Received: 9 January 2021 | Revised: 14 May 2021 | Accepted: 17 May 2021

DOI: 10.1002/cso2.1026

COMPUTATIONAL AND SYSTEMS ONCOLOGY Open Access

WILEY

ORIGINAL ARTICLE

Predicting time to relapse in acute myeloid leukemia through stochastic modeling of minimal residual disease based on clonality data

Khanh N. Dinh[1] | **Roman Jaksik**[2] | **Seth J. Corey**[3] | **Marek Kimmel**[2,4]

[1] Irving Institute of Cancer Dynamics, Columbia University, New York, New York, USA

[2] Department of Systems Biology and Engineering, Silesian University of Technology, Gliwice, Poland

[3] Departments of Pediatric Hematology/Oncology and Stem Cell Transplantation and Cancer Biology, Cleveland Clinic, Cleveland, Ohio, USA

[4] Department of Statistics, Rice University, Houston, Texas, USA

Correspondence
Marek Kimmel, Department of Systems Biology and Engineering, Silesian University of Technology, Gliwice, Poland.
Email: kimmel@rice.edu
Seth Corey, Departments of Pediatric Hematology/Oncology and Stem Cell Transplantation and Cancer Biology, Cleveland Clinic, Cleveland, OH, USA.
Email: coreys2@ccf.org

Khanh N. Dinh and Roman Jaksik are first authors.

Abstract

Event-free and overall survival remain poor for patients with acute myeloid leukemia. Chemoresistant clones contributing to relapse arise from minimal residual disease (MRD) or newly acquired mutations. However, the dynamics of clones comprising MRD is poorly understood. We developed a predictive stochastic model, based on a multitype age-dependent Markov branching process, to describe how random events in MRD contribute to the heterogeneity in treatment response. We employed training and validation sets of patients who underwent whole-genome sequencing and for whom mutant clone frequencies at diagnosis and relapse were available. The disease evolution and treatment outcome are subject to stochastic fluctuations. Estimates of malignant clone growth rates, obtained by model fitting, are consistent with published data. Using the estimates from the training set, we developed a function linking MRD and time of relapse with MRD inferred from the model fits to clone frequencies and other data. An independent validation set confirmed our model. In a third dataset, we fitted the model to data at diagnosis and remission and predicted the time to relapse. As a conclusion, given bone marrow genome at diagnosis and MRD at or past remission, the model can predict time to relapse and help guide treatment decisions to mitigate relapse.

KEYWORDS
acute myeloid leukemia, clonal evolution, minimal residual disease

1 | INTRODUCTION

Acute myeloid leukemia (AML) is the most common myeloid malignancy with over 21,000 cases diagnosed annually in the United States [1]. Rates of event-free and overall survival are poor. Despite stem cell transplantation and new drug approvals, chemoresistance and relapse remain major obstacles to survival for those with AML [2, 3]. Poor outcome factors include older age, complex cytogenetics, prior history of chemotherapy or myelodysplastic syndromes, and specific gene mutations [4]. Minimal residual disease (MRD) is emerging as an important predictor of relapse and decreased survival for individuals with AML [5, 6]. The genetic and immunophenotypic

STATEMENT OF SIGNIFICANCE

MRD remains the chief obstacle to curing AML. Stochastic, not deterministic, events produce small clones constituting MRD. Our model predicts time to relapse and mandates intervention, such as change in treatment.

diversity of AML makes testing for MRD challenging [7] and requires novel approaches.

MRD measurement must predict relapse better than existing risk stratification with thresholds of MRD established empirically. Different treatments and different leukemogenic drivers contribute to the empirical definition. The currently accepted definition for MRD in AML is 10^{-4} for multiparameter flow cytometry or 10^{-5} for qPCR. Estimated 5-year overall survival was 68% for patients without MRD versus 34% for those with MRD, for an average hazard ratio of 0.36 (95% credible interval = 0.33–0.39) [2].

Different mathematical models were proposed for the emergence relapse and treatment outcomes of AML and other hematologic malignancies [8–17]. Two types of modeling are commonly used: deterministic and stochastic [10]. Opinions are split as to whether the stochastic models of leukemogenesis and evolution of leukemia pre- and post-treatment are more relevant than deterministic ones [10, 18, 19]. Deterministic models do not account for the unpredictability of clonal heterogeneity in the genomic landscapes [9], although some assign different parameters to each individual [14]. Unpredictability of AML development may be due to randomness in the times at which driver mutations arise and from which aggressive cell clones originate [12, 20]. Hence, our specific objective is to characterize the range of treatment outcomes by taking into account that when leukemic cell population is reduced to very low numbers, survival of any particular clone is stochastic. Even with the same parameters of disease state such as blast percentages in bone marrow (BM) and peripheral blood, leukemic clonal percentages at diagnosis, and treatment responses such as chemotherapy-induced cell death rate (Table S1), a small residual population of AML cells may or may not regrow. As MRD is defined as the presence of leukemic cells detected by molecular assays, such as flow cytometry or PCR, below the sensitivity of morphologic measurement, it is of clinical interest to verify if its level is a predictor of the time to relapse (TTR). Stochastic modeling seems a relevant tool.

We present a stochastic model of proliferation and differentiation of hemopoietic stem cells (HSC) that predicts TTR following AML chemotherapy, informed by individual patient's leukemic BM clonal structure at diagnosis and relapse, and BM and peripheral blood cellularity, and blast fraction. Performance of the model vis-à-vis relapse data is measured by model-derived MRD–TTR characteristic, in independent training and validation datasets. The model demonstrates that the scenarios of effects of treatment depend to large extent on chance. Finally, a recent independent third dataset, in which BM clonal structure was observed at diagnosis, remission, and relapse, is used to predict the relapse time and BM clonal structure at relapse, based on observations at diagnosis and remission.

2 | MATERIALS AND METHODS

2.1 | Patient data and informed consent

We used de-identified patient data from Ding et al. [21] and Shlush et al. [22] for the training and validation study, respectively. The model's ability to predict TTP is tested using a third dataset from Ediriwickrema et al. [23]. Written informed consent was obtained for the approved protocols by the Washington University Medical School Institutional Review Board, the Research Ethics Board of the University of Toronto Health Network, and the Stanford Institutional Review Board of the Stanford Medical Center. Information about the patient data is detailed in Table S3 and Figure S1.

2.2 | Mathematical model

Our stochastic model of proliferation and competition of normal and leukemic cells is a multitype age-dependent Markov branching process [24]. Its structure is schematically depicted in Figure 1A,B. The model is designed so that its expected value trajectories are approximately identical to the deterministic model of [14]. Proliferation of cells in each clone occurs in linearly ordered compartments of mitotic and mature cells. This is an approximation of more complex hierarchies in normal and malignant BM [14]. Each event in cell's lifetime (division, differentiation, or death) is characterized by independent exponentially distributed waiting times. Because the waiting times are random, the order in which events occur differs among simulations. This results in variable outcomes under the same conditions and may contribute to interpatient heterogeneity. Identically as in [14], the model also involves growth regulation feedbacks, amounting to growth slowdown of both normal and malignant cells if BM becomes crowded. Parameter fitting for the stochastic model was performed by first approximating it with expected value ordinary differential equations (ODEs).

