Preprints of the 19th World Congress
The International Federation of Automatic Control
Cape Town, South Africa. August 24-29, 2014

Application of the stochastic Moran Model of population genetics to understanding the timing of a driver mutation in Myelodysplastic Syndrome (MDS)

Tomasz Wojdyła*, Taly Glaubach, Seth Corey** and Marek Kimmel*****

** Silesian University of Technology, Gliwice, 44-100, Poland (e-mail: tpwojdyla@polsl.pl)*
*** Northwestern University School of Medicine, Chicago, IL, 60611 (e-mail: s-corey@northwestern.edu, tglaubach@luriechildrens.org)*
**** Rice University, Houston, TX 77005, USA (e-mail: kimmel@rice.edu)*

Abstract: We hypothesize that the GCSFR mutants introduce into those pathways changes that will lead to impaired differentiation and enhanced survival based on single cell behaviors as modulated by noise in intracellular proximal signaling and distal gene regulatory networks. We will show that GCSFR truncation mutants become fixed in the granulocyte progenitor population because of an *incremental* growth advantage. We will also show by modeling and experimentation that these GCSFR mutations alter the signaling pathways to further perturb growth, impair differentiation, and enhance survival.

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

1. INTRODUCTION

How a blood stem cell develops into a highly specialized cell over several cell divisions is a profound question involving determinism and stochasticity and broadly applicable to human diseases and therapeutics (Glaubach and Corey 2012). Hematopoiesis provides the best-characterized system for cell fate decision-making in both health and disease. Yet, the precise roles of external cues, intracellular signaling, gene regulatory networks, and homeostatic mechanisms have been elusive because of their complexity and inobservability. The most common white blood cell in humans, the granulocyte is absolutely essential for host defense and survival. Its pathophysiological 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 (GCSF). However, SCN often transforms into secondary myelodysplastic syndrome (sMDS) or acute myeloid leukemia (sAML). A great unresolved clinical question is whether chronic, pharmacologic doses of GCSF contribute to this transformation. Human clinical and experimental data strongly suggest that linkage, whereas none had been predicted in mouse models. Firstly, a number of epidemiological clinical trials have demonstrated a strong association between exposure to GCSF and sMDS/sAML. Secondly, mutations in the distal domain of the GCSF Receptor (GCSFR, gene name CSF3R) have been isolated from SCN patients who developed sMDS/sAML or patients with de novo MDS. Recently, clonal evolution over ~20 years was documented in an SCN patient who developed sMDS/sAML (Beekman et al., 2012). What is particularly striking is that out of five different mutations in the GCSFR gene, some persisted into the leukemic clone but others were lost.

Multiple lines of evidence have demonstrated clonal evolution in cancer. Two thorny questions have emerged: how to account for the large number of mutations isolated in each tumor and how to identify which ones as "drivers" or "passengers". Studying the SCN→sMDS/sAML model will simplify resolving the complexity of clonal evolution. This model is simpler because i) there is one known founding genetic mutation (e.g. ELANE), ii) fewer mutations have been isolated, and iii) biological and clinical evidence strongly supports driver mutations in GCSFR. An important and interesting question is whether mutations exist that promote instability and contribute to the malignant phenotype but are lost beforehand. Existence of these "lost driver mutations" can be inferred using drift and selection mathematical models of population genetics, augmented to include genome rearrangements. Another important question is which of the variants identified by genome sequencing are truly deleterious. We developed an original algorithm based on capture-recapture paradigm, which may help establishing the likelihood of deleteriousness in absence of gold standard (Hicks et al., 2011).

The role of stochastic events in hematopoiesis has been discussed for the past 60 years since the beginnings of experimental hematology by Till and McCulloch (Whichard et al, 2010). The two opposing paradigms, deterministic hematopoiesis based on the firm regulation of peripheral

blood cell populations, and stochastic hematopoiesis based on variability observed in seeded bone marrow cells, are still awaiting a grand synthesis. This is in spite of the existence of substantial experimental findings, particularly those in the recent decade, using techniques of single-cell measurements. Disease-accompanying dynamics have been over the years variously modeled as deterministic or stochastic. Examples of stochastic phenomena observed in hematopoiesis include, but are not limited to:

- Stochastic fluctuations in the number of HSC making self-renewal versus commitment decisions result in high variability in the magnitude of the response to infection.

- The same stochastic fluctuations may lead to depletion of the HSC compartment when facing massive infection such as in neonatal sepsis.

- Presence of variant proteins in molecular switches responding to hematopoietic growth factors such as G-CSF leads to aberrant proliferation and leukemia, again with an important chance component.

