Analysis of two mechanisms of telomere maintenance based on the theory of g-Networks and stochastic automata networks

Kyung Hyun Lee1† and Marek Kimmel1,2*†

From The Sixth International Workshop on Computational Network Biology: Modeling, Analysis, and Control (CNB-MAC 2019)
Niagara Falls, NY, USA. 07 September 2019

Abstract

Background Telomeres, which are composed of repetitive nucleotide sequences at the end of chromosomes, behave as a division clock that measures replicative senescence. Under the normal physiological condition, telomeres shorten with each cell division, and cells use the telomere lengths to sense the number of divisions. Replicative senescence has been shown to occur at approximately 50–70 cell divisions, which is termed the Hayflick’s limit. However, in cancer cells telomere lengths are stabilized, thereby allowing continual cell replication by two known mechanisms: activation of telomerase and Alternative Lengthening of Telomeres (ALT). The connections between the two mechanisms are complicated and still poorly understood.

Results In this research, we propose that two different approaches, G-Networks and Stochastic Automata Networks, which are stochastic models motivated by queueing theory, are useful to identify a set of genes that play an important role in the state of interest and to infer their previously unknown correlation by obtaining both stationary and joint transient distributions of the given system. Our analysis using G-Network detects five statistically significant genes (CEBPA, FOXM1, E2F1, c-MYC, hTERT) with either mechanism, contrasted to normal cells. A new algorithm is introduced to show how the correlation between two genes of interest varies in the transient state according not only to each mechanism but also to each cell condition.

Conclusions This study expands our existing knowledge of genes associated with mechanisms of telomere maintenance and provides a platform to understand similarities and differences between telomerase and ALT in terms of the correlation between two genes in the system. This is particularly important because telomere dynamics plays a major role in many physiological and disease processes, including hematopoiesis.

Keywords: Queueing network theory, G-networks, Stochastic automata networks, Correlation analysis, Gene regulatory networks, Telomeres

*Correspondence: kimmel@rice.edu
Kyung Hyun Lee is the first author.
†Kyung Hyun Lee and Marek Kimmel contributed equally to this work.
1Department of Statistics, Rice University, 6100 Main Street, Houston 77057, TX, USA
2Department of Systems Biology and Engineering Silesian University of Technology, Akademicka 16, 44-100 Gliwice, Poland

© The Author(s). Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated in a credit line to the data.
Background

The introduction of Stochastic Automata Networks (SANs) [1] has led researchers to explore an efficient solution of finite multi-component Markov chains of potentially very high dimension. The main idea of the SAN formalism is to employ Kronecker algebra operations for generating the infinitesimal generator (also known as an intensity matrix or transition rate matrix) of a Markov chain in a product form in order to ease the problem of dimensionality and the complexity of the vector-matrix product [2, 3]. Specifically, a compact representation using tensor product and sums of matrices of the form, which is formally proved in [1]

\[ Q = \bigoplus_{i=1}^N L_i + \sum_{r=1}^R \left( \bigotimes_{i=1}^N M_{r,i} + \bigotimes_{i=1}^N N_{r,i} \right), \tag{1} \]

provides the transition rates of a time-continuous Markov chain of the type just described. Notation: \( N \) is the number of automata, \( R \) is the number of synchronizations, and \( L_i \) and \( M_{r,i} \) are matrices containing information about the local transitions and the effect of the synchronization \( r \) on the automaton \( i \), respectively. \( N_{r,i} \) is the normalizing matrix of \( M_{r,i} \), and \( \otimes \) and \( \oplus \) denote, respectively, the (generalized) tensor product and tensor sum [4]. Notation will be discussed in detail in the following section.

For a decade after SAN introduction, many outstanding analytical results have been proved. For example, Plateau and Stewart [5] proved in year 2000 that a product-form steady-state distribution for SANs without synchronizations exists, as long as some numerical conditions related to local balance are satisfied. Prior to that, Boujdaine et al. in 1997 [6] considered a special class of SAN having limited synchronization, and proved a sufficient condition for existence of stationary distribution, by applying properties of Kronecker sum. Note that classic queuing networks such as Jackson’s networks and G-Networks with positive and negative customers [7] fall under this special class.

Most analyses and applications of SANs focus on finding steady-state distribution of queuing system models. To our knowledge, correlation analysis in SANs is rarely examined, although correlation quantifies the degree of inter-relatedness of two automata. It contributes to the understanding of how the association of two automata changes with time according to the state of interest, with the behavior of other automata in the system simultaneously considered. In the current study, we first examine the exact-form of an infinitesimal generator for G-Networks with positive and negative customers. Then it is applied to a gene regulatory network (GRN) having five genes related to the onset of cancer in order to show the time-dependent dynamics of transition rates and correlation between two particular genes of interest. Throughout this paper, our analysis is based on time-continuous Markov chains, but we note that most results are valid in the time-discrete case provided complications such as periodicity are excluded.

Stochastic automata network

The SAN consists of a number of automata, which are correlated by synchronizing transitions. Each automaton has states and transitions. The state space of the system is the Cartesian product of the states of the automata [6]. There exist two types of transitions: local and synchronizing.

1. Transition in one automaton may initiate a new simultaneous transition in another automaton. Such transitions are collectively referred to as synchronizing transitions. In one synchronizing transition, two automata are paired. The first is called a master automaton; it affects the state of another automaton as its state changes. The second is called a slave automaton; it is affected by its master automaton [8].

2. In any given automaton, transitions not classified as synchronizing transitions are local transitions. Such transitions only change the state of one automaton.

In the SAN, two assumptions are made. First, the time to transition is exponentially distributed, and second, that a multidimensional Markov chain represents the SAN, although a particular automaton may not be Markovian itself. Then the (global) infinitesimal generator (the matrix of time derivatives at 0 of the transition probabilities), which reflects a change in transition probability from one state to another, is given by Eq. 1, with notation explained in Table 1.

Specifically, each matrix in \( Q \) is defined as follows:

- \( L_i \) is a local normalized transition rate matrix of automaton \( i \); thus

\[ L_i[ m, n ] \geq 0 \text{ if } m \neq n \text{ and } \sum_{n=1}^N L_i[ m, n ] = 0 \]

| Notation | Definition |
|----------|------------|
| \( N \)  | Number of automata in the network |
| \( K \)  | Number of states, where the state space of each automaton is \( \{ 0, 1, \ldots, K - 1 \} \) |
| \( R \) and \( r \) | Number of synchronizing events and an index of each event, respectively |
| \( L_i \) | Local transition rate matrix (normalized) of the automaton \( i \) |
| \( M_{r,i} \) | Effect matrix corresponding to the \( r^{th} \) synchronizing event acting on the automaton \( i \) |
| \( N_{r,i} \) | Normalizing matrix needed for the tensor product of \( M_{r,i} \) to constitute a transition rate matrix |
• For each synchronizing transition, for all 
\( i \in \{1, 2, \cdots, N\}\), \( M_{r,i} = I_K \), where \( I_K \) is an identity
matrix of size \( K \), except for the case when \( r \) is two
indices corresponding to master and slave
automaton. Therefore

\[
\bigotimes_{i=1}^{N} M_{r,i} = I_K \otimes \cdots \otimes I_K \otimes (D_{r,msr(r)} - D_{r,msr(r)pr}) \\
\otimes I_K \otimes \cdots \otimes I_K \otimes E_{r,sl(r)} \otimes I_K \otimes \cdots \otimes I_K
\]

where \( msr(r) \) and \( sl(r) \) are indices for master
automaton and slave automaton, respectively, in the
\( r^{th} \) synchronization. Here, \( D_{r,msr(r)} \) denotes the
transition rate matrix due to the \( r^{th} \) synchronizing
event on the master automaton, and \( D_{r,msr(r)pr} \) is its
diagonal matrix. \( E_{r,sl(r)} \), the transition probability
matrix, is the effect of the \( r^{th} \) synchronizing
event on the automaton \( sl(r) \).