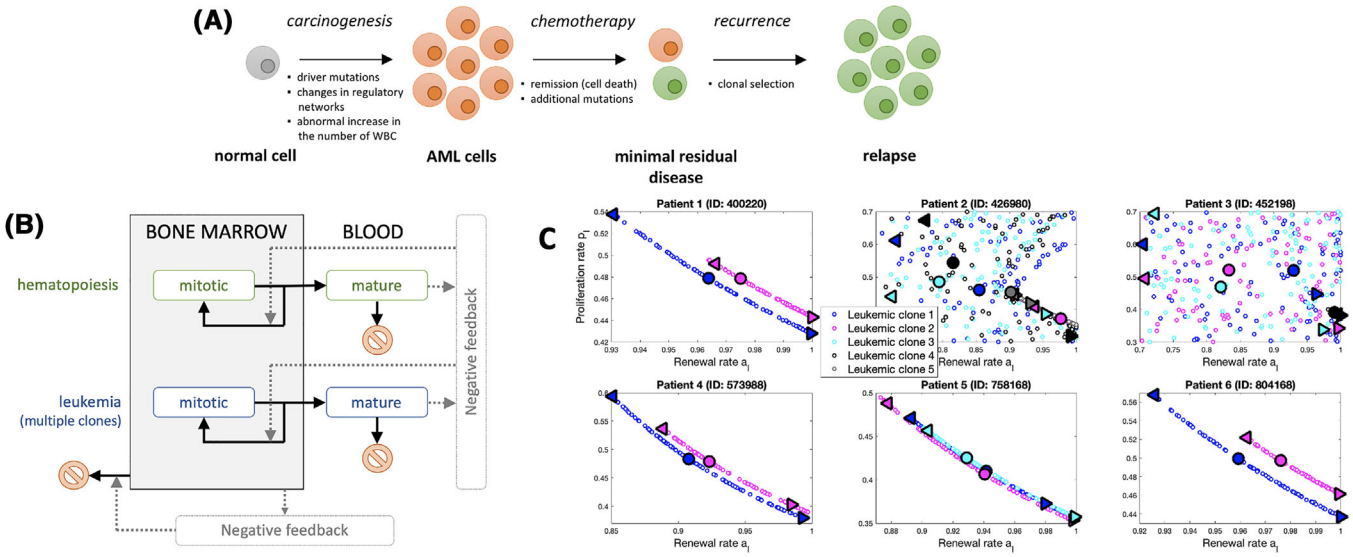

FIGURE 1 (A) Sequence of events in acute myeloid leukemia, including carcinogenesis, diagnosis followed by chemotherapy, and relapse. Leukemic clones arise and compete against each other and healthy hematopoietic clones. Our mathematical model starts at diagnosis and does not include carcinogenesis. (B) Schematic of the model. Normal and malignant myeloid cells are dividing and maturing, as described by a stochastic population process (see Supporting Information Data). Clones of malignant cells are considered separate subpopulations. Negative feedback controlling the total myeloid cell population in blood downregulates all clones' self-renewal rates if blood is overpopulated. Another negative feedback controlling the total myeloid population in bone marrow upregulates the death rates of all compartments in bone marrow if it is overpopulated. Feedback configuration assumed follows closely [14]. (C) Parameter estimates for all patients in the training set. One hundred parameter sets were found for each patient starting from randomly selected initial conditions of the fitting procedure. Parameter sets corresponding to clones present at both diagnosis and relapse form curves in the parameter space, while those corresponding to clones absent at relapse display a more complicated pattern. For each patient, three parameter sets were chosen for stochastic simulations, corresponding to low, medium, and high renewal rates (left triangles, circles, and right triangles)

2.3 | Stochastic model for proliferation and competition of normal and leukemic cells

Proliferation of cells in each clone is represented by an ordered sequence of different compartments; Figure 1B. The two-compartment model for the hematopoietic clones that we adopted was established in [25]. The healthy clone is divided into mitotic compartment $c_1(t)$ and mature compartment $c_2(t)$. The mitotic cell compartment, representing the more complex multistage differentiation process of hematopoietic stem cells (HSCs), hematopoietic progenitor cells (HPCs), and precursor cells, is located in the BM. Mitotic cells can divide into two daughter cells at the proliferation rate p^c, and each of the daughter cells is either a new mitotic cell or a mature cell. The fraction of daughter cells returning to the mitotic cell compartment is called the self-renewal rate a^c. The mature cell compartment consisting of neutrophil granulocytes, a major subtype of white blood cells, is located in the blood. Mature cells die at a constant rate d^c. Each leukemic clone, as detected by sequencing data, also consists of two compartments: mitotic population $l_1(t)$ in BM, and mature population $l_2(t)$ in blood. The rules determining its divisions, differentiations, and deaths are similar as in the hematopoietic clone, with proliferation rate p^l, renewal rate a^l, and death rate d^l. There are two feedback systems governing the populations in blood and BM. The first feedback system reacts to overpopulating in the blood by downregulating the self-renewal rates of the hematopoietic and all leukemic clones by a factor of

$$s(t) = \frac{1}{1 + k \cdot (c_2(t) + l_3(t))}. \quad (1)$$

The second feedback system controls the total population in the BM:

$$x(t) = c_1(t) + l_1(t) + l_2(t) \quad (2)$$

and if this population is too high, the death rates of all compartments in bone are increased by

$$\mathrm{d}(x(t)) = A_1 \cdot \max\left(0, x(t) - A_2 \cdot \overline{c_1}\right). \quad (3)$$

Finally, treatment drugs used in many chemotherapy protocols are characterized by increased killing of cells in the DNA synthesis phase of the cell cycle. Therefore, during treatment, the death rates of mitotic cells are increased

by factors proportional to their proliferation rates:

$$d^{c} = \ldots + k_{c} \cdot p^{c}, \tag{4}$$

$$d^{l} = \ldots + k_{l} \cdot p^{l}. \tag{5}$$

If any clonal mitotic population decreases below one cell during the time course, the clone is marked as dead and remains 0 until relapse. We assume that chemotherapy kills leukemic clones at a higher rate than the normal clone, and therefore $k_l > k_c$. Furthermore, we simplify the model by assuming a fixed ratio between the killing rates for all patients. The ratio $k_l : k_c = 5 : 1$ was found to result in realistic behavior of the disease trajectories. The additional death rates during treatments are therefore $k_l = 5 \cdot \alpha$ and $k_c = \alpha$, where α is the strength of chemotherapy. There are two requirements for α: (a) the total leukemic population must be repressed to <5% of total cell population in BM, as all patients in the datasets achieve complete remission; and (b) the leukemic population cannot be depleted, as the patients eventually relapse. We chose the parameter α for each patient before parameter fitting by sampling several values and tested if proliferation rates and self-renewal rates can be fitted and satisfy the two conditions within a realistic timeframe.

Further details of the model are given in the online Supporting Information Methods, in the *Expected value approximation for fitting the stochastic model* section, and subsequent sections.

2.4 | Hill function depicting dependence of TTR on MRD

The MRD–TTR dependence summarizes predictions of our model. Hill function can fit a wide spectrum of "sigmoidal" (switching from concave to convex) or "non-sigmoidal" (convex) functions tending to 0 at infinite time and is among the most parsimonious functions of this type. We used the shifted form $\mathrm{TTR} = A/(B + (\log(\mathrm{MRD}) - C)^{n})$ with constants A, B, C, and n estimated from data. The MRD–TTR Hill function that is used to predict TTR based on MRD level is estimated based on the *training dataset* from Ding et al. [21]. We then used the *validation dataset* from Shlush et al. [22] to test the prediction, by simulating the MRD (not included in the data) and used it to see how the TTR fits the Hill curve. Finally, we used the *dataset* from Ediriwickrema et al. [23], which includes both clinically measured MRD and observed TTR, to compare these to the Hill function developed in the training step.

Further details of methods used are found in the online Supporting Information Methods.

2.5 | Code availability

The codes used in this study, written in Fortran and Matlab, are available at https://github.com/dinhngockhanh/stochasticMRDinAML.

3 | RESULTS

We consider the time interval between diagnosis and initial relapse that includes cytotoxic chemotherapy, chemotherapy-induced myelosuppression and decrease in leukemic cells, nonleukemic marrow recovery, and growth of the leukemic clones due to refractory or relapsed AML. Figure 1A depicts a simplified sequence of events.

In this work, we assumed that no additional clones arise by mutation after diagnosis, as was suggested in [6, 26, 27]. It does not preclude new mutations arising in these clones, but given that clones were identified based on bulk sequencing and variant clustering in the original publications, this assumption seems justifiable as a first-order approximation. Underlying our model are assumptions on growth, differentiation, and competition of normal and leukemic clones [14], as depicted in Figure 1B and detailed in *Methods*.

To train the model, we employed clonal percentages at diagnosis and relapse from mutational data of six patients recorded in The Cancer Genome Atlas (TCGA) and the Genotypes and Phenotypes (dbGaP) databases [21]. Paucity of serial, paired sequencing data is a serious limitation. In about 200 TCGA cases of AML sequenced at diagnosis, 20 patients were also sequenced at relapse, and only six had BM and peripheral blood cell counts. To supplement these numbers, we included a validation set, a subset of five patients from Shlush et al. [22]. We also used the model to predict TTR based on cases from a third dataset [23].

3.1 | Mathematical model of heterogeneity in clonal evolution

We developed a stochastic model of evolution of leukemic clones and the nonneoplastic hematopoietic clones. Parameters were estimated by fitting the expected value model to the patient's clinical data (*Methods*). Because the fits are not unique, three sets of estimated parameters for stochastic simulations were acquired from the list of expected value fits, with low, average, and high renewal rates (Parameter Sets 1, 2, and 3, respectively). Supporting Information Methods, Tables S1 and S2, and section *Inputs of the model: Clinical data and clonal landscapes at diagnosis and relapse*, contain further information.

3.2 | Estimation of coefficients

For each patient in the training set, 100 parameter sets were found using parameter fitting as presented in Figure 1C. There were several observations. For the clones at relapse (Figure S1), the parameter pairs (a^1, p^1), probability of self-renewal (as opposed to maturation), and the proliferation rate formed curves in the parameter space, as in [28]. A given clone curve was above others if the clone was more competitive (higher renewal probabilities or higher proliferation rates; see Patients 400220, 573988, 758168, and 804168 in Figure 1C).

