A mathematical model of subpopulation kinetics for the deconvolution of leukaemia heterogeneity

María Fuentes-Garí[1], Ruth Misener[2], David García-Munzer[1], Eirini Velliou[3], Michael C. Georgiadis[4], Margaritis Kostoglou[4], Efstratios N. Pistikopoulos[5], Nicki Panoskaltsis[6], Athanasios Mantalaris[1]*

1. Biological Systems Engineering Laboratory, Centre for Process Systems Engineering, Department of Chemical Engineering, Imperial College London, South Kensington Campus, SW7 2AZ, London, UK.
2. Department of Computing, Imperial College London, South Kensington Campus, SW7 2AZ, London, UK.
3. Department of Chemical and Process Engineering, University of Surrey, GU2 7XH, Guildford, Surrey, UK.
4. Aristotle University of Thessaloniki, Thessaloniki, Greece.
5. Artie McFerrin Department of Chemical Engineering, Texas A&M, College Station TX, USA
6. Department of Hematology, Imperial College London, Northwick Park & St. Mark's Campus, HA1 3UJ, Harrow, Middlesex, UK

*Corresponding author: a.mantalaris@imperial.ac.uk

Abstract: Acute Myeloid Leukaemia is characterized by marked inter- and intra-patient heterogeneity, the identification of which is critical for the design of personalized treatments. Heterogeneity of leukaemic cells is determined by mutations which ultimately affect the cell cycle. We have developed and validated a biologically-relevant, mathematical model of the cell cycle based on unique cell cycle signatures, defined by duration of cell cycle phases and cyclin profiles as determined by flow cytometry, for 3 leukaemia cell lines. The model was discretised for the different phases in their respective progress variables (cyclins and DNA), resulting in a set of time-dependent ordinary differential equations. Cell cycle phase distribution and cyclin concentration profiles were validated against population chase experiments. Heterogeneity was simulated in culture by combining the 3 cell lines in a blinded experimental set-up. Based on individual kinetics, the model was capable of identifying and quantifying cellular heterogeneity. When supplying the initial conditions only, the model predicted future cell population dynamics and estimated the previous heterogeneous composition of cells. Identification of heterogeneous leukaemia clones at diagnosis and post-treatment using such a mathematical platform has the potential to predict multiple future outcomes in response to induction and consolidation chemotherapy as well as relapse kinetics.

Keywords: AML, cell cycle, leukemia heterogeneity, mathematical model, population balance model

Introduction

Leukaemia heterogeneity develops as a result of compound genetic and micro-environmental modifications and is reflected by increased cell proliferation, block of differentiation pathways and reduction of apoptotic signals [1, 2]. Cooperative effects of several mutations, as well as their order of appearance, result in cell sub-populations that acquire unique traits [3, 4]. The resulting disease- and patient-specific heterogeneity is one of the main sources of variation in treatment response [1, 5]. In Acute Myeloid Leukaemia (AML), current gold-standard treatment depends on the use of cell cycle specific (CCS) chemotherapy. Delivery of truly personalized chemotherapy remains a challenge since most patients relapse due to clonal resistance or the emergence of more aggressive leukaemic sub-clones as a result of the initial chemotherapy used to treat the disease [5, 6].

We have previously shown that assessment of proliferation kinetics and cell cycle times in AML may provide optimized chemotherapy protocols in order to improve tumour cell kill yet minimize toxicity for the patient [7, 8]. Defining the changes that occur in the cell cycle is central to these optimized treatment protocols since many chemotherapeutic drugs are cell cycle phase-specific and heterogeneity of patient responses may be broadly defined in terms of cell cycle phase distributions and timings [9]. The four phases of the cell cycle (G1, S, G2 and M) are governed by the periodic fluctuation in the concentration of cyclins (Figure 1), a set of proteins that bind to cyclin-dependent kinases (CDKs), which are present in non-limiting concentrations [10]. Given favourable conditions, cells exit G0 and enter G1 in response to increasing cyclin D concentration. Cyclin E is produced during G1 and peaks at the transition to S phase, when it binds primarily to CDK-2, activating DNA duplication mechanisms. G2 starts when DNA synthesis is complete, and is reflected by increased cyclin B concentration, a protein that accumulates as G2 tasks are completed [11]. Once cyclin B threshold concentration is reached, it binds to CDK-1 resulting in the transition to mitosis (M). AML patients can have discrepant phase durations [9] and characteristically overexpressed/unscheduled cyclin patterns [12, 13]. Together, cyclin fluctuation patterns and cell cycle times can serve as a unique signature to define patient-specific disease characteristics mathematically, which is essential to the design of personalized treatments.

Prior work has modelled cell cycle phase compartments using ordinary differential equations [14]; these models quantify changes in cell number in each phase as a function of time and capture macroscopic responses of cell populations. A more detailed analysis of the system requires a description of the cell cycle phase distribution and experimental validation of the intra-phase kinetics. A first approach could be to consider the system from the point of view of a single cell as it moves within and between phases as a function of time (discrete cells, continuous phase progression); however, this definition would be computationally expensive, unnecessary and experimentally infeasible given the need to obtain data for and model millions of cells [15]. A more efficient and tangible approach is to observe the passage of cells in time at a specific point within the phase through a population function (continuous cells, continuous phase progression). Such a model is dependent on time, but also on a state variable that distributes each phase into a continuum of populations, known as a population balance model (PBM) [16]. Solving PBMs explicitly is practically impossible in most cases; the complex task of state variable space discretisation is often required. PBMs are relevant to the cell cycle in that they enable tracking of the movement of cells inside each phase through the increase in the state variable and interpret the transition between phases as the summation of exiting populations from different parts of the phase. Typically, PBMs of the cell cycle consist of three stages, namely an aggregated G0/G1 phase, S phase and an aggregated G2/M phase [17]. Traditionally, properties such as cell age, size or volume have been used as state variables. Age-based PBMs cannot be directly validated as age is not a biological property; it can be used as a mathematical artefact theoretically correlating phase coordinate to phase time; however, perturbations in the transition state cannot be explained by time alone [18]. Size- and volume- based PBMs have perhaps more relevant qualitative biologically-meaningful state vari-

ables; high variability in the experimental data renders validation of these models extremely difficult requiring the estimation of the necessary parameters [19]. Finally, whereas DNA content is a very good state variable choice for S phase, where it is synthesized and can be measured, it bears no relevance to the progress of cells in other phases. Consequently, traditional PBM models cannot inform experiments with regards to the expected state variable level (since it is either a "virtual" state variable or very difficult to measure); only models that specifically predict cellular properties (for a single cell) explicitly quantify these levels, but at the expense of macroscopic growth [20, 21]. Hence, there is a need for cell cycle PBMs that can be seamlessly validated experimentally and for explicit models of cell cycle checkpoints which can incorporate growth kinetic behaviour.

Herein, we have developed a mathematical model of the cell cycle using PBM that captures kinetics at both levels: micro (intracellular protein concentration) and macro (cell growth). The model is then used to predict heterogeneous proliferation kinetics, such as the ones presented by three leukemic cell lines (K-562, MEC-1, MOLT-4) with very disparate origins (AML, CLL and ALL respectively), given a reduced set of experimental parameters.

Results

Development of a multi-stage PBM based on cyclin concentration

The model is composed of 3 compartments according to DNA content: G (DNA content of 1), S (increasing DNA content 1-2) and M (DNA content of 2), similar to other models available [17, 18]. The novelty resides in the fact that cyclin E and cyclin B are used as state variables (defined as C_E and C_B in the model) for G0/G1 and G2/M respectively, as these are the phases where their concentration actively increases linked to phase progression. DNA content (defined as DNA in the model) is used as in previous models [19] for the representation of progress in S phase. Each of the three phases is modelled by a PBM equation (Equations 1–3; refer to Table S2 for variable definitions), and these equations are linked by the transfer of cells from phase to phase (Figure 1) [22, 23].

$$\frac{\partial G(\mathrm{C}_E,t)}{\partial t}+\frac{\partial\left(G(\mathrm{C}_E,t)\cdot\frac{d\mathrm{C}_E}{dt}\right)}{\partial \mathrm{C}_E}=-r_{G\to S}(\mathrm{C}_E)\cdot G(\mathrm{C}_E,t) \tag{1}$$

$$\frac{\partial S(\mathrm{DNA},t)}{\partial t}+\frac{\partial\left(S(\mathrm{DNA},t)\cdot\frac{d\mathrm{DNA}}{dt}\right)}{\partial \mathrm{DNA}}=0 \tag{2}$$

$$\frac{\partial M(\mathrm{C}_B,t)}{\partial t}+\frac{\partial\left(M(\mathrm{C}_B,t)\cdot\frac{d\mathrm{C}_B}{dt}\right)}{\partial \mathrm{C}_B}=-r_{M\to G}(\mathrm{C}_B)\cdot M(\mathrm{C}_B,t) \tag{3}$$

$G(\mathrm{C}_E,t)$ represents the number of cells in G0/G1 at time t that have a cyclin content C_E; similarly, $S(\mathrm{DNA},t)$ and $M(\mathrm{C}_B,t)$ represent the number of cells in S and G2/M that have a DNA content *DNA* and a cyclin B content C_B respectively at time t. $r_{M\to G}(\mathrm{C}_B)$ and $r_{G\to S}(\mathrm{C}_E)$ represent the transition rates from G2/M to G0/G1 and from G0/G1 to S respectively (both dependent on the particular phase state variable). Biologically, growth rates account for the speed at which the accumulation or production of cyclin/DNA occurs in a cell in the relevant phases. Mathematically, growth rates represent how quickly cells progress through the phase. Phase durations (T_G, T_S and T_M) are defined as the average time a cell spends in a phase. Cyclin minima represent the baseline expression at the start of the phase ($\mathrm{C}_{E,\min}$ and $\mathrm{C}_{B,\min}$ for G0/G1 and G2/M) while cyclin thresholds ($\mathrm{C}_{E,\mathrm{thr}}$ and $\mathrm{C}_{E,\mathrm{thr}}$) account for the average cyclin level at which cells move to the next phase. A constant cyclin E production rate (r_G) is used for G1 (Eq. 4) [24]. DNA production (r_S) is approximated as a lineal function (Eq. 5) based on normalized results [25]. Constant cyclin B production (r_M) occurs during G2 with a concentration plateau at

transition [26]; since transition occurs rapidly, we assume a constant cyclin B production throughout G2 (Eq. 6).