\[
D_{r,msr(r)}[m,n] \geq 0 \text{ if } m \neq n \quad \text{and} \quad \sum_{n=1}^{N} D_{r,msr(r)}[m,n] = 0, \forall m \\
E_{r,sl(r)}[m,n] \geq 0 \quad \text{and} \quad \sum_{n=1}^{N} E_{r,sl(r)}[m,n] = 1, \forall m
\]

• A normalizing matrix \( \bigotimes_{i=1}^{N} N_{msr} \) is a
\((K^N \times K^N) \)-dimensional diagonal matrix, whose
\( r^{th} \) diagonal element is the negative sum of the \( l^{th} \) row
of \( \bigotimes_{i=1}^{N} M_{r,i} \).

G-Networks with positive and negative customers

Gelenbe- or G-Networks [9], devised by Erol Gelenbe, are
Markovian stochastic models grounded in queueing
network theory [10]. They were extensively applied to study
probability models of various objects [11–18]. However they
are particularly appropriate to construct GRNs, as
they introduce a novel notion called a ‘negative customer’,
which can be biologically interpreted as a repression
signal [19]. G-Networks model the number of positive
customers, which is mRNA expression level biologically, in an
automaton or a gene having an infinite state space. Positive
and negative customers synchronously move from a
master gene to a slave gene within the system with transition
probabilities, \( p_{msr(r),sl(r)}^+ \) and \( p_{msr(r),sl(r)}^- \), respectively. It implies that there are two types of the synchronizing transition. Note that the behavior of positive customers identical to that of positive customers in Jackson networks [20]. In contrast, negative customers function distinctively as follows: they are not accumulated and instantaneously leave a queue after the completion of their tasks. If a negative customer arrives at a non-empty queue, it destroys one current positive customer. However if the queue is empty, then the negative customer disappears without locally affecting the queue. Both types of customers arrive at the \( i^{th} \) gene from outside of the system at rates \( \lambda_i^+ \) and \( \lambda_i^- \), respectively. It is assumed that service disciplines
are the same for all queue [21], and the service times for
each queue \( i \) are independent and identically exponentially
distributed with rates \( \mu_i \in (0, \infty) \) for \( i = 1, 2, \cdots, N \).

Figure 1 visually illustrates four activities of mRNA and/or
protein molecules, and Table 2 suggests how the biological
terms that refer to these activities can be identified with
those used in G-Networks.

G-Networks have an advantageous property of being
analytically tractable because of the existence of product-
form stationary distributions under the typical Markov
Chain assumptions as shown in Eq. 2 [7,22].

Let \( x = \{x_1, x_2, \cdots, x_N\} \) be the vector of non-
negative integers representing the state of the
network. Then the time-dependent vector, \( x(t) : t \geq 0 \), is a
continuous-time Markov chain, which satisfies the system
of Chapman-Kolmogorov equations [23]. Element \( x_i(t) \),
\( 1 \leq i \leq N \), of vector \( x(t) = (x_1(t), x_2(t), \cdots, x_N(t)) \), is the
number of customers (or the mRNA expression level) of
gene \( i \) at time \( t \). It has been proved that the joint steady-
state distribution \( \pi(x) \) of \( x(t) \), has the form of a product
of stationary distributions of each queue,

\[
\pi(x) = \lim_{t \to \infty} P(X(t) = x) = \prod_{i=1}^{N} (1 - q_i) \cdot q_i^x
\]

each satisfying the balance equation of the G-Networks
[23–25], i.e.

\[
q_i = \lim_{t \to \infty} P(x_i(t) > 0) = \frac{\lambda_i^+ + \sum_{j=1}^{N} q_j \cdot \mu_j \cdot p_{ji}^+}{\mu_i + \lambda_i^- + \sum_{j=1}^{N} q_j \cdot \mu_j \cdot p_{ji}^-} > 0,
\]

for \( i = 1, 2, \cdots, N \). It implies that the stationary
probability for each positive recurrent state is expressed in
the terms of a product of functions depending solely on the
state of a single queue.

According to [8], matrices constituting the global
infinitesimal generator \( Q \) are defined as follows in the

![Fig. 1 Four activities for gene regulation in a G-network model](image-url)


Table 2 Vocabulary and notation for application of G-Networks to Gene Regulatory Networks

| Biology                                      | G-Networks                                      | Notation |
|---------------------------------------------|-------------------------------------------------|----------|
| Gene                                        | Node where customers are stored                 |          |
| Signals/notch in the scale of the gene activity information [19] | Positive/negative customers                     |          |
| mRNA expression level                       | The number of positive $x_i$ customers          |          |
| Translation / protein bursting              | Arrival rate of positive customers $\lambda^+_i$ from the outside of the system |          |
| Degradation                                 | Arrival rate of negative customers $\lambda^-_i$ from the outside of the system |          |
| Activation / transcription                  | Transition probabilities of $p^+_i$ positive customers |          |
| Repression                                  | Transition probabilities of $p^-_i$ negative customers |          |
| Signals not influencing the gene activity  | Customers that exit the $i$ system              |          |
| Protein-protein interaction                 | Service rate $\mu_i$                           |          |

The first and second columns contain biological and G-Networks terminologies of gene regulation, respectively. The third column includes the corresponding notations used in this study. The subscript $i$ indicates the $i$th gene.

terms of customer-related rates:

$$L_i = \lambda^+_i \left( Upp - \mu_i \right) + \mu_i \cdot \left( Low - \lambda^+_i \right) + \lambda^-_i \left( Low - \lambda^-_i \right), \forall i \in \mathbb{N}$$

$$D_{r, \text{mar}}(r) = \begin{cases} 
\mu_i \cdot \lambda^+_i \left( \mu_i \cdot \lambda^+_i \right), & \text{if a customer enters a master automaton and moves to a slave automaton as a positive customer} \\
\mu_i \cdot \lambda^-_i \left( \mu_i \cdot \lambda^-_i \right), & \text{if a customer enters a master automaton and moves to a slave automaton as a negative customer} 
\end{cases}$$

$$E_{r, \text{mar}}(r) = \begin{cases} 
Upp, & \text{if a positive customer is received at a slave automaton} \\
Low, & \text{if a negative customer is received at a slave automaton} 
\end{cases}$$

$$N_{r, \text{mar}} = \begin{cases} 
I_K, & \text{if a slave automaton receives a positive customer} \\
I_K, & \text{if a slave automaton receives a negative customer} 
\end{cases}$$

where

- $I$ : Identity matrix
- $Upp$ : Matrix with entries 0 except the main upper diagonal which is 1
- $Low$ : Matrix with entries 0 except the main lower diagonal which is 1
- $I^0_1$ : Identity matrix except the first diagonal element which is 0
- $I^0_K$ : Identity matrix except the $K^{th}$ diagonal element which is 0

Results

This section consists of two parts. The first part shows the mathematical result that derives the correlation between two genes utilizing Kronecker algebra. The second part contains the biological results based on the application of the mathematical result to the gene regulatory network associated with telomere biology.