For the clones that, in the model, became extinct at relapse, there was no pattern to the pairs (a^1, p^1) (Patients 426980 and 452198 in Figure 1C). The pairs well below the curves corresponded to Scenario 1, in which the clones died because they could not compete against the other leukemic clones or the normal clone. The pairs either above the curves or slightly below belonged to Scenario 2, in which the competitive clones died during chemotherapy because of high proliferation rates.

Figure 2 shows single expected trajectory for Patient 3 corresponding to a particular parameter set. For the same patient, different parameter sets resulted in different trajectories, however, with the same clonal percentages at diagnosis and relapse. Under this parameter set, the model predicted that leukemic Clone 1 was eliminated during treatment, Clone 2 escaped treatment but grew very slowly and was undetectable at relapse, Clone 3 escaped treatment but subsequently became outcompeted by other clones, and Clone 4 grew steadily from a low level after remission and was the only leukemic clone detected at relapse. Figure S2A–E displays the expected trajectories for other cases in the training set.

We performed an identical analysis for patients in the validation set. The results of parameter estimation based on validation cases are in agreement with those based on the training set (Figure S4A).

3.3 | Stochastic simulations

For each patient in the training set, three parameter sets are chosen to represent the range of outcomes under different assumptions of mitotic cell behavior. Figure 3A shows the expected value and stochastic outcomes for Patient 5. Under Parameter Set 1, leukemic clones had high proliferation rates and low renewal rates and were highly affected by chemotherapy as mitotic cell death was assumed proportional to proliferation rate. In the expected value model, the disease persisted with a very small population at remission, which eventually gave rise to relapse. Because the simulated leukemic population was small at remission, random fluctuations led to eradication of one or more leukemic clones after treatment. For Parameter Set 3, with low proliferation rates and high renewal rates, chemotherapy had a smaller impact on the leukemic populations. The disease was more frequent at remission and increased in size until relapse. Parameter Set 2 presented a middle ground between Parameter Sets 1 and 3; the disease was almost eradicated by chemotherapy, as in Parameter Set 1, but remission contained a larger leukemic population and therefore 64.1% of stochastic simulations led to the same outcome as the expected value simulation.

Figure 3B shows the expected and stochastic outcomes for Patient 2. The expectations showed that only Clones 2 and 5 were present at relapse, and 42.4%, 92.3%, and 0% of the stochastic simulations under Parameter Sets 1, 2, and 3 led to this outcome, respectively. Under Parameter Set 1, all other leukemic clones were eradicated by chemotherapy. However, random fluctuations also might lead to either Clone 2 or 5 being eliminated even after treatment. The stochastic simulations under Parameter Set 3 followed a different route. No leukemic clone was entirely eradicated by the time of complete remission, but Clones 1 and 4 were gradually outcompeted. Furthermore, Clone 3 was present at relapse in all stochastic simulations, but at population sizes smaller than the detection level. Similarly as for Parameter Set 1, Clones 2 and 5 could each be erased due to stochastic fluctuations. Figure S3A–D displays the stochastic outcomes for the remaining four cases in our training set.

Stochastic results for the validation set are shown in Figure S4B–H. Similarly to the training set, different parameter sets for the same patients may lead to significant changes in the potential paths to relapse and their probabilities. For instance, under clonal tree 2 of Patient 2 (Figure S4C), Parameter Set 1 resulted in a wide range of possible relapse outcomes, the most probable of which was only of frequency 16.9%. However, under Parameter Set 3 in the same patient, all stochastic simulations led to the same clonal composition observed in the patient's data. Overall, our model showed that stochastic fluctuations played an important role in deciding which leukemic clones survived chemotherapy and progressed to relapse.

3.4 | Comparison to independent estimates of MRD, clonal growth rates, and WBC

Ivey et al. [5] studied the dynamics of mutant levels in BM and peripheral blood at several time points following remission to detect MRD as early as possible. All patients had leukemic cells harboring mutated *NPM1*, one of the most commonly affected genes in AML [29]. Under the

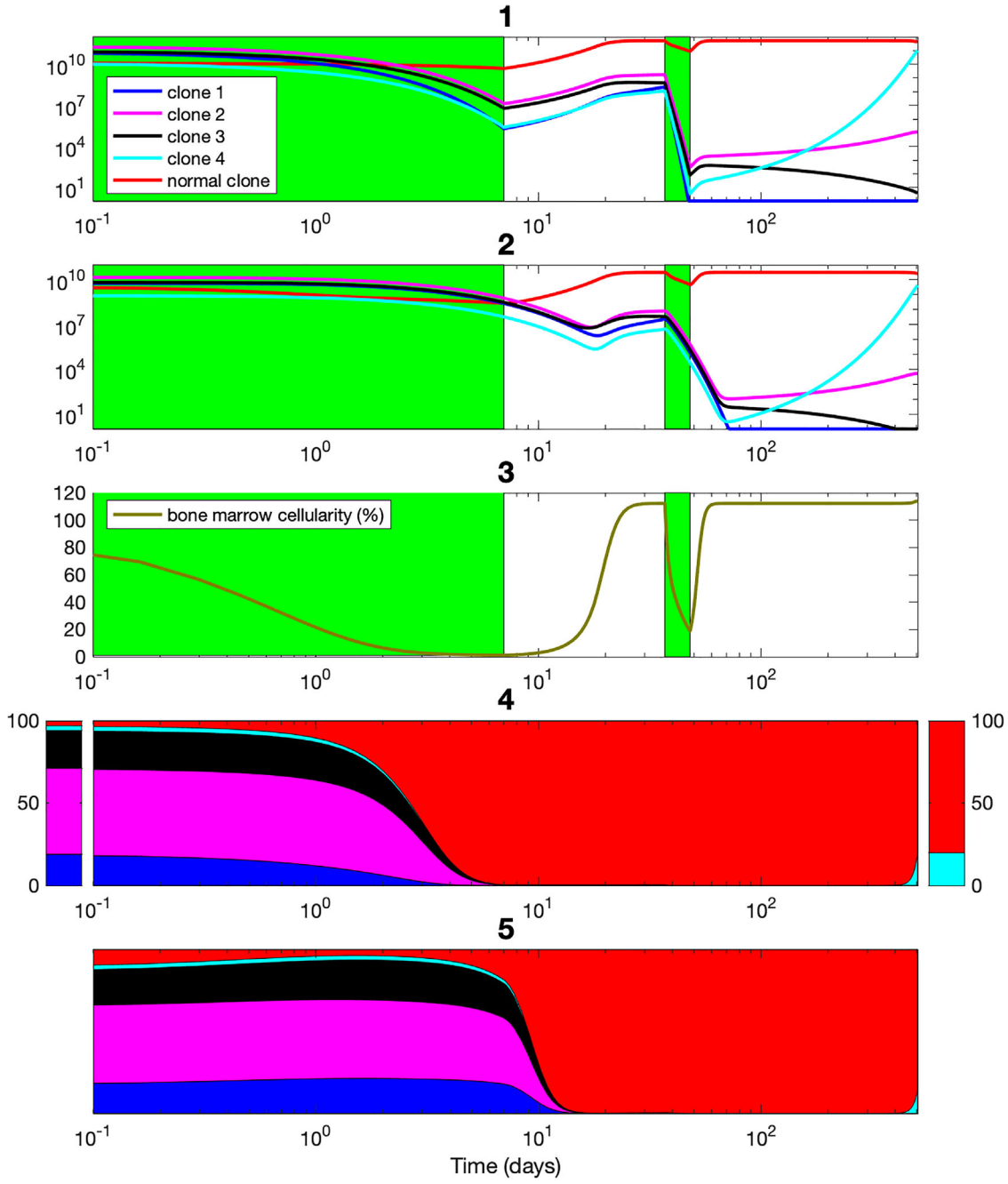

FIGURE 2 Results of the fitting procedure for Patient 3 (ID: 452198). Example of fitting the expected trajectories using a single parameter set (time in logarithmic scale). (A) Evolution of mitotic populations in the bone marrow of all clones, with cell counts in logarithmic scale. (B) Evolution of the mature populations in peripheral blood. Green bars indicate chemotherapy. (C) Evolution of bone marrow cellularity. (D) Evolution of clonal percentages in bone marrow. The bar chart on the left depicts clone percentages at diagnosis, and the one on the right depicts clone percentages at relapse (both based on sequencing data). The parameter sets are chosen to fit the clonality data in these two bar charts, as shown in the middle chart. Color codes for clones in (B) and (D) are the same as in (A), as described in the inset. (E) Evolution of clonal percentages in peripheral blood

assumption that the regrowth of the malignant clone was exponential, the resulting growth rates varied from 0.3 to 2.0 $\log_{10}$ per month, as reproduced in Figure 4A. We compared the growth rates of the clones in the TCGA AML patients to verify if they fitted into this range (assuming clones with different mutations had similar growth rates to the *NPM1*-mutant clones). Only the clones still present in the patients' clinical data at relapse were considered. Their growth rates were estimated based on the parameters resulting from model fitting (see Supporting Information Methods section *Minimal residual disease*). When compared to [5], the growth rates for all patients in our

