



# Stochastic hypothesis of transition from inborn neutropenia to AML: interactions of cell population dynamics and population genetics

Marek Kimmel<sup>1,2\*</sup> and Seth Corey<sup>3,4</sup>

<sup>1</sup> Department of Statistics, Rice University, Houston, TX, USA

<sup>2</sup> Department of Bioengineering, Rice University, Houston, TX, USA

<sup>3</sup> Department of Pediatrics, Feinberg School of Medicine, Northwestern University, Chicago, IL, USA

<sup>4</sup> Department Cell and Molecular Biology, Feinberg School of Medicine, Northwestern University, Chicago, IL, USA

## Edited by:

Heiko Enderling, Tufts University School of Medicine, USA

## Reviewed by:

Stephan Von Gunten, University of Bern, Switzerland

Rainer K. Sachs, University of California at Berkeley, USA

## \*Correspondence:

Marek Kimmel, Department of Statistics, Rice University, 6100 Main Street, Houston, TX 77005, USA.  
e-mail: kimmel@rice.edu

We present a stochastic model of driver mutations in the transition from severe congenital neutropenia to myelodysplastic syndrome to acute myeloid leukemia (AML). The model has the form of a multitype branching process. We derive equations for the distributions of the times to consecutive driver mutations and set up simulations involving a range of hypotheses regarding acceleration of the mutation rates in successive mutant clones. Our model reproduces the clinical distribution of times at diagnosis of secondary AML. Surprisingly, within the framework of our assumptions, stochasticity of the mutation process is incapable of explaining the spread of times at diagnosis of AML in this case; it is necessary to additionally assume a wide spread of proliferative parameters among disease cases. This finding is unexpected but generally consistent with the wide heterogeneity of characteristics of human cancers.

**Keywords:** severe congenital neutropenia, myelodysplastic syndrome, acute myeloid leukemia, branching process, driver mutations, clonal evolution

## INTRODUCTION

Granulocytes are essential for host defense and survival. Their importance is apparent in severe congenital neutropenia (SCN). Life-threatening infections in children with SCN can be avoided through the use of recombinant granulocyte colony-stimulating factor (G-CSF). However, SCN often transforms into secondary myelodysplastic syndrome (sMDS) and then into secondary acute myeloid leukemia (sAML). A great unresolved clinical question is whether chronic, pharmacological doses of G-CSF contribute to this transformation (Glaubach and Corey, 2012). A number of epidemiological clinical trials have demonstrated a strong association between exposure to G-CSF and sMDS/sAML (Dong et al., 1995; Donadieu et al., 2005; Rosenberg et al., 2006; Germeshausen et al., 2007; Carlsson et al., 2012). Mutations in the distal domain of the G-CSF Receptor (G-CSFR) have been isolated from patients with SCN who developed sMDS/sAML or patients with *de novo* MDS (Beekman and Touw, 2010). Most recently, clonal evolution over approximately 20 years was documented in a patient with SCN who developed sMDS/sAML (Beekman et al., 2012). Clonal evolution of a sick hematopoietic progenitor cell in SCN involves perturbations in proximal and distal signaling networks triggered by a mutant G-CSFR. Transition from SCN → sMDS → sAML involves chance mechanisms such as mutations, drift and transcription, and receptor noise, which require that stochastic models are needed (Whichard et al., 2010).

In the present paper we use stochastic modeling to understand the wide range of times at which the transition to sAML occurs. We develop a model in the form of a multitype branching process, which allows one tying population genetics and

population dynamics aspects of the transition from SCN to sMDS to sAML, and validating it against existing evidence. Branching processes have been used widely to model mutation, selection, and drift processes in populations of variable size, to which the classical Wright–Fisher model does not apply (Cyran and Kimmel, 2010). We adopted approach similar to that developed in Nowak's group (Bozic et al., 2010), modified to bring out stochastic time intervals between successive driver mutations.

The model we developed allows predicting the time at transition to sAML given the probability of each successive driver mutation, the number of mutations needed, and the proliferative potential of each successive mutated clone of hemopoietic stem cells. We can then compare these times to observed distribution of times at transition. As documented in the paper, the outcome is intriguing: stochasticity inherent in the mutation process is insufficient to explain the wide distribution of times at transition (ranging from 1 to 38; **Table 1**). Additional factors are required, one of which may be a wide interpatient spread of proliferative potential of the mutant clones.

## POPULATION GENETICS AND POPULATION DYNAMICS MODEL OF THE SCN → sMDS → sAML TRANSITION

Missense, nonsense, and frameshift mutations, and dysregulated alternative splicing in G-CSFR have been isolated in patients with MDS/AML. In the study of Beekman et al. (2012), nonsense and missense mutations in G-CSFR arose during the course of the disease. In the model we envision, population genetics, and population dynamics of proliferating bone marrow cells are closely intertwined.

