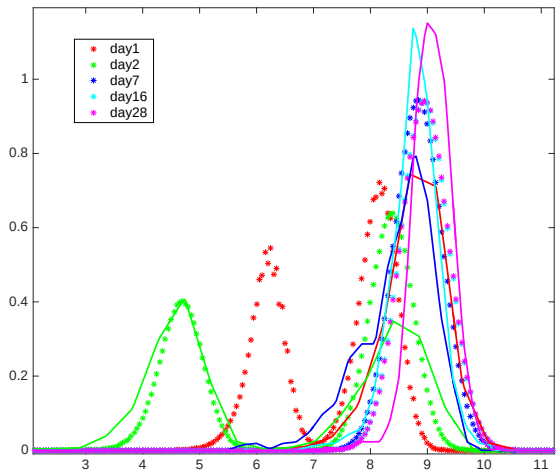

Figure S. 10: Results obtained by simulating the two Fokker-Planck equations describing Model 2 and the mixture using optimal parameters as explained in text. The low CD10 fraction (CD10-low) is used as initial data such that the two sub-populations are initially in proportion $a_1 : (1 - a_1)$. The second peak on Day 2 is well fitted by the model, at the price of introducing a spurious peak on Day 1.

($x_2^* \approx -10$). On Day 1 the two sub-populations are about evenly mixed (3 : 2) and only at around the final time of observation does the first start to take over ($a_2 \approx 28$). In a qualitative sense the dynamics of Model 2 is simple: initially one finds a mixture of both sub-populations, and over time the second sub-population moves away from the parental distribution towards its own attractor (this is true also for CD10-high). We are therefore able to reproduce the second peak on Day 2, at the price of also having a second peak at an intermediate position on Day 1 (not observed in the data).

The sub-populations of Model 2 deduced from the CD10-high and CD10-low sub-populations agree between themselves and with the parental distribution. For both sets of data Model 2 predicts a second sub-population, but on opposite positions with respect to the parental distribution. Model 2 provides a good description of the second peak in the distribution observed with CD10-low data on Day 2, but not for the change in CD10-high data.

The conclusion from Model 2 is hence that a conceptually simple model of two sub-populations developing independently without transitions can reproduce part of the data, but cannot consistently reproduce all parts of the data.

2.4.3 Model 3

In this model we assume there to be three underlying sub-populations in the observed population. We call these sub-populations respectively S_L(low), S_C (center) and S_H (high), and their three distributions f_L, f_c and f_H. Each of the three sub-population obeys FPE with different parameters $[B, k, x^*]$, and in addition S_L and S_H can turn into S_c. In other words, a cell in a given sub-population gravitates to its respective attractor by different rates for different attractors, but if is in S_L or S_H it may also switch to S_c and start to gravitate to that attractor. The main assumption, similar to Model 2, is that we can only observe how many cells there are in total from the sub-populations having the same value of the cell surface marker.

Mathematically Model 3 is then described as fol-

lows:

$$\begin{aligned}
\frac{\partial f_C(x,t)}{\partial t} =& -\frac{\partial}{\partial x}(A_C(x)f_C(x,t)) \\
&+\frac{\partial^2}{\partial x^2}(B_C(x)f_C(x,t)) \\
&+p_H * f_H(x,t) + p_L * f_L(x,t) \\
A_C(x) =& -k_C(x-x_C), \quad B_C(x) = B_C \\
\frac{\partial f_H(x,t)}{\partial t} =& -\frac{\partial}{\partial x}(A_H(x)f_H(x,t)) \\
&+\frac{\partial^2}{\partial x^2}(B_H(x)f_2(x,t)) - p_H * f_H(x,t) \\
A_H(x) =& -k_H(x-x_H), \quad B_H(x) = B_H \\
\frac{\partial f_L(x,t)}{\partial t} =& -\frac{\partial}{\partial x}(A_L(x)f_L(x,t)) \\
&+\frac{\partial^2}{\partial x^2}(B_L(x)f_L(x,t)) - p_L * f_L(x,t) \\
A_L(x) =& -k_L(x-x_L), \quad B_L(x) = B_L
\end{aligned} \tag{10}$$

