Supplementary Material for:

A mathematical model of sub-population kinetics for the deconvolution of leukaemia heterogeneity

María Fuentes-Garí, Ruth Misener, David García-Münzer, Eirini Velliou,
Margaritis Kostoglou, Michael Georgiadis,
Efstratios N. Pistikopoulos, Athanasios Mantalaris

Contents

1 Approximation of cumulative distribution function

The cumulative distribution function is an integral function that can be numerically calculated using integrals; however, this can consume time during simulations, therefore for applications where extreme accuracies are not necessary (such as here, because the experimental inaccuracy would be greater anyway), the function is approximated using a mathematical formula that is very quickly computed by the simulator.

The cumulative distribution function, also called error function:

$$\Gamma_G\left(\mathrm{C}_E\right) = \frac{1}{\sigma_E\ \sqrt{2\ \pi}} \int_{\mathrm{C}_{E,\min}}^{\mathrm{C}_E} e^{-\left(\mathrm{C}_E - \mathrm{C}_{E,\mathrm{thr}}\right)^2 / \left(2 \cdot \sigma_E^2\right)} d\mathrm{C}_E \tag{1}$$

$$\Gamma_M\left(\mathrm{C}_B\right) = \frac{1}{\sigma_B\ \sqrt{2\ \pi}} \int_{\mathrm{C}_{B,\min}}^{\mathrm{C}_B} e^{-\left(\mathrm{C}_B - \mathrm{C}_{B,\mathrm{thr}}\right)^2 / \left(2 \cdot \sigma_B^2\right)} d\mathrm{C}_B \tag{2}$$

is approximated using Press' method [1], with x being a normally distributed variable with a mean of 0 and a standard deviation equal to 1:

$$\mathrm{erf}(x) = \begin{cases} 1 - \tau & \text{for} \quad x \geqslant 0 \\ \tau - 1 & \text{for} \quad x < 0 \end{cases} \tag{3}$$

where:

$$\begin{aligned} \tau = t \cdot \exp(&-x^2 - 1.26551223 + 1.00002368 \cdot t + 0.37409196 \cdot t^2 \\ &+ 0.09678418 \cdot t^3 - 0.18628806 \cdot t^4 + 0.27886807 \cdot t^5 \\ &- 1.13520398 \cdot t^6 + 1.48851587 \cdot t^7 - 0.82215223 \cdot t^8 + 0.17087277 \cdot t^9) \end{aligned} \tag{4}$$

and

$$t = \frac{1}{1 + 0.5 \cdot |x|} \tag{5}$$

$\mathrm{C}_{B,b}$ and $\mathrm{C}_{E,e}$ are converted to standard normally distributed variables as follows:

$$x_{E,e} = \frac{\mathrm{C}_{E,e} - \mathrm{C}_{E,\mathrm{thr}}}{\sigma_E} \qquad \forall \quad e \in \{0, ..., n_E\} \tag{6}$$

[1] Press WH, Flannery BP, Teukolsky SA, Vetterling WT (1992), *Numerical Recipes in Fortran 77: The Art of Scientific Computing* (Cambridge University Press) pp. 214.

