Cell Population Dynamics: Its Relationship with Finite State Markov Chain and its Asymptotic Behavior

Da-Quan Jiang* Yue Wang† Da Zhou‡

Abstract

We consider a generalized model of cell population dynamics with $n$ different phenotypes. Both the Markov branching process and the ODE system are presented, and exploited to investigate the dynamics of the phenotypic proportions. We will give a sufficient and necessary condition under which the phenotypic proportions satisfy the Kolmogorov forward equations of an $n$-state Markov chain. In general cases, we demonstrate that these proportions will tend to constants regardless of initial population states under weak conditions. These results explain the experimental phenomenon reported in Gupta et al.'s paper [9]. As an application, we will also give sufficient and necessary conditions under which the proportion of one phenotype tends to 0 or 1.

KEY WORDS: population dynamics, Markov chains, asymptotic behavior, branching processes

2010 Mathematics Subject Classification: 60J85, 92D25, 34D05

1 Introduction

With the same genetic background, cell population may have different cellular phenotypes. This has been one of the major topics in the research of cell population dynamics [1, 15]. Very recently much attention has been paid to the stochastic conversions between different phenotypes [3, 9, 27]. Generally, we can use a branching process (stochastic model) [10, 14, 23, 24, 25] or an ODE system (deterministic model) [16] to describe the dynamics of such cell population with multiple phenotypes. However, in many experimental settings, it is difficult or even impossible to count the total cell population [3, 24, 25]. Thus in the last fifty years, people began to consider the proportions of cell individuals with distinct phenotypes instead of the absolute numbers of cells of various phenotypes [10].

*LMAM, School of Mathematical Sciences & Center for Statistical Science, Peking University, Beijing 100871, P.R. China. Email address: jiangdq@math.pku.edu.cn
†Department of Applied Mathematics, University of Washington, Seattle, WA 98195, USA. Email address: yuewang@uw.edu
‡School of Mathematical Sciences, Xiamen University, Xiamen 361005, P.R. China. Email address: zhouda@xmu.edu.cn
In the experiments on breast cancer cell lines, Gupta et al. [9] found that the proportion of each phenotype will always tend to a certain constant regardless of the initial population states. They assumed that the evolution of the phenotypic proportions satisfies an n-state Markov chain, and used the ergodicity of the Markov chain to explain this phenomenon [9]. However, we find that this assumption needs some special conditions to be valid. So we try to remove this assumption and explain the experimental phenomenon in [9] under more general context.

In Section 2, we will give the mathematical description of our models, which is based on [14] and [11]. In Section 3, we will check under what conditions the assumption in [9] is valid. In Section 4, we will prove that under some mild conditions, the experimental phenomenon in [9] mentioned above will always happen in both the deterministic model and the stochastic model. In Section 5, as an application of our conclusions, we will investigate under what conditions one of the phenotypes will die out or dominate.

2 Model description

Assume that the population of cells have n phenotypes: $X_1, X_2, \ldots, X_n$. Assume that all the cells evolve independently. We can present the generalized cell divisions, death and phenotypic conversions as the following reaction form:

$$X_i \overset{\alpha_i}{\rightarrow} d_{i1}X_1 + d_{i2}X_2 + \cdots + d_{in}X_n,$$

where $\alpha_i$ is the reaction rate, $d_{i1}, d_{i2}, \ldots, d_{in}$ are nonnegative integers. For example, $i = 1, d_{11} = 2, d_{12} = \cdots = d_{1n} = 0$ corresponds to the symmetric division: $X_1 \overset{\alpha_1}{\rightarrow} 2X_1$;

$i = 1, d_{i1} = d_{i2} = 1, d_{i3} = \cdots = d_{in} = 0$ corresponds to the asymmetric division: $X_1 \overset{\alpha_1}{\rightarrow} X_1 + X_2$;

$i = 1, d_{i1} = d_{i2} = \cdots = d_{in} = 0$ corresponds to the death: $X_1 \overset{\alpha_1}{\rightarrow} \emptyset$; and

$i = 1, d_{i1} = d_{i3} = \cdots = d_{in} = 0, d_{i2} = 1$ corresponds to the phenotypic conversion: $X_1 \overset{\alpha_1}{\rightarrow} X_2$.