**Table 1 | Summary of life histories of patients transitioning from severe congenital neutropenia (SCN) to secondary myelodysplastic syndrome (sMDS) to secondary acute myeloid leukemia (sAML) (Walter et al., 2012).**

Phase of disease	Age at diagnosis (years)	Number of co-existing mutations
SCN	0–0.5	1*
MDS	1–12	1–3 ± chromosomal loss or gain**
AML	2–38	1–9 ± chromosomal loss or gain***

\**ELANE, HAX1, G6PC, WAS, CSF3R.*

\*\**GCSF3R, ZC3H18, LLLGL2; RAS ± monosomy 7.*

\*\*\**RUNX1, ASXL1, p300, CEBA, CSF3R, MGA, SUZ12, LAMB, FBXO18, CCDC15, ± monosomy 7, trisomy 21.*

### POPULATION GENETICS PERSPECTIVE

Proliferating healthy cells in the bone marrow mutate at random times, possibly influenced by super-pharmacological doses of GCSF. A summary of possible mutations and their consequences for proliferation dynamics of granulocyte precursors is depicted in **Figure 1**. GCSF signaling occurs through its cognate receptor, GCSFR. It involves both proximal signaling networks consisting of signaling molecules such as Lyn, Jak, STAT, Akt, and ERK, and distal gene regulatory networks consisting of transcription factors. Together, these signaling networks promote proliferation, survival, and differentiation. In patients with SCN, the earliest known mutation to contribute to transformation to secondary MDS or AML is a nonsense mutation in the GCSFR gene. This mutation leads to a truncated receptor, GCSR delta 715 (Glaubach and Corey, 2012, and reference therein).

It follows from a simple calculus of mutation events that as long as the cell population size is kept in check, the rate at which new mutant clones appear in the population is rather low. When the population expands, new mutant clones arise faster (see further on).

In our model we take the view that carcinogenesis is driven by a succession of small-scale (e.g., point) mutations in specific loci. Other viewpoints (epigenetic effects, karyotypic alterations, intercellular interactions, etc.) have been suggested. In treatment-related MDS some drugs (e.g., many alkylating agents) induce t-MDS primarily via large scale alterations that lead to karyotypic instability (Bhatia, 2011).

### POPULATION DYNAMICS PERSPECTIVE

Limited mutation load at the SCN phase causes neutropenia and fluctuations of cell population size. With time, accumulation of driver mutations causes expansion of mutant clones, which however are not yet expanding at a dramatic rate. At some point in time, mutations accumulate sufficiently to cause a major change in the proliferation law and the now malignant cell population starts rapidly expanding.

Our model is based on the following hypotheses (**Figure 2**):

1. At the time of diagnosis of SCN, GCSF therapy is initiated, which induces an initial series of  $X$  driver mutations, occurring at random times.

2. The  $X$ -th mutation causes transition to the MDS, during which further  $Y$  mutations occur.
3. After  $X + Y$  mutations, the AML stage begins, during which the subsequent mutant clone grow at increasing rate, which in turn shortens times at which still new mutations appear.

In the model, the increasing proliferation rate of successive mutant clones causes acceleration of growth of the malignant bone marrow stem cell population, which shortens the time interval to appearance of new clones, which in turns increases proliferation rate, and so forth; this results in a positive feedback. As we will see, the stochastic nature of the process (the times to appearance of each next mutant are random) causes a spread of the timing of the subsequent mutations, particularly the first  $X$  mutations during the SCN phase. This may result in the transition to MDS not manifesting itself for a very long time in a fraction of cases.

### ROLE OF STOCHASTIC DYNAMICS IN THE MODEL

We explain some other intuitions underlying the model. For a new subclone, stochastic theory is used to estimate extinction probability, with extinction after more than a few cell generations being negligible in view of the growth advantage of the new clone. However the time at which the next mutation occurs in a cell clone is also stochastic and it is as a rule more dispersed for the slower-growing clones. Therefore the time to reaching the threshold number of bone marrow stem cells (which in our model defines the time at sAML diagnosis), is a random variable. One of the questions we ask is if dispersion of this time matches the wide distribution of the times at diagnosis (Rosenberg et al., 2010).