- Molecular switches under stochastic fluctuations in molecular pathways and receptor noise may become reversible, which results in reversibility and plasticity at the level of the hematopoetic stem cell (HSC) and early committed cell level.

Recently, a third approach is emerging, which may be termed the molecular determinism (term coined based on ideas in Snijder and Pelkmans 2001 and Pelkmans 2012). According to molecular determinism, stochastic variability of the proliferating bone marrow cells can be reduced to complicated series of deterministic events including molecular switches, which are multistable by nature and which trigger proliferation and/or maturation decisions. This is distinct from older proposals involving chaotic dynamics (Raue et al. 2010, Laurent et al. 2010).

Mathematical, and in particular stochastic, principles have been used to explain the balance of factors contributing to behavior of a cell population as a whole. However, new techniques for gathering data and probing biological processes at a molecule and cell level continuously provide unprecedented amounts of new information, which leads to re-examination of these models. This has led to a renewed skepticism concerning stochastic modeling as a paradigm. As argued by Snijder and Pelkmans (2011), deterministic approach (or, what was called "molecular determinism" earlier on in the current paper) can resolve apparently stochastic phenomena with deterministic variability. They argue that cell-state parameters, such as cell size, growth rate, and cell cycle state, can be used to explain cell-to-cell variability, similarly as spatial cell population context parameters such as local cell density and location on cell colony edges. Tracing back cell-to-cell variability in time over multiple cell cycles may identify inherited, predetermining factors in cells of the same lineage. Snijder and Pelkmans (2011) also advocate repeated stimulation of the same cells to help identify the presence of deterministic factors in seemingly stochastic cell-to-cell variability. Complicated dynamics leading to chaotic (and sometimes indistinguishable from stochastic) behavior has been appreciated for some time. For example, existing mathematical models of cell cycle regulation (cf. e.g., Kimmel et al. 1984 and references therein) relies on nonlinear regulatory functions to control cell population distribution. However, these models also include a very real phenomenon of uneven allocation of constituents to progeny cells, which arguably is either truly stochastic or is indistinguishable from stochastic. Moreover, the idea of "backtracking" complicated (chaotic) trajectories seems to be doubtful from mathematical viewpoint. Schroeder (2011) discussed the need for long-term continuous follow-up on individual cells in order to understand the specific rules of proliferation and differentiation. This paper also touches upon issues such as influence of imaging techniques on cell behavior and difficulty with cell-tracking using existing software.

Returning to molecular determinism, a very good example of this approach seems to be the paper by Takizawa et al. (2012), concerning a purely deterministic and demand-driven integrated model of regulation of early hematopoiesis. This models is very complex and it involves "view of how cytokines, chemokines, as well as conserved pathogen structures, are sensed, leading to divisional activation, proliferation, differentiation, and migration of hematopoietic stem and progenitor cells, all aimed at efficient contribution to immune responses and rapid reestablishment of hematopoietic homeostasis". Takizava et al. (2012) paper is too involved physiologically to be discussed at length here. Let us notice that it contrasts with the simpler (and stochastic) models of Ogawa (1999) and Abkowitz et al. (1996). In these latter, the branching process paradigm is used at its simplest, with cells depicted as independent individuals, splitting at random and possibly interacting with a limited number of smaller entities.

Another current concept is that of non-genetic variability as a substrate for natural section, as espoused by Huang's group (Brock et al. 2009). For example, slow fluctuations in mammalian cells are the expression of heritability (memory) of protein abundance in successive generations of normal or cancer cells (Cohen et al. 2008, Sigal et al. 2006). One example is the non-inherited form of drug resistance in cancer. Theoreticians have been suggesting this for several decades, because of similar experimental evidence. The memory of protein abundance and dynamic homeostasis, which implied slow fluctuations in individual cells, were important constituents of many of the cell cycle regulation and unequal division models (Webb 1987, Arino and Kimmel 1987, Tyson and Hannsgen 1986). Development of resistance to chemotherapy by gene amplification (genetic, but non-mutation driven) have been pondered by theorists equally long ago (Harnevo and Agur 1991, Kimmel and Axelrod 1990).

Questions about the dynamics of hematopoiesis are resurfacing due to new experimental studies concerning lineage-specific growth factors, morphogens, the microenvironment, and the plasticity of stem cells. These

new findings allow a re-examination of two long-standing questions whether hematopoiesis is stochastic or deterministic and whether it is discrete or continuous. These issues exist for other non-hematopoietic stem cell systems; however, hematopoiesis serves as the most informative and accessible mammalian tissue system to look for answers (Whichard et al. 2010). Since quantitative systems analysis based on multi-scale modeling is needed to understand the complexity and dynamics of hematopoiesis, therefore determining the correct approach to this modeling is of more than academic interest.