From the stochastic viewpoint, the equation $X_i \overset{\alpha_i}{\rightarrow} d_{i1}X_1 + d_{i2}X_2 + \cdots + d_{in}X_n$ means that for an $X_i$ cell, it will live an exponential time with expectation $1/\alpha_i$ and turn into $d_{i1}$ $X_1$ cells, $d_{i2}$ $X_2$ cells, $\ldots$, $d_{in}$ $X_n$ cells, where $d_{i1}, d_{i2}, \ldots, d_{in}$ are nonnegative random integer variables. $d_{i1}, d_{i2}, \ldots, d_{in}$ are not necessarily independent, but they are assumed to be independent of the exponential reaction time. In fact this is a continuous-time branching process with state space $(\mathbb{Z}^*)^n$, each component of which represents the population of a phenotype. For example, if one $X_1$ cell splits symmetrically, the process will move from the state $(s_1, s_2, \ldots, s_n)$ to the state $(s_1 + 1, s_2, \ldots, s_n)$. We require that each second order moment $\mathbb{E}d_{ij}^2 < \infty$, then this process will not explode in finite time with probability one [3, Section V.7.1, (3)–(4)].

Now we consider the expectation of the populations of the $n$ phenotypes at time $t$, $(x_1(t), x_2(t), \ldots, x_n(t))$. Based on [3, Section V.7.2, (5)–(9)], we have the deterministic
model, namely the following ODE system:

\[
\begin{aligned}
\frac{dx_1}{dt} &= a_{1,1}x_1 + a_{1,2}x_2 + \cdots + a_{1,n}x_n, \\
\frac{dx_2}{dt} &= a_{2,1}x_1 + a_{2,2}x_2 + \cdots + a_{2,n}x_n, \\
&\quad \vdots \\
\frac{dx_n}{dt} &= a_{n,1}x_1 + a_{n,2}x_2 + \cdots + a_{n,n}x_n.
\end{aligned}
\]  

(1)

where \(a_{i,i} = a_i(E_{d_{ii}} - 1) \geq -\alpha_i, \ a_{i,j} = a_jE_{d_{ji}} \geq 0 \ (i \neq j)\). Define \(A = \begin{bmatrix} a_{1,1} & \cdots & a_{1,n} \\ \vdots & \ddots & \vdots \\ a_{n,1} & \cdots & a_{n,n} \end{bmatrix}\), the coefficient matrix of (1).

3 The relation between the \(n\)-state Markov chain and the deterministic model

In this section, we will discuss when the deterministic model can be equivalently captured by the Kolmogorov forward equations of an \(n\)-state Markov chain.

First we consider the proportions of each expected subpopulation \(x_1(t), x_2(t), \ldots, x_n(t)\) in (1) among the expected whole population \(x_1(t) + \cdots + x_n(t)\).

Define

\[
p_1(t) = \frac{x_1(t)}{x_1(t) + x_2(t) + \cdots + x_n(t)}, \quad \cdots, \quad p_n(t) = \frac{x_n(t)}{x_1(t) + x_2(t) + \cdots + x_n(t)}.
\]

Using \(p_n(t) = 1 - \sum_{i=1}^{n-1} p_i(t)\), we can get the differential equations of \(p_1(t), \cdots, p_{n-1}(t)\) from (1):

\[
\begin{aligned}
\frac{dp_1}{dt} &= \sum_{i=1}^{n-1} A_ip_ip_i + \sum_{i=1}^{n-1} B_{1,i}p_i + a_{1,n}, \\
\frac{dp_2}{dt} &= \sum_{i=1}^{n-1} A_ip_2p_i + \sum_{i=1}^{n-1} B_{2,i}p_i + a_{2,n}, \\
&\quad \vdots \\
\frac{dp_{n-1}}{dt} &= \sum_{i=1}^{n-1} A_ip_{n-1}p_i + \sum_{i=1}^{n-1} B_{n-1,i}p_i + a_{n-1,n}.
\end{aligned}
\]  

(2)

where \(A_i = -\sum_{j=1}^{n} a_{j,i} + \sum_{j=1}^{n} a_{j,n}, \ B_{i,i} = a_{i,i} - a_{i,n} - \sum_{i=1}^{n} a_{i,n}, \ B_{i,j} = a_{i,j} - a_{i,n}(i \neq j)\).