Mathematical result: derivation of joint transient distribution of two automata

If $Q$ is the infinitesimal generator and $P(t)$ is the transition probability matrix of a finite Markov chain, we can derive a set of differential equations called Kolmogorov’s forward equations, in the form of $P'(t) = P(t) \cdot Q$. Solution of the forward equation is given in this case by the power series expansion that converges for any square matrix $Q$:

$$P(t) = e^{Qt} = \sum_{n=0}^{\infty} \frac{(Qt)^n}{n!}. \quad (4)$$

The resulting $(1 \times K^N)$-dimensional time-dependent state probability vector at time $t$ has the form

$$\pi(t) = \pi(0) \cdot e^{Qt} \quad (5)$$

where $\pi(t)$ is a state probability vector at time $t$ [26]. In order to find a joint state probability vector $\pi_{ij}(t)$ of two automata, say $i$ and $j$, we introduce two matrices $C_{ij}$ and $C_{ij}^*$ which are $(K^N \times K^2)$- and $(K^2 \times K^N)$-dimensional, respectively. After several steps of calculation, we obtain

$$\pi_{ij}(t) = \pi(0) \cdot \frac{C_{ij}}{1 \times K^N \cdot K^N \times K^2} \cdot e^{Qt} \cdot C_{ij}^*, \quad (6)$$

where, for all $i \neq j$, $i, j = 1, 2, \ldots, N$, $C_{ij}$ and $C_{ij}^*$ are defined by

$$C_{ij} = \sum_{n=1}^{N} U_{i,j,n} \cdot V_{i,j,n} = \begin{cases} 
I_K, & \text{if } n = i \text{ or } n = j \\
1_K, & \text{if } n \neq i \text{ and } n \neq j 
\end{cases}$$

$$C_{ij}^* = \sum_{n=1}^{N} V_{i,j,n} \cdot U_{i,j,n} = \begin{cases} 
I_K, & \text{if } n = i \text{ or } n = j \\
1_K, & \text{if } n \neq i \text{ and } n \neq j 
\end{cases}$$

Note that the $(1 \times K)$-dimensional matrix $1^T_K = [1 \ 1 \ \cdots \ 1]$ indicates the matrix transpose of $1_K$. This results in

$$C_{ij}^* \cdot C_{ij} = \left( \sum_{n=1}^{N} V_{i,j,n} \right) \cdot \left( \sum_{n=1}^{N} U_{i,j,n} \right) = K^{N-2} \cdot I_{K^2}, \quad (7)$$

where

$$V_{i,j,n} \cdot U_{i,j,n} = \begin{cases} 
I_K, & \text{if } n = i \text{ or } n = j \\
K, & \text{if } n \neq i \text{ and } n \neq j. 
\end{cases}$$
Therefore Eq. 6 finally becomes
\[ e^{Q_i \cdot t} = \frac{1}{K^{N-2}} \cdot C_{ij} \cdot e^{Q_i \cdot t} \cdot C_{ij} \cdot \] (8)

Equation 8 implies that a time-dependent joint state probability vector \( \pi_{ij}(t) \) can be easily calculated, once the infinitesimal generator, \( Q \), of the entire system is obtained.

\[ \pi_{ij}(t) = \pi_{ij}(0) \cdot e^{Q_i \cdot t} \]

\[ = \frac{1}{K^{N-2}} \cdot \pi_{ij}(0) \cdot C_{ij}^* \cdot e^{Q_i \cdot t} \cdot C_{ij} \]

\[ = \left\{ p_{0,0}(t), p_{0,1}(t), p_{0,2}(t) \cdots, p_{K-1,K-2}(t), p_{K-1,K-1}(t) \right\} \] (9)

We can then form a \((K \times K)\)-dimensional table that describes the joint probabilities for each time point \( t \). Based upon it, the association between two genes for all \( t \) is evaluated using the Pearson product-moment correlation coefficient [27]. It implies the correlation of a pair of genes in the regulatory system varies by each time (each generation of cells) as shown in Fig. 2.

Application to the gene regulatory networks

GRNs and parameters from g-Networks

In this research, we use a GRN provided in Fig. 3. It consists of 5 statistically significant genes associated with telomere maintenance, which are discovered by (modified) Abnormal Pathway Detection Algorithm (APDA) based on G-Networks [18]. (See Methods for detailed information on the modification). We explore the correlation between a pair of genes, e.g., CEBPA and hTERT, which is not apparent from Fig. 3. On the biology side, Kirwan et al. in 2009 [28] reported that mutations in hTERT within 4 of 20 families were identified in familial acute myeloid leukemia (AML). It is known that the mutation of CEBPA is one of the important factors in AML and its prognosis [29]. However the direct relationship between CEBPA and AML is still not satisfactorily understood. The joint state probability vector of two genes, which can be calculated from Eq. 9, is expected to show the flow of probabilities for the \( K^2 \)-tuples of mRNA expression levels at each fixed time \( t \), and thereby illustrates the relationship between any two genes.

In this case, the number of automata is equal to \( N = 5 \), while the number of synchronizing events is equal to \( R = 7 \). We limit the number of gene expression levels to \( K = 7 \) for two reasons. The first is that, because the data are normalized, there are few genes with a mRNA expression level of 7 or higher. Furthermore, the memory limitations enforce it. In this paper, we assume that pairs of genes are highly positively correlated at the initial time point; that is

\[ p_{ij}(0) = \begin{bmatrix}
\frac{5}{7} & \frac{1}{7} & 0 & 0 & 0 & 0 & 0 \\
\frac{1}{7} & \frac{5}{7} & 0 & 0 & 0 & 0 & 0 \\
0 & \frac{1}{7} & \frac{5}{7} & 0 & 0 & 0 & 0 \\
0 & 0 & \frac{1}{7} & \frac{5}{7} & 0 & 0 & 0 \\
0 & 0 & 0 & \frac{1}{7} & \frac{5}{7} & 0 & 0 \\
0 & 0 & 0 & 0 & \frac{1}{7} & \frac{5}{7} & \frac{1}{7} \\
0 & 0 & 0 & 0 & 0 & \frac{1}{7} & \frac{5}{7} \\
\end{bmatrix} \]

Note that this matrix needs to be transformed into a \((1 \times K^2)\)-dimensional joint state probability vector as
Fig. 3 A network with 5 automata (genes) and 7 synchronizing transitions

$\pi_i(t)$ in Eq. 9. Tables 3 and 4 exhibit the values of parameters that are required to build the infinitesimal generator $Q$. They are estimated and optimized according to the APDA [18] based on G-Networks. Specifically, values in Table 3 allow to calculate a joint probability vector $\pi_{ij}(t)$, under the normal condition, with the Table 4 contains information for malignant (ALT or active telomerase) cells. The three families of parameters, $\lambda_i^+$, $\lambda_i^-$ and $\mu_i$ for all $i$, are identical for both normal and ALT cells, while transition probabilities, $p_{ij}^+$, $p_{ij}^-$ and $d_i$, are different between the cell types. Transition probabilities are deciding factors in correlation coefficients based on Eq. 9 and the rates of convergence to stationarity in G-Networks.