$$r_G = \frac{d\mathrm{C}_E}{dt} = \frac{\mathrm{C}_{E,\mathrm{thr}} - \mathrm{C}_{E,\mathrm{min}}}{T_G} \qquad \text{(G0/G1)} \tag{4}$$

$$r_S = \frac{d\mathrm{DNA}}{dt} = \frac{2-1}{T_S} \qquad \text{(S)} \tag{5}$$

$$r_M = \frac{d\mathrm{C}_B}{dt} = \frac{\mathrm{C}_{B,\mathrm{thr}} - \mathrm{C}_{B,\mathrm{min}}}{T_M} \qquad \text{(G2/M)} \tag{6}$$

A cell in G0/G1 or G2/M can either move to the next phase or move to the next cyclin level. Biologically, the higher the cyclin concentration and the more likely it is for the cell to transition. Mathematically, the transition probability ($\Gamma(\mathrm{C}_E)$ and $\Gamma(\mathrm{C}_B)$) accounts for the likelihood of a cell at a particular position in the phase moving to the next phase, which is explicitly calculated as the ratio between transition happening ($r_{G\to S}$ or $r_{M\to G}$) and all of transition and growth happening ($r_{G\to S} + r_G$ or $r_{M\to G} + r_M$).

$$\Gamma_G(\mathrm{C}_E) = \frac{r_{G\to S}(\mathrm{C}_E)}{r_{G\to S}(\mathrm{C}_E) + r_G} \tag{7}$$

$$\Gamma_M(\mathrm{C}_B) = \frac{r_{M\to G}(\mathrm{C}_B)}{r_{M\to G}(\mathrm{C}_B) + r_M} \tag{8}$$

A recent study of different transition rate functions in cell cycle PBMs has indicated that the particular function used had little impact on the ability of the model to fit the experimental data [27]. We assumed a normal cumulative distribution function for the transition probabilities $\Gamma(\mathrm{C}_E)$ and $\Gamma(\mathrm{C}_B)$ (see Section 1 in SI).

The boundary conditions address the discontinuities between phases, where cells from a different phase enter a new phase. Since cells transitioning to S or to G0/G1 arrive with different cyclin concentrations from G0/G1 (Eq 9) or G2/M (Eq 10) respectively, the total number of cells is calculated by taking the integral of the transition term over all cyclin levels. For G0/G1, incoming cells are doubled to account for cell division; for S phase, all the cells with a doubled DNA content are considered to transition to the start of G2 (Eq 11):

$$r_G \cdot G(\mathrm{C}_E = \mathrm{C}_{E,\mathrm{min}}, t) = 2 \int_{\mathrm{C}_{B,\mathrm{min}}}^{\mathrm{C}_{B,\mathrm{max}}} r_{M\to G}(\mathrm{C}_B) \cdot M(\mathrm{C}_B, t)\ d\mathrm{C}_B \tag{9}$$

$$r_S \cdot S(\mathrm{DNA} = 1, t) = \int_{\mathrm{C}_{E,\mathrm{min}}}^{\mathrm{C}_{E,\mathrm{max}}} r_{G\to S}(\mathrm{C}_E) \cdot G(\mathrm{C}_E, t)\ d\mathrm{C}_E \tag{10}$$

$$r_M \cdot M(\mathrm{C}_B = \mathrm{C}_{B,\mathrm{min}}, t) = r_S \cdot S(\mathrm{DNA} = 2, t) \tag{11}$$

Two assumptions were made: (i) the G0/G1 phase is aggregated: leukaemic cell lines are highly proliferative and therefore only a small percentage of cells with DNA content 1 will actually be quiescent (Figure 2) and, (ii) the G2/M phase is aggregated: the duration of the M phase is short enough compared to that of G2, such that it does not affect significantly the overall cell cycle progress [28].

Given the complexity of the equations (partial differential and integral terms), discretisation of the state variable space is required. Phase domains start at $\mathrm{C}_{E,\mathrm{min}}$ (G0/G1), 1 (S) and $\mathrm{C}_{B,\mathrm{min}}$ (G2/M), and are truncated at $\mathrm{C}_{E,\mathrm{max}}$ (G0/G1), 2 (S) and $\mathrm{C}_{B,\mathrm{max}}$ (G2/M). Compartments are subdivided into n_i bins $\forall\, i \in \{E; D; B\}$; each of the bins representing a range of state variable levels which correlates to the bin number according to the following

equations:

$$\nu_E = n_E/(\mathrm{C}_{E,\max} - \mathrm{C}_{E,\min}) \tag{12}$$

$$\nu_D = n_D/(\mathrm{DNA}_{\max} - \mathrm{DNA}_{\min}) = n_D/(2-1) = n_D \tag{13}$$

$$\nu_B = n_B/(\mathrm{C}_{B,\max} - \mathrm{C}_{B,\min}) \tag{14}$$

Each bin corresponds to state variable levels that span: $1/\nu_E$ (G0/G1 phase), $1/\nu_D$ (S phase) , and $1/\nu_B$ (G2/M phase). The state variable levels for each bin become:

$$\mathrm{C}_{E,e} = (e-0.5)/\nu_E + \mathrm{C}_{E,min} \qquad \forall e \in \{1,\ldots,n_E\} \tag{15}$$

$$\mathrm{DNA}_d = (d-0.5)/\nu_D + 1 \qquad \forall d \in \{1,\ldots,n_D\} \tag{16}$$

$$\mathrm{C}_{B,b} = (b-0.5)/\nu_B + \mathrm{C}_{B,\min} \qquad \forall b \in \{1,\ldots,n_B\} \tag{17}$$

In the discrete form, Equations 7 and 8 become $\Gamma_{G,e} = \Gamma_G(\mathrm{C}_{E,e})$ and $\Gamma_{M,b} = \Gamma_M(\mathrm{C}_{B,b})$. The model now considers n_i subpopulations $\forall\, i \in \{E;\ D;\ B\}$ corresponding to each state variable level in the compartment (Figure 1), which are defined as a vector of length n_i. Following discretisation of cyclins and DNA via a fully stable upwind scheme [29, 30], the model equations are simplified into n_i ODEs per compartment as follows:

$$\frac{dG_e}{dt} = (G_{e-1}(t) - G_e(t)) \cdot \nu_E \cdot r_G - G_e(t) \cdot \nu_E \cdot r_{G\to S,e} \qquad \forall e \in \{2,\ldots,n_E\} \tag{18}$$

$$\frac{dS_d}{dt} = S_{d-1}(t) \cdot r_S \cdot \nu_D - S_d(t) \cdot r_S \cdot \nu_D \qquad \forall d \in \{2,\ldots,n_D\} \tag{19}$$

$$\frac{dM_b}{dt} = M_{b-1}(t) \cdot r_M \cdot \nu_B - M_b(t) \cdot \nu_B \cdot \left(r_M + r_{M\to G,b}\right) \qquad \forall b \in \{2,\ldots,n_B\} \tag{20}$$

The discretised counterpart of $\partial\left(G(\mathrm{C}_E,t)\cdot\frac{d\mathrm{C}_E}{dt}\right)/\partial \mathrm{C}_E$ (Eq 1) is $(G_{e-1}(t) - G_e(t)) \cdot \nu_E \cdot r_G$ (Eq 18); the transition term $r_{G\to S}(\mathrm{C}_E)\cdot G(\mathrm{C}_E,t)$ corresponds to:$G_e(t)\cdot\nu_E\cdot r_{G\to S,e}$ in the discretised form; change with time is converted from $\partial G(\mathrm{C}_E,t)/\partial t$ to dG_e/dt. In addition, the boundary conditions become:

$$\frac{dG_1}{dt} = 2\cdot\sum_{b=1}^{n_B} M_b\cdot\nu_B\cdot r_{M\to G,b} - G_1(t)\cdot\nu_E\cdot(r_G + r_{G\to S,1}) \tag{21}$$

$$\frac{dS_1}{dt} = \sum_{e=1}^{n_E} G_e\cdot\nu_E\cdot r_{G\to S,e} - S_1(t)\cdot\nu_D\cdot r_S \tag{22}$$

$$\frac{dM_1}{dt} = S_{n_D}\cdot\nu_D\cdot r_S - M_1(t)\cdot\nu_B\cdot(r_M + \cdot r_{M\to G,1}) \tag{23}$$

Experimental validation of the PBM

EdU is a thymidine analogue that can be incorporated in the DNA of cells during S phase [9]. Cells undergoing DNA duplication are effectively labelled but not G0/G1 or G2/M cells, resulting in two separate populations that can later be tracked by flow-cytometry. This provides a suitable method to generate cell cycle "movies" from which cell cycle times can be extracted. EdU exposure does not significantly affect cell proliferation so long as the uptake is short and in low concentrations [31]. Regardless, only information from the unlabelled population was utilized. Subsequent phase deconvolution can be performed by DNA staining and the concentration of cyclins E and B is monitored by simultaneous antibody staining.

The experimental system is composed of thousands of cells, each characterized by a fluorescence intensity per channel. Global phase behaviour is obtained by normalizing the geometric mean fluorescence of individual cells [32]. A similar approach can be used in the case of the discretised model, except the system is composed of groups of cells with similar characteristics instead of single cells. The equivalence between the data analysis procedure used to aggregate flow-cytometric data and the mathematical procedure to combine the simulation data of each of the bins (refer to Figure S5 in SM for more details) is as follows:

$$\text{Geometric mean}_{exp} = \left(\prod_{i=1}^{N_{\text{cells}}} f_i \right)^{1/N_{\text{cells}}} \tag{24}$$

$$\text{Geometric mean}_{PBM} = \left(\prod_{j=1}^{N_{\text{bins}}} \mathrm{C}_j^{N_{\text{cells},\,j}} \right)^{1/N_{\text{cells},tot}} \tag{25}$$

$$\text{where} \quad N_{\text{cells},tot} = \sum_{j=1}^{N_{\text{bins}}} N_{\text{cells},\,j} \tag{26}$$

where i indicates the cell index and j the bin index; f_i and f_j are the normalized fluorescence of a cell and of a bin, respectively. Because Eq 24 and 25 are adapted formulas for each of the systems to calculate a common variable, the resulting values are equivalent and thus comparable (Figure S5).