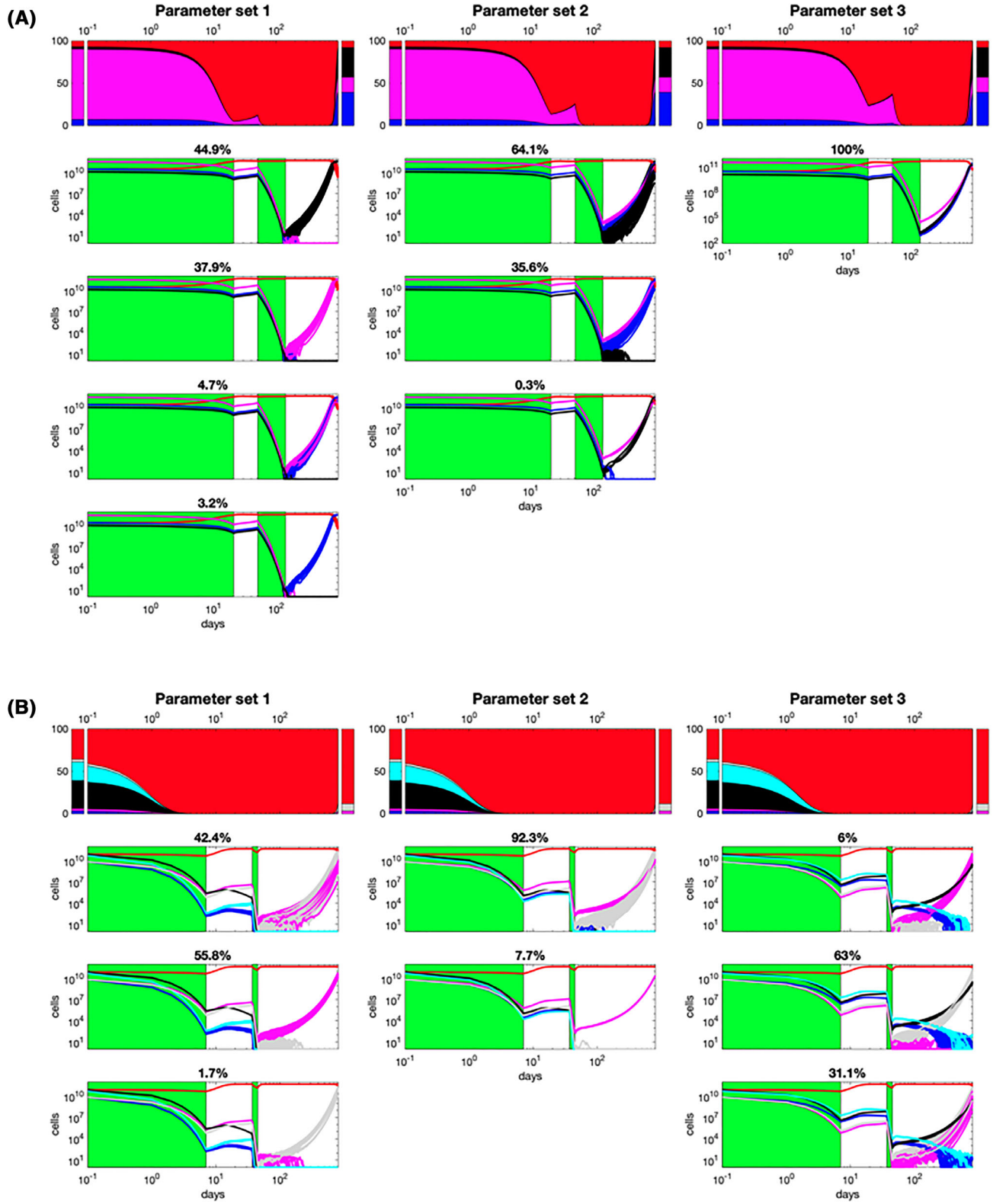

FIGURE 3 (A) Results of stochastic modeling of Patient 5 (ID: 758168). Columns correspond to different parameter sets. Row 1: Evolution of clonal percentages in the expected value model. Rows 2 and on: Outcomes of the stochastic model, with cell counts in logarithmic scale, with corresponding frequencies of occurrence listed. Stochastic simulations agree with the expected value results in Parameter Set 3. In Parameter Set 1, no simulation leads to the same outcome as the expected value simulation, where all three leukemic clones detected at diagnosis are present at relapse. Agreement of stochastic and expected value outcomes is 64.1% in Parameter Set 2 and 100% in Parameter Set 3. (B) Results of the stochastic model for Patient 2 (ID: 426980). Columns correspond to parameter sets. Row 1: Evolution of clonal percentages in the expected value model. Rows 2 and on: Outcomes of the stochastic model, in logarithmic scale with frequencies of occurrence listed

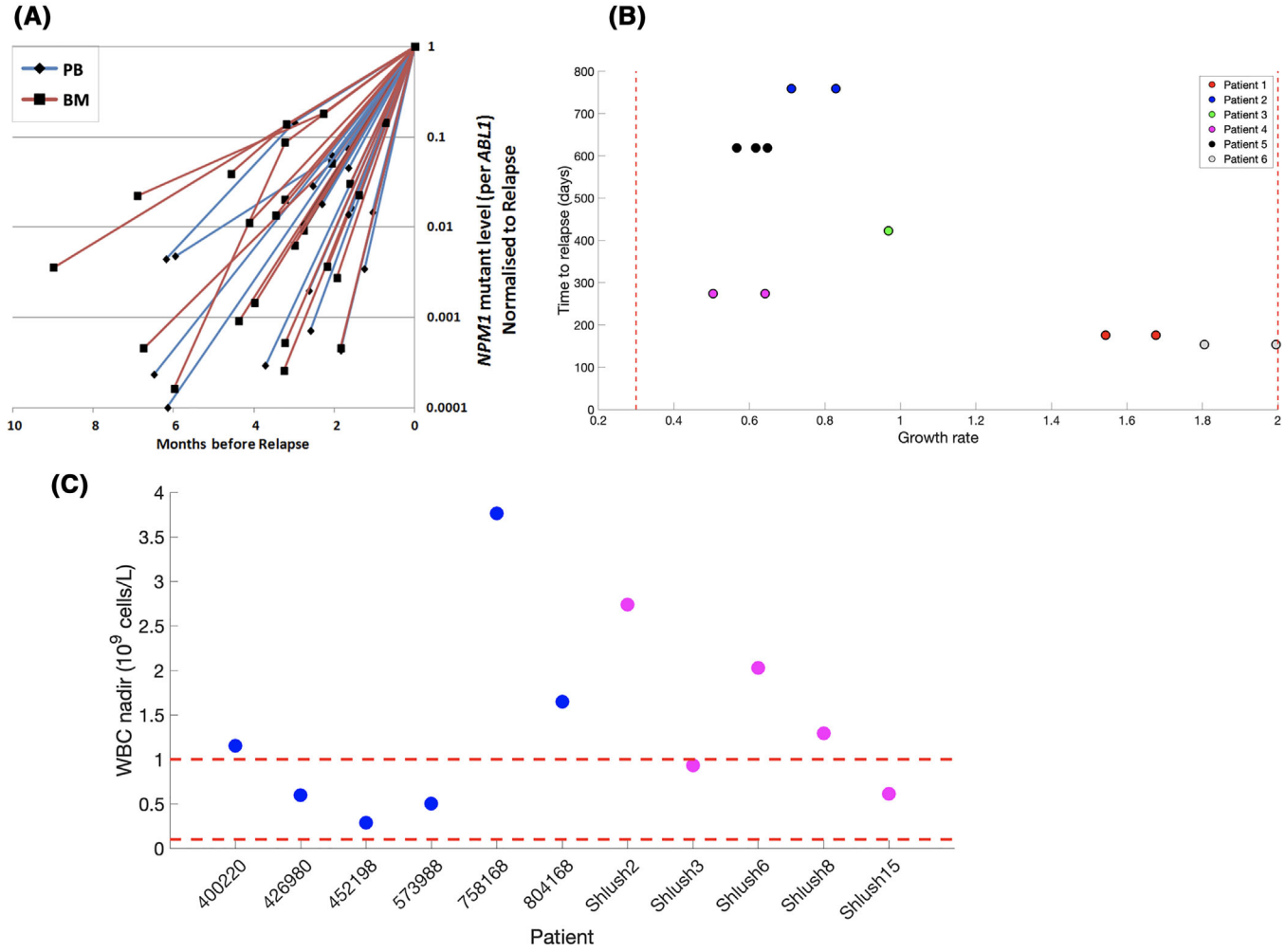

FIGURE 4 (A) Kinetics of relapse of NPM1-mutated AML (reproduced from [5] with permission). Based on sequential monitoring of samples obtained in human patients after the end of chemotherapy until molecular or hematological relapse. BM, bone marrow; PB, peripheral blood. (B) Inferred growth rates of different clones in training and validation patients. Only the clones that still exist at relapse in the patients' data were considered. The red lines are bounds for growth rates from 0.3 to 2.0 log10 per month, based on the range of growth rates of NPM1 mutant clones in [5] (see Fig. 7). Growth rates computed from our model (circles) fit into this range. (C) Nadir values of the WBC count during chemotherapy. The WBC nadirs for patients in the training and validation dataset were computed using the expected value fits. WBC computations assumed that an average patient has 5 L of blood. Blue dots: training dataset. Purple dots: validation dataset. Red lines: range of observed WBC nadirs of 14 patients in [22]

analysis fit in the range of experimentally observed data (Figure 4B).