Table 3 Values of the parameters needed to determine the infinitesimal generator (Q) using 4 normal cell lines, where $d_i = 1 - \sum_{j=1}^{N} (p_{ij}^+ + p_{ij}^-)$

| Gene | Translation ($\lambda_i^+$) | Degradation ($\lambda_i^-$) | $\mu_i$ | $d_i$ |
|------|--------------------------|--------------------------|--------|------|
| CEBPA | 4 | 1.18489 | 3 | 0.333 |
| E2F1 | 4 | 1.79118 | 3 | 0.333 |
| FOXM1 | 4 | 2.78510 | 2 | 0.5 |
| c-MYC | 5 | 6.70809 | 3 | 0.333 |
| hTERT | 5 | 6.07682 | 1 | 1 |

Normal: Activation/transcription processes ($p_{ij}^+$)

| Gene | CEBPA | E2F1 | FOXM1 | c-MYC | hTERT |
|------|-------|------|-------|-------|-------|
| CEBPA | 0 | 0 | 0 | 0 | 0 |
| E2F1 | 0 | 0 | 0 | 0.333 | 0 |
| FOXM1 | 0 | 0 | 0 | 0.5 | 0 |
| c-MYC | 0 | 0 | 0 | 0 | 0.333 |
| hTERT | 0 | 0 | 0 | 0 | 0 |

Normal: Repression process ($p_{ij}^-$)

| Gene | CEBPA | E2F1 | FOXM1 | c-MYC | hTERT |
|------|-------|------|-------|-------|-------|
| CEBPA | 0 | 0.333 | 0.333 | 0 | 0 |
| E2F1 | 0 | 0 | 0 | 0 | 0.333 |
| FOXM1 | 0 | 0 | 0 | 0 | 0 |
| c-MYC | 0.333 | 0 | 0 | 0 | 0 |
| hTERT | 0 | 0 | 0 | 0 | 0 |

The “service” (or firing) rate of gene $i$, denoted by $\mu_i$, represents the protein–protein interactions, e.g., phosphorylation and ubiquitination. Gene $i$ activates and inhibits gene $j$ with probability $p_{ij}^+$ and $p_{ij}^-$, respectively. Genes in the rows correspond to the “starting” genes, while those in columns to the ending genes.
### Table 4 Values of the parameters that are needed to determine the infinitesimal generator ($Q$) using 18 ALT cell lines

| Parameter | Translation ($\lambda^+$) | Degradation ($\lambda^-$) | $\mu_i$ | $d_i$ |
|-----------|---------------------------|---------------------------|--------|-------|
| CEBPA     | 4                         | 1.18489                   | 3      | 0.5433|
| E2F1      | 4                         | 1.79118                   | 3      | 0.1258|
| FOXM1     | 4                         | 2.78510                   | 2      | 0.2417|
| c-MYC     | 5                         | 6.07809                   | 3      | 0.0280|
| hTERT     | 5                         | 6.07682                   | 1      | 1     |

ALT: Activation/transcription processes ($p^+_ij$)

|                  | CEBPA | E2F1 | FOXM1 | c-MYC | hTERT |
|------------------|-------|------|-------|-------|-------|
| CEBPA            | 0     | 0    | 0     | 0     | 0     |
| E2F1             | 0     | 0    | 0     | 0     | 0.5574|
| FOXM1            | 0     | 0    | 0     | 0     | 0.7583|
| c-MYC            | 0     | 0    | 0     | 0     | 0.3855|
| hTERT            | 0     | 0    | 0     | 0     | 0     |

ALT: Repression process ($p^-_{ij}$)

|                  | CEBPA | E2F1 | FOXM1 | c-MYC | hTERT |
|------------------|-------|------|-------|-------|-------|
| CEBPA            | 0     | 0.2268| 0.2230| 0     | 0     |
| E2F1             | 0     | 0    | 0     | 0     | 0     |
| FOXM1            | 0     | 0    | 0     | 0     | 0     |
| c-MYC            | 0.5865| 0    | 0     | 0     | 0     |
| hTERT            | 0     | 0    | 0     | 0     | 0     |

The definitions and representations of $d_i$, $\mu_i$, $p^+_ij$, and $p^-_{ij}$ are the same as in Table 3.

### Simulation study using estimated parameters

Estimated and/or assumed values of the parameters listed in Tables 3 and 4 determine the infinitesimal generator, $Q$, and the steady-state distribution of each gene. Developed on the global balance equation of G-Networks [23] and the exponentially-distributed holding times of Markov chains [30], the empirical cumulative density function (ECDF) was exploited to confirm that parameter estimation via the APDA is appropriate aligning with the theory of G-Networks. We compare the ECDF of simulated data utilizing values in Tables 3 and 4 and the theoretical cumulative density function (CDF) based on $q_i$ in Eq. 3. The algorithm for the simulation is explained in detail in Methods.

The ECDF, usually denoted by $\hat{F}_n(x)$, is associated with cumulative frequency of observations (empirical sampled data). It is a non-parametric estimator of the underlying CDF of a random variable of interest and is formally defined as follows:

$$\hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}_{[x_i \leq x]} ,$$

where $\mathbb{I}_{[\cdot]}$ is an indicator function. That is, the ECDF is a step function that increases by $\frac{1}{n}$ at each datum. The stationary probability distribution of each gene in G-Networks is in the geometric product-form as stated in Eq. 2, which results in the following tail of the theoretical CDF, $T(x_i)$:

$$T(x_i) = 1 - F(x_i) = 1 - \left(1 - q_i^{x_i+1}\right) = q_i^{x_i+1} .$$

Figure 4 (and Fig. 9) illustrates how the ECDF of hTERT utilizing estimated/assumed values of the parameters (Tables 3 and 4) converges to the theoretical CDF regardless of cell types. Similar patterns are obtained for all 5 statistically significant genes in this research.

### Correlation between a pair of genes

We discuss the similarities or differences in the correlation of two genes between normal and malignant cells with either active telomerase or ALT. Figure 5 based on Eq. 9 illustrates the following:

- First, Fig. 5 demonstrates the change of correlation between a pair of genes in the given system over time in normal cells and ALT (malignant) cells. We can then infer both the trend of correlation between a pair of genes and the speed to the transient state, i.e., before it reaches a steady state. Due to the product-form stationary probability distribution of network state in the theory of G-Networks, the correlation coefficient becomes 0 when the system becomes stationary [31]. Assuming that the pair of genes are positively correlated at time 0, the positive correlation persists over time in both types of cells (normal and ALT) and practically reaches 0 (the steady state) approximately at $t = 0.8$. As shown in the figure, the patterns of correlation between each pair of genes do not noticeably differ among the types of cells (normal or malignant with either ALT or
active telomerase shown), although they are dissimilar from one pair to another within the same type of cells. For example, the correlation between CEBPA and FOXM1 in a transient state is stronger than that of c-Myc and hTERT, and it reaches a steady state more slowly in all cell types.