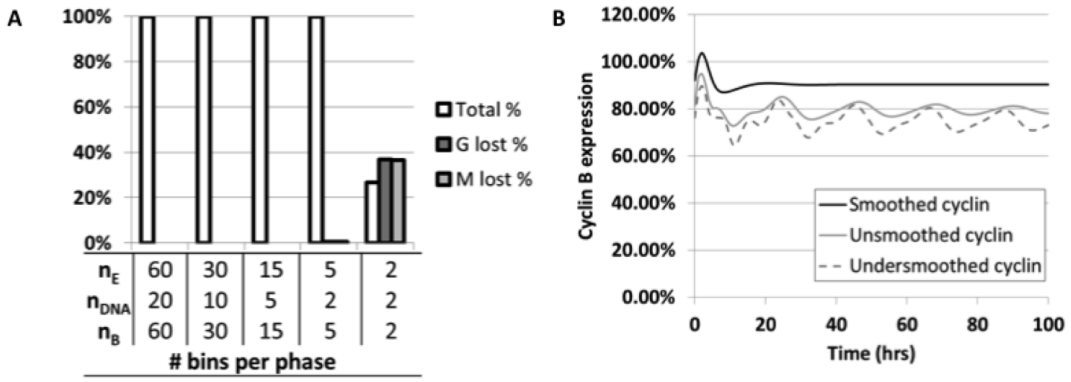

Figure S1: Analysing the number of discretisation intervals needed by the PBM is useful to define the most computationally efficient conditions satisfying the required resolution in the variables. **(A)** Evolution of cell loss with decreasing number of intervals in each phase (Total %: percentage of total cells remaining after 1000h, G lost %: percentage of total cells that were lost on the last bin of G0/G1 (cumulative), M lost %: percentage of total cells that were lost on the last bin of G2/M (cumulative)). **(B)** Changes in Cyclin B expression patterns with increase in bin numbers (Smoothed: nB=5; Unsmoothed: nB=60; Undersmoothed: nB=200). Examples for K-562 under even phase distribution conditions

and

$$x_{B,b} = \frac{C_{B,b} - C_{B,\text{thr}}}{\sigma_B} \qquad \forall \quad b \in \{0, ..., n_B\} \tag{7}$$

$x_{E,e}$ or $x_{B,b}$ can be replaced in equations 4 and 5 and the resulting $\text{erf}(x_{E,e})$ or $\text{erf}(x_{B,b})$ can be calculated.

2 Derivation of the even phase distribution equations

Even phase distributions occur when cells are asynchronously moving from phase to phase, which over time leads to a steady state in the phase distributions. A steady state is defined by the absence of change in a variable over time. Mathematically, it is achieved by setting the derivative (change) of the particular variable to zero and calculating its constant value. Below are the calculations for the PBM case.

Let us assume the model reaches steady-state (a situation found when the cell cycle distri-

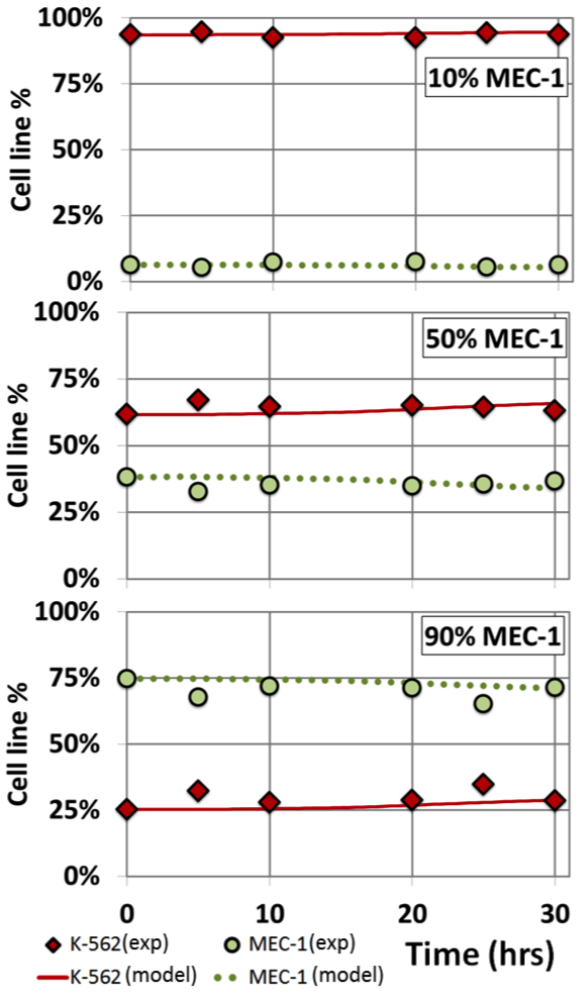

Figure S2: K-562, MEC-1 cells in each co-culture experiment (circles and diamonds: experimental results; lines: model output). Results are shown after adjustment of kinetic parameters in co-culture.