3.5 | Predicting relapse from MRD

We next examined the model's ability to predict the chance of relapse, TTR, and the clonal landscape at relapse. We employed the data of four patients in [23], which contained clonal cell counts at diagnosis, remission, and relapse from single-cell DNA sequencing. Relapse occurred in all four cases.

The data at diagnosis and remission were used to fit the model, then the resulting stochastic trajectories were extrapolated to predict relapse. Figure 5A showed the expected value and stochastic outcomes for Patient SU067. The model accurately predicted that relapse would occur. Moreover, it also predicted that relapse would be caused by the NPM1/WT1 clone alone, in agreement with patient data. The three parameter sets differed slightly in the rate to relapse. Parameter Set 3 (low proliferation rate and high renewal rate) appeared closest in blast percentage to the clinical data.

Figure 5B shows the results for Patient SU291. The model predicted that relapse would not occur, contrary to the clinical data. This was because sequencing at two remissions (days 26 and 359) detected no blasts in BM, although the clonal landscape at relapse consisted of both

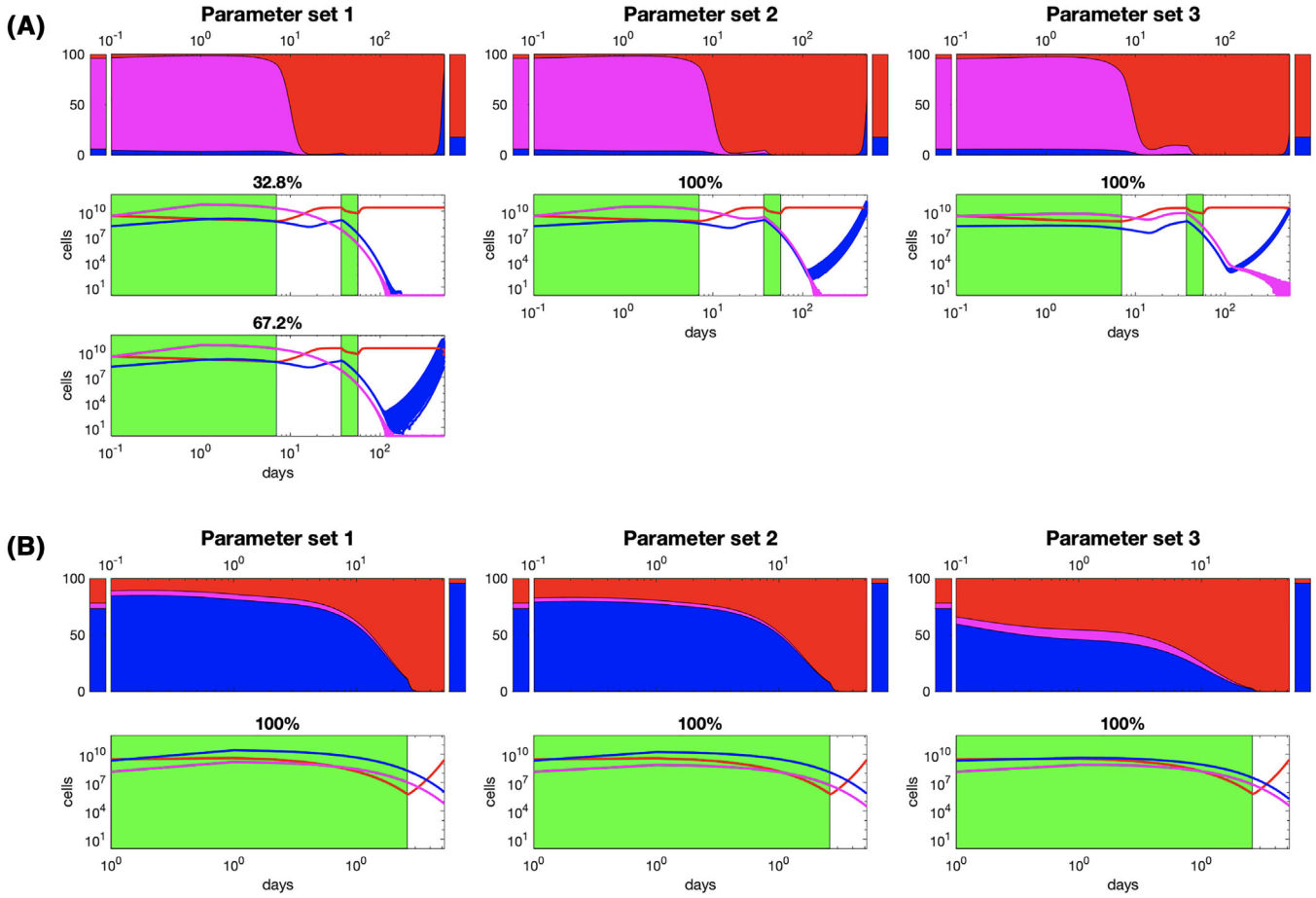

FIGURE 5 (A) Results of stochastic modeling of Patient SU067. Columns correspond to different parameter sets. Row 1: Evolution of clonal percentages in the expected value model. The model is fitted to the clonal data at diagnosis (left bar) and at remission (MRD; end of treatment). The predictions for the clonal composition at relapse are then compared against actual measurements (right bar). Rows 2 and on: Outcomes of the stochastic model, with cell counts in logarithmic scale, with corresponding frequencies of occurrence listed. Stochastic simulations can predict relapse from MRD data. (B) Results of stochastic modeling of Patient SU291. Stochastic simulations predict that relapse will not occur, contrary to real data. This is because the single-cell sequencing at remission did not detect MRD

clones detected at diagnosis (NPM1 and NPM1/IDH2) and a new clone (NPM1/IDH1/FLT3D835Y). This suggests that there was MRD after treatment, undetected by single-cell sequencing of around 1000 cells.

The results for Patients 320 and 372 are shown in Figure S5. The model predicted relapse in Parameter Set 1 for Patient 372, in agreement with clinical data. It also predicted the correct clone to drive relapse, and the blast percentage at relapse was close to the clinical observations. Parameter Sets 2 and 3 predicted relapse, which implies that the leukemic clones in this case were on the aggressive side of the parameter spectrum. For Patient 320, the model predicted relapse in all parameter sets. However, it predicted the wrong clone to relapse. All clones detected at relapse in this patient were already found at diagnosis but not at remission, once again suggesting that the MRD was below the sensitivity level of single-cell sequencing.

3.6 | Relationship between MRD and TTR

We next studied the connection between MRD (the leukemic clonal percentage in BM at 4 months after remission) and TTR (the period from remission to the time blasts exceed 5% of BM) inferred from the stochastic model as explained earlier on. A sigmoidal Hill function [30] was fitted to the predictions. Since patient 5 (ID: 758168) was the only patient with a cytogenetic abnormality, another sigmoidal Hill function was fitted excluding this patient to study sensitivity of the relationship between MRD and TTR to the change (Figure 6A). Data from our validation study were superimposed for comparison (Figure 6B). We observed that including or excluding patient 5 did not significantly change the sigmoidal Hill fit. Finally, we included the data of eight relapsed AML cases from [23],