Now consider an \(n\)-state continuous-time Markov chain with Q-matrix \(\{q_{i,j}\}\). We can describe it by the Kolmogorov forward equations:

\[
\begin{aligned}
\frac{dP_1(t)}{dt} &= q_{1,1}P_1(t) + q_{2,1}P_2(t) + \cdots + q_{n,1}P_n(t), \\
\frac{dP_2(t)}{dt} &= q_{1,2}P_1(t) + q_{2,2}P_2(t) + \cdots + q_{n,2}P_n(t), \\
&\quad \vdots \\
\frac{dP_n(t)}{dt} &= q_{1,n}P_1(t) + q_{2,n}P_2(t) + \cdots + q_{n,n}P_n(t).
\end{aligned}
\]  

(3)
where \( P_i(t) \) is the probability of the Markov chain being in state \( i \) at time \( t \). Using \( \sum_{i=1}^{n} P_i(t) = 1 \) to remove \( P_n(t) \), we can rewrite (3) as:

\[
\begin{align*}
\frac{dP_1(t)}{dt} &= (q_{1,1} - q_{n,1})P_1(t) + (q_{2,1} - q_{n,1})P_2(t) + \cdots + (q_{n-1,1} - q_{n,1})P_{n-1}(t) + q_{n,1}, \\
\frac{dP_2(t)}{dt} &= (q_{1,2} - q_{n,2})P_1(t) + (q_{2,2} - q_{n,2})P_2(t) + \cdots + (q_{n-1,2} - q_{n,2})P_{n-1}(t) + q_{n,2}, \\
\vdots \\
\frac{dP_{n-1}(t)}{dt} &= (q_{1,n-1} - q_{n,n-1})P_1(t) + \cdots + (q_{n-1,n-1} - q_{n,n-1})P_{n-1}(t) + q_{n,n-1}.
\end{align*}
\]

(4)

In order that (2) has the same form of (1), all second-order coefficients \( A_i \) in (2) should be 0, namely

\[
K := \sum_{i=1}^{n} a_{i,1} = \sum_{i=1}^{n} a_{i,2} = \cdots = \sum_{i=1}^{n} a_{i,n-1} = \sum_{i=1}^{n} a_{i,n}.
\]

(5)

If so, we can rewrite (2) as:

\[
\begin{align*}
\frac{dp_1(t)}{dt} &= (a_{1,1} - K)p_1(t) + a_{1,2}p_2(t) + \cdots + a_{1,n}p_n(t), \\
\frac{dp_2(t)}{dt} &= a_{2,1}p_1(t) + (a_{2,2} - K)p_2(t) + \cdots + a_{2,n}p_n(t), \\
\vdots \\
\frac{dp_n(t)}{dt} &= a_{n,1}p_1(t) + a_{n,2}p_2(t) + \cdots + (a_{n,n} - K)p_n(t).
\end{align*}
\]

(6)

Notice (5) and that \( a_{i,j} (i \neq j) \) is nonnegative. (6) has the same form as (3). Thus we have

**Theorem 1.** Equation (2) is the sufficient and necessary condition for that the proportions of different phenotypes in the deterministic model (1) satisfy the Kolmogorov forward equations of an \( n \)-state Markov chain.

It should be pointed out that the condition (5) has an interesting biological explanation. In fact, \( \sum_{i=1}^{n} a_{i,j} \) is the contribution of the phenotype \( X_j \) to the increase of the total population. So this condition means that the contributions of \( X_1, X_2, \ldots, X_n \) to the total population increase are the same.