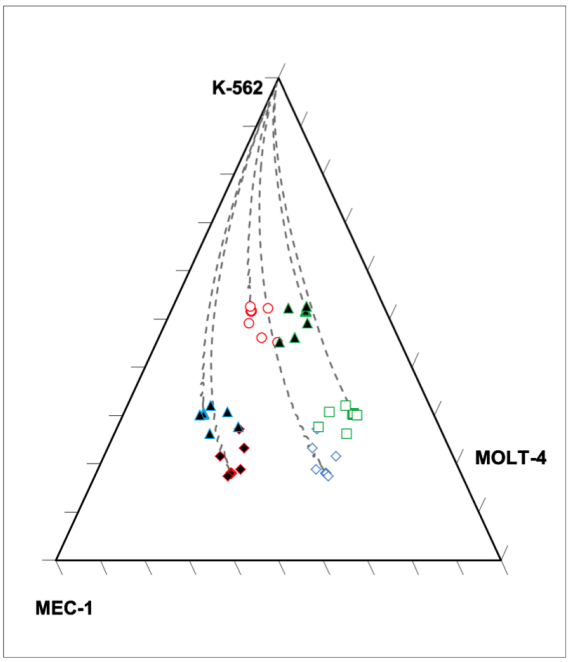

Figure S3: Representing the co-culture contents in a ternary plot allows the use of one (the Euclidean distance) vs two (two of the cell line percentages) goodness of fit indicators. In this case, we apply this strategy to define which of the six scenarios is closest to model predictions: e
xample of a model calculation for all 6 scenarios combining 10%/30%/60% mixtures (co-culture T6). Points denote experimental values arranged for that particular scenario, while dashed lines represent the simulation profile starting from each of the initial 6 scenario.

bution is constant). It follows that:

$$\frac{dG_{e,\%}}{dt} = -r_G \cdot G_{e,\%} + r_G \cdot G_{e-1,\%} - r_{G\to S,e} \cdot G_{e,\%} \qquad = 0 \quad \forall \quad e \in \{2,\ldots,n_E\} \tag{8}$$

$$\frac{dS_{d,\%}}{dt} = -r_S \cdot S_{d,\%} + r_S \cdot S_{d-1,\%} \qquad = 0 \quad \forall \quad d \in \{2,\ldots,n_D\} \tag{9}$$

$$\frac{dM_{b,\%}}{dt} = -r_M \cdot M_{b,\%} + r_M \cdot G_{b-1,\%} - r_{M\to G,b} \cdot M_{b,\%} \qquad = 0 \quad \forall \quad b \in \{2,\ldots,n_B\} \tag{10}$$

Reorganizing:

$$G_{e,\%} = \frac{r_G \cdot G_{e-1,\%}}{r_{G\to S,e} + r_G} = (1-\Gamma_{G,e}) \cdot G_{e-1,\%} \qquad \forall \quad e \in \{2,...,n_E\} \tag{11}$$

$$S_{d,\%} = S_{d-1,\%} \qquad \forall \quad d \in \{2,...,n_D\} \tag{12}$$

$$M_{b,\%} = \frac{r_M \cdot M_{b-1,\%}}{r_{M\to G,b} + r_M} = (1-\Gamma_{M,b}) \cdot M_{b-1,\%} \qquad \forall \quad b \in \{2,...,n_B\} \tag{13}$$

Summing across all populations:

$$\sum_{e'=1}^{n_E} G_{e',\%} = \left(\sum_{e'=2}^{n_E} \prod_{e''=2}^{e'} (1-\Gamma_{G,e''}) \right) \cdot G_{1,\%} + G_{1,\%} \tag{14}$$

$$\sum_{d'=1}^{n_D} S_{d'} = n_D \cdot S_{1,\%} \tag{15}$$

$$\sum_{b'=1}^{n_B} M_{b',\%} = \left(\sum_{b'=2}^{n_B} \prod_{b''=2}^{b'} (1-\Gamma_{M,b''}) \right) \cdot M_{1,\%} + M_{1,\%} \tag{16}$$

Since $\Gamma_{G,1} = 0$ and $\Gamma_{M,1} = 0$:

$$\sum_{e'=1}^{n_E} G_{e',\%} = \left(\sum_{e'=1}^{n_E} \prod_{e''=1}^{e'} (1-\Gamma_{G,e''}) \right) \cdot G_{1,\%} \tag{17}$$

$$\sum_{b'=1}^{n_B} M_{b',\%} = \left(\sum_{b'=1}^{n_B} \prod_{b''=1}^{b'} (1-\Gamma_{M,b''}) \right) \cdot M_{1,\%} \tag{18}$$