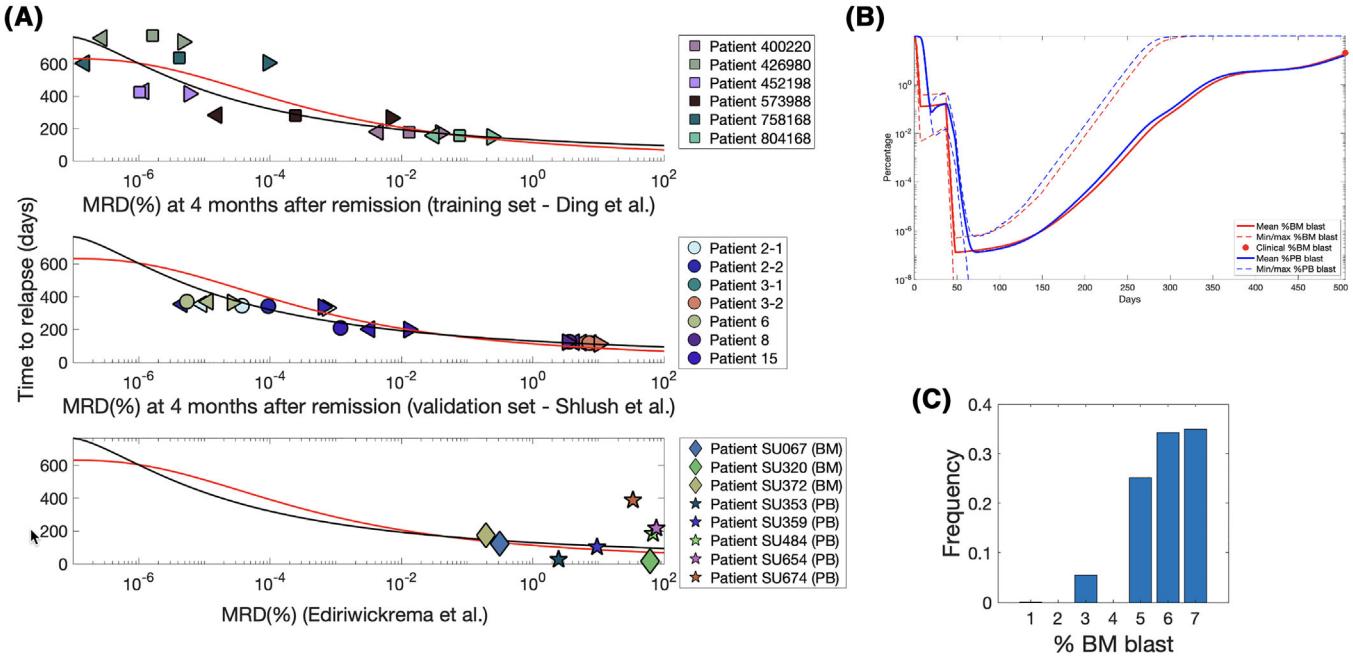

FIGURE 6 (A) Relationship between MRD and TTR (time to relapse) for all patients. Estimates of MRD in BM and TTR are mean values from 1000 stochastic simulations (Top and Middle) or are clinically measured (Bottom). Colors identify patients; left triangles correspond to Parameter Set 1, circles, squares, diamonds or stars correspond to Parameter Set 1, and right triangles correspond to Parameter Set 3. Top: A sigmoidal Hill function of the form $\text{TTR} = A/(B + (\log(\text{MRD}) - C)^n)$ is fitted to the training dataset [21], with (red line) or without (black line) patient 5 (ID: 758168). Middle: MRD and TTR estimates from [22] agree with this MRD–TTR relationship, even though these data points were not fitted. Bottom: Clinical measurements of MRD in BM or PB, and observations of TTR from [23] are compared against the sigmoidal Hill function. The data points with MRD in BM largely agree with the predictions, whereas the data points with MRD in PB do not. Parameters for the Hill fit with patient 5: $A = 14036.3$, $B = 22.2$, $C = -7$, $n = 2.38$. Parameters for the Hill fit without patient 5: $A = 2836.6$, $B = 3.7$, $C = -7$, $n = 1.49$. (B and C) Analysis of MRD for Patient 3 (ID: 452198). (B) Evolution of %BM blast (red) and %PB blast (blue) from diagnosis to relapse. Solid lines: mean values; broken lines: minimum/maximum values. The mean %BM blast and the mean %PB blast are largely close from complete remission to relapse. The minimum %BM blast and %PB blast indicate that some simulations lead to the disease being eradicated. (C) Frequency graph of %BM blast when %PB blast reaches 5%

where the MRD was clinically examined in either BM (three cases) or PB (five cases).

MRD at 4 months was typically smallest in Parameter Set 1 and largest in Parameter Set 3. This was in agreement with our previous observation from the stochastic model; Parameter Set 1 showed that leukemic clones were affected the most by chemotherapy and therefore the disease level at remission was very small (some clonal populations decreased to $\sim 10^1$–10^2 cells; see Figures 1–3). Comparing all patients, it is clear that larger MRD is associated with shorter TTR and smaller MRD is associated with higher variance in relapse time.

Even though the sigmoidal Hill functions were only fitted to data from the six cases in our model training set, they accurately predicted the TTR for the five patients in the validation study. The fit was somewhat better if the Hill function fitted without Patient 5 in the training set was used. A possible reason was that no patients in the validation study had any cytogenetic abnormality. Comparing the results from two different clonal trees for Patients 2 and 3 revealed slight changes in MRD level and virtually no difference in TTR.

The data from [23] included the clinically observed MRD and TTR. The data appeared to agree with the sigmoidal Hill functions if the MRD came from BM, whereas the data with MRD from PB did not. One possible reason was that our model was trained for MRD in BM. On the other hand, there was no obvious pattern to the MRD–TTR in PB, suggesting that MRD in PB is a poor predictor of TTR.

We also carried out a comparative analysis of MRD in BM and peripheral blood, which led to results consistent with the usual definition of the TTR (Supporting Information Methods section *Comparison to independent estimates of the MRD, clonal growth rates, and WBC*; also see Figure 6C). In addition, we also examined data from the two patients from [21] who underwent autologous BM transplant. *Autologous transplant patients* section and Figure S6 in Supporting Information Methods online contain methodology details. These patients' simulated MRD and observed TTR fit the Hill function relationship too.

See *Conclusions from the statistical analysis* in the Supporting Information Methods for calculations supporting validity of the Hill function approximation.

4 | DISCUSSION

Identification of MRD as an independent prognostic factor, stratification of patients by their status, and early intervention with change in therapy improved outcomes in acute lymphoblastic leukemia [6, 22, 31–35]. A challenge remains how to define MRD in AML and use that information to intervene and improve outcome [7]. The threshold of what constitutes MRD in AML is not well established for cytogenetics, or leukemia-associated immunophenotype, or mutational allelic frequency [34]. Half of AML patients with MRD absent relapse [2]. Each of these assays has a different level of sensitivity. Currently accepted definition of MRD at 0.01% is based on flow cytometry leukemia-associated immunophenotypes. Different subtypes of AML, for example, *NPM1* or *FLT3* mutations, may have different MRD dynamics. Relapse can be associated with loss of founder mutations [36]. Flow cytometry false-negative results occur at almost 19% [6]. Nonetheless, a growing number of clinical trials document the prognostic value of MRD in AML [34].

To estimate our model parameters, we included treatment protocol, blood counts, and BM biopsy results from which the clonal landscapes at diagnosis and relapse were determined. The model was calibrated to fit the clonal frequencies, while satisfying the biological requirements as observed in the data. We first fitted the model based on a *training set* of six cases from Ding et al. [21] and then *validated* the finding using an independent set of five cases from Shlush et al. [22]. Three findings emerge:

1. Depending on the number of malignant cell clones and other parameters, all estimated based on clinical and genomic data, the evolution of disease and the outcome of treatment were subject to stochastic fluctuations that might radically alter the course of disease.
2. The estimates of malignant clone growth rates, obtained by fitting model to data, were consistent with those by Ivey et al. [5] based on a study of dependence of MRD and TTR.
3. Using estimates from the training set, we developed a function linking MRD and TTR with MRD unobserved, but inferred from the model fits to the data. The relationship was consistent with MRD estimates based on the validation set. Further, this characteristic predicted the TTR given MRD in patients from a third dataset [23].

An important issue is whether the clone growth rates estimated using the model are systematically different for patients with different AML subtypes. In our data, the single APML Patient 758168 had the longest TTR. However, as depicted in Figure 4, most rates identified in patients in the training and validation study fall within the NPM1 range from [5]. Mutations in genes such as FLT3-ITD and DNMT3A, are associated in the literature with increased risk of relapse. However, in the pooled dataset of 15 cases, the average times to relapse of cases with and without the high-risk mutations FLT3-ITD or DNMT3A equaled 274 and 247 days (medians 278 and 176 days, respectively). The difference in means is not significant and the trend in medians is opposite to that expected by the differential-risk hypothesis.

At least one publication described clonal heterogeneity by modifying the ODE model to accommodate new leukemic clones arising from mutations [35]. We did not include mutations in our model because relapse more likely results from MRD than new clones during chemotherapy [22, 37]. The ODE models were used to track the evolution of different leukemic clones from diagnosis to TTR [14, 28, 35]. This is appropriate when populations approach equilibrium. However, during and after chemotherapy, leukemic populations experience bottlenecks and fluctuate stochastically. One conclusion from our analysis is that it is necessary to distinguish between the division rate and self-renewal fraction. Various combinations of these two parameters lead to the same growth rates but different treatment effects. This was noticed before [38], but does not seem to be universally appreciated.