- Second, Eq. 9 uncovers the interactions between (1) CEBPA and hTERT, (2) FOXM1 and hTERT and (3) FOXM1 and E2F1, which are unknown in Fig. 3. The pattern in Fig. 5, which is analogous to Fig. 7, delineates that the relationship of the three pairs of genes is similar regardless of the mechanisms by which telomeres maintain their sufficient lengths. However, within the same condition of cells, the magnitude of correlation and the speed to the steady state vary by interactions. It is supported by the interaction of CEBPA and hTERT, which is relatively weaker than that of other pairs, even though its trend over time is indistinguishable by the cell types.

Note that if we conversely assume that a pair of genes are highly negatively correlated at the initial time point, then the negative correlation prevails over time as shown in Fig. 8.

Discussion

One important advantage of using mathematical modeling to understand dynamics of a biological system is that the model can be a cost-effective and time-saving supplement or even a substitute for laboratory experiments in which patterns of the system are delicate and complex. Especially the stochastic approach is useful to quantify the role of fluctuations in the behavior of the system of interest. In this paper, we propose that a stochastic paradigm involving G-Networks and Stochastic Automata Networks both stemming from queuing theory, can contribute to the analysis of similarities and differences between telomerase activation and ALT using genes related to telomere maintenance. However, there are the following two caveats.

- First, correlation coefficient, whose sign relies upon the initial joint state probability vector $\pi_y(0)$ in our example, evaluates the strength of the evidence for a relationship of two genes but does not determine causal relationships. In other words, even if a positive correlation between CEBPA and hTERT in ALT cells is revealed, it is not known which gene activates another one.

![Fig. 4 ECDF plots of hTERT under the normal condition (top) and ALT (malignant) condition (bottom). Orange and green straight and thick lines represent theoretical (logarithmically transformed) CDF of normal and ALT conditions, respectively. Purple, blue (thin) and red (thick) lines serve as the logarithmically transformed ECDF of hTERT after 10 steps, 100 steps and 500 steps, respectively.](image)

![Fig. 5 Trend of correlation between each pair of genes in the system over time in ALT (malignant) cells. The general relationship of two genes in the system is similar regardless of the mechanism by which telomeres maintain their sufficient lengths. However, within the same condition of cells, the magnitude of correlation and the speed to the steady state vary by interactions.](image)
Second, the dimension of the infinitesimal generator, \( Q \), which is a function of tensor products, can be enormously large, with the corresponding matrix being sparse depending on the number of states of each queue. The state space used in this research is relatively small. Computation in larger dimensions can be made practical by using Compressed Sparse Row and a number of methods for the approximation of the matrix exponential [32] such as Krylov-type techniques.

**Conclusion**

In spite of the caveats discussed in the previous section, the APDA based on G-Networks initially introduced in [18] can suggest a set of statistically significant genes (CEBP, FOXM1, E2F1, c-MYC, hTERT) in cells with either telomerase or ALT compared to normal cells from preexisting GRNs. We further confirmed that the APDA satisfactorily estimates the parameters, which decide the rates of convergence to stationarity as the ECDF of the simulated data using the estimated parameters approaches to theoretical CDF using \( q_{i|1,2,\ldots,N} \) as time progresses. Further, our correlation analysis based on SANs helps to infer the link between genes in various conditions that has not yet been discovered through experiments. We prove that our mathematical expression (Eq. 9) allows analytically calculating the correlation coefficients between any pair of genes in the given system at every time point and compare them in various types of cells for an order relation. Specifically in our example, the trend of correlation among genes in the system is not influenced by the mechanism (telomerase vs. ALT) by which telomeres maintain their sufficient lengths, even though the trend differs from one pair to another within the same type of cells. To conclude, this study provides a platform to detect significant genes and infer the previously unknown connection among them by applying modeling techniques borrowed from queueing theory. This is particularly important because telomere dynamics plays a major role in many physiological and disease processes, including hematopoiesis.

**Methods**

**Detailed proof of Eq. 9**

If \( Q \) is the infinitesimal generator and \( P(t) \) is the transition probability matrix of a finite Markov chain, we can derive a set of differential equations called Kolmogorov's forward equations, in the form of

\[
\frac{d}{dt} P(t) \equiv P'(t) = P(t) \cdot Q,
\]

\( P(0) \) initial condition is \( P(0) = I \), where \( I \) is an identity matrix having the same dimension as \( Q \) [33], consistent with

\[
p_{ij}(0) = \begin{cases} P(X(0) = i \mid X(0) = i) = 1, & \text{if } i = j \\ P(X(0) = j \mid X(0) = i) = 0, & \text{if } i \neq j \end{cases}
\]

Solution of the forward equation for finite Markov chains is given by the power series expansion that always converges for any square matrix \( Q \):

\[
P(t) = e^{Qt} = \sum_{n=0}^{\infty} \frac{(Qt)^n}{n!} 
\]

\[
= I + Qt + \frac{(Qt)^2}{2!} + \frac{(Qt)^3}{3!} + \frac{(Qt)^4}{4!} + \ldots 
\]

\[
\approx I + Qt + \frac{(Qt)^2}{2!} + \frac{(Qt)^3}{3!} + ||H(t)||. 
\]

where \( ||·|| \) denotes \( L_1 \)-norm, and \( H(t) \) is is a linear operator with \( ||H(t)|| = o(t) \), as \( t \to 0 \). When the infinitesimal generator, \( Q \), is irreducible, its transition probability matrix, \( P(t) \) defined in Eq. 11, is strictly positive for each time \( t > 0 \). It signifies that as the trajectory length \( N \) approaches to the limit, at least one transition between any pair of states will almost surely happen without regard to the sparsity of \( Q \) [34]. Note that \( 0 = \pi \cdot Q \) is a set of \( |S| \) linear equations, which is usually used to detect the stationary distribution \( \pi \), if such one exists [26].

A formal solution to the time dependent state probability vector is defined as

\[
\pi(t) = \pi(0) \cdot e^{Qt} \tag{12}
\]

where \( \pi(t) \) is a state probability vector at time \( t \) [35, 36]. In order to find a joint state probability vector \( \pi_{ij}(t) \) of two automata, say \( i \) and \( j \), we introduce two matrices called \( C_{ij} \) and \( C^*_{ij} \), which are \( (KN \times K^2) \)- and \( (K^2 \times KN) \)-dimensional matrices, respectively. Then based on Eq. 12, we achieve

\[
\pi_{ij}(t) = \pi_0 \cdot C_{ij} \tag{13}
\]

\[
\Rightarrow \pi_{ij}(0) \cdot e^{Q_{ij}t} = \pi_0 \cdot e^{Q_{ij}t} \cdot C_{ij} 
\]

\[
\Rightarrow \pi_0 \cdot C_{ij} \cdot e^{Q_{ij}t} = \pi_0 \cdot e^{Q_{ij}t} \cdot C_{ij} 
\]

\[
\Rightarrow C_{ij} \cdot e^{Q_{ij}t} = e^{Q_{ij}t} \cdot C_{ij} 
\]

\[
\Rightarrow C^*_{ij} \cdot e^{Q_{ij}t} = C^*_{ij} \cdot e^{Q_{ij}t} \cdot C_{ij} 
\]