The choice of having S_L, S_H turning into S_C, but not vice versa is partly for practical reasons, because it reduces the number of parameters, and partly because the "middle attractor" will reconstitute the parental distribution and hence should be more stable. Furthermore we have to introduce two parameters to model the initial conditions as a mixture of the three sub-population, *i.e.*:

$$\begin{aligned}
f_C(x,0) &= (1-a_H-a_L) * W_0(x) \\
f_H(x,0) &= a_H * W_0(x) \\
f_L(x,0) &= a_L * W_0(x)
\end{aligned}$$

Model 3 hence has in all 13 parameters: $B_C, k_C, x_C, B_H, k_H, x_H, B_L, k_L, x_L, p_H, p_L, a_H, a_L$, and further simplifications discussed below.

CD10-high Our main objective is to reproduce the shift from Day 0 to Day 1 without introducing additional qualitative errors, as this could not be captured in Model 1 or Model 2. We first observe that by optimizing parameters starting from the high DC10 fraction data (CD10-high) on Day 0 as initial conditions and using the distributions on Day 1, 2, Day 7, Day 16 and Day 28 for the fitting we find a_L to be almost zero and a_H to be almost one. We can therefore make the simplifying assumption that in fact at Day 0 all the cells are in the high-value attractor S_H meaning that a_L is set to zero and a_H is set to one. Model 3 is then (for the CD10-high fraction) simplified to a model describing only the interplay of S_H and S_C:

$$\begin{aligned}
\frac{\partial f_C(x,t)}{\partial t} =& -\frac{\partial}{\partial x}(A_C(x)f_C(x,t)) \\
&+\frac{\partial^2}{\partial x^2}(B_C(x)f_C(x,t)) + p_H * f_H(x,t) \\
A_C(x) =& -k_C(x-x_C), \quad B_C(x) = B_C \\
\frac{\partial f_H(x,t)}{\partial t} =& -\frac{\partial}{\partial x}(A_H(x)f_H(x,t)) \\
&+\frac{\partial^2}{\partial x^2}(B_H(x)f_2(x,t)) - p_H * f_H(x,t) \\
A_H(x) =& -k_H(x-x_H), \quad B_H(x) = B_H \\
f_C(x,0) =& 0 \\
f_H(x,0) =& W_0(x)
\end{aligned} \tag{11}$$

This reduced Model 3 has 7 parameters: $B_C, k_C, x_C^*, B_H, k_H, x_H^*, p_H = [0.0373, 0.288, 9.12, 0.136, 9.38, 11.1, 1.68]$ describing two attractors not too far apart ($x_c^* \approx 9.1, x_H^* \approx 11.1$) where both the noise and the rate of attraction are much larger in S_H ($B_H \approx 0.4, k_H \approx 9$ than in S_c ($B_c \approx 0.04, k_c \approx 0.3$). The variance of the distribution for large times is again $\sigma_{c,high} = \sqrt{\frac{B_C}{k_C}} = 0.359568$, which matches well the parental distribution.

Qualitatively, reduced Model 3 describes a situation where the high attractor (S_H) relaxes to its asymptotic shape on a time scale of hours ($1/k_H\sigma \approx 6h$) and then leaks over to the center attractor on the time scale of a day ($1/p_H \approx 15h$). After several days the fraction of cells left in S_H is therefore small. The central attractor (S_c) on the other hand relaxes on a much longer time scale ($1/k_c\sigma \approx 10$day) and so its shape will change relatively little while all the cells move over from S_H. It is therefore consistent that

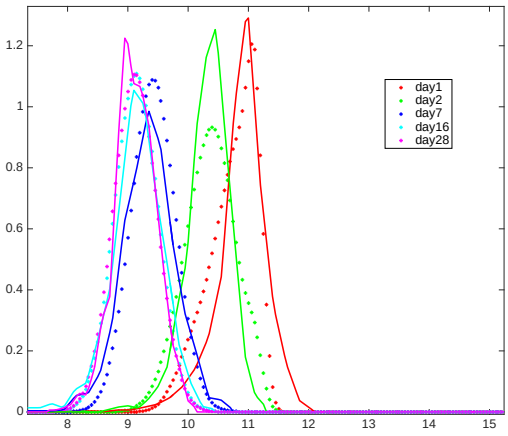

Figure S. 11: Results of reduced Model 3 for the CD-high fraction, initial data taken at Day 0 as in Fig.S5. Initially on Day 1 (red curve) the distribution moves out, after which it relaxes back to the parental distribution through Day 2 (green), Day 7 (navy blue), Day 16 (turquoise) and Day 28 (magenta). This model reproduces this data quite well both qualitatively and quantitatively.