When chance events such as extinction or "crowding out" of small clones have a sufficiently high probability or, in our model, when the leukemic cell population is small, stochastic effects may lead to "coin flip" outcomes, as demonstrated by our simulations. Accordingly, the model demonstrates that the effects of treatment are affected by random events such as extinction versus persistence and subsequent expansion of small clone present at remission. This leads to uncertainty and inevitable prediction errors. Mathematical control theory [39] argues that processes such as this can be controlled in a way that increases the chance of eventual success. What is needed is a periodic evaluation of the MRD, with period adjusted to the estimated growth rates of the clones.

Identification of MRD as early as possible and changing the treatment plan will enhance long-term disease control and curability. One challenge is the identification of high-value target in relapsing clones and effective pharmacologic targeting. Studying the success of MRD in ALL instructs us that time points and thresholds can be identified with well-controlled clinical trials.

Longitudinal monitoring of variant allele frequency coupled with multitype branching process simulation provides a precision medicine approach to improve clinical management of patients with AML. Tracking an individual's leukemic clones and predicting when relapse might occur may prompt the oncologist to change therapies, prepare for a stem cell transplant should there be acceptable comorbidities, provide chemotherapy as a bridge to transplant, and proceed to transplant.

ACKNOWLEDGMENTS

Marek Kimmel and Seth Corey were supported by NIH grant R01 HL128173 and the Alex's Lemonade Stand Foundation for Childhood Cancer Innovation Award. Marek Kimmel was supported in part by NIH grant R01 HL136333 (Katherine Y. King, PI) and by grant 2018/29/B/ST7/02550 from the National Science Center (Poland) (Marek Kimmel, PI). Roman Jaksik was supported by the Polish National Science Centre grant No. 2016/23/D/ST7/03665. Khanh Dinh was supported by the Department of Statistics and Herbert and Florence Irving Institute for Cancer Dynamics, Columbia University. We thank Drs. John E. Dick, Amanda Mitchell, and Mark D. Minden of the Princess Margaret Cancer Centre, University of Toronto for access to the data used for the validation dataset.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

AUTHOR CONTRIBUTIONS

Marek Kimmel conceived and supervised the study. Roman Jaksik assembled data. Marek Kimmel and Khanh Dinh wrote code, ran the model, and analyzed the output. Seth J. Corey provided expertise on leukemia biology and clinical relevance. Marek Kimmel, Roman Jaksik, Seth J. Corey, and Khanh Dinh wrote the paper.

DATA AVAILABILITY STATEMENT

The data supporting the findings of this study are available upon request from the corresponding author (Marek Kimmel).

ORCID

Khanh N. Dinh https://orcid.org/0000-0002-0010-4251
Roman Jaksik https://orcid.org/0000-0003-1866-6380
Marek Kimmel https://orcid.org/0000-0001-8161-890X

REFERENCES

1. R. M. Shallis, R. Wang, A. Davidoff, X. Ma, and A. M. Zeidan, *Epidemiology of acute myeloid leukemia: Recent progress and enduring challenges*, Blood Rev. **36** (2019), 70–87.
2. N. J. Short, and F. Ravandi, *How close are we to incorporating measurable residual disease into clinical practice for acute myeloid leukemia?* Haematologica. **104** (2019), 1532–1541.
3. H. Dohner, D. J. Weisdorf, and C. D. Bloomfield, *Acute myeloid leukemia*, N. Engl. J. Med. **373** (2015), 1136–1152.
4. H. Döhner, E. Estey, D. Grimwade, S. Amadori, F. R. Appelbaum, T. Büchner, H. Dombret, B. L. Ebert, P. Fenaux, R. A. Larson, R. L. Levine, F. Lo-Coco, T. Naoe, D. Niederwieser, G. J. Ossenkoppele, M. Sanz, J. Sierra, M. S. Tallman, H. -. F. Tien, A. H. Wei, B. Löwenberg, and C. D. Bloomfield, *Diagnosis and management of AML inadults: 2017 ELN recommendations from an international expertpanel*, Blood. **129** (2017), no. 4, 424–447.
5. A. Ivey, R. K. Hills, M. A. Simpson, J. V. Jovanovic, A. Gilkes, A. Grech, Y. Patel, N. Bhudia, H. Farah, J. Mason, K. Wall, S. Akiki, M. Griffiths, E. Solomon, F. McCaughan, D. C. Linch, R. E. Gale, P. Vyas, S. D. Freeman, N. Russell, A. K. Burnett, and D. Grimwade, *Assessment of minimal residual disease in standard-risk AML*, N. Engl. J. Med. **374** (2016), no. 5, 422–433.
6. M. Jongen-Lavrencic, T. Grob, D. Hanekamp, F. G. Kavelaars, A. al Hinai, A. Zeilemaker, C. A. J. Erpelinck-Verschueren, P. L. Gradowska, R. Meijer, J. Cloos, B. J. Biemond, C. Graux, M. van Marwijk Kooy, M. G. Manz, T. Pabst, J. R. Passweg, V. Havelange, G. J. Ossenkoppele., M. A. Sanders., G. J. Schuurhuis, B. Löwenberg, and P. J. M. Valk, *Molecular minimal residual disease in acute myeloid leukemia*, N. England J. Med. **378** (2018), no. 13, 1189–1199.
7. F. Ravandi, R. B. Walter, and S. D. Freeman, *Evaluating measurable residual disease in acute myeloid leukemia*, Blood Adv. **2** (2018), no. 11, 1356–1366.
8. L. K. Andersen, and M. C. Mackey, *Resonance in periodic chemotherapy: A case study of acute myelogenous leukemia*, J. Theor. Biol. **209** (2001), no. 1, 113–130.
9. F. Michor, T. P. Hughes, Y. Iwasa, S. Branford, N. P. Shah, C. L. Sawyers, and M. A. Nowak, *Dynamics of chronic myeloid leukaemia*, Nature. **435** (2005), no. 7046, 1267–1270.
10. Z. L. Whichard, C. A. Sarkar, M. Kimmel, and S. J. Corey, *Hematopoiesis and its disorders: a systems biology approach*, Blood. **115** (2010), no. 12, 2339–2347.
11. M. Horn, I. Glauche, M. C. Müller, R. Hehlmann, A. Hochhaus, M. Loeffler, and I. Roeder, *Model-based decision rules reduce the risk of molecular relapse after cessation of tyrosine kinase inhibitor therapy in chronic myeloid leukemia*, Blood. **121** (2013), no. 2, 378–384.
12. M. Kimmel, and S. Corey, *Stochastic hypothesis of transition from inborn neutropenia to AML: Interactions of cell population dynamics and population genetics*, Front. Oncol. **3** (2013), 89.
13. E. Pefani, N. Panoskaltsis, A. Mantalaris, M. C. Georgiadis, and E. N. Pistikopoulos, *Design of optimal patient-specific chemotherapy protocols for the treatment of acute myeloid leukemia (AML)*, Comput. Chem. Eng. **57** (2013), 187–195.
14. T. Stiehl, N. Baran, A. D. Ho and A. Marciniak-Czochra, *Clonal selection and therapy resistance in acute leukaemias: mathematical modelling explains different proliferation patterns at diagnosis and relapse*, J. R. Soc., Interface. **11** (2014), no. 94, 20140079.
15. M. Fuentes-Garí, R. Misener, M. C. Georgiadis, M. Kostoglou, N. Panoskaltsis, A. Mantalaris, and E. N. Pistikopoulos, *Selecting a differential equation cell cycle model for simulating leukemia treatment*, Ind. Eng. Chem. Res. **54** (2015), no. 36, 8847–8859.