Table S1: Real cell line contents of blind tests

	K-562	MEC-1	MOLT-4
T1	20%	40%	40%
T2	17%	17%	66%
T3	50%	0%	50%
T4	0%	20%	80%
T5	10%	90%	0%
T6	10%	30%	60%
T7	80%	10%	10%
T8	60%	40%	0%
T9	20%	50%	30%

Reorganizing:

$$\frac{G_{1,\%}}{\sum_{e'=1}^{n_E} G_{e',\%}} = \frac{1}{\left(\sum_{e'=1}^{n_E} \prod_{e''=1}^{e'} (1-\Gamma_{G,e''})\right)} \tag{19}$$

$$\frac{S_{1,\%}}{\sum_{d'=1}^{n_D} S_{d'}} = \frac{1}{n_D} \tag{20}$$

$$\frac{M_{1,\%}}{\sum_{b'=1}^{n_B} G_{b',\%}} = \frac{1}{\left(\sum_{b'=1}^{n_B} \prod_{b''=1}^{b'} (1-\Gamma_{M,b''})\right)} \tag{21}$$

Using equations 11– 13 we finally obtain the percentage of cells in each bin among all cells in the phase:

$$\frac{G_e}{\sum_{e'=1}^{n_E} G_{e'}} = \frac{\prod_{e'=1}^{e} (1-\Gamma_{G,e'})}{\sum_{e'=1}^{n_E} \prod_{e''=1}^{e'} (1-\Gamma_{G,e''})} \qquad \forall e \in \{1,\ldots,n_E\} \tag{22}$$

$$\frac{S_d}{\sum_{d'=1}^{n_D} S_{d'}} = \frac{1}{n_D} \qquad \forall d \in \{1,\ldots,n_D\} \tag{23}$$

$$\frac{M_b}{\sum_{b'=1}^{n_B} M_{b'}} = \frac{\prod_{b'=1}^{b} (1-\Gamma_{M,b'})}{\sum_{b'=1}^{n_B} \prod_{b''=1}^{b'} (1-\Gamma_{M,b''})} \qquad \forall b \in \{1,\ldots,n_B\} \tag{24}$$

3 Cell cycle oscillations and cyclin concentration

Experimentally, the fluorescence of thousands of single cells are recorded using flow-cytometry. This data has to be gathered in a significant manner by aggregating the fluorescences of all cells and applying statistical analyses to derive significant information. As a downside, if an event is very fast in time, it is very unlikely that many cells will be found at that stage, and as a result, the behaviour during that short period of time will be lost in the global behaviour of the rest of the population. In the next paragraph, we explain how this occurs with cyclin expression and the impact it has on the experimental results (vs the model output).

Oscillations in cyclin concentration appear when the period of oscillation (the phase time) is larger than both the sampling frequency and the dispersion in the phase. In the case of cyclin E, since G1 phase represents a substantial portion of the total cell cycle time, oscillations were more evident. However, due to a generally shorter G2/M phase, the oscillations for cyclin B can only be seen in the beginning, when the EdU^- population was more narrowly distributed. For instance, a cell line with a faster G2/M phase, such as MOLT-4, will only reveal the cyclin B peak at the very beginning, while K-562, which has much slower G2/M kinetics, displays the whole cyclin B trajectory. Cyclin concentration, both *in vitro* and *in silico*, is the result of population-wide averaging (geometric mean). If two different populations appear in the same phase, an averaged cyclin concentration level is reached, which might not be representative of any of the two. This was the case for cyclin B in K-562 cells (Figure 4 in the manuscript)

4 Table of variables

A table of variables is required to define all the terms used in the equations, their significance and the corresponding units.