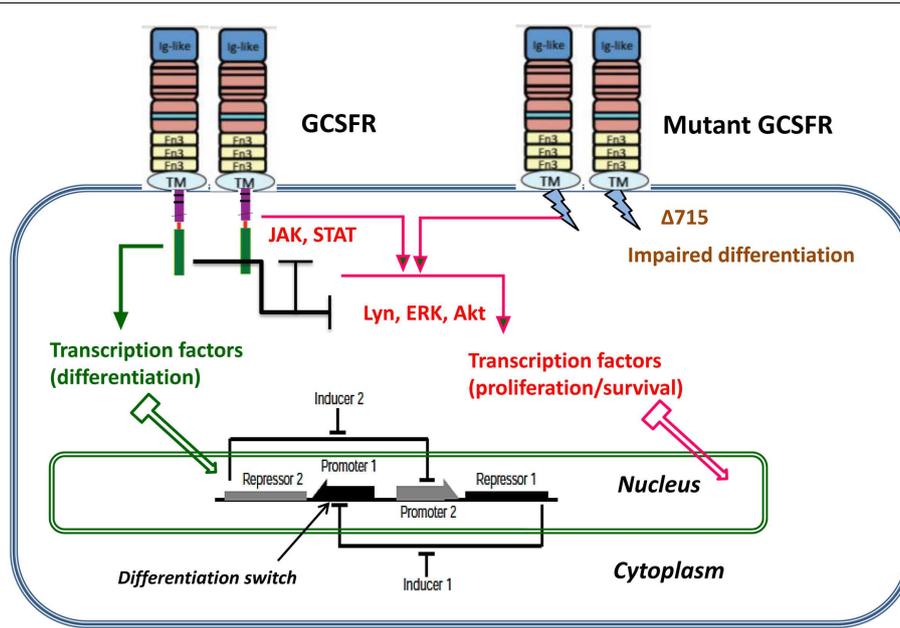
## MATERIALS AND METHODS

### MATHEMATICS OF THE MODEL

The population-genetic effect of population size-dependent accumulation of mutations occurs as a natural consequence of the proliferation law in the form of a multitype Galton–Watson branching process:

1. Consecutively arising surviving mutant clones are numbered with the index  $k$ , ranging from 1 to  $K$ ; time interval between the appearance of the  $k$ -th and  $k + 1$ -st surviving mutant clones is denoted by  $\tau_k$ .  $k$ -th mutant cells have accumulated  $k$  driver mutations (assuming the clone in SCN bone marrow at diagnosis has a single cell with one driver mutation, which seems a defensible idealization).
2. All clones expand as Galton–Watson branching processes (see further on). Cell life length is constant and equal to  $T$ , and at that time the cell either produces two progeny with probability  $b_k$  (cell type  $k$ ) or dies (or becomes quiescent or differentiated, which does not make a difference for disease dynamics) with probability  $1 - b_k$ .
3. A cell of type  $k$  can mutate upon its birth (for definiteness) to type  $k + 1$  with probability  $u$ .

These three rules allow one derive the probability distributions of time intervals  $\tau_k$ , probabilities of survival of each clone, and expected growth laws of each clone. Mathematical details follow from the theory of Galton–Watson branching process; see for example the monograph by Kimmel and Axelrod (2002).



**FIGURE 1 | Dynamic stochastic model of impaired differentiation in granulocyte precursors.** GCSF signaling occurs through its cognate receptor, GCSFR. It involves both proximal signaling networks consisting of signaling molecules such as Lyn, Jak, STAT, Akt, and ERK, and distal gene regulatory networks consisting of transcription factors. Together,

these signaling networks promote proliferation, survival, and differentiation. In patients with severe congenital neutropenia, the earliest known mutation to contribute to transformation to secondary MDS or AML is a nonsense mutation in the GCSFR gene. This mutation leads to a truncated receptor, GCSFR delta 715.

We assume that cell division is effective with probability  $b$ , i.e., the probability generating function (pgf) of the number of progeny cells per parent cell has the form  $f(s) = bs^2 + (1 - b)$ . The extinction probability  $q$  is the smaller solution of the equation  $q = f(q)$ , which is less than 1 if the process is supercritical. In our case,

$$q = bq^2 + (1 - b) \Rightarrow q = (b^{-1} - 1); b \in (0.5, 1]. \quad (1)$$

Similarly, the expected number of progeny of a cell is equal to  $f'(1) = 2b$ , hence the expected number of cells at time  $t$  is equal to  $N(t) = (2b)^{(t/T)}$ , which yields the value of  $\lambda$

$$\exp(\lambda t) = (2b)^{(t/T)} \Rightarrow \lambda = \ln(2b) / T. \quad (2)$$

We will use “continuous” time  $t$  for notational convenience. However, we consider generations of cells dividing at discrete times  $t_i = iT$ , where  $T$  is the average cell cycle time. As it is known, the expected (mean) growth law in the Galton–Watson process has the form

$$E[\# \text{ cells, at time } t, \text{ in a clone started at time } t_0] \stackrel{\text{def}}{=} [N(t) = \exp(\lambda(t - t_0))], \text{ as } t \rightarrow \infty. \quad (3)$$

To determine the distribution of time to a mutation creating a new non-extinct clone, we consider a newborn cell. In this cell, mutation may occur with probability  $u$ , and if the extinction probability of the mutant clone is  $q'$ , then the probability that the cell does not produce a new mutant clone is equal

to  $1 - u(1 - q')$ . Until time  $t_i = iT$ , approximately  $\sum_{j=0}^i N(t_j) = N(t_0) (\exp(\lambda T)^{i+1} - 1) (\exp(\lambda T) - 1)^{-1}$  new cells are born, and assuming independence, we obtain

$$\begin{aligned} & \Pr [\text{no mutant initiating nonextinct clone appears until time } t_i = iT] \\ &= \Pr [\tau =: \text{time to nearest nonextinct mutant clone} > t_i = iT] = (1 - u(1 - q'))^{\sum_{j=0}^i N(t_j)} \quad (4) \\ &= (1 - u(1 - q'))^{N(t_0) \frac{\exp(\lambda T)^{i+1} - 1}{\exp(\lambda T) - 1}} = a^{d(\exp(\lambda T)^{i+1} - 1)}. \end{aligned}$$