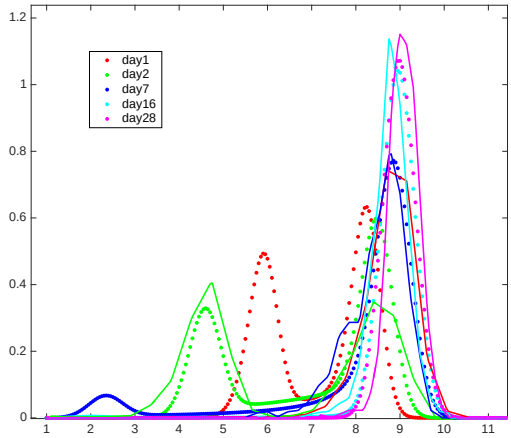

Figure S. 12: Results of second reduced Model 3 for the CD-low fraction, initial data taken at Day 0 as in Fig.S5. The second peak on Day 2 (green) is well reproduced, while at the same time a spurious second peak on Day 1 (red) is observed. In addition, a small spurious second peak is also observed on Day 7 (navy blue). The final relaxation to the parental distribution is well observed. This model reproduces this data quite about as well as Model 2, the major problem being the spurious peak on Day 1.

the parameters B_C, k_C, x_C^* are quite similar to the parameters of the simpler one-attractor Model 1 *i.e.* $[B, k, x^*] = [0.0512, 0.367, 9.15]$ as determined above.

CD10-low We proceed similarly as to the CD10-high fraction and assume that the process starts with only two sub-population S_C (center) and S_L; however now we cannot assume that their proportions are known, which brings us to 8 parameters $Y = [B_C, k_C, x_C^*, B_L, k_L, x_L^*, p_L, a_L] = [0.060, 0.437, 8.98, 0.060, 0.409, 2.00, 0.521, 0.306]$.
The predicted width of the parental distribution is $\sigma_{c,low} = \sqrt{\frac{B_C}{k_C}} = 0.370497$. On a qualitative level we find that reduced Model 3 can reproduce the second peak in Day 2, again at the price of introducing a similar but smaller second peak in the data on Day 1 (not present in the data). We find that the parameters describing S_C are similar to the CD10-high fraction, while the parameters obtained for S_L are different from those describing S_H. The major qualitative differences are that the predicted center of S_L is more different from the parental distribution $x_L^* \approx 2$ and the time scale of S_L is as slow as S_C ($k_L \sim k_C \sim 0.4\text{day}^{-1}$).

We can test the robustness of the model in different ways. For instance, we can fix the three parameters

$$[B_C, k_C, x_C] = [0.0373, 0.2885, 9.1177]$$

obtained from the CD10-high fraction and optimize the other parameters using the CD10-low fraction. This leads to a model with only five parameters

$$[B_{02}, k_2, x_{02}, a_1, p_2] = [0.060, 0.412, 2.00, 0.575, 0.327]$$

which fits the experimental data from Day 1 to Day 28 as well as the eight-parameter model, particularly

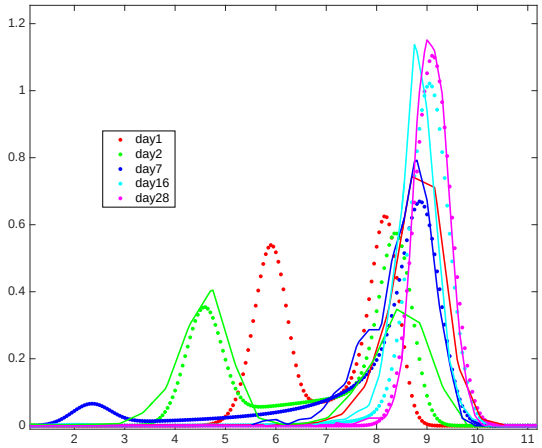

Figure S. 13: Results of second reduced Model 3 for the CD-low fraction, where part of the parameters are determined from the DC10-high fraction and the rest are found from initial data taken on Day 0. The results are practically indistinguishable from those of Fig.S12 showing that the properties of the center attractor are coherently recovered from the two data sets.