Table S2: Definitions

	Symbol	Description	Units	Source
Cell Lines	K562	Chronic Myeloid Leukemia; ATCC Cat. #CCL243		
	MEC1	B-Cell Chronic Lymphocytic Leukemia; DSMZ Cat. #ACC495		
	MOLT4	Acute Lymphocytic Leukemia; ATCC Cat. #CRL1582		
Cell Cycle Phases	G	Aggregated phase including G_0 (resting) & G_1 (growth preparing for DNA synthesis) phases		
	S	Synthesis phase; cells replicate their DNA		
	M	Aggregated phase incorporating G_2 (growth preparing for mitosis) & M (mitosis) phases		
State Variables	C_B	Cyclin B; Increased expression during G2 phase	normalized cyclin	
	C_E	Cyclin E; Increased expression during $G_1 \rightarrow$ S transition	normalized cyclin	
	DNA	Deoxyribonucleic acid; Doubles during S phase	[copies]	
	t	Time	[hr]	
Measurable Model Parameters	$C_{E,\min}$	Minimum value of cyclin E		
	$C_{B,\min}$	Minimum value of cyclin B		
	$C_{E,\text{thr}}$	Transition threshold of cyclin E		
	$C_{B,\text{thr}}$	Transition threshold of cyclin B		
	$C_{E,\max}$	Maximum value of cyclin E		
	$C_{B,\max}$	Maximum value of cyclin B		
	G_0	Initial cellular fraction in G	[–]	
	S_0	Initial cellular fraction in S	[–]	
	M_0	Initial cellular fraction in M	[–]	
	$S_{0,1}$	Fraction of S phase cells in first quarter of S phase	[–]	
	$S_{0,2}$	Fraction of S phase cells in second quarter of S phase	[–]	
	$S_{0,3}$	Fraction of S phase cells in third quarter of S phase	[–]	
	$S_{0,4}$	Fraction of S phase cells in last quarter of S phase	[–]	
	σ	Standard deviation of cyclin distribution		
	T_G	Cell cycle time in G phase	[hr]	
	T_S	Cell cycle time in S phase	[hr]	
	T_M	Cell cycle time in M phase	[hr]	
Derived Model Parameters	r_G	dC_E/dt; Increasing rate of C_E expression during G phase	[normalized cyclin/ hr]	Eq 4
	r_S	$d\text{DNA}/dt$; Increasing rate of DNA during S phase	[copies / hr]	Eq 5
	r_M	dC_B/dt; Increasing rate of C_B expression during M phase	[normalized cyclin/ hr]	Eq 6
	$r_{G\rightarrow S}(C_E)$	Transition rate from G to S phase	[normalized cyclin/ hr]	Eq 1
	$r_{M\rightarrow G}(C_B)$	Transition rate from M to G phase	[normalized cyclin/ hr]	Eq 3
	$\Gamma_G(C_E)$	Probability distribution for transition according to cyclin levels in G	[–]	Eq 7
	$\Gamma_M(C_B)$	Probability distribution for transition according to cyclin levels in M	[–]	Eq 8
Discretised Model Parameters	n_E	Number of bins in the G phase		
	n_D	Number of bins in the S phase		
	n_B	Number of bins in the M phase		
	G_e	Cells in the e^{th} bin of G; $e \in \{1,\ldots,n_E\}$		
	S_d	Cells in the d^{th} bin of S; $d \in \{1,\ldots,n_D\}$		
	M_b	Cells in the b^{th} bin of M; $b \in \{1,\ldots,n_B\}$		
	v_E	Cyclin E to bin width conversion factor	[–]	Eq 12
	v_D	DNA to bin width conversion factor	[–]	Eq 13
	v_B	Cyclin B to bin width conversion factor	[–]	Eq 14
	$r_{G\rightarrow S,e}$	Transition rate from G to S phase	[normalized cyclin/ hr]	Eq 18
	$r_{M\rightarrow G,b}$	Transition rate from M to G phase	[normalized cyclin/ hr]	Eq 20
	$\Gamma_{G,e}$	Probability distribution for transition according to cyclin levels in G	[–]	
	$\Gamma_{M,b}$	Probability distribution for transition according to cyclin levels in M	[–]	