where, for the  $k$ -th mutant population

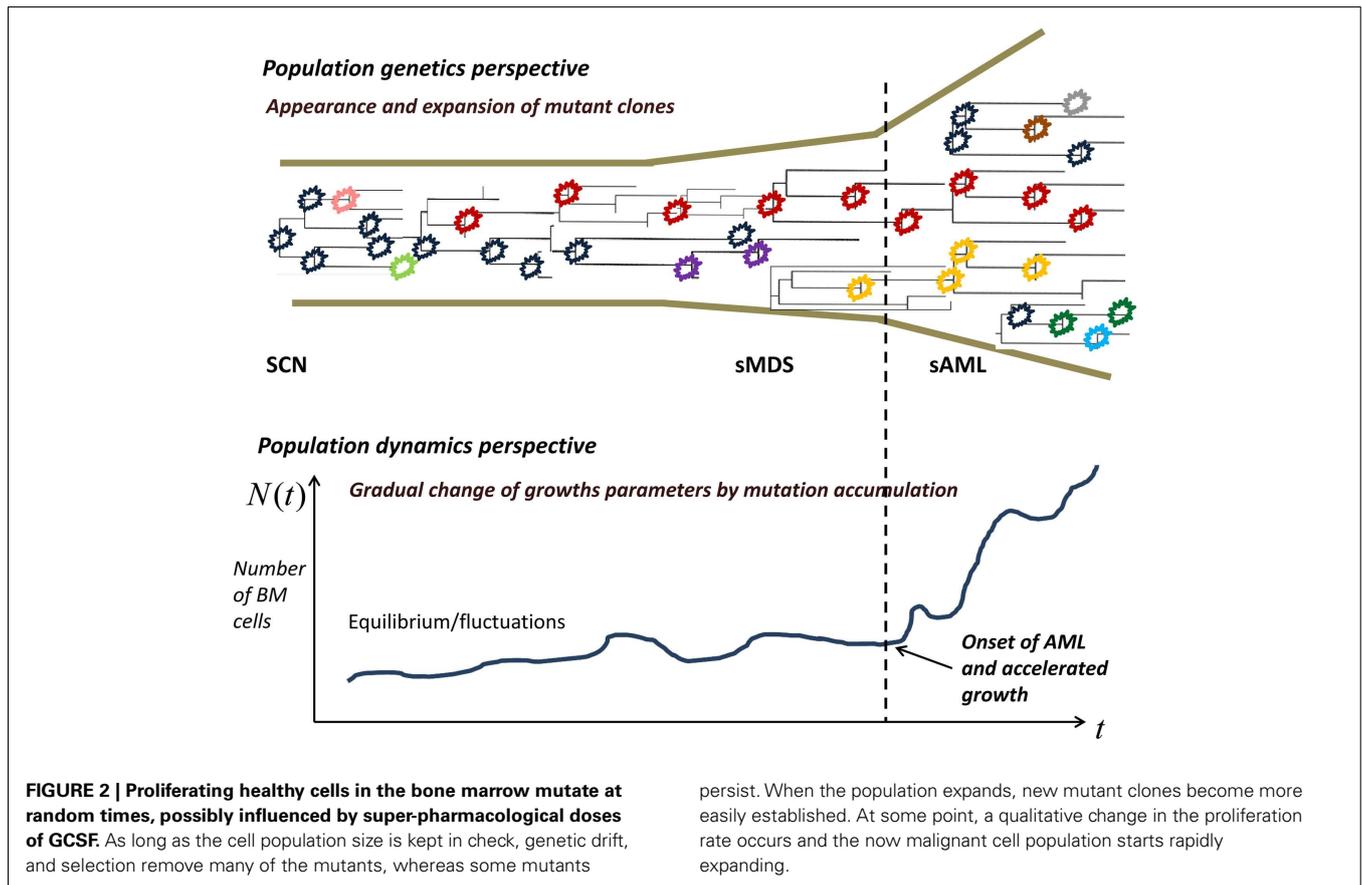
$$\begin{aligned} a &= (1 - (1 - q_{k+1}) u_k) = (1 - u_k (2b_{k+1} - 1) / b_{k+1}), \\ c &= \lambda_k = \ln(2b_k) / T, \\ d &= (\exp(\lambda_k T) - 1)^{-1} = (2b_k - 1)^{-1}. \end{aligned}$$

Since the distribution tail of random variable  $\tau_k$  has the form

$$\Pr [\tau_k > \tau] = a^{d(\exp(c\tau) - 1)} = \exp(-\ln(a^{-1}) d (\exp(c\tau) - 1)).$$

it can be algorithmically generated using the inverse tail method

$$\tau_k = c^{-1} \ln(\ln r / (d \ln a) + 1), \quad (5)$$



**FIGURE 2 | Proliferating healthy cells in the bone marrow mutate at random times, possibly influenced by super-pharmacological doses of G-CSF.** As long as the cell population size is kept in check, genetic drift, and selection remove many of the mutants, whereas some mutants

persist. When the population expands, new mutant clones become more easily established. At some point, a qualitative change in the proliferation rate occurs and the now malignant cell population starts rapidly expanding.

where  $r$  is a pseudo-random number uniformly distributed from 0 to 1. In this framework, a sample path of the number of cells in the  $k$ -th mutant clone (which contains cells with  $k$  mutations accumulated) is equal to

$$N_k(t) = \begin{cases} 0 & t \leq \sum_{j=1}^{k-1} \tau_j \\ \exp(\lambda_k (t - \sum_{j=1}^{k-1} \tau_j)) & \text{otherwise} \end{cases} \quad (6)$$

The derivations presented are quite similar to those of Bozic et al. (2010), except that in that paper, expected times  $E(\tau_k)$  to the next mutation have been used. Here, we are interested in exposing stochastic variability in the time course of the SNC  $\rightarrow$  sMDS  $\rightarrow$  sAML transition. Another refinement would be to use distributions of cell counts instead of expected values  $N_k(t)$ . This would result in serious computational problems, arguably without much impact on the results.