2. MORAN MODEL OF GENETIC EVOLUTION UNDER DRIFT AND SELECTION

We will use the Moran model with selection (Durret, 2008). In this model, the population of granulocyte precursors is considered to be constant, with variable in time proportion of mutants and time runs in discrete units (e.g. days). We consider a population of N biological cells, which at time 0 contains i mutant cells. The mutant has selective advantage expressed by the relative fitness r, equal to the ratio of average progeny count of the mutant to that of the WT. For an advantageous mutant, $r > 1$. Under Moran Model, the probability of fixation of the mutant is equal to $P[T_0 > T_N] = (1 - 1/r^i)/(1 - 1/r^N)$, where T_0 and T_N are times to extinction fixation of the mutant. For large N, the expected time to fixation, given fixation occurs, is equal to $E[T_0 \mid T_0 > T_N] \sim 2\log(N)/(1\text{-}1/r)$ (Durrett, 2008). To obtain the exact expression for the expected time to fixation we can use the general expression for the probability that in a Markov chain, the first instance of hitting state j at step n starting from state i equals to f_n^i, the i-th entry of the column vector f_n

$$f_n = (P^{(j)})^{n-1} f_1$$

with entries $(f_n^0, \ldots, f_n^N)$ and $P^{(j)}$ is the transition probability matrix with the j-th column replaced by a column of 0's. Vector f_1 is set as the j-th column of the transition probability matrix P. For technical purposes, instead of matrix P, we use matrix $\tilde{P}$, which has $\tilde{P}_{00} = \tilde{P}_{NN} = 0$. This does not change the outcome of the iteration, but allows inversion of the matrices in expressions that follow. Finally, iteration yields

$$\sum_{n\geq 1} f_n = (I - \tilde{P}^{(j)})^{-1} f_1,$$

$$\sum_{n\geq 1} n f_n = (I - \tilde{P}^{(j)})^{-2} f_1.$$

Since in first entrance into N is equivalent to absorption, then the N-th entry of vector $P[T_0 > T_N] = \sum_n f_n^N$ and $E[T_N; T_0 > T_N] = \sum_n n f_n^N$. Application of these expressions requires inversion of very large matrices, but this is accomplished efficiently because of the matrices' band structure.

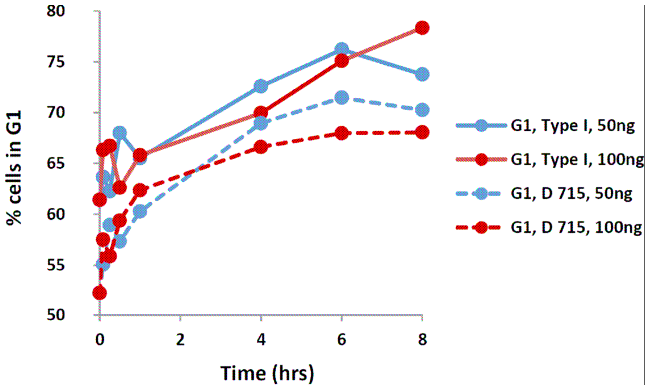

Figure 1 Summary of the dynamics of cell cycle distribution of BaF3 GR1 or D715 cells, following release from starvation block. Dose and receptor-type dependence of the %G1 cells. %G1 in D715 mutants is lower by about 5%, which considering the noise translates into 1-5% growth rate advantage ($\alpha = 0.95 - 0.99$; see Table 1).

3. FIXATION OF THE TRUNCATION MUTANT GCSFR D715.

Recognizing that ~70% of SCN patients who developed sMDS will express a truncation mutant GCSFR, we wish to compute the probability of fixation and the expected time to fixation of the D715 mutant, which has a growth rate advantage over the wild-type (WT) type I GCSFR. In proliferating cells, we assume that fitness advantage is conferred by difference of cell generation times of WT cells and that of mutant cells. If we denote by τ_w and τ_m the respective generation times, then $r = \exp[\Delta t\,(1/\tau_m - 1/\tau_w)] \approx 1 + (\Delta t/\tau_w)(1 - \alpha)/\alpha$ where $\alpha = \tau_m/\tau_w$ (Fig. 1). In other words, α measures the selective advantage of the mutant.