16. M. Craig, A. R. Humphries, and M. C. Mackey, *A mathematical model of granulopoiesis incorporating the negative feedback dynamics and kinetics of G-CSF/neutrophil binding and internalization*, Bull. Math. Biol. **78** (2016), 2304–2357.
17. A. Silva, M. C. Silva, P. Sudalagunta, A. Distler, T. Jacobson, A. Collins, T. Nguyen, J. Song, D. -. T. Chen, Lu Chen, C. Cubitt, B. Rachid, L. Perez, Rebatchouk Dmitri, W. Dalton, J. Greene, R. Gatenby, R. Gillies, E. Sontag, M. B. Meads, and K. H. Shain, *An ex vivo platform for the prediction of clinical response in multiple myeloma*, Cancer Res. **77** (2017), no. 12, 3336–3351.
18. J. Xu, Y. Wang, P. Guttorp, and J. L. Abkowitz, *Visualizing hematopoiesis as a stochastic process*, Blood Adv. **2** (2018), no. 20, 2637–2645.
19. M. Kimmel, *Stochasticity and determinism in models of hematopoiesis*, Adv. Exp. Med. Biol. **844** (2014), 119–152.
20. T. Wojdyla, H. Mehta, T. Glaubach, R. Bertolusso, M. Iwanaszko, R. Braun, S. J. Corey, and M. Kimmel, *Mutation, drift and selection in single-driver hematologic malignancy: Example of secondary myelodysplastic syndrome following treatment of inherited neutropenia*, PLoS Comput. Biol. **15** (2019), no. 1, e1006664.
21. Li Ding, T. J. Ley, D. E. Larson, C. A. Miller, D. C. Koboldt, J. S. Welch, J. K. Ritchey, M. A. Young, T. Lamprecht, M. D. McLellan, J. F. McMichael, J. W. Wallis, C. Lu, D. Shen, C. C. Harris, D. J. Dooling, R. S. Fulton, L. L. Fulton, K. Chen, H. Schmidt, J. Kalicki-Veizer, V. J. Magrini, L. Cook, S. D. McGrath, T. L. Vickery, M. C. Wendl, S. Heath, M. A. Watson, D. C. Link, M. H. Tomasson, W. D. Shannon, J. E. Payton, S. Kulkarni, P. Westervelt, M. J. Walter, T. A. Graubert, E. R. Mardis, R. K. Wilson, and J. F. DiPersio, *Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing*, Nature. **481** (2012), no. 7382, 506–510.
22. L. I. Shlush, A. Mitchell, L. Heisler, S. Abelson, S .W. K. Ng, A. Trotman-Grant, J. J. F. Medeiros, A. Rao-Bhatia, I. Jaciw-Zurakowsky, R. Marke, J. L. McLeod, M. Doedens, G. Bader, V. Voisin, C. J. Xu, J. D. McPherson, T. J. Hudson, J. C. Y. Wang, M. D. Minden, and J. E. Dick, *Tracing the origins of relapse in acute myeloid leukaemia to stem cells*, Nature. **547** (2017), no. 7661, 104–108.
23. A. Ediriwickrema, A. Aleshin, J. G. Reiter, M. R Corces, T. Köhnke, M. Stafford, M. Liedtke, B. C. Medeiros, and R. Majeti, *Single-cell mutational profiling enhances the clinical evaluation of AML MRD*, Blood Adv. **4** (2020), no. 5, 943–952.
24. M. Kimmel, and D. E. Axelrod, *Branching Processes in Biology (Interdisciplinary Applied Mathematics)*, 2nd ed., Springer, 2015.
25. A. Marciniak-Czochra, T. Stiehl, A. D. Ho, W. Jäger, and W. Wagner, *Modeling of asymmetric cell division in hematopoietic stem cells—regulation of self-renewal is essential for efficient repopulation*, Stem Cells Dev. **18** (2009), no. 3, 377–386.
26. E. Pefani, N. Panoskaltsis, A. Mantalaris, M. C. Georgiadis, and E. N. Pistikopoulos, *Design of optimal patient-specific chemotherapy protocols for the treatment of acute myeloid leukemia (AML)*, Comput. Chem. Eng. **57** (2013), 187–195.
27. P. A. Greif, L. Hartmann, S. Vosberg, S. M. Stief, R. Mattes, I. Hellmann, K. H. Metzeler, T. Herold, S. A. Bamopoulos, P. Kerbs, V. Jurinovic, D. Schumacher, F. Pastore, K. Bräundl, E. Zellmeier, B. Ksienzyk, N. P. Konstandin, S. Schneider, A. Graf, S. Krebs, H. Blum, M. Neumann, C. D. Baldus, S. K. Bohlander, S. Wolf, D. Görlich, W. E. Berdel, B. J. Wörmann, W. Hiddemann, and K. Spiekermann, *Evolution of cytogenetically normal acute myeloid leukemia during therapy and relapse: An exome sequencing study of 50 patients*, Clin. Cancer Res. **24** (2018), no. 7, 1716–1726.
28. T. Stiehl, N. Baran, A. D. Ho, and A. Marciniak-Czochra, *Cell division patterns in acute myeloid leukemia stem-like cells determine clinical course: A model to predict patient survival*, Cancer Res. **75** (2015), no. 6, 940–949.
29. E. M. Heath, S. M. Chan, M. D. Minden, T. Murphy, L. I. Shlush, and A. D. Schimmer, *Biological and clinical consequences of NPM1 mutations in AML*, Leukemia. **31** (2017), no. 4, 798–807.
30. U. Alon, *An introduction to systems biology: Design principles of biological circuits*, 2nd ed., CRC Press, Boca Raton, FL, 2019. 31.
31. R. Pieters, H. de Groot-Kruseman, V. Van der Velden, M. Fiocco, H. van den Berg, E. de Bont, R. M Egeler, P. Hoogerbrugge, G. Kaspers, E. Van der Schoot, V. De Haas, and J. Van Dongen, *Successful therapy reduction and intensification for childhood acute lymphoblastic leukemia based on minimal residual disease monitoring: Study ALL10 from the Dutch Childhood Oncology Group*, J. Clin. Oncol. **34** (2016), no. 22, 2591–2601.
32. A. von Stackelberg, F. Locatelli, G. Zugmaier, R. Handgretinger, T. M. Trippett, C. Rizzari, P. Bader, M. M. O'Brien, B. Brethon, D. Bhojwani, P. G. Schlegel, A. Borkhardt, S. R. Rheingold, T. M. Cooper, C. M. Zwaan, P. Barnette, C. Messina, G. Michel, S. G. DuBois, K. Hu, M. Zhu, J. A. Whitlock, and L. Gore, *Phase I/Phase II study of blinatumomab in pediatric patients with relapsed/refractory acute lymphoblastic leukemia*, J. Clin. Oncol. **34** (2016), no. 36, 4381–4389.
33. C. A. Miller, J. McMichael, Ha X. Dang, C. A. Maher, Li Ding, T. J. Ley, E. R. Mardis, and R. K. Wilson, *Visualizing tumor evolution with the fishplot package for R*, BMC Genomics. **17** (2016), no. 1, 1–3, http://doi.org/10.1186/s12864-016-3195-z.
34. G. J. Schuurhuis, M. Heuser, S. Freeman, M. -. C. Béné, F. Buccisano, J. Cloos, D. Grimwade, T. Haferlach, R. K. Hills, C. S. Hourigan, J. L. Jorgensen, W. Kern, F. Lacombe, L. Maurillo, C. Preudhomme, A. van der Reijden Bert, C. Thiede, A. Venditti, P. Vyas, B. L. Wood, R. B. Walter, K. Döhner, G. J. Roboz, and G .J. Ossenkoppele, *Minimal/measurable residual disease in AML: a consensus document from the European LeukemiaNet MRD Working Party*, Blood. **131** (2018), no. 12, 1275–1291.
35. T. Stiehl, C. Lutz, and A. Marciniak-Czochra, *Emergence of heterogeneity in acute leukemias*, Biol. Direct. **11** (2016), no. 1, 1–9.
36. R. Beekman, M .G. Valkhof, M. A. Sanders, M. H. van Strien Paulette, J. R. Haanstra, L. Broeders, M. Geertsma-KleinekoortWendy, A. J. P. Veerman, P. J. M. Valk, R. G. Verhaak, B. Löwenberg, and I. P. Touw, *Sequential gain of mutations in severe congenital neutropenia progressing to acute myeloid leukemia*, Blood. **119** (2012), no. 22, 5071–5077.
37. H.-C. Bhang, Ruddy D. A., R. V. Krishnamurthy, J. X. Caushi, R. Zhao, M. M. Hims, A. P. Singh, I. Kao, D. Rakiec, P. Shaw, M. Balak, A. Raza, E. Ackley, N. Keen, M. R. Schlabach, M. Palmer, R. J. Leary, D. Y. Chiang, W. R. Sellers, F. Michor, V. G. Cooke, J. M. Korn, F. Stegmeier, *Studying clonal dynamics in response to cancer therapy using high-complexity barcoding*, Nat. Med. **21** (2015), no. 5, 440–448.
38. T. Stiehl, and A. Marciniak-Czochra, *Stem cell self-renewal in regeneration and cancer: Insights from mathematical modeling*, Curr. Opin. Syst. Biol. **5** (2017), 112–120.
39. A. Swierniak, M. Kimmel, and J. Smieja, *Mathematical modeling as a tool for planning anticancer therapy*, Eur. J. Pharmacol. **625** (2009), 108–121.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: K. N. Dinh, R. Jaksik, S. J. Corey, M. Kimmel, *Predicting time to relapse in acute myeloid leukemia through stochastic modeling of minimal residual disease based on clonality data*, Comp. Sys. Onco. **1** (2021), e1026. https://doi.org/10.1002/cso2.1026