**MODELING THE SNC  $\rightarrow$  sMDS  $\rightarrow$  sAML TRANSITION**

Equations 1 and 2 allow generating realizations of times to successive driver mutations under different values of mutation rates and proliferative characteristics of the mutant clones. We make the following assumptions:

1. Transition to sMDS requires one or two somatic driver mutations, whereas the transition to sAML requires at least three somatic driver mutations (cf. Table 1).

2. Diagnosis of sAML requires presence of  $10^4$  leukemic HSC. For details of computations leading to this estimate, see further on.
3. Successive mutant clones have increasing proliferative potential. We assume a power law for the coefficients  $b_i$ , which seems to lead to fits that do not contradict data:

$$b_i = \min(0.5 + A(\epsilon + (i - 1)^\kappa), 1), \quad (7)$$

where coefficients  $A$ ,  $\epsilon$ , and  $\kappa$  are considered further on.

4. As it will be seen, it is necessary to assume that the coefficients  $A$  be generated from a probability distribution instead of assuming a constant value. We assume the distribution function  $F_A(a)$  selected so that the times of at diagnosis of sAML fit available statistics (for details see further on).

**ESTIMATE OF THE NUMBER OF LEUKEMIC CELLS**

We carried out computations based on two literature sources and then used rounding to the nearest order of magnitude to obtain a working threshold number of the leukemic initiating cells (LIC) (Bonnet and Dick, 1997). In both cases we assume that the volume of human bone marrow is equal to  $V = 1700$  ml and that LIC cells constitute a fraction  $\psi = 10^{-6}$  of leukemic bone marrow mono-nucleated cells (BMMNC). We also assume that in sAML, fraction  $\rho = 0.8$  of BMNNC is constituted by leukemic cells.

**Estimate 1**

Dedeepiya et al. (2012) provide an estimate of the number of BNNMC per 1 ml  $B = 3.67 \times 10^6$ . This results in an estimate of the number  $L$  of LIC cells in the entire bone marrow  $L = \rho \times \psi \times V \times B = 4991$  cells.

**Estimate 2**

Bender et al. (1994) provide estimates of  $B$  in the range from  $3.02 \times 10^6$  to  $4.71 \times 10^6$ . This results in  $L = 4107 \div 5535$  cells. These estimates are remarkably consistent. Rounding to the nearest order of magnitude results in a working estimate of  $L = 10^4$  cells.

**TIME AT DIAGNOSIS OF sAML AND DISTRIBUTION OF PARAMETER A**

Under given values of parameters  $\kappa$  and  $\epsilon$  as well as mutation rates  $u_k$ , the time at diagnosis of sAML, defined as the time  $T$  from initiation of GCSF treatment such that

$$\sum_k N_k(T) = L$$

depends on parameter  $A$  according to an approximate power law

$$T = f(A) = \exp(\alpha) A^\beta,$$

where  $\beta < 0$ . This dependence, which was obtained via simulation studies (not shown), allows finding the distribution of  $A$  that leads to a clinically observed distribution of the time of sAML diagnosis according to the following expression for distribution tails

$$\bar{F}_A(a) = 1 - \bar{F}_T(f(a)),$$

where  $\bar{F}_T(t) = \Pr[T > t]$  is the tail of the distribution of time  $T$ . This in turn allows generating pseudo-random realizations of  $A$  according to the expression

$$A = f^{-1}(\bar{F}_T^{-1}(R)) = (\exp(-\alpha) \bar{F}_T^{-1}(R))^{1/\beta}, \tag{8}$$

where  $R$  is a pseudo-random number from the uniform distribution on the  $(0, 1)$  interval.

We need to approximate the tail of the distribution of the time at diagnosis of sAML. A recent source is the paper by Rosenberg et al. (2010). These authors reported results of a prospective study of 374 SCN patients, and included estimates of hazard rates and cumulative probability of sMDS/sAML as a function of time after GCSF treatment. Hazard rate grows for the first 5 years and then plateaus. To simplify computations we adopted a piecewise constant estimate of the hazard rate  $h_T(t)$

$$h_T(t) = \begin{cases} 0.01 & t \in [0, 3) \\ 0.02 & t \in [3, \infty) \end{cases}$$

with time in years. Comparing with Figure 1A in Rosenberg et al. (2010) we see that  $h_T(t)$  remains within the confidence band computed based on the prospective study. Using the expression

$$\bar{F}_T(t) = \exp\left(-\int_0^t h_T(\tau) d\tau\right)$$

and inverting the tail function  $\bar{F}_T(t)$  we complete the derivation of expression Eq. 8 (elementary details not shown).