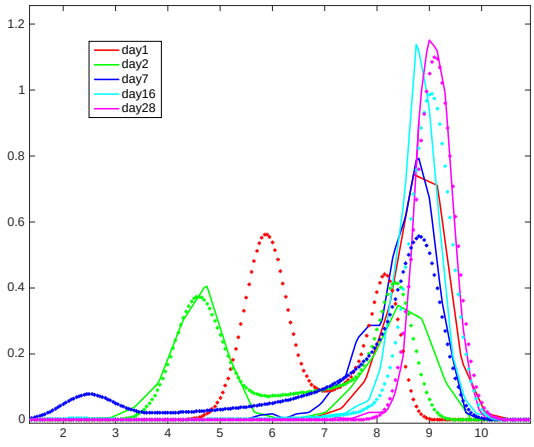

Figure S. 14: Results of second reduced Model 3 for the CD-low fraction, where data on Day 1 is not used. The spurious peak on Day 1 is then not penalized by the optimization and is more massive than in Figs.S12 and Fig.S13. The observed second peak on Day 2 is on the other hand reproduced more accurately.

for the return to the parental distribution on Day 28, see Fig.S13.

We can also disregard Day 1, to investigate the possibility that there is something wrong with the Day 1 data. The five-parameter model is then

$$[B_{02}, k_2, x_{02}, a_1, p_2] = [0.100, 0.435, 2.187, 0.714, 0.303]$$

which improves the fit for Day 2, which was to be expected, as shown in Fig.S14.

2.5 General Discussion

In this Supplementary Material we have shown that the underlying dynamics of a cell attractor can be reconstructed from snap-shots. Several variations of the basic model have been investigated to exemplify that the complex dynamics of outliers, representing cells that could be in one out of several attractors, can be captured more or less well by including different model details. The reconstructed dynamics of central attractor, which reproduces the parental distribution at long times, is well captured in all cases except the simplest model and one data set, demonstrating the robustness of the approach.

References

[1] N.G. Van Kampen. *Stochastic Processes in Physics and Chemistry*. Personal Library. North Holland; 3 edition, May 2007.

[2] John Guckenheimer and Philip Holmes. *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields*, volume 42 of *Applied Mathematical Sciences*. Springer, February 1998.

[3] Vladimir I. Arnold. *Geometrical Methods in the Theory of Ordinary Differential Equations*, volume 250 of *Grundlehren der mathematischen Wissenschaften*. Springer; 2nd ed, January 1997.

[4] Mark I. Freidlin and Alexander D. Wentzell. *Random Perturbations of Dynamical Systems*, volume 260 of *Grundlehren der mathematischen Wissenschaften*. Springer; 3rd ed, June 2012.

[5] Mariana Haragus and Gérard Iooss. *Local Bifurcations, Center Manifolds, and Normal Forms in Infinite-Dimensional Dynamical Systems*. Universitext. Springer; 2011 Edition, December 2010.

[6] C. Elphick, E. Tirapegui, M. E. Brachet, P. Coullet, and G. Iooss. A simple global characterization for normal forms of singular vector fields. *Phys. D*, 29(1-2):95–127, November 1987.

[7] K.P. Burnham and D.R. Anderson. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach (2nd ed.)*. Springer-Verlag, 2002.

[8] Gideon E. Schwarz. Estimating the dimension of a model. *Annals of Statistics*, 6(2):461–464, 1978.

[9] W. Press, B. Flannery, S. Teukolsky, and W. Vetterling. *Numerical Recipes*. Cambridge University Press, 1988.