The transition probability function chosen implies that cells statistically never reach the maximum cyclin value in the model. It is assumed that a maximum probability of 99.99% can be achieved. Therefore, the maximum value of cyclin is theoretically obtained as the cyclin value at which 99.99% of the cells would have transitioned, which is equivalent to solving $\Gamma_G\left(\mathrm{C}_{E,max}\right) = 0.9999$ and $\Gamma_M\left(\mathrm{C}_{B,max}\right) = 0.9999$. A conservation analysis (detailed next) confirms this doesn't result in significant cell loss while providing enough flexibility for "outlier" cyclin expression events to take place (Figure S1). By setting the duplication factor to one, cell numbers in the model are constant over time; the numerical solutions are tested to fulfil this property at two different levels: total cell number and final phase bins ($G_{n_E}(t)$ and M_{n_B}). For the total cell number, the maximum loss recorded was $1.2{\cdot}10^{-5}$% in G phase and $2{\cdot}10^{-6}$% in M phase (K-562). The gPROMS solver used was DASOLV with $\varepsilon = 1{\cdot}10^{-5}$; the cell loss is within the error of the numerical solver therefore it can be assumed to be zero. The test for the final phase bins lead to even smaller cell losses ($\cdot 10^{-37}$%). The model was additionally tested for cell conservation based on the number of discretisation intervals allowed. The duplication factor was again set to 1 and the model was run for 5 different scenarios with n_E, $n_D NA$ and n_B set to decreasing numbers, for a total of 1000h. The results in terms of total cells remaining after 1000h compared to initial cell number (represented as Total %) and percentage of cells in G0/G1 and G2/M phases exiting at the last bin are presented in Figure S1A. Discretisation intervals must be reduced to very few to push the model into conservation issues. However, since we are relying on bin numbers for averaged cyclin expression, it is still important to keep a wide distribution for a good resolution in cyclin expression (otherwise situations like Figure S1B can occur).

Because transitions are modelled according to a normal cumulative distribution, the probability of transition in the last bins of G and M (G_{n_E} and M_{n_B}) is very high (we have assumed 99.99%). When converted via Eq 7 and 8, the resulting transition rates ($r_{G\to S,e}$ and $r_{M\to G,b}$) become very high as compared to the growth rates (r_G and r_M). Therefore, the cell number that is lost through the passage to the next (nonexistent) bin through growth

processes is minimal, satisfying the conservation requirements:

$$\lim_{n_E \to \infty} -G_{n_E}(t) \cdot r_G \cdot \nu_{n_E} = 0 \tag{27}$$

$$\lim_{n_B \to \infty} -M_{n_B}(t) \cdot r_M \cdot \nu_{n_B} = 0 \tag{28}$$

In order to calculate cyclin thresholds, $\mathrm{EdU}^{-}\mathrm{C}_B^{+}$ at time 0 was considered to contain cells in late G2/M phase. $\mathrm{C}_{B,\mathrm{thr}}$ and $\mathrm{C}_{E,\mathrm{thr}}$ were set to the value at the peak, occuring when $G_\%$ and $M_\%$ are minimum, respectively (therefore the cells remaining are towards the end of the phase and have a cyclin content closer to the threshold). For $G_\%$, it occurs after at least $T_G - t_{exposure}$ hours, while for $M_\%$ it is recorded at the very beginning (after $T_M - t_{exposure}$, which is usually close to zero because T_M tends to be in the range of EdU exposure durations). The minimum cyclin values were set as the minimum values reached in the phase of interest throughout the experiment.

Regarding the intra-phase model initialization, two situations were encountered: even or biased cell distribution. Even phase distributions occur when cell populations are continuously coming in from the previous phase and leaving to the next phase. At steady-state, the percentage of cells (with respect to the total cells in a phase) found in a bin is proportional to the complementary probability of cells transitioning from that bin (given constant intra-phase growth). A more detailed derivation can be found in Section 2 of SM.

$$\frac{G_e}{\sum_{e'=1}^{n_E} G_{e'}} = \frac{\prod_{e'=1}^{e} (1 - \Gamma_{G,e'})}{\sum_{e'=1}^{n_E} \prod_{e''=1}^{e'} (1 - \Gamma_{G,e''})} \qquad \forall e \in \{1, \ldots, n_E\} \tag{29}$$

$$\frac{S_d}{\sum_{d'=1}^{n_D} S_{d'}} = \frac{1}{n_D} \qquad \forall d \in \{1, \ldots, n_D\} \tag{30}$$

$$\frac{M_b}{\sum_{b'=1}^{n_B} M_{b'}} = \frac{\prod_{b'=1}^{b} (1 - \Gamma_{M,b'})}{\sum_{b'=1}^{n_B} \prod_{b''=1}^{b'} (1 - \Gamma_{M,b''})} \qquad \forall b \in \{1, \ldots, n_B\} \tag{31}$$

Even phase distributions are used for the initial cell cycle distribution in every phase when modelling the total population. Biased phase distributions correspond to the situation when no cells are entering a phase but cells in the phase keep progressing and exiting to the next phase, accumulating towards the end of the phase (until the phase is depleted completely). As a result of EdU exposure, the EdU^{-} population consists of G0/G1 and G2/M cells only. Specifically, G2/M cells includes only those cells that were not in S phase at the start of EdU exposure. This means the EdU^{-} G2/M subpopulation is composed of cells that have been in the phase at least for the duration of the exposure, and have progressed through the phase. Therefore, the G2/M phase of the EdU^{-} population at time 0 can be modelled as a biased phase distribution. To capture this behaviour, the model was initialized for the duration of the EdU exposure and cell entrance to G2/M is blocked, resulting in the new boundary condition:

$$dM_1/dt = -M_1(t) \cdot r_M \cdot \nu_B - M_1(t) \cdot r_{M \to G,1} \cdot \nu_B \tag{32}$$

The resulting G0/G1 and G2/M intra-phase distribution at the end of the simulation of the exposure is used to initialize the model with the experimental data at time 0.

The model at this point included 283 variables and 400 parameters (measured and derived, see Section 6 in SM). Global sensitivity analysis (GSA) identifies the parameters that have an impact on model output, by observing the change in the outputs when parameter values are varied [33]. It assesses which experimental values critically need to be determined with experimental accuracy, and which others can be estimated or kept at their nominal values for model validity. Three groups of significant parameters were identified (Figure 3, see Materials and Methods for details): (i) $\mathrm{C}_{B,\mathrm{min}}$, $\mathrm{C}_{B,\mathrm{thr}}$, $\mathrm{C}_{E,\mathrm{min}}$ and $\mathrm{C}_{E,\mathrm{thr}}$ for cyclin E & B concentrations; (ii) T_G, T_S and T_M

for cell cycle phase kinetics; and (iii) G_{ini}, S_{ini} and M_{ini} (the percentage of total cells in G1 , S and G2/M at time zero) for the cell cycle distribution. Groups (i) and (ii) were important throughout the culture, although phase times (ii) were more significant in their respective phases when a majority of cells were exiting that phase, or in their subsequent phases, when a majority of cells were entering the phase. However, group (iii) was only significant in the first few hours. Only in one of the cell lines (MEC-1) was phase kinetics affected by cyclin threshold/minimum values (maximum observed sensitivity index value for $C_{E,\mathrm{thr}}$ was 0.35). Conversely, cyclin concentration was only affected by cell cycle times during the initial hours. Finally, since σ_E and σ_B were not identified as significant parameters, they were both fixed at 20% of the difference $C_{E,\mathrm{thr}} - C_{E,\mathrm{min}}$ and $C_{B,\mathrm{thr}} - C_{B,\mathrm{min}}$. In summary, an accurate determination of the cell cycle times (T_G, T_S and T_M) for phase kinetics and cyclin thresholds and minima ($C_{B,\mathrm{min}}$, $C_{B,\mathrm{thr}}$, $C_{E,\mathrm{min}}$ and $C_{E,\mathrm{thr}}$) for cyclin expression is essential to fully characterize the model.

The cell cycle times (T_G, T_S and T_M) were obtained by following the entrance/exit times of the EdU^- population ($\pm$ 2h) to and from each phase for the first cycle. The initial EdU^- cell cycle distribution was used to initialize and run the model. The agreement between the model and the experiments was remarkably good for all three phases in each cell line (Figure 3), with the model prediction falling within the 95% confidence region for most time points. For the geometric mean of cyclin concentration, the trends and magnitudes were captured; although the experimental data were inherently noisy (due to the need for normalization steps against isotypes or other phases, which increased the number of sources for data uncertainty, and the absence of replicates), a reasonable fit was achieved. In the particular case of cyclin B concentration profile in K-562, the DNA deconvolution of the initial S-phase cell cycle distribution places a fraction of the cells that were in the left-most section of the second DNA peak at the right-most part of S phase. Indeed, this caused a fraction of the S-phase population to enter G2/M at 2-4h, completely shadowing the high cyclin B concentration of the remaining G2/M population at that time (Figure 4A). If corrected for no cells at the end of S phase initially, the model matched the experimental data (Figure 4B) in all but one point as determined by the residual sum of squares (Figure 4C). Cell numbers analysed in this region are significantly lower since cells keep exiting the phase resulting in experimental data becoming less robust thus explaining the mismatch observed at 6h.

GSA revealed that the most significant parameters for phase kinetics were the cell cycle times. Furthermore, the analysis showed that initial cell cycle distribution values were not necessary for longer analyses (over 5h). We hypothesized that co-culture conditions would have an effect on individual kinetics. A preliminary K-562/MEC-1 co-culture experiment was carried out at 3 set ratios: 10%, 50% and 90% MEC-1. MEC-1 kinetics was clearly slower, so its time parameters were readjusted to fit the experimental data in one of the cultures (50%), and the new model results were validated against the 10% and 90% experiments (Figures 5 and S2). Observe how the relative RSS is higher towards the end of the culture (Figure 5C), especially in the 90% and 10% MEC-1 co-cultures, and much lower in all other time points.

Forward and backward heterogeneous cell population dynamics can be predicted using the PBM

Nine different co-culture mixtures of the cell lines were prepared by operator 1 (Table E1). Operator 2, blinded to the nature of the samples, performed the analysis at time 0 and determined the initial percent of each cell type in order to run the model. Subsequently, the rest of the samples were analyzed, after which all the data was gathered and compared to model simulations. In the last three tests, the model was also evaluated for its backwards prediction capability. A pre-run plot of model predictions (see Table S4 for details on model parameters used) was used to estimate the evolution of cell line contents (Figure 6A) together with the total cell number. The experimental

trends for the total cell kinetics in the last three tests (T7-T9) were correctly predicted (Figure 6B). Later points were over-estimated; the reason for this is two-fold: (i) nutrient depletion and metabolite accumulation can have a negative effect on growth which the model does not currently incorporate; and (ii) MOLT-4 and MEC-1 have maximum recommended densities of 1.5-2M cells/mL while K-562 cannot grow at such high cell densities: T9, with a higher content of MEC-1 and MOLT-4, was correctly predicted, while for T7 and T8 (which have a majority of K-562 cells) the model overestimates the last 2-3 points where the cell density was over 1-1.2M cells/mL. The model accurately predicted the evolution of all cell populations for the duration of the second part of the experiment (0-30h, Figure 6C). Additionally, the contents of the original mixture (at -48h) were found correctly in 2 out of the 3 cases (Figures 6A and 6D).