**OVERVIEW OF PARAMETER ESTIMATION**

The form of expression Eq. 8 and plausible estimates of parameters  $\kappa$  and  $\epsilon$  as well as of mutation rates  $u_k$ , are difficult to be uniquely determined with the data available at the present time. We used the following heuristic procedure:

1. Driver mutation rates increase from the reference value by a factor of 5, starting mutation 3, so that  $u_{1,2} = u$  but  $u_{3,4,5} = 5u$ . The increase is needed for the later mutations to occur in quick succession, so that mutation 3 occurs before  $\sum_k N_k(T) > L$ , with  $L = 10^4$  being a relatively low value.
2. Reference driver mutation rate had to be set equal to 0.00034, 10 times higher than the value estimated by Bozic et al. (2010). This is required for enough mutations to accumulate before the threshold time  $T$ .
3. Proliferation rate increases as power  $\kappa$  of the mutation number, value  $\kappa = 2$  provides sufficient acceleration to explain relative rapidity of the AML stage. The offset parameter  $\epsilon = 0.02$  keeps proliferation rate before mutation 1 sufficiently low.
4. Once estimates of parameters  $u_k$ ,  $\kappa$ , and  $\epsilon$  are obtained, estimates of the power law parameters  $\alpha$  and  $\beta$  are determined by a simulation study, and the generator of random parameter  $A$  is obtained via expression Eq. 8.

**RESULTS**

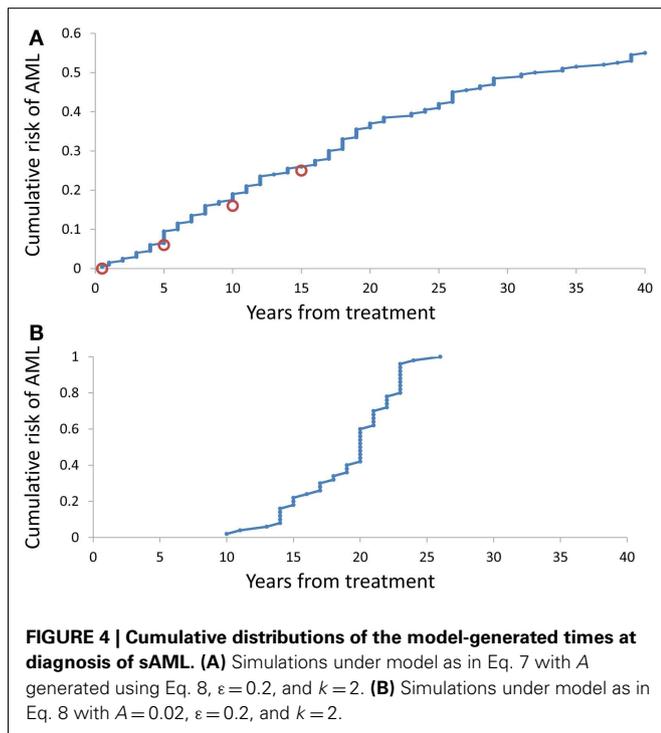
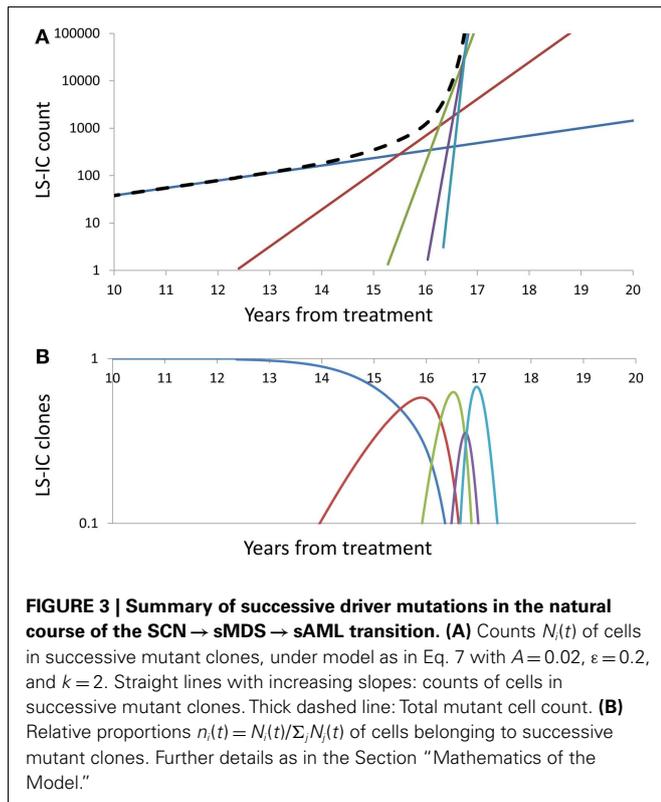
**SIMULATED COURSE OF DISEASE**

Figure 3 depicts the impact of successive driver mutations on the natural course of the SCN  $\rightarrow$  sMDS  $\rightarrow$  sAML transition. Figure 3A depicts counts  $N_i(t)$  of cells in successive mutant clones as a function of time, under model as in Eq. 7 with  $A = 0.005$ ,  $\epsilon = 0.02$ , and  $\kappa = 2$ . Straight lines with increasing slopes are counts of cells in successive mutant clones. We observe that the time intervals separating the origins of successive clones are decreasing with each mutation event. Thick dashed line represents the total mutant cell count. It is also interesting to observe that clones with increasing numbers of mutations dominate transiently, until they are replaced by other clones with higher proliferative capacity (selective value). Figure 3B depicts relative proportions  $n_i(t) = N_i(t) / \sum_j N_j(t)$  of cells belonging to successive mutant clones.