Here, \( C_{ij} \) and \( C^*_{ij} \) are defined as follows respectively:

\[
C_{ij} = \bigotimes_{n=1}^N U_{i,j,n}, \quad \forall i \neq j, \: i, j = 1, 2, \ldots, N 
\]

where

\[
U_{i,j,n} = \begin{cases} I_K, & \text{if } n = i \text{ or } n = j \\ 1_K, & \text{if } n \neq i \text{ and } n \neq j \end{cases}
\]
Therefore Eq. 13 finally becomes

\[ C_{ij} = \sum_{n=1}^{N} V_{ij,n} \times \prod_{n=1}^{N} U_{ij,n} \]

where

\[ V_{ij,n} = \left\{ \begin{array}{ll} I_{k} & \text{if } n = i \text{ or } n = j \\ 1 & \text{if } n \neq i \text{ and } n \neq j \end{array} \right. \]

Note that \( I_{k}^T \) indicates the matrix transpose of \( I_{k} \), and we have

\[ I_{k} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \quad \text{and} \quad I_{k}^T = \begin{bmatrix} 1 & 1 & \cdots & 1 \end{bmatrix}. \]

This results in

\[ C_{ij} \cdot C_{ij} = \left( \sum_{n=1}^{N} V_{ij,n} \right) \times \left( \prod_{n=1}^{N} U_{ij,n} \right) \]

\[ = \sum_{n=1}^{N} V_{ij,n} \cdot U_{ij,n} = K^{N-2} \cdot I_{k}^2 \]

Therefore Eq. 13 finally becomes

\[ K^{N-2} \cdot I_{k}^2 \cdot e^{Q_{ij}^t} = C_{ij}^* \cdot e^{Q_{ij}^t} \cdot C_{ij} \]

\[ \Rightarrow e^{Q_{ij}^t} = \frac{1}{K^{N-2} \cdot I_{k}^2} \cdot C_{ij}^* \cdot e^{Q_{ij}^t} \cdot C_{ij} \]

(14)

Equation 14 implies that a time-dependent joint state probability vector \( \pi_{ij}(t) \) can be easily calculated, once the (global) infinitesimal generator of the entire system is obtained.

\[ \pi_{ij}(t) = \frac{\pi_{ij}(0) \cdot e^{Q_{ij}^t}}{1 \times K} \cdot \frac{1}{K^2} \cdot C_{ij}^* \cdot e^{Q_{ij}^t} \cdot C_{ij} \]

\[ \Rightarrow \pi_{ij}(t) = \frac{\pi_{ij}(0) \cdot e^{Q_{ij}^t} \cdot C_{ij}^* \cdot e^{Q_{ij}^t} \cdot C_{ij}}{K^{N-2} \cdot I_{k}^2} \]

(15)

Once \( \pi_{ij}(t) \) of two genes is obtained, we then can form a \((K \times K)\)-dimensional contingency table that describes the joint probability for each time point \( t \). Based upon it, the association between two genes for all \( t \) is evaluated using the Pearson product-moment correlation coefficient.

**Application to telomere maintenance**

**Data description**

Gene expression data (GSE14553) were obtained from the National Center for Biotechnology Information Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo/). Data includes expression gene levels of 3 types of cells: 18 ALT cell lines, 16 telomerase-positive cell lines and 4 normal (fibroblast) cell lines. Initially, the gene expression levels of 29 genes (hTERT, CRY2, FOXM1, c-MYC, SOCS1, AHNAK, HDAC5, TK1, S100A4, LMO4, PRKD1, PER2, PRKCA, CEBPA, E2F1, RBPJ, TOMM20, AIP, EHD1, TGOLN2, TTC17, LAMP2, ATP5D, ADGRL1, LASP1, RPRD1A, HNRNPA3, KAT2A, STK24) were extracted from the data according to [37]. However, only 5 significant genes (CEBPA, E2F1, FOXM1, c-MYC, hTERT) were extracted and scaled with mean 3 and variance 1 [18, 38]. Table 5 and Fig. 6 outline the data in detail.

**Modification of the aPDA**

We adopt APDA introduced originally in [18], but slight modification of assumptions on \( \lambda_i^+ \) and \( \lambda_i^- \) is added in this research as follows:

**Table 5** Mean mRNA expression levels (normalized and scaled) of genes in normal and ALT cells

| Gene       | Normal | ALT   | Mean diff | P-value |
|------------|--------|-------|-----------|---------|
| CEBPA      | 4.724  | 2.617 | 2.107     | 0.003*  |
| E2F1       | 2.475  | 3.117 | -0.642    | 0.029*  |
| FOXM1      | 2.484  | 3.115 | -0.631    | 0.027*  |
| c-MYC      | 1.957  | 3.232 | -1.275    | 0.000*  |
| hTERT      | 2.662  | 3.075 | -0.414    | 0.128   |

\( P \)-values were obtained from the Student’s t-test based on the mean differences of the expression levels. An asterisk indicates statistical significance of \( p < 0.05 \).
• Define initiating genes as genes which do not receive any customer from other genes but send a customer to others within the system.

• The arrival rate of negative customers from the external system, denoted by $\lambda_i^+$, of the initiating genes is assumed to be 0. This assumption enables the initiating genes to maintain their customers; thereby to be consistently activated.

• The arrival rate of positive customers from the external system is defined as follows:

$$\lambda_i^+ = \begin{cases} D_i^{in} + 3, & \text{if } i \in \{\text{Indices of non-initiating genes}\} \\ \frac{\lambda_i}{x_i + \mu_i} \times \mu_i, & \text{if } i \in \{\text{Indices of initiating genes}\} \end{cases}$$

where $x_i$ is the average mRNA expression level of the gene $i$ in normal cell lines. Here, $D_i^{in}$ represents the in-degree of gene $i$, which is the sum of the number of edges incoming to node $i$. The purpose of adding 3 to the non-initiating genes is to prevent the numerator of $q_i$ from becoming 0. The definition of $\lambda_i^+$ for the initiating genes stems from the following logic: the in-degree, arrival rate of the negative customers from the external system and transition probabilities of an initiating gene are all 0; that is,

$$D_i^{in} = \lambda_i^+ = p_i^+ = p_i^- = 0 \text{ for } i \in \{\text{Indices of initiating genes}\}.$$

In such cases, the parameters of stationary distribution in Eq. 3 are $q_i = \frac{\lambda_i^+}{\mu_i}$, resulting in $\lambda_i^+ = q_i \times \mu_i$ for all $i \in \{\text{Indices of initiating genes}\}$. Because the numerical value of $q_i$ is unknown yet, we estimate it based on the property of the geometric distribution with parameter $(1 - q_i)$ as follows:

$$\bar{x}_i = \frac{1 - (1 - q_i)}{1 - q_i} = \frac{q_i}{1 - q_i}$$

$$\Rightarrow \bar{q}_i = \frac{\bar{x}_i}{1 + \bar{x}_i} \quad (16)$$

Equation 16 finally leads to the definition of $\lambda_i^+$ for $i \in \{\text{Indices of initiating genes}\}$. Note that, according to Eq. 3, $\bar{q}_i$ describes the estimate of the steady-state probability that there is at least one mRNA of the gene $i$ [18].

Rules and assumptions of other parameters (shown in Table 2) in this research are identical to those in [18].

Algorithm: simulation study

According to the global balance equation of G-Networks [23–25], a movement of mRNA expression levels of a gene in small time $\Delta t$ is decided by one of the 4 cases as shown in Table 6.