Unknown heterogeneous leukaemic populations can be deconvoluted using the PBM platform

A second strategy was implemented where operator 2 pre-ran the model assigning each initial percent to each of the cell lines (resulting in 6 different possible scenarios per blind experiment, Figure S3). The whole experimental panel was then revealed (unassigned to specific cell lines) and compared to the model output of each scenario in terms of Euclidian distance in a ternary plot: the lower the distance (relative to the other 5 scenario), the likelier to be a good match. The ranking of the scenarios likeliness given by the model for each mixture is shown in Figure 7. In mixtures T1, T2, T3 and T7, the model's highest ranking candidate matched the true experimental content, while in a further 3 mixtures (T6, T8 and T9) the actual content was found in one of the top 3 candidates. Overall, the correct solution, as a sum of the Euclidian distances in all 9 tests, scored lowest. Only in two cases (T4 and T5) did the model fail to identify the actual mixture as a real candidate, partly because other scenarios had extremely close values (within the experimental error margins). Of note, cells were co-cultured in non-standard conditions (cell densities and cell culture media used), challenging the ability of the PBM to work under uncertain conditions. Finally, the cell cycle times of the three cell lines used were relatively close; primary leukaemic cells may have more disparate populations, with variation of days [9]. In this case, the differences would be largely sufficient for the model to identify and quantify heterogeneous cell populations based on cell cycle kinetics.

Discussion

A PBM of the cell cycle based on cyclin concentration and DNA content was developed and utilized to deconvolute leukaemia population kinetics. GSA established which model parameters were critically required (and which did not need to be identified) so that they were obtained experimentally for three leukaemic cell lines by following a synchronous cell population over time, after which the model was run and compared against the actual cell cycle distribution and cyclin concentration. The agreement between the model and the experimental data was excellent, for micro (intracellular growth) as well as macro (population growth) kinetics, with most predictions falling within the 95% experimental confidence area. Additionally, residual sum of squares identified the time points where the model deviated from the experiments. The most sensitive regions included the last hours in culture (over 20h), where cells experience exposure to significant concentration gradients in terms of nutrient depletion and metabolite accumulation (Figure 5C); and regions where the number of cells analysed is not sufficient to provide an accurate representation of the phenomena occurring in culture (Figure 4C). Next, the model was successfully used for two different strategies in the context of co-culture kinetics: (i) the prediction of backward and forward culture evolution given known cell line-specific cell cycle kinetics and initial conditions and (ii) the

identification of cell type and content given an unknown experimental panel.

Clonal heterogeneity of AML and competitive outgrowth of more "fit" clones in a "Darwinian" model renders the successful treatment of this disease particularly challenging [34]. It is currently unclear which factors within a subclone or in the microenvironment make certain clones out-compete others and how this dynamic can be altered by chemotherapy schedule [9, 34, 35]. Recently, it has been proposed that tumour-specific cell-autonomous and non-cell-autonomous factors (e.g. cytokines) can alter growth kinetics and properties of tumours and that these qualities may be targeted to manipulate tumour growth [36]. Although there are many other properties relevant to leukaemia and the treatment of leukaemia, most if not all of these factors ultimately affect the cell cycle. Hence, this work describes a model of the cell cycle that can be combined or upgraded to capture additional phenomena such as differentiation (affecting traditional chemotherapy, but also affected by novel agents targeting specific types of haematopoietic cells) or environmental cues (metabolism, cytokines, contact inhibition etc). The key is capturing the most important phenomena at a sufficient level of detail in order to maintain the fidelity and relevance to the biological system. Defining which are the additional phenomena that are key in response to chemotherapy is definitely a critical step. Pharmacokinetics / pharmacodynamic (PK/PD) models capture body processes that are relevant to the drug distribution and transport to the bone marrow, as well as its effect, for example. Our model may be able to capture clonal heterogeneity which, combined with PK/PD, may lead to improved and more effective therapies.

The ability to assess both backward and forward evolution of cell clonal content using the PBM model presented herein provides a quantitative tool for the estimation of population dynamics and the study of leukaemia progression in both treated and untreated patients [6, 37, 38]. This tool also has the potential to evaluate clonal models of leukaemogenesis [4, 6, 39] in order to better understand the pathogenesis of disease. In the context of CCS drug dosage and scheduling, the PBM could give a narrower window of action than what is currently used in treatment regimens, thereby limiting drug toxicity yet improving efficacy [8, 7]. Drugs, such as small molecule inhibitors, currently being tested that specifically target leukaemia sub-clones (e.g. those expressing FLT3-ITD, [40]) could be dosed and scheduled to optimize cell kill during therapy according to anticipated sub-clonal composition. Ultimately, balancing CCS chemotherapy with more advanced gene-targeted treatments will require detailed knowledge of sub-population evolution, kinetics and dynamics during treatment calculated and manipulated using computational methods such as those presented herein for designer therapies. The availability of individual clone sensitivities to each drug (EC50) and the use of accurate models simulating drug distribution in the human body will be key for simulating cell death by chemotherapy. The PBM's high resolution in describing intra-phase events allows incorporating cell death not only in a phase-specific, but also in a phase coordinate-specific manner (data not shown).

Genome-based methods are currently employed in order to define mutational heterogeneity in AML subclones, thereby identifying evolution of disease throughout treatment [6]. In order to optimize treatments in AML, we pursued a complementary strategy of clonal identification by using cell cycle kinetics to effectively represent heterogeneity. Cells may acquire proliferative mutations that modify cell cycle times through bypassing of cell cycle controls; overexpression of p21 enables cells with DNA damage to continue to the next phase prematurely [41], while mutations in Cdc20 delay the timing of cyclin B degradation, and pRb or E2F mutations promote early entry into S phase by premature increase in the concentration of cyclin E [42]. A plethora of sub-populations, each with specific cell cycle signatures arises; the PBM can estimate cell cycle evolution for each of them and capture overall behaviour. Since cell cycle duration has been shown to correlate with prognosis in response to treatment [43], an *a priori* classification of sub-populations according to slow/intermediate/fast cell cycle kinetics may be initially developed. A more sophisticated extrapolation for patient cell populations is currently being studied by

the use of steady state parameters [44], rather than the dynamic counterparts (used here), which could potentially be measured in the diagnostic bone marrow or peripheral blood sample at disease presentation. Technology now exists wherein patient AML mutations and subclones are identified routinely [45]. Put together with cell cycle data (which could be done at diagnosis and throughout treatment if needed by assessment of S-Phase or cyclin expression by flow cytometry), PK/PD combined models are being developed to optimise chemotherapy dose and schedule according to anticipated clonal output [8, 44]. With this method, there is the potential to tailor treatments by manipulating leukaemic heterogeneity and instilling long-term control of tumor growth, limiting expansion of more pathogenic clones [36, 46].

Heterogeneous cyclin production patterns originate as a result of cell cycle mutations (e.g., cyclins D1 and E1, KIP1, INK4B and INK4A, CDK4, Rb [11]), making them leukaemia-specific. Predicting the intensity and timing of cyclin concentration increase can become critical in defining the optimal dose and schedule of current as well as novel treatment strategies [8, 47]. Due to their critical role in the control of cell cycle progression, cyclin-blockade strategies arrest proliferation, without the risk of additional mutagenesis inherent in traditional chemotherapeutics. Specific cell cycle drug targets, such as cyclin D and CDK1-cyclin B complexes [48, 49], may be identified and tracked using the PBM platform developed herein and used to identify the best dose and schedule pre-clinically, expediting drug development trajectories.

The development of detailed models of the cell cycle that are experimentally validated is critical in the implementation of more advanced pharmacokinetic/pharmacodynamic models [50, 51]. Additionally, linking a small subset of measurable variables to unique characteristics of the individual is necessary for the development of personalized treatment. The PBM we have developed paves the way in connecting both. Ultimately, the application of such a tool could inform not only optimal timing and type of personalized treatment for improved outcomes, but also provide a platform for pre-clinical assessment of novel targeted therapies for leukemia and other cancers.

Materials and methods

Computational tools

The computational tools used for carrying out the simulations in this manuscript are detailed below, as well as the sensitivity analysis methods. The model was implemented on gPROMS ModelBuilder 3.5.3 and simulated on a 64-bit Windows with an Intel Core 2 Quad CPU 2.67 GHz and 4GB RAM. GSA was performed on MATLAB for the discretised model by connecting gPROMS through gO:MATLAB. Sobol's method [52] was used, screening the influence of 24 parameters with 20,000 intervals each on the G1-S-G2/M phase percents, cyclin expression, % cell loss, transition rates and total cell number. The inputs were varied $\pm$ 40%, built onto a matrix in MATLAB® and a call was sent sequentially via gO:MATLAB® for every set of parameter values. The outputs at 0h, 1h, 2h, 5h, 10h, 15 and 20h for each parameter were recorded and the sensitivity indexes were subsequently calculated on GUI-HDMR [53] MATLAB package. The analysis was made on the model including the reinitialization period for EdU exposure (biased phase distribution).

Leukemia cell lines

Cell lines are lab-adapted populations that give reproducible experimental results and require basic culture methods for their maintenance. Three leukaemia cell lines were used: K-562 (chronic myeloid leukaemia in blast crisis; has unscheduled cyclin E production and overexpressed cyclin B [54]; MEC-1 (chronic lymphoid leukemia, which expresses CD19) and MOLT-4 (acute lymphoid leukaemia, has scheduled cyclin E and B [13]). K-562

(ATCC, MD, USA) was cultured in Isocove Modified Dulbecco's Medium (IMDM; Invitrogen, CA, USA) with 10% heat inactivated Fetal Bovine Serum (FBS; Invitrogen) and 1% penicillin-streptomycin (P/S; Invitrogen). MEC-1 (DSMZ, Germany) cells were cultured in IMDM + 20% FBS + 1% P/S for 1 week, and in IMDM + 10% FBS + 1% P/S thereafter. MOLT-4 (ATCC) cells were cultured in RPMI-1640 (Invitrogen) + 10% FBS + 1% P/S. All cells lines were used at $p<20$.

Labelling cells with EdU

In vitro labelling of cells with EdU is necessary to segregate the global population into two approximately synchronous populations for dynamics to be observed. The AF647 5-ethynyl-2'-deoxyuridine (EdU) kit (Invitrogen) was reconstituted and used according to manufacturer's instructions. Cells were resuspended in fresh medium and pre-cultured for 12h in T175 flasks (Corning, NY, USA) at 300,000 cells/mL (K-562), 400,000 cells/mL (MEC-1) or 350,000 cells/mL (MOLT-4). EdU (10nM) was added (optimal concentration was determined by preliminary experiments, data not shown) and after 1h (MEC-1) or 2h (K-562, MOLT-4), cells were washed twice and re-suspended in fresh medium for culture under standard conditions in 6 well-plates (Corning; 4-6mL/well). Two cultures were exposed with a difference of 11-14 hrs; samples were collected every 2hrs for intervals of up to 14hrs consecutively and the data was merged (overlap in cell cycle distribution values was confirmed at matching points). An unexposed culture was used as a control.