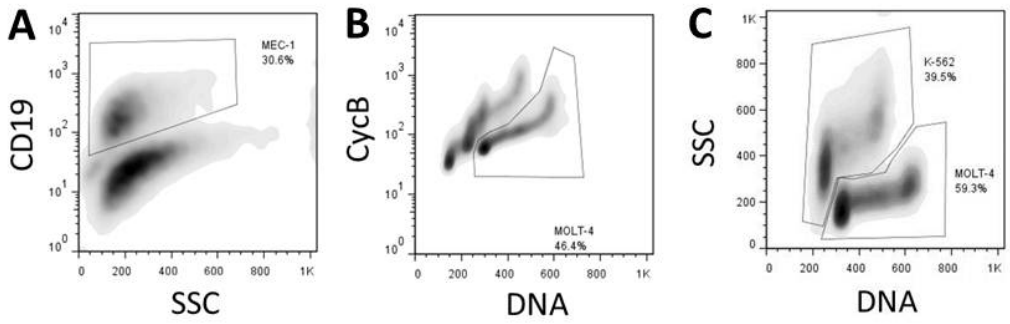

Figure S4: Flow cytometry plots of co-cultures. All axes are expressed in fluorescence intensity. SSC accounts for cellular complexity; DNA appears as a more pronounced group on the left (G1 phase cells), a continuous shade (S phase cells) leading to a second, smaller group on the right (G2/M phase cells); CD19 is a MEC-1 specific marker. The plots shown above are the ones used to determine the percentages of each cell line in co-culture, according to the different patterns in CD19, SSC, CycB and DNA displayed by each of them. Regions where a particular cell line is identified are shown along with the corresponding percentage. **(A)** FITC CD19 vs SSC plot of a K-562, MEC-1 and MOLT-4 mixture - CD19 positive region identified based on a separate MEC-1 single culture and then applied to the co-culture plot to obtain the MEC-1 percents; **(B)** FITC Cyclin B vs DNA plot of a K-562, MEC-1 and MOLT-4 mixture - MOLT-4 region identified based on a separate Cyclin B vs DNA plot of MOLT-4 single culture; **(C)** DNA vs SSC plot of a K-562 and MOLT-4 mixture - K-562 and MOLT-4 regions defined based on single cultures.

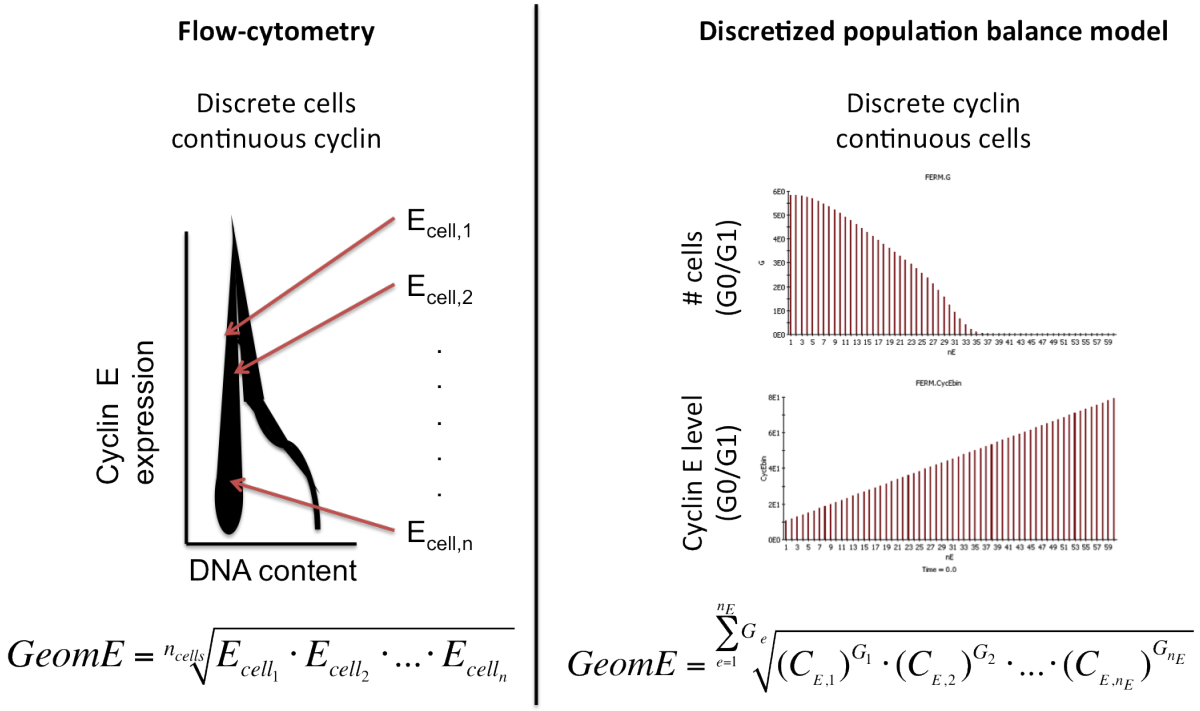

Figure S5: Comparison of calculation paths for processing flow-cytometry and mathematical data to obtain the same geometric mean of cyclin parameter (example case for G0/G1 phase and cyclin E). In flow-cytometry, the individual fluorescences of cells are multiplied and the nth root of the product is calculated (n being the total number of cells). In the PBM, the cyclin values in each bin (C_e) to the power of the number of cells in that bin (G_e) are multiplied, and then the nthroot of the whole is calculated (n being the sum of cells in all bins).