Algorithm 1 contains the details of the simulation assessing the estimated parameter values from the APDA.

**Algorithm 1** Simulation assessing the estimated parameter values from the APDA

1: Calculate the transition rates for each case. For Case 1 and Case 2, the dimension of the transition rate matrix will be $1 \times n$, while it is $n \times n$ for the others. $I$ represents an indicator function.

- Case 1: $\alpha_i^- = \lambda_i^+ + \mu_i \cdot d_i + \sum_{j=1}^{5} (\mu_i \cdot p_{ij}^-) \cdot I_{[x_j=0]}$

- Case 2: $\alpha_i^+ = \lambda_i^+ \cdot I_{[x_i>0]}$

- Case 3: $\alpha_{ji}^- = \mu_i \cdot p_{ij}^+ \cdot I_{[x_j=0]}$

- Case 4: $\alpha_{ji}^+ = \mu_i \cdot p_{ji}^-$

2: Track the time to the next movement, $T$, which follows the exponential distribution [30]:

$$T \sim \text{Exponential}(\alpha),$$

where

$$\alpha = \sum_{i=1}^{n} \left( \alpha_i^- + \alpha_i^+ \right) + \sum_{j=1}^{n} \left( \alpha_{ji}^- + \alpha_{ji}^+ \right).$$

3: Obtain the transition probabilities for each case as follows:

$$\text{Case 1: } p(x_i) = \frac{\alpha_i^-}{\alpha} \quad \text{Case 2: } p(x_i) = \frac{\alpha_i^+}{\alpha}$$

$$\text{Case 3: } p(x_i) = \frac{\alpha_{ji}^-}{\alpha} \quad \text{Case 4: } p(x_i) = \frac{\alpha_{ji}^+}{\alpha}$$

4: Set the mean mRNA expression level of each gene as initial points, $x_0 = \{\bar{x}_1, \bar{x}_2, \cdots, \bar{x}_N\}$.

5: Select one case from 4 total cases according to the appropriate transition probabilities obtained in Step 3.

6: If either Case 1 or Case 2 is selected, then randomly choose a number from $\{1, \cdots, n\}$. If not, then choose two numbers without replacement.

7: Update the mRNA expression levels according to the case chosen.

- Case 1: $x_{i,1} = \max(0, x_{i,0} - 1)$

- Case 2: $x_{i,1} = \max(0, x_{i,0} + 1)$

- Case 3: $x_{i,1} = \max(0, x_{i,0} - 1)$ and $x_{j,1} = \max(0, x_{j,0} + 1)$

- Case 4: $x_{i,1} = \max(0, x_{i,0} - 1)$ and $x_{j,1} = \max(0, x_{j,0} - 1)$

8: Iterate Step 5 to Step 7 as many as needed. In this research, 1000 times of movement were made, and the iteration was set to 800 in order to obtain the empirical distribution for each movement.
Table 6: Possible movements of mRNA expression levels

| Case       | Description                                                                 |
|------------|------------------------------------------------------------------------------|
| Case 1     | When Gene \(i\) loses one customer                                            |
| \(x_i^-\)  | \((x_1, \cdots, x_i-1, x_{i+1}, \cdots, x_n)\)                               |
| Case 2     | When Gene \(i\) gains one customer                                           |
| \(x_i^+\)  | \((x_1, \cdots, x_i+1, x_{i+1}, \cdots, x_n)\)                               |
| Case 3     | When Gene \(i\) loses one customer, while Gene \(j\) gains one              |
| \(x_{ij}^{++}\) | \((x_1, \cdots, x_i-1, x_{i+1}, \cdots, x_{j-1}, x_{j+1}+1, \cdots, x_n)\) |
| Case 4     | When both Gene \(i\) and Gene \(j\) lose one customer                        |
| \(x_{ij}^{--}\) | \((x_1, \cdots, x_i-1, x_{i+1}, \cdots, x_{j-1}, x_{j+1}, \cdots, x_n)\) |

Appendix: supplementary figures/tables

Table 7: Correlation coefficients at \(t = 0.2\) for each type of cells (normal, ALT, and telomerase-positive cells)

| Gene 1 and Gene 2 | Normal | ALT | Telom. |
|-------------------|--------|-----|--------|
| 0.3321            | 0.3219 | 0.3414 |
| 0.3258            | 0.3140 | 0.3317 |
| 0.2190            | 0.2097 | 0.2046 |
| 0.2905            | 0.2271 | 0.2234 |
| 0.2943            | 0.3095 | 0.3055 |
| 0.2028            | 0.2047 | 0.1951 |
| 0.2398            | 0.2454 | 0.2260 |
| 0.1995            | 0.2007 | 0.1936 |
| 0.2254            | 0.2289 | 0.2188 |
| 0.1583            | 0.1519 | 0.1396 |

Table 8: Mean mRNA expression levels (normalized and scaled) of genes in normal and telomerase-active (Telom.) cells

| Gene          | Normal | Telom. | Mean diff | P-value |
|---------------|--------|--------|-----------|---------|
| CEBPA (Gene 1)| 4.643  | 2.589  | 2.054     | 0.003*  |
| E2F1 (Gene 2) | 3.428  | 2.893  | 0.535     | 0.190   |
| FOXM1 (Gene 3)| 2.918  | 3.020  | -0.102    | 0.754   |
| c-MYC (Gene 4)| 1.842  | 3.290  | -1.448    | 0.000*  |
| hTERT (Gene 5)| 1.874  | 3.282  | -1.408    | 0.000*  |

P-values were obtained from the Student’s t-test based on the mean differences of the expression levels. An asterisk indicates statistical significance of \(p < 0.05\)

Table 9: Parameters of stationary distribution, \(q_i(1, 2, \cdots, N)\) in Eq. 3, of normal and malignant (either ALT or telomerase) cells

|        | CEBPA | E2F1 | FOXM1 | c-MYC | hTERT |
|--------|-------|------|-------|-------|-------|
| Normal (for ALT) | 0.8253 | 0.7122 | 0.7130 | 0.6618 | 0.7269 |
| ALT    | 0.7235 | 0.7571 | 0.7570 | 0.7637 | 0.7546 |
| Normal (for telomerase) | 0.8228 | 0.7741 | 0.7448 | 0.6481 | 0.6521 |
| Telomerase | 0.7478 | 0.7431 | 0.7513 | 0.7669 | 0.7429 |

Note that the APDA [18] differently estimates \(\lambda^-\) (degradation) of normal cells depending on the cell type to be compared.
Fig. 7 Trend of correlation between a pair of genes in the system over time in (top) normal cells and (bottom) in malignant cells from 16 telomerase-positive cell lines.