Table 1 Probabilities of fixation and expected times to fixation of the D715 variant of the GCSFR gene, under the Moran Model

N	α	i	$P[T_0>T_N]$	$E[T_N \mid T_0>T_N]$ (yr)	
				Asympt. time cont. Moran Model	Exact time discr. Moran Model
300,000	.60	7	.72	0.41	1.35
300,000	.80	19	.71	1.10	3.42
300,000	.95	91	.70	5.25	14.17
300,000	.98	235	.70	13.53	32.59
300,000	.99	476	.70	27.35	59.87
30,000	.99	476	.70	22.35	39.80
3,000,000	.99	476	.70	34.34	NA

Table 1 depicts outcomes of computations with several variants of parameter values. We estimate that the number of granulocyte progenitors in a large-volume aspirate (50 ml) is approximately 300,000. This is about 1/10 of the human bone marrow volume. We accept an upper bound value of 3,000,000, and for comparison a lower bound value of 30,000. We accept a cell cycle time of 96 hours in WT cells and 96α in mutants. We adjust the "initial" number i of mutant cells needed to obtain mutant fixation with about 70% probability, and then calculate the corresponding expected (mean) time to fixation. Let us note that at the arguably realistic range of α (0.95-0.99, see Figure 1) we obtain the "initial" counts of mutant cells in the range 91 – 476, and time to fixation range is 5.25 – 27.35 years. Altering the cell cycle time results in proportional change of time to fixation. Similarly, multiplying the total number N of cells results in change of the time to fixation.

4. INTERPRETATION OF RESULTS AND DISCUSSION

The expected times to fixation of the GCSFR D715 mutant in the range of (5-27) years are consistent with the timing of the sMDS onset. (From European SCN Registry data, age at diagnosis of SCN with sMDS and GCSFR mutation is 13 ± 9 years.) The 70% fixation probability requires 91 – 476 "initial" cells harboring the mutation. This might be resulting from the recurrent nature of the mutation; if this is the case, the "initial" time represents the initial period of SCN after administration of GCSF. To experimentally validate our model, we will measure the growth advantage of the D715 cells as supported by gene expression analysis.

A characteristic feature of human cancers is their very wide heterogeneity with respect to extent of involvement, genotype and rate of progression and spread (Hanahan and Weinberg, 2011). This is in contrast to induced animal tumors, which are relatively uniform. Secondary AML, resulting from a transition from severe congenital neutropenia via myelodysplastic syndrome, is not an exception, with onset varying from 1 to 38 years of age and with wide variability of mutational background. 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 clone 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.

Here, we concentrated on a single aspect of the disease, mutation leading to the MDS stage. We previously presented a paper (Kimmel and Corey, 2013) which addressed some aspects of the SNC → sMDS → sAML transition. Among other, although the framework we used allowed deriving 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 did not have at our disposal sequencing data to validate such an expression. Also, we did not attempt to fit the distribution of the age at diagnosis of the sMDS, since we were missing data on individual life histories, which would involve somatic mutation as well as sequencing data. These and other questions were postponed to a future study. From the mathematical point of view, the model of Kimmel and Corey (2013) was extremely simplified. It considered each new mutation to provide more selective advantage to the arising clone. Therefore, it was incapable of explaining the observation of Beekman et al. (2012), of mutation which appeared at the sMDS stage and disappeared at the sAML stage. This linear structure of mutation confered desirable simplicity to modeling but was 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 might be gauged.

ACKNOWLEDGEMENTS

TW and MK were supported by the grant DEC-2012/04/A/ST7/00353 from the National Science Center (Poland). TG and SC were supported by the NIH grant RO1-CA108992.

REFERENCES

Abkowitz, Janis L., Sandra N. Catlin, and Peter Guttorp. "Evidence that hematopoiesis may be a stochastic process in vivo." Nature medicine 2, no. 2 (1996): 190-197.

Arino, Ovide, and Marek Kimmel. "Asymptotic analysis of a cell cycle model based on unequal division." SIAM Journal on Applied Mathematics 47, no. 1 (1987): 128-145.

Beekman, R., Valkhof, M.G., Sanders, M.A., van Strien, P.M., Haanstra, J.R., Broeders, L., Geertsma-Kleinekoort, W.M., Veerman, A.J., Valk, P.J., Verhaak, R.G., Löwenberg, B., Touw, I.P. (2012). Sequential gain of mutations in severe congenital neutropenia progressing to acute myeloid leukemia. *Blood*, 119: 5071-5077.

Bozic, I., Antal, T., Ohtsuki, H., Carter, H., Kim, D., Chen, S., Karchin, R., Kinzler, K.W., Vogelstein, B., Nowak, M.A. (2010). Accumulation of driver and passenger mutations during tumor progression. *Proc Natl Acad Sci USA*, 107: 18545–18550.