5 Table of parameters

A table of parameters is required to define which entities are given experimentally and what values they take.

Table S3: Measured Parameter Values

	K-562	MEC-1	MOLT-4
$C_{E,\min}$ (%)	10	10	0
$C_{B,\min}$ (%)	22	20	60
$C_{E,\mathrm{thr}}$ (%)	50	30	80
$C_{B,\mathrm{thr}}$ (%)	180	135	125
$G_{EdU^-,0,\%}$ (%)	69	74	66
$S_{EdU^-,0,\%}$ (%)	9	14	31
$M_{EdU^-,0,\%}$ (%)	22	12	3
$S_{0,1}$ (% over S)	20	42	67
$S_{0,2}$ (% over S)	44	4.8	10
$S_{0,3}$ (% over S)	32	4.2	8
$S_{0,4}$ (% over S)	4	35.9	14
T_G (h)	9	8.5	12
T_S (h)	10	7.5	15
T_M (h)	5	3	2

6 Summary of PBM variables and parameters

The following measured and derived parameters are required to run the discretised PBM:

- $n_E \times \mathrm{C}_E$ levels, $n_B \times \mathrm{C}_B$ levels and $n_D \times$ DNA levels (derived)
- $n_E \times \Gamma_{\mathrm{C}_{E,e}}$ and $n_B \times \Gamma_{\mathrm{C}_{B,b}}$ transition probabilities (derived)
- $n_E \times r_{G\to S,e}$ and $n_B \times r_{M\to G,b}$ transition rate coefficients (derived)
- growth rates r_G, r_S and r_M (derived)
- discretisation factors ν_E, ν_D and ν_B (derived)
- initial conditions $G_{EdU^-,0,\%}$, $S_{EdU^-,0,\%}$ and $M_{EdU^-,0,\%}$ (measured)
- cyclin minima $\mathrm{C}_{E,\mathrm{min}}$ and $\mathrm{C}_{B,\mathrm{min}}$ (measured)
- cyclin thresholds $\mathrm{C}_{E,\mathrm{thr}}$ and $\mathrm{C}_{B,\mathrm{thr}}$ (measured)
- cyclin maxima $\mathrm{C}_{E,\mathrm{max}}$ and $\mathrm{C}_{B,\mathrm{max}}$ (derived)
- cell cycle phase durations T_G, T_S and T_M (measured)
- intra S phase distributions $\mathrm{S}_{0,1}$, $\mathrm{S}_{0,2}$ and $\mathrm{S}_{0,3}$ (measured)
- transition variabilities σ_E and σ_B (theoretically measured, but not significant according to GSA)

The following variables are calculated when running the discretized PBM:

- $n_E \times G_e$, $n_B \times M_b$, $n_D \times S_d$ and T cell numbers
- $n_E \times G_{e,\%}$, $n_B \times M_{b,\%}$ and $n_D \times S_{d,\%}$ phase percentages
- geometric mean of cyclin E and B: Geom_E and Geom_B

Since we have set $n_E = 60$, $n_D = 20$ and $n_B = 60$, the total number of measured and derived parameters is 400 and the total number of variables is 283.

7 Co-culture assumptions

Table S4: Co-culture parameter values

	Co-culture #	K-562	MEC-1	MOLT-4
$C_{E,\min}$ (%)		10	10	0
$C_{B,\min}$ (%)		22	20	60
$C_{E,\text{thr}}$ (%)		50	30	80
$C_{B,\text{thr}}$ (%)		180	135	125
T_G (h)		9	12.1	12
T_S (h)		10	10.7	15
T_M (h)		5	4.3	2
Measured cell line %	T1	18	30	52
	T2	10	17	73
	T3	40	0	60
	T4	0	12	88
	T5	31	69	0
	T6	13	17	70
	T7	81	6	13
	T8	64	36	0
	T9	15	38	47