With Theorem 1 we can analyze the asymptotic behavior of the deterministic model of population dynamics exploiting the \( n \)-state Markov chain. From the Markov chain theory [17], we know that if \( A \) is irreducible, then the solution of (6) will converge to the unique invariant distribution, no matter what the initial values are. This is just the mathematical basis of Gupta et al.’s work [9]. It has been reported that the condition in Theorem 1 is satisfied not only in the breast cancer cell lines in Gupta et al’s experiments, but also in colon cancer cell lines [22, 27].

4 Asymptotic behavior in general cases

In general cases, (5) is not satisfied since different phenotypes may differ in cell cycling time [19, 8], then the \( n \)-state Markov chain simplification is invalid. Thus we need other methods.
to study the asymptotic behavior of the population dynamics. In this section, we will prove that under some mild conditions, the proportions of different phenotypes will tend to some constants regardless of initial population states.

In this section, we will use the following theorem:

**Theorem 2** (Perron-Frobenius Theorem). For a real square matrix $M$ with nonnegative off-diagonal elements, there is a real eigenvalue $\lambda$ (called Perron eigenvalue), such that for any eigenvalue $\mu \neq \lambda$, $\text{Re} \, \mu < \lambda$. The Perron eigenvalue has a right eigenvector (called Perron eigenvector) with nonnegative (positive if this matrix is irreducible) components.

**Remark 1.** The result in the irreducible case is Theorem 1.5 in [20]. The result in the reducible case is Appendix Theorem 2.4 in [12]. (Notice that for sufficiently large $k$, $M + kI$ is nonnegative.)

Denote the Perron eigenvalue of the coefficient matrix $A$ of (1) by $\lambda_1$ and its Perron eigenvector by $\vec{u} = (u_1, u_2, \cdots, u_n)$ (which is made nonnegative and normalized, i.e. $\sum_{i=1}^{n} u_i = 1$). When $\lambda_1$ is simple, such $\vec{u}$ is unique. To show how universal of Perron eigenvalue being simple, we recall the following theorem (see Appendix C of [26] for the proof):

**Theorem 3.** The set of all $n$-order real square matrices with repeated eigenvalue has measure 0 (as a subset of $\mathbb{R}^{n^2}$).

Thus almost surely any small perturbation on $A$ will let it leave this zero-measure set and have simple Perron eigenvalue. Since population dynamics in real environments always have small fluctuations, it is reasonable to assume that $\lambda_1$ is simple.

4.1 deterministic model

We have proved the following theorem in Appendix B of [26].

**Theorem 4.** Assume that $\lambda_1$ is simple. Starting from any initial value except for the point in some zero-measure set, we have $(x_1(t), x_2(t), \cdots, x_n(t))/\exp(\lambda_1 t) \to c \vec{u}$ as $t \to \infty$, where $c > 0$ is a constant. In this case, the solution to (2) will tend to $\vec{u}$ as $t \to \infty$. Thus (3) has one and only one stable fixed point $\vec{u}$ and no stable limit cycle.

In some cases, (2) has unstable fixed points or unstable limit cycles, thus the above theorem cannot be valid for all initial values [26]. However, since real environments always have fluctuations, unstable fixed points or limit cycles will not hold. So the proportions will almost always tend to this stable fixed point. This gives a satisfactory deterministic explanation of the phenotypic equilibrium phenomenon reported in Gupta et al.’s paper [9].

**Remark 2.** If $\lambda_1$ is not simple, then the convergence result may not hold. Consider $A$ with $a_{i,j} = 0, \forall i, j$. Here $\lambda_1 = 0$ is not simple, and the system will never move. Convergence to a common point will never occur.
4.2 stochastic model

Let $\vec{X}(t) = (X_1^*(t), X_2^*(t), \ldots, X_n^*(t))$ be the population of $n$ phenotypes at time $t$.