Brock, Amy, Hannah Chang, and Sui Huang. "Non-genetic heterogeneity—a mutation-independent driving force for the somatic evolution of tumours." Nature Reviews Genetics 10, no. 5 (2009): 336-342.

Cohen, Ariel A., et al. "Dynamic proteomics of individual cancer cells in response to a drug". Science, 2008, 322.5907: 1511-1516.

Durrett, R. (2008). Probability Models for DNA Sequence Evolution (Probability and Its Applications). *Springer* 2008.

Glaubach, Taly, and Seth J. Corey. "From famine to feast: sending out the clones." Blood 119, no. 22 (2012): 5063-5064.

Griffiths, R.C., Pakes, A.G. (1988) An infinite-alleles version of the simple branching process. *Adv Appl Probab,* 20: 489–524.

Hanahan, D., Weinberg, R.A. (2011) Hallmarks of cancer: the next generation. *Cell,* 144: 646-674.

Harnevo, Linda E., and Zvia Agur. "Drug resistance as a dynamic process in a model for multistep gene amplification under various levels of selection stringency." Cancer chemotherapy and pharmacology 30, no. 6 (1992): 469-476.

Hicks, S., Wheeler, D.A., Plon, S.E., Kimmel, M. (2011). Prediction of missense mutation functionality depends on both the algorithm and sequence alignment employed. *Human Mutation*, 32: 661-668.

Kimmel, Marek, and David E. Axelrod. "Mathematical models of gene amplification with applications to cellular drug resistance and tumorigenicity."Genetics 125, no. 3 (1990): 633-644.

Kimmel, M., Corey, S. (2013). Stochastic hypothesis of transition from inborn neutropenia to AML: interactions of cell population dynamics and population genetics. *Frontiers in Oncology 3,* 89.

Kimmel, Marek, Zbigniew Darzynkiewicz, Ovide Arino, and Frank Traganos. "Analysis of a cell cycle model based on unequal division of metabolic constituents to daughter cells during cytokinesis." Journal of theoretical biology 110, no. 4 (1984): 637-664.

Kimmel, M., Mathaes, M. (2010). Modeling neutral evolution of Alu elements using a branching process. *BMC Genomics*, 11: Suppl 1: S11.

Laurent, Michel, Jean Deschatrette, and Claire M. Wolfrom. "Unmasking chaotic attributes in time series of living cell populations." PloS ONE 5, no. 2 (2010): e9346.

Ogawa, M. "Stochastic model revisited." International journal of hematology 69, no. 1 (1999): 2.

Parkin, Brian, Peter Ouillette, Yifeng Li, Jennifer Keller, Cindy Lam, Diane Roulston, Cheng Li, Kerby Shedden, and Sami N. Malek. "Clonal evolution and devolution after chemotherapy in adult acute myelogenous leukemia." Blood121, no. 2 (2013): 369-377.

Pelkmans, Lucas. "Using cell-to-cell variability—a new era in molecular biology." Science 336, no. 6080 (2012): 425-426.

Raue, A., V. Becker, U. Klingmüller, and J. Timmer. "Identifiability and observability analysis for experimental design in nonlinear dynamical models."Chaos: An Interdisciplinary Journal of Nonlinear Science 20, no. 4 (2010): 045105-045105.

Schroeder, Timm. "Asymmetric cell division in normal and malignant hematopoietic precursor cells." Cell stem cell 1, no. 5 (2007): 479-481.

Sigal, Alex, et al. "Variability and memory of protein levels in human cells."Nature 444.7119 (2006): 643-646.

Snijder, Berend, and Lucas Pelkmans. "Origins of regulated cell-to-cell variability." Nature Reviews Molecular Cell Biology 12, no. 2 (2011): 119-125.

Takizawa, Hitoshi, Steffen Boettcher, and Markus G. Manz. "Demand-adapted regulation of early hematopoiesis in infection and inflammation." Blood 119, no. 13 (2012): 2991-3002.

Tyson, John J., and Kenneth B. Hannsgen. "Cell growth and division: a deterministic/probabilistic model of the cell cycle." Journal of mathematical biology 23, no. 2 (1986): 231-246.

Webb, G. F. "Random transitions, size control, and inheritance in cell population dynamics." Mathematical biosciences 85, no. 1 (1987): 71-91.

Whichard, Zakary L., Casim A. Sarkar, Marek Kimmel, and Seth J. Corey. "Hematopoiesis and its disorders: a systems biology approach." Blood 115, no. 12 (2010): 2339-2347.