Co-culture experiments

Co-culture indicates an experimental setup where two or more separate cell lines are physically put together in the same culture medium. In order to distinguish the populations, specific analysis methods detecting each of them individually have to be developed. K-562 and MEC-1 cells were mixed at the appropriate ratios and pre-cultured in IMDM+10%FBS, After 48h, they were resuspended in fresh medium at a density of $0.5 \cdot 10^6$ cells/mL and cultivated in 6-well plates (Corning), 4mL/well. Triplicate samples of 10^6 cells were taken at 0h, 5h, 10h, 20h, 25h and 30h. For the blinded experiments, operator 1 prepared unknown mixtures of K-562, MEC-1 and MOLT-4 cell lines (different cell types and ratios), which were pre-cultured for 48h. Samples were taken every 5-10h for a total of 30h. Triplicate samples were collected and split into two tests: samples to be stained for CD19 and fixed in para-formaldehyde 4% and samples to be fixed in ethanol for later detection of cyclin B concentration. In the segmentation of co-cultured samples, the percentage of MEC-1 cells were inferred from the population with CD19 expression, the MOLT-4 content from the population with high cyclin B concentration (Figure S4); K-562 was the % remaining. For samples where only K-562 and MOLT-4 were present, SSC vs DNA analyses sufficed to gate each cell type. Note that this was impossible when all three cell lines were together as SSC vs DNA of MEC-1 overlaps with both K-562 and MOLT-4 (only approximate values of percent of each cell line could be deduced).

Flow cytometry

Flow cytometry is a technique that allows translating single cell properties (DNA content, protein expression) to fluorescence intensity by labelling them with antibodies specific to the protein or substances that bind to DNA strands and fluoresce when exposed to a laser beam of a particular wavelength. The cyclin B IgG1κ antibody kit (clone GNS-1; BD, NJ, USA) was used according to manufacturer's instructions. Anti-human cyclin E antibody (clone HE12), together with the isotype control IgG1 (clone ICIGG1) and the FITC secondary antibody to IgG were prepared as per manufacturer's protocol (all three from Abcam, UK). The V450 Ki-67 IgG1κ (clone B56,

BD) antibody and V450 IgG1κ isotype control (clone MOPC-21, BD) were both diluted 4x. A propidium iodide (PI; Sigma, MO, USA) solution was prepared by dissolving 10mg/L PI with 100mg/L DNase-free RNase (Sigma) in phosphate buffered saline (PBS; Invitrogen), as previously described (51). Triplicate samples were collected regularly for each condition (EdU: every 2h; No EdU: every 6h) and fixed in 70% ethanol at -20°C for a maximum of 2 days prior to acquisition. Cells were then labelled with antibodies after permeabilisation in the following order: (i) exposure to EdU reagent, (ii) staining with either cyclin B or E antibodies, and (iii) resuspended in PI.

For CD19 labelling in co-culture experiments, labelled or control antibody was added to each of the three replicate samples and incubated for 30min at 4°C in the dark, fixed in para-formaldehyde at 4°C, and data was acquired within 2 days with a Guava flow-cytometer (easyCyte 8HT, Millipore, MA, USA); a Fortessa flow-cytometer (LSRFortessa, BD) was used for all other data acquisition. FACSDiva software was used during acquisition of data on Fortessa, while easyCyte was used on the Guava; for all samples, 20,000 events were acquired. Data analysis, gating and geometric mean calculations were performed in FlowJo 8.7 (TreeStar Inc, OR, USA).

In order to process this data, several steps have to be performed for every sample: (i) gating out the debris on the FSC vs SSC; (ii) gating out the doublets on the PI (area) vs PI (width) signal; (iii) deconvoluting the cell cycle distribution (Dean-Jett-Fox method) resulting in G0/G1, S (split in 4 equal gates at time 0 only) and G2/M gates; (iv) gating the EdU positive from the EdU negative populations (from control which was unexposed to EdU but exposed to click reaction cocktail); (v) gating the cyclin positive from the cyclin negative populations (from isotype control exposed to EdU and reaction cocktail). From this analysis, 4 different populations are identified: EdU^+ C^+; EdU^+ C^-; EdU^- C^+; EdU^- C^-, in addition to the ungated total population. For each sample: (i) the geometric mean of the fluorescence in the C^+ population in the phase of interest divided by the geometric mean of the fluorescence in the total population in the phase where the cyclin expression is minimal gives the normalised sample expression; (ii) for the isotype samples, the geometric mean of the fluorescence in the phase of interest divided by the geometric mean of the fluorescence in the total population in the phase where the cyclin expression is minimal gives the baseline expression of the phase; (iii) finally, the normalized values obtained in (1) are divided by the baseline values from (2) giving the normalized cyclin expression of the phase of interest for the sample (Equations 33 and 34).

$$\mathrm{C}_{E,norm} = \frac{f_{\text{sample}}\left(\mathrm{C}^{+}_{E,G1}\right) / f_{\text{sample}}\left(\mathrm{C}^{tot}_{E,G2/M}\right)}{f_{\text{isotype}}\left(\mathrm{C}^{tot}_{E,G1}\right) / f_{\text{isotype}}\left(\mathrm{C}^{tot}_{E,G2/M}\right)} \tag{33}$$

$$\mathrm{C}_{B,norm} = \frac{f_{\text{sample}}\left(\mathrm{C}^{+}_{B,G2/M}\right) / f_{\text{sample}}\left(\mathrm{C}^{tot}_{B,G1}\right)}{f_{\text{isotype}}\left(\mathrm{C}^{tot}_{B,G2/M}\right) / f_{\text{isotype}}\left(\mathrm{C}^{tot}_{B,G1}\right)} \tag{34}$$

Data accessibility statement

The experimental data presented in the manuscript is provided as an Excel attachment.

Competing interests statement

We have no competing interests.

Acknowledgements

MFG thanks C. Simpson and J. Srivastava for flow-cytometry advice, N. Diangelakis and M. Papathanasiou for assistance in sensitivity analysis procedures.

Funding statement

This work is supported by ERC-BioBlood (#340719; MFG, EV), ERC-Mobile Project (#226462; EV), by the EU 7th Framework Programme [MULTIMOD Project FP7/2007-2013, #238013; MFG, DGM] and by the Richard Thomas Leukaemia Research Fund. RM is thankful for a Royal Academy of Engineering Research Fellowship.

Author Contributions

MFG wrote the manuscript and carried out the experiments, data processing, model development, parameter estimation, GSA and model validation. RM carried out the experiments and participated in the model development, as well as providing feedback throughout. DGM provided the initial version of the PBM and the idea of using cyclins as state variables, as well as input in the process of adaptation for leukaemia. EV gave feedback on experimental results. MK and MCG developed and provided the method for discretising the PBM. ENP gave feedback on model development and validation. NP reviewed the manuscript and provided advice on clinical relevance of the results presented here. AM reviewed the manuscript and suggested the idea of a co-culture blinded experiment for model validation. All authors gave final approval for publication.

1 References

[1] Ley TJ, Miller C, Ding L, Raphael BJ, Mungall AJ, Robertson AG, et al. Genomic and Epigenomic Landscapes of Adult De Novo Acute Myeloid Leukemia. New England Journal of Medicine. 2013;368(22):2059–2074.

[2] Kelly LM, Gilliland DG. Genetics of myeloid leukemias. Annu Rev Genomics Hum Genet. 2002;3:179–98.

[3] Warner JK, Wang JC, Hope KJ, Jin L, Dick JE. Concepts of human leukemic development. Oncogene. 2004;23(43):7164–77.

[4] Klco JM, Spencer DH, Miller CA, Griffith M, Lamprecht TL, O'Laughlin M, et al. Functional heterogeneity of genetically defined subclones in acute myeloid leukemia. Cancer Cell. 2014;25(3):379–92.

[5] Meacham CE, Morrison SJ. Tumour heterogeneity and cancer cell plasticity. Nature. 2013;501(7467):328–37.

[6] Ding L, Ley TJ, Larson DE, Miller CA, Koboldt DC, Welch JS, et al. Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing. Nature. 2012;481(7382):506–10.

[7] Pefani E, Panoskaltsis N, Mantalaris A, Georgiadis MC, Pistikopoulos EN. Design of optimal patient-specific chemotherapy protocols for the treatment of acute myeloid leukemia (AML). Computers & Chemical Engineering. 2013;57:187–195.

[8] Pefani E, Panoskaltsis N, Mantalaris A, Georgiadis MC, Pistikopoulos EN. Chemotherapy Drug Scheduling for the Induction Treatment of Patients With Acute Myeloid Leukemia. Ieee Transactions on Biomedical Engineering. 2014;61(7):2049–2056.

[9] Preisler HD, Raza A, Gopal V, Ahmad S, Bokhari J. Distribution of Cell-Cycle Times Amongst the Leukemia-Cells within Individual Patients with Acute Myelogenous Leukemia. Leukemia Research. 1995;19(10):693–698.

[10] Pietras EM, Warr MR, Passegue E. Cell cycle regulation in hematopoietic stem cells. Journal of Cell Biology. 2011;195(5):709–720.

[11] Malumbres M, Barbacid M. To cycle or not to cycle: A critical decision in cancer. Nature Reviews Cancer. 2001;1(3):222–231.

[12] Iida H, Towatari M, Tanimoto M, Morishita Y, Kodera Y, Saito H. Overexpression of cyclin E in acute myelogenous leukemia. Blood. 1997;90(9):3707–3713.

[13] Gong JP, Ardelt B, Traganos F, Darzynkiewicz Z. Unscheduled Expression of Cyclin-B1 and Cyclin-E in Several Leukemic and Solid Tumor-Cell Lines. Cancer Research. 1994;54(16):4285–4288.

[14] Daukste L, Basse B, Baguley BC, Wall DJN. Mathematical Determination of Cell Population Doubling Times for Multiple Cell Lines. Bulletin of Mathematical Biology. 2012;74(10):2510–2534.

[15] Singhania R, Sramkoski RM, Jacobberger JW, Tyson JJ. A Hybrid Model of Mammalian Cell Cycle Regulation. Plos Computational Biology. 2011;7(2).

[16] Ramkrishna D. Population balances: Theory and applications to particulate systems in engineering. San Diego, CA: Academic Press; 2000.

[17] Sidoli FR, Asprey SP, Mantalaris A. A coupled single cell-population-balance model for mammalian cell cultures. Industrial & Engineering Chemistry Research. 2006;45(16):5801–5811.