**TIME AT sAML DIAGNOSIS**

It is somewhat surprising that under any combination of coefficients  $A$  and  $k$ , the range of simulated times at sAML diagnosis is rather narrow. Figure 4B depicts ranked simulated times at sAML diagnosis under model as in Eq. 7 with  $A = 0.005$ ,  $\epsilon = 0.02$ , and  $\kappa = 2$ . Spread of these values is narrow, with interquartile range between 15 and 21. Systematic simulation experiments demonstrate that this is the case for a wide range of  $A$  and  $\kappa$  parameter values. This outcome is in contrast to the wide spread of times at diagnosis summarized in Table 1 and that based on Rosenberg et al. (2010).

Simulation-estimation experiment outlined in the Methods demonstrates that distribution of simulated times (counting from initiation of CGSF treatment) at sAML (Rosenberg et al., 2010) is reproduced by our model. Figure 4A cumulative distribution



of the times at sAML diagnosis under model as in Eq. 7 with  $\kappa=2$ ,  $\epsilon=0.02$ , and  $A$  generated from the distribution in Eq. 8 with  $\alpha=-0.655$  and  $\beta=-0.912$ .

## CONCLUSION

The process of development and replacement of leukemic clones is influenced by the processes of genetic drift and selection (Walter et al., 2012). These forces are usually analyzed by geneticists in the framework of the Wright–Fisher or coalescent model (see Discussion and references in Cyran and Kimmel, 2010). However, in the case of expanding cell clones, the more appropriate population process seems to be one of the types of branching processes; in our case, the Galton–Watson process (Kimmel and Axelrod, 2002). In the particular version of the multitype Galton–Watson process that we use, genetic drift’s mechanism is the loss of variants through extinction and selection is embodied in the principle that each next surviving clone is proliferating faster (has greater fitness).

A characteristic feature of human cancers is very wide heterogeneity with respect to extent of involvement, genotype and rate of progression, and spread (Michor et al., 2004; Hanahan and Weinberg, 2011). This is in contrast to induced animal tumors, which are relatively uniform. Secondary AML, resulting from a transition from SCN via myelodysplastic syndrome, is not an exception, with onset varying from 1 to 38 years of age and with wide variability of mutational background (Table 1). It is interesting, and we consider it a major result, that such spread of the age of onset is not due solely to stochastic nature of mutation-driven transitions, but it requires a large variability in proliferative potential from one disease case to another. Also, this distribution of coefficient  $A$ , which parameterizes the proliferative potential, is right-skewed, with slowly evolving (low- $A$ ) clones prevailing. This provides a testable hypothesis about distribution of proliferating rates in leukemic stem cell clones.

The model presented in this paper addresses certain aspects of the SNC → sMDS → sAML transition. Among other, although we might derive an expression relating the number of driver (selective) mutations to the corresponding count of accumulated passenger (neutral) mutations (similarly as it was done in Bozic et al. (2010), we do not have at our disposal sequencing data to validate such an expression. Also, we do not attempt here to fit the distribution of the age at diagnosis of the sMDS, since we are missing data on individual life histories, which would involve somatic mutation as well as sequencing data.

From the mathematical point of view, the current model is also somewhat simplified. It considers each new mutation to provide more selective advantage to the arising clone. This is in apparent disagreement with the recent observation of Beekman et al. (2012), of mutations which appear at the sMDS stage and disappear at the sAML stage. The linear structure of mutation confers desirable simplicity to modeling but is not necessarily realistic. In the framework of multitype branching processes and special processes such as Griffiths and Pakes branching infinite allele model (Griffiths and Pakes, 1988; Kimmel and Mathaes, 2010), more complicated scenarios can be gaged.

## ACKNOWLEDGMENTS

Marek Kimmel was supported by the CPRIT grant RP101089 and by DEC-2012/04/A/ST7/00353 grant from the National Science Center (Poland). Seth Corey was supported by RO1-CA108992.