Fig. 8 Trend of correlation between a pair of genes in the system over time in normal cells with the assumption that a pair of genes are initially negatively correlated.
Fig. 9 ECDF plots of E2F1 under the normal condition (top) and ALT (malignant) condition (bottom). Orange and green straight and thick lines represent theoretical (logarithmically transformed) CDF of normal and ALT conditions, respectively. Top: Orange, blue, purple, red (thick) lines serve as the logarithmically transformed ECDF of E2F1 after 10 steps, 100 steps, 300 steps and 500 steps, respectively. Bottom: Orange, blue, purple, red (thick) lines serve as the logarithmically transformed ECDF of E2F1 after 10 steps, 100 steps, 500 steps and 1000 steps, respectively.
Abbreviations
ALT: Alternative lengthening of telomeres; SAN: Stochastic automata network; GRN: Gene regulatory network; APDA: Abnormal pathway detection algorithm; AML: Acute myeloid leukemia; ECDF: Empirical cumulative density function; CDF: Cumulative density function

Authors’ contributions
All authors contributed to the development and implementation of the mathematical result. KL analyzed the data for the application part and produced the numerical results/plots using R. She wrote the first draft of the paper. MK revised the draft for submission. All authors read and approved the final manuscript.

Acknowledgments
We thank Dr. Khanh Dinh of the department of Statistics at Rice University (currently at Columbia University) for help with Fortran programming.

About this supplement
This article has been published as part of BMC Genomics Volume 21 Supplement 9: 2020 Selected original articles from the Sixth International Workshop on Computational Network Biology: Modeling, Analysis, and Control (CNB-MAC 2019): genomics. The full contents of the supplement are available online https://bmcgenomics.biomedcentral.com/articles-supplements/volume-21-supplement-9.

Funding
This research was supported by National Institute of Health grant 5R01HL128173. The funding agency played no part in neither any aspect of the study, nor the decision to publish. Publication costs were funded in part by the Rice University Libraries Open Access Publishing Fund.

Availability of data and materials
Data used in this research can be downloaded from the National Center for Biotechnology Information Gene Expression Omnibus. Further details are in the Data description section.

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Published: 9 September 2020

References
1. Plateau B. On the stochastic structure of parallelism and synchronization models for distributed algorithms. In: ACM SIGMETRICS Performance Evaluation Review, vol 13. New York: ACM; 1985. p. 147–54.
2. Fourneau JM, Plateau B, Stewert W. Product form for stochastic automata networks. In: Proceedings of the 2nd international conference on Performance evaluation methodologies and tools. Citeeseer, 2007. p. 32.
3. Fernandes P, Plateau B, Stewert WJ. Optimizing tensor product computations in stochastic automata networks. RAID-Oper Res. 1998;32(3):325–51.
4. Alexander G. Konecker products and matrix calculus with applications. New York: Halsted; 1981.
5. Plateau B, Stewert WJ. Stochastic automata networks: Springer; 2000. p. 24.
6. Boudaife F, Fourneau J, Mikou N, et al. Product Form Solution for Stochastic Automata Networks with synchronization. In: 5th Process Algebra and Performance Modeling Workshop. Twente: Citeeseer, 1997.
7. Gelenbe E. Product-form queueing networks with negative and positive customers. J Appl Probab. 1991;28(3):656–63.
8. Thi TD, Fourneau JM. Stochastic automata networks with master/slave synchronization: Product form and tensor. In: International Conference on Analytical and Stochastic Modeling Techniques and Applications. Springer; 2009. p. 279–93.
9. Fourneau JM, Gelenbe E. Flow equivalence and stochastic equivalence in G-networks. CMS. 2004;12(2):179–92.
10. Gelenbe E, Muntz RR. Probabilistic models of computer systems Part I (exact results). Acta Informatica. 1976;7(1):35–60.
11. Matyaltski MA. Forecasting anticipated incomes in the Markov networks with positive and negative customers. Autom Remote Control. 2017;78(5):815–25.
12. Fourneau JM. G-networks of unreliable nodes. Probab Eng Informational Sci. 2016;30(3):361–78.
13. Yin Y. Optimum energy for energy packet networks. Probab Eng Informational Sci. 2017;31(4):516–39.
14. Aguilar J, Gelenbe E. Task assignment and transaction clustering heuristics for distributed systems. Inf Sci. 1997;97(1-2):199–219.
15. Basterrech S, Rubino G. Echo State Queueing Network: A new reservoir computing learning tool. In: 2013 IEEE 10th Consumer Communications and Networking Conference (CCNC). IEEE; 2013. p. 118–123.
16. Serrano W, Gelenbe E. The deep learning random neural network with a management cluster. In: International Conference on Intelligent Decision Technologies. Springer; 2017. p. 185–95.
17. Chakravarty SR. A catastrophic queueing model with delayed action. Appl Math Model. 2017;46:631–49.
18. Kim H. Modelling and analysis of gene regulatory networks based on the G-network. Int J Adv Intell Paradigms. 2014;6(1):28–51.
19. Arazí A, Ben-Jacob E, Yechiali U. Bridging genetic networks and queueing theory. Physica A Stat Mech Appl. 2004;332:585–616.
20. Jackson JR. Jobshop-like queueing systems. Manag Sci. 1983;30(1):131–42.
21. Gelenbe E, Fourneau JM. G-networks with resets. Perform Eval. 2002;49(1-4):179–91.
22. Gelenbe E. Random neural networks with negative and positive signals and product form solution. Neural Comput. 1989;1(4):502–10.
23. Gelenbe E. G-networks by triggered customer movement. J Appl Probab. 1993;30(3):742–8.
24. Gelenbe E. Analytical solution of gene regulatory networks. In: 2007 IEEE International Fuzzy Systems Conference. IEEE; 2007. p. 1–6.
25. Gelenbe E, Schassberger R. Stability of product form g-networks. Probab Eng Informational Sci. 1992;6(3):271–6.
26. Ross SM. Introduction to probability models. Cambridge: Academic press; 2014.
27. Galton F. Regression towards mediocrity in hereditary stature. J Anthropol Inst G B Irel. 1886;15:246–63.
28. Kirvan M, Vuillamy T, Marrone A, Walne AJ, Beswick R, Hillmen P, et al. Defining the pathogenic role of telomerase mutations in myelodysplastic syndrome and acute myeloid leukemia. Hum Mutat. 2009;30(11):1567–73.
29. Lin LJ, Chen CY, Lin DT, Tsay W, Tang JL, Yeh YC, et al. Characterization of CEBPA mutations in acute myeloid leukemia: most patients with CEBPA mutations have biallelic mutations and show a distinct immunophenotype of the leukemic cells. Clin Cancer Res. 2005;11(4):1372–9.
30. Grimmett G, Grimmett GR. Strzaker D. Probability and random processes: Oxford university press.
31. Gelenbe E. G-networks: a unifying model for neural and queueing networks. Ann Oper Res. 1994;48(3):433–61.
32. Moeller C, Van Loan C. Nineteen dubious ways to compute the exponential of a matrix, twenty-five years later. SIAM Rev. 2003;45(1):3–49.
33. Feller W. On boundaries and lateral conditions for the Kolmogorov differential equations. Ann Math. 1957;65:S27–70.
34. McGibbon RT, Pande VS. Efficient maximum likelihood parameterization of continuous-time Markov processes. J Chem Phys. 2015;143(3):034109.
35. Harrison PG, Patel NM. Performance modelling of communication networks and computer architectures (International Computer Science, Engineering & Mathematics). Addison-Wesley Longman Publishing Co., Inc.; 1992.
36. Bellman R, Sudderth WE. Dynamic programming and Markov processes. Addison-Wesley; 1970.
37. Thattai M, Van Oudenaarden A. Intrinsic noise in gene regulatory networks. Proc Natl Acad Sci. 2001;98(15):8641–9.

Publisher’s Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.