[18] Sherer E, Hannemann RE, Rundell A, Ramkrishna D. Analysis of resonance chemotherapy in leukemia treatment via multi-staged population balance models. Journal of Theoretical Biology. 2006;240(4):648–661.

[19] Liu YH, Bi JX, Zeng AP, Yuan JQ. A population balance model describing the cell cycle dynamics of myeloma cell cultivation. Biotechnology Progress. 2007;23(5):1198–1209.

[20] Tyson JJ. Modeling the Cell-Division Cycle - Cdc2 and Cyclin Interactions. Proceedings of the National Academy of Sciences of the United States of America. 1991;88(16):7328–7332.

[21] Weis MC, Avva J, Jacobberger JW, Sreenath SN. A Data-Driven, Mathematical Model of Mammalian Cell Cycle Regulation. PLoS One. 2014;9(5).

[22] García Münzer DG, Kostoglou M, Georgiadis MC, Pistikopoulos EN, Mantalaris A. A Cyclin Distributed Cell Cycle Model in GS-NS0. 24th European Symposium on Computer Aided Process Engineering, Pts a and B. 2014;33:19–24.

[23] García Münzer DG, Kostoglou M, Georgiadis MC, Pistikopoulos EN, Mantalaris A. Cyclin and DNA Distributed Cell Cycle Model for GS-NS0 Cells. PLoS Comput Biol. 2015 02;11(2):e1004062. Available from: http://dx.doi.org/10.1371%2Fjournal.pcbi.1004062.

[24] Ohtsubo M, Theodoras AM, Schumacher J, Roberts JM, Pagano M. Human Cyclin-E, a Nuclear-Protein Essential for the G(1)-to-S Phase-Transition. Molecular and Cellular Biology. 1995;15(5):2612–2624.

[25] Avva J, Weis MC, Sramkoski RM, Sreenath N, Jacobberger JW. Dynamic Expression Profiles from Static Cytometry Data: Component Fitting and Conversion to Relative, "Same Scale" Values. PLoS One. 2012;7(7).

[26] Kang Q, Pomerening JR. Punctuated cyclin synthesis drives early embryonic cell cycle oscillations. Molecular Biology of the Cell. 2012;23(2):284–296.

[27] Chaffey GS, Lloyd DJB, Skeldon AC, Kirkby NF. The Effect of the G(1) - S transition Checkpoint on an Age Structured Cell Cycle Model. PLoS One. 2014;9(1).

[28] Morgan D. The cell cycle: principles of control. London, UK: New Science Press Ltd; 2007.

[29] Kostoglou A, Karabelas AJ. Evaluation of numerical methods for simulating an evolving particle size distribution in growth processes. Chemical Engineering Communications. 1995;136:177–199.

[30] Lapidus L, Pinder GF. Numerical Solution of Partial Differential Equations in Science and Engineering. New York: Wiley-Interscience; 1982.

[31] Cappella P, Gasparri F, Pulici M, Moll J. A novel method based on click chemistry, which overcomes limitations of cell cycle analysis by classical determination of BrdU incorporation, allowing multiplex antibody staining. Cytometry Part A. 2008;73A(7):626–636.

[32] Münzer DGG, Kostoglou M, Georgiadis MC, Pistikopoulos EN, Mantalaris A. Developing a cyclin blueprint as a tool for mapping the cell cycle in GS-NSO. Biochemical Engineering Journal. 2013;81:97–107.

[33] Kiparissides A, Koutinas M, Kontoravdi C, Mantalaris A, Pistikopoulos EN. 'Closing the loop' in biological systems modeling - From the in silico to the in vitro. Automatica. 2011;47(6):1147–1155.

[34] Greaves M, Maley CC. Clonal evolution in cancer. Nature. 2012;481(7381):306–313.

[35] Stiehl T, Baran N, Ho AD, Marciniak-Czochra A. Clonal selection and therapy resistance in acute leukaemias: mathematical modelling explains different proliferation patterns at diagnosis and relapse. Journal of the Royal Society Interface. 2014;11(94).

[36] Marusyk A, Tabassum DP, Altrock PM, Almendro V, Michor F, Polyak K. Non-cell-autonomous driving of tumour growth supports sub-clonal heterogeneity. Nature. 2014;514(7520):54–+.

[37] Walter RB, Laszlo GS, Lionberger JM, Pollard JA, Harrington KH, Gudgeon CJ, et al. Heterogeneity of clonal expansion and maturation-linked mutation acquisition in hematopoietic progenitors in human acute myeloid leukemia. Leukemia. 2014;28(10):1969–77.

[38] Wiseman DH, Greystoke BF, Somervaille TC. The variety of leukemic stem cells in myeloid malignancy. Oncogene. 2014;33(24):3091–8.

[39] Shackleton M, Quintana E, Fearon ER, Morrison SJ. Heterogeneity in cancer: cancer stem cells versus clonal evolution. Cell. 2009;138(5):822–9.

[40] Serve H, Krug U, Wagner R, Sauerland MC, Heinecke A, Brunnberg U, et al. Sorafenib in combination with intensive chemotherapy in elderly patients with acute myeloid leukemia: results from a randomized, placebo-controlled trial. J Clin Oncol. 2013;31(25):3110–8.

[41] Viale A, De Franco F, Orleth A, Cambiaghi V, Giuliani V, Bossi D, et al. Cell-cycle restriction limits DNA damage and maintains self-renewal of leukaemia stem cells. Nature. 2009;457(7225):51–6.

[42] Schnerch D, Yalcintepe J, Schmidts A, Becker H, Follo M, Engelhardt M, et al. Cell cycle control in acute myeloid leukemia. Am J Cancer Res. 2012;2(5):508–28.

[43] Raza A, Preisler HD, Day R, Yasin Z, White M, Lykins J, et al. Direct Relationship between Remission Duration in Acute Myeloid-Leukemia and Cell-Cycle Kinetics - a Leukemia Intergroup Study. Blood. 1990;76(11):2191–2197.

[44] Fuentes-Garí M, Misener R, Pefani E, García-Münzer DG, Kostoglou M, Georgiadis MC, et al. Cell cycle model selection for leukemia and its impact on chemotherapy outcomes. 24th European Symposium on Computer Aided Process Engineering. 2015;accepted.

[45] Buccisano F, Maurillo L, Del Principe MI, Del Poeta G, Sconocchia G, Lo-Coco F, et al. Prognostic and therapeutic implications of minimal residual disease detection in acute myeloid leukemia. Blood. 2012;119(2):332–341.

[46] Zhao BY, Hemann MT, Lauffenburger DA. Intratumor heterogeneity alters most effective drugs in designed combinations. Proceedings of the National Academy of Sciences of the United States of America. 2014;111(29):10773–10778.

[47] Sawai CM, Freund J, Oh P, Ndiaye-Lobry D, Bretz JC, Strikoudis A, et al. Therapeutic Targeting of the Cyclin D3:CDK4/6 Complex in T Cell Leukemia. Cancer Cell. 2012;22(4):452–465.

[48] Lapenna S, Giordano A. Cell cycle kinases as therapeutic targets for cancer. Nature Reviews Drug Discovery. 2009;8(7):547–566.

[49] Musgrove EA, Caldon CE, Barraclough J, Stone A, Sutherland RL. Cyclin D as a therapeutic target in cancer. Nature Reviews Cancer. 2011;11(8):558–572.

[50] Velliou E, Fuentes-Garí M, Misener R, Pefani E, Rende M, Panoskaltsis N, et al. A framework for the design, modeling and optimization of biomedical systems. Proceedings of the 8th International Conference on Foundations of Computer-Aided Process Design. 2014;34:225–236.

[51] Fuentes-Garí M, Velliou E, Misener R, Pefani E, Rende M, Panoskaltsis N, et al. A systematic framework for the design, modeling and optimization of personalized healthcare: making and healing blood. Computers & Chemical Engineering. 2015;accepted. Available from: http://dx.doi.org/10.1016/j.compchemeng.2015.03.008.

[52] Sobol IM. Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. Mathematics and Computers in Simulation. 2001;55(1-3):271–280.

[53] Ziehn T, Tomlin AS. GUI-HDMR - A software tool for global sensitivity analysis of complex models. Environmental Modelling & Software. 2009;24(7):775–785.

[54] Viallard JF, Lacombe F, Dupouy M, Ferry H, Belloc F, Reiffers J. Flow Cytometry Study of Human Cyclin B1 and Cyclin E Expression in Leukemic Cell Lines: Cell Cycle Kinetics and Cell Localization. Experimental Cell Research. 1999;247(1):208 – 219.