## REFERENCES

- Beekman, R., and Touw, I. P. (2010). G-CSF and its receptor in myeloid malignancy. *Blood* 115, 5131–5136.
- Beekman, R., Valkhof, M. G., Sanders, M. A., van Strien, P. M., Haanstra, J. R., Broeders, L., et al. (2012). Sequential gain of mutations in severe congenital neutropenia progressing to acute myeloid leukemia. *Blood* 119, 5071–5077.
- Bender, J. G., Unverzagt, K., Walker, D. E., Lee, W., Smith, S., Williams, S., et al. (1994). Phenotypic analysis and characterization of CD34+ cells from normal human bone marrow, cord blood, peripheral blood, and mobilized peripheral blood from patients undergoing autologous stem cell transplantation. *Clin. Immunol. Immunopathol.* 70, 10–18.
- Bhatia, S. (2011). Role of genetic susceptibility in development of treatment-related adverse outcomes in cancer survivors. *Cancer Epidemiol. Biomarkers Prev.* 20, 2048–2067.
- Bonnet, J., and Dick, J. (1997). Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell. *Nat. Med.* 3, 730–737.
- Bozic, I., Antal, T., Ohtsuki, H., Carter, H., Kim, D., Chen, S., et al. (2010). Accumulation of driver and passenger mutations during tumor progression. *Proc. Natl. Acad. Sci. U.S.A.* 107, 18545–18550.
- Carlsson, G., Fath, A., Berglöf, E., Lagerstedt-Robinson, K., Nordenskjöld, M., Palmblad, J., et al. (2012). Incidence of severe congenital neutropenia in Sweden and risk of evolution to myelodysplastic syndrome/leukaemia. *Br. J. Haematol.* 158, 363–369.
- Cyran, K. A., and Kimmel, M. (2010). Alternatives to the Wright-Fisher model: the robustness of mitochondrial Eve dating. *Theor. Popul. Biol.* 78, 165–172.
- Dedeepiya, V. D., Rao, Y. Y., Jayakrishnan, G. A., Parthiban, J. K., Baskar, S., Manjunath, S. R., et al. (2012). Index of CD34+ cells and mononuclear cells in the bone marrow of spinal cord injury patients of different age groups: a comparative analysis. *Bone Marrow Res.* 2012, 787414.
- Donadieu, J., Leblanc, T., Bader, M., eunierB., Barkaoui, M., Fenneteau, O., Bertrand, Y., et al. (2005). Analysis of risk factors for myelodysplasias, leukemias and death from infection among patients with congenital neutropenia. Experience of the French Severe Chronic Neutropenia Study Group. *Haematologica* 90, 45–53.
- Dong, F., Brynes, R. K., Tidow, N., Welte, K., Lowenberg, B., and Touw, I. P. (1995). Mutations in the gene for the granulocyte colony-stimulating-factor receptor in patients with acute myeloid leukemia preceded by severe congenital neutropenia. *N. Engl. J. Med.* 333, 487–493.
- Germeshausen, M., Ballmaier, M., and Welte, K. (2007). Incidence of CSF3R mutations in severe congenital neutropenia and relevance for leukemogenesis: results of a long-term survey. *Blood* 109, 93–99.
- Glaubach, T., and Corey, S. J. (2012). From famine to feast: sending out the clones. *Blood* 119, 5063–5064.
- Griffiths, R. C., and Pakes, G. (1988). An infinite-alleles version of the simple branching process. *Adv. Appl. Probab.* 20, 489–524.
- Hanahan, D., and Weinberg, R. A. (2011). Hallmarks of cancer: the next generation. *Cell* 144, 646–674.
- Kimmel, M., and Axelrod, D. E. (2002). *Branching Processes in Biology*. New York: Springer.
- Kimmel, M., and Mathaes, M. (2010). Modeling neutral evolution of Alu elements using a branching process. *BMC Genomics* 11(Suppl. 1), S11. doi:10.1186/1471-2164-11-S1-S11
- Michor, F., Iwasa, Y., and Nowak, M. A. (2004). Dynamics of cancer progression. *Nat. Rev. Cancer* 4, 197–205.
- Rosenberg, P. S., Alter, B. P., Bolyard, A. A., Bonilla, M. A., Boxer, L. A., Cham, B., et al. (2006). The incidence of leukemia and mortality from sepsis in patients with severe congenital neutropenia receiving long-term G-CSF therapy. *Blood* 107, 4628–4635.
- Rosenberg, P. S., Zeidler, C., Bolyard, A. A., Alter, B. P., Bonilla, M. A., Boxer, L. A., et al. (2010). Stable long-term risk of leukaemia in patients with severe congenital neutropenia maintained on G-CSF therapy. *Br. J. Haematol.* 150, 196–199.
- Walter, M. J., Shen, D., Ding, L., Shao, J., Koboldt, D. C., Chen, K., et al. (2012). Clonal architecture of secondary acute myeloid leukemia. *N. Engl. J. Med.* 366, 1090–1098.
- Whichard, Z. L., Sarkar, C. A., Kimmel, M., and Corey, S. J. (2010). Hematopoiesis and its disorders: a systems biology approach. *Blood* 115, 2339–2347.

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 15 January 2013; accepted: 02 April 2013; published online: 29 April 2013.

Citation: Kimmel M and Corey S (2013) Stochastic hypothesis of transition from inborn neutropenia to AML: interactions of cell population dynamics and population genetics. *Front. Oncol.* 3:89. doi: 10.3389/fonc.2013.00089

This article was submitted to *Frontiers in Molecular and Cellular Oncology*, a specialty of *Frontiers in Oncology*. Copyright © 2013 Kimmel and Corey. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in other forums, provided the original authors and source are credited and subject to any copyright notices concerning any third-party graphics etc.