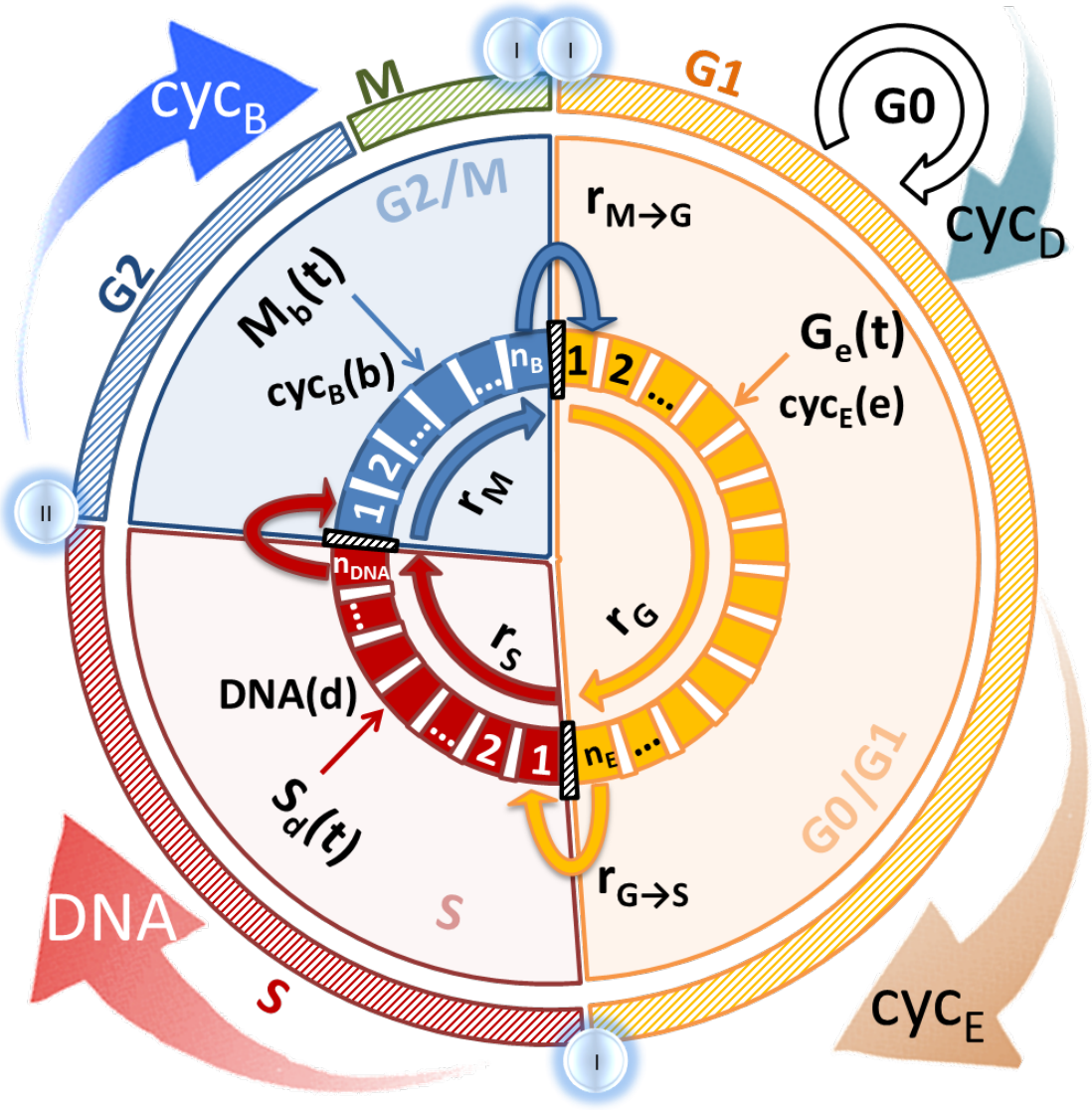


Figure 1: The cell cycle. Outer circles represent biological events (succession of phases and phase-specific cell cycle protein expression); inner "pie" circles summarize model simplifications and the discretisation strategy based on state variable level. State variables are chosen according to relevance to their phase: cyclin E for G0/G1, DNA for S and cyclin B for G2/M. Bins in each phase indicate discrete state variable levels: $\mathrm{C}_E(e)$, $\mathrm{DNA}(d)$ and $\mathrm{C}_B(b)$ correspond to the state variable levels of bin number e, d and b, while cell numbers in each bin are represented by $G_e(t)$, $S_d(t)$ and $M_b(t)$ in G0/G1, S and G2/M respectively. Transition rates ($r_{G\to S,e}$ for G0/G1 and $r_{M\to G,b}$ for G2/M) account for the likelihood of cells moving to the next phase and growth rates (r_G, r_S and r_M) reflect progress within the phase.

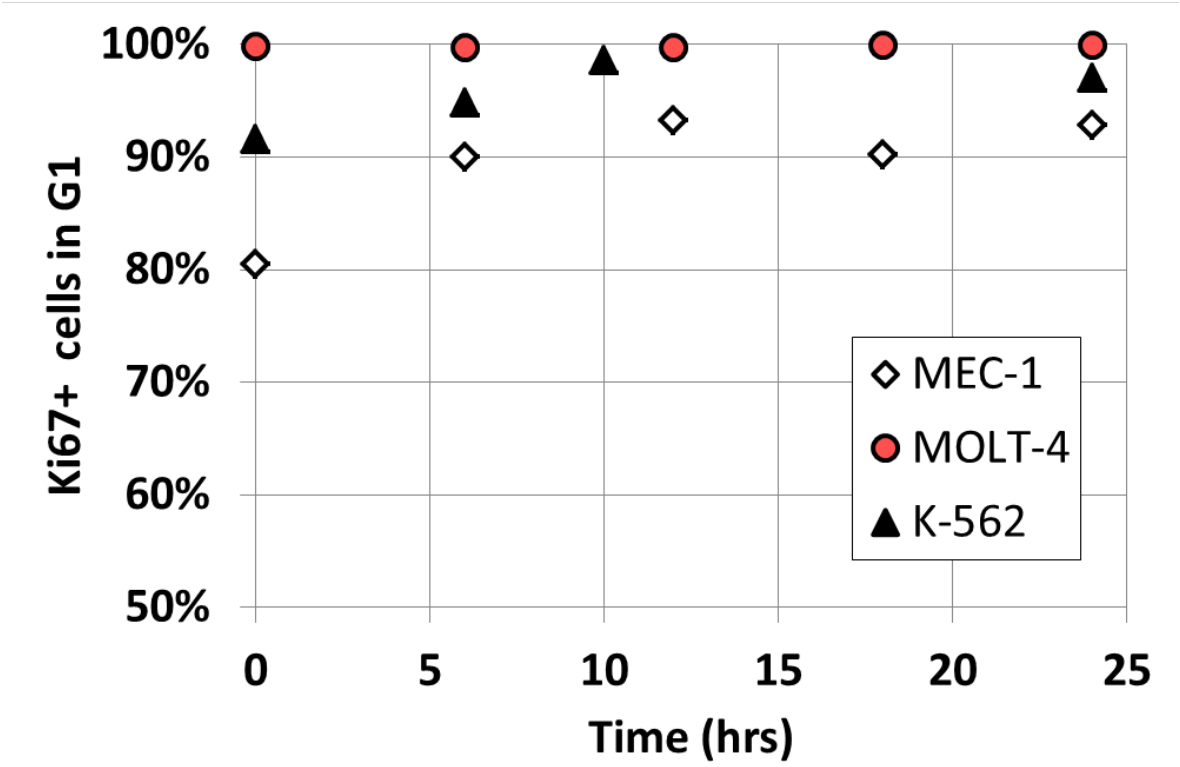


Figure 2: Evolution of the %Ki-67 positive cells among G0/G1 cells over time (control cells) as determined by flow cytometry. Ki-67 is a protein expressed by cells out of quiescence; G0 cells are thus identified by their lack of Ki-67 expression. In the experiments performed with EdU, control cultures were monitored for Ki-67 levels to validate the assumption that only a small percentage of the cells is quiescent at any time. Indeed, Ki-67 was expressed by at least 90% of the cells overall, with the exception of MEC-1 at time 0h which was found to be 80% (this is believed to be an effect of the washing steps stress and not an ubiquitous condition).

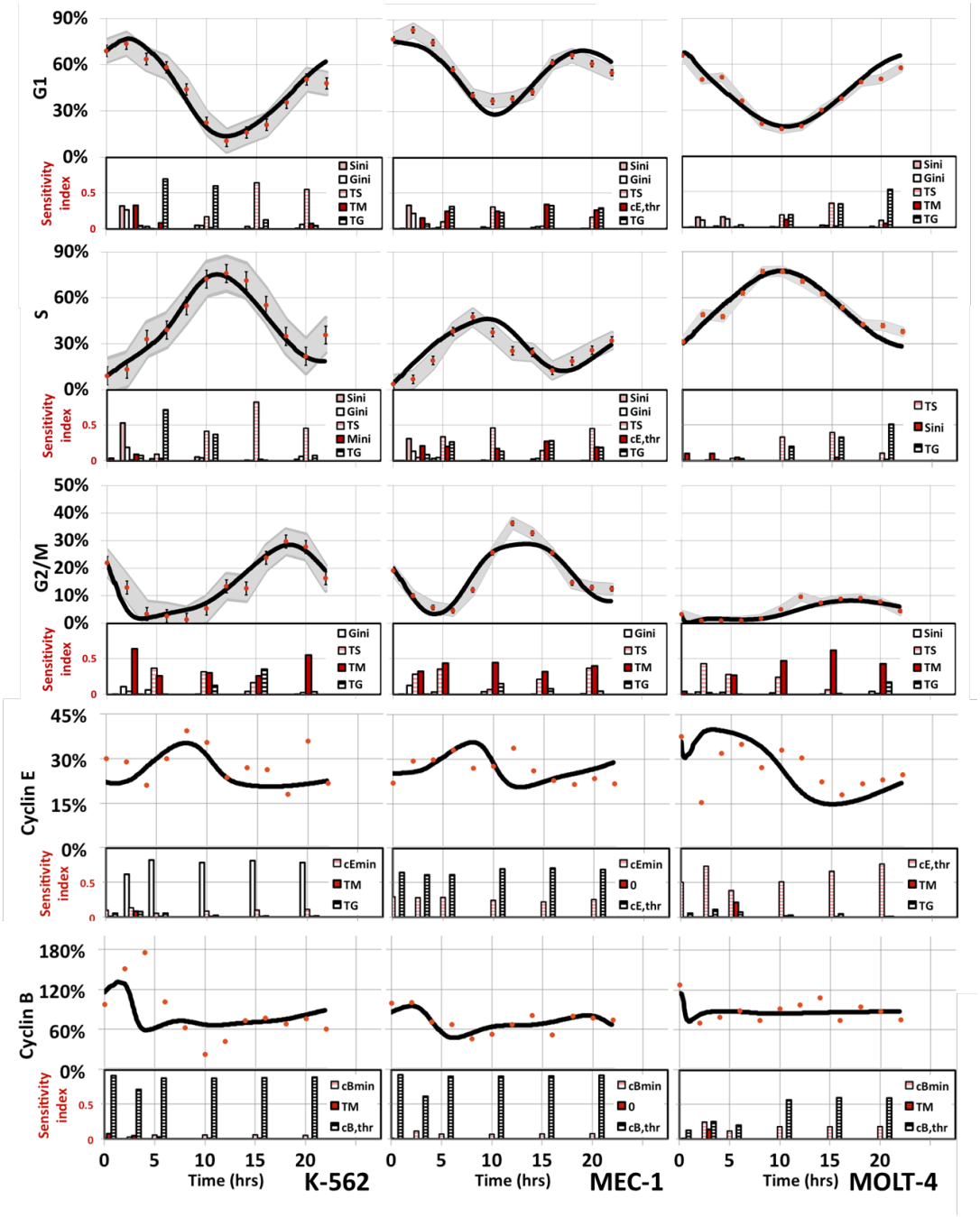


Figure 3: Cell cycle phase kinetics. Comparison of experimental results (red dots) to model output (black lines) in cell cycle phase percent (top 3 panels: G0/G1, S, G2/M phases) or normalized cyclin E & B expression (2 bottom panels) over time. 95% confidence areas are shown according to the standard deviation of G0/G1, S and G2/M (n=2/3). Sensitivity indexes for the most significant parameters (>0.1) on each output are shown beneath. For K-562, T_G is significant in G1 at the time cells are exiting the phase (5-10h), in S phase at the time they are entering the phase (5-10h), and in G2/M when cells are entering the phase (15h). Similarly, T_S is significant in G1 when cells are entering the phase again (15-20h), in S when cells are exiting the phase (10-20h) and in G2/M when cells are entering the phase (5-15h). T_M is only significant in G2/M throughout and in the first 2-3h for G1. For MEC-1, T_G is significant in G1 and S (to a lesser extent), T_S in S and G2/M (in G1 only at 10h); T_M in G2/M only. For MOLT-4, T_G is significant after 10h in both G1 and S; T_S in S (10h and 15), G1 (45h) and G2/M (0-10h); T_M in G2/M only from 5h. For cyclin E, $C_{E,\mathrm{thr}}$ is the most significant parameter, followed by $C_{E,\mathrm{min}}$. The same pattern is seen for cyclin B, with $C_{B,\mathrm{thr}}$ and $C_{B,\mathrm{min}}$ (to a lesser extent) appearing significant throughout. Initial conditions (G_{ini} and S_{ini}) only appear significant in the first few hours (0-3h) in K-562 and MEC-1.

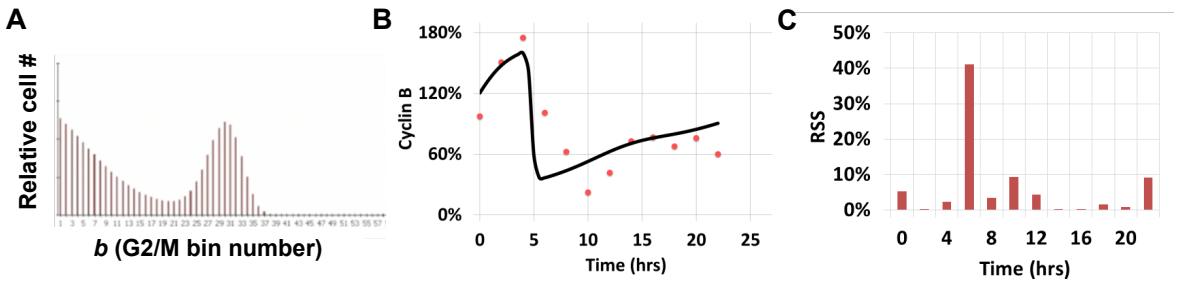


Figure 4: Recalculation of cyclin B trajectory when accounting for amended initial cell cycle distribution (EdU$^-$). (A) Entrance of displaced S phase cells at 2-4h into G2/M phase when a small, previously present population of G2/M reaches the bin where the threshold cyclin B concentration occurs. Effect due to experimental DNA deconvolution errors.
(B) Geometric mean of cyclin B evolution with corrected initial conditions (no cells in the second half of S phase at time zero). Observe how the match with the model becomes closer.
(C) Residual Sum of Squares (RSS) of the experimental and computational results of (B) over time. All but one data point fall within the experimental error region (10%).

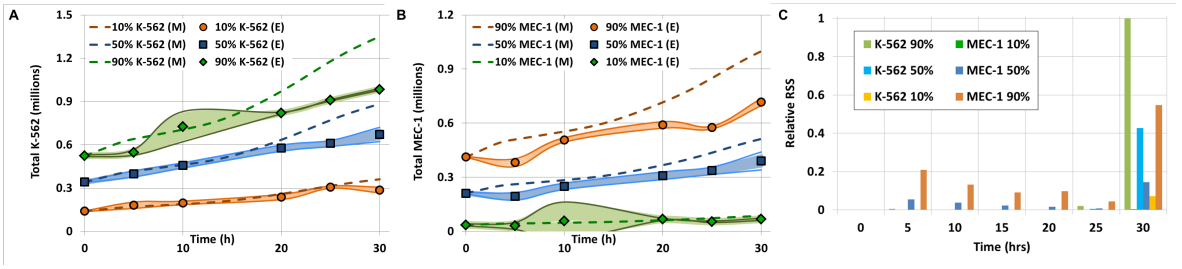
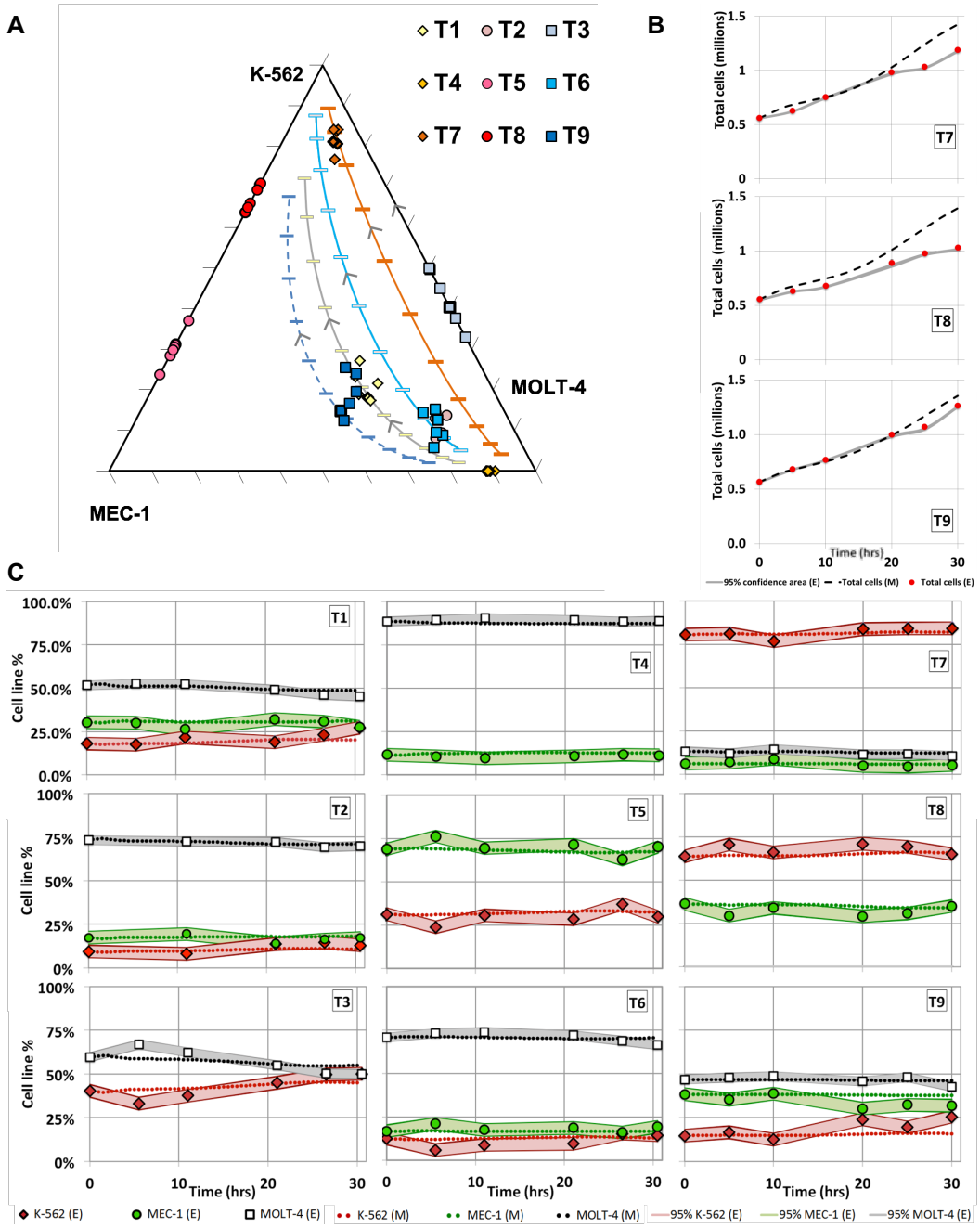


Figure 5: Live cell density over time. Comparison of experimental results (squares, diamonds, and circles) to modelling output (dashed lines) for
(A) K-562
(B) MEC-1
(C) Relative residual sum of squares (RSS)
95% confidence regions are shown according to experimental standard deviation (n=3). Simulations performed after adjustment of kinetic parameters in co-culture.



D

	K-562		MEC-1		MOLT-4	
	E	M	E	M	E	M
T7	80.9%	77.8%	7.3%	6.3%	11.8%	15.9%
T8	60.5%	59.8%	39.5%	40.2%	0.0%	0.0%
T9	23.5%	11.8%	35.8%	37.4%	40.7%	50.8%

Figure 6: Forward and backward culture evolution. (A) ternary plot of predicted vs experimental culture evolution: lines represent model simulation of culture evolution while points denote experimental values
(B) comparison of experimental total cell counts (red dots) vs model output (dashed lines) for T7, T8 and T9
(C) comparison of experimental cell line content (%) to model output given only the initial conditions in 9 different blind tests (T1-T9) over time
(D) backward prediction of population dynamics in T7, T8 and T9 cultures. E: experimental value; M: model prediction. Envelopes in (B) and (C) represent 95% confidence areas according to experimental standard deviation, n=3.

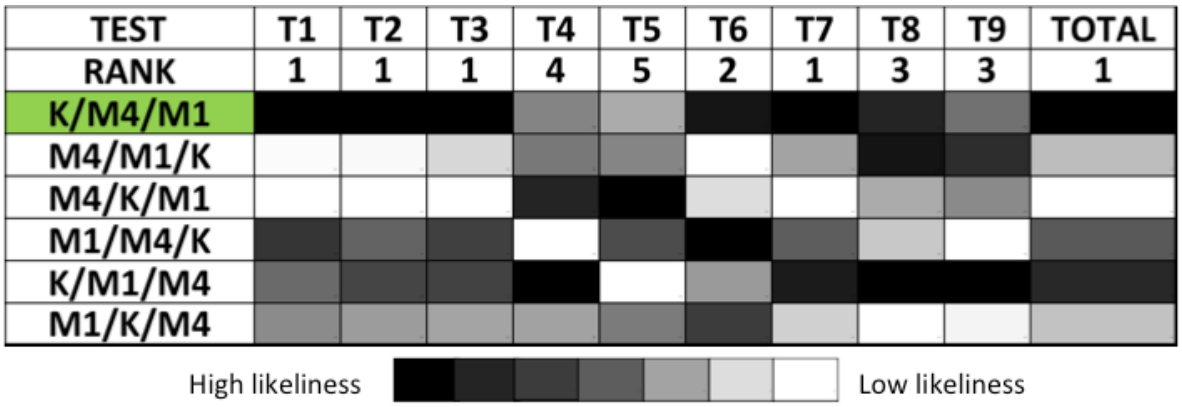


Figure 7: Ranking of scenarios likeliness according to Euclidian distance between experimental data and PBM output. K/M4/M1 is the right answer in all cases, as correctly guessed by the model. (K: K-562; M4: MOLT-4; M1: MEC-1).