Since 1970s, probabilists proved that $(X_1^*(t), X_2^*(t), \ldots, X_n^*(t))/e^{\lambda_1 t} \to W\vec{u}$ under different conditions, where $W$ is a nonnegative random variable. In [13, 2, 3] and [21], it is required that $\lambda_1 > 0$ and $A$ is irreducible (this implies $\lambda_1$ is simple). In [3] it is proved that $W = 0$ or $W > 0$ according to whether the population will become extinct. In [24, 25] it is required that the initial population tends to infinity. Janson [11] requires that $\lambda_1 > 0$, $\lambda_1$ is simple, and assumes a special condition about communicating classes structure (see Remark[3]). So far [11] is the best result about this problem. Based on [11] and [3], we will prove the above result without Janson’s last assumption. We can see the benefit of this improvement in Section 5.

In this section, we assume that $\lambda_1$ is simple and positive. $\lambda_1 > 0$ means that the total cell population is increasing.

We can divide the $n$ phenotypes into several communicating classes according to $A$. If necessary, we can order the classes and rearrange the phenotypes suitably to make $A$ block-triangular. (Each block corresponds to a communicating class.) Thus the eigenvalues of $A$ consist of all eigenvalues of blocks on the diagonal. Every eigenvalue corresponds to a communicating class. (See [11] and [13] for details.)

Denote the communicating class corresponding to the Perron eigenvalue $\lambda_1$ by $T$.

For example, consider matrix $A =$

$$
\begin{bmatrix}
D_1 & 0 & 0 & 0 \\
X & D_2 & 0 & 0 \\
X & 0 & D_3 & 0 \\
0 & X & X & D_4
\end{bmatrix},
$$

where each $X$ represents different nonnegative matrices (not 0). Assume that $D_3$ has the Perron eigenvalue $\lambda_1$, then $D_3$ corresponds to the communicating class $T$. Denote the other three communicating classes by $C_1, C_2, C_4$.

For two communicating classes $C_i$ and $C_j$, we write $C_i \Rightarrow C_j$ if there exist phenotype $X_{k_i} \in C_i$ and $X_{k_j} \in C_j$ such that $a_{k_i, k_j} > 0$. For two communicating classes $C$ and $D$, we write $C \rightarrow D$ if there exist communicating classes $C = C_1, C_2, \ldots, C_m = D$ such that $C_i \Rightarrow C_{i+1}, \forall 1 \leq i < m$. Stipulate that $C_i \Rightarrow C_1$ and $C_1 \rightarrow C_i$.

Then we can illustrate the communicating classes in the example above as

```
C1
  ↙
  ↙
C2
  ↙
  ↙
  ↗
C4
  ↗
T
```

For a communicating class $C$, define $\hat{C} = \{X_i | X_i \in C, C_j \rightarrow C\}$. In other words, $\hat{C}$ is the set of all phenotypes that can produce (directly or indirectly) phenotypes in $C$. In the example above $\hat{T} = C_1 \cup T$.

For a communicating class $C$, define $\bar{C} = \{X_i | X_i \in C, C \rightarrow C_j\}$. In other words, $\bar{C}$ is
the set of all phenotypes that can be produced (directly or indirectly) by phenotypes in $C$. In the example above $\bar{T} = T \cup C_4$. 

For the Markov branching process $\bar{X}(\cdot)$, we say that a cell $Y$ with phenotype in $\bar{T}$ becomes “essentially extinct” if at some time no cell of any phenotypes in $\bar{T}$ is $Y$ or its descendants. In other words, $Y$ and its descendants become extinct inside $\bar{T}$. We say that a trajectory of the branching process $\bar{X}(\cdot)$ becomes “essentially extinct” if at some time no cell of any phenotypes in $\bar{T}$ remains. This means that we can never get a cell with phenotypes in $\bar{T}$ any more. Let the branching process $\bar{X}(\cdot)$ start at any initial population $\bar{X}(0)$ as long as it has some cells with phenotypes in $\bar{T}$.

Let $P_1^*(t) = \frac{X_1^*(t)}{\sum_{i=1}^n X_i^*(t)}$ be the proportion of phenotype $X_i$, as long as the denominator is not zero.

We now state the main result in this section and then give the proof of it.

**Theorem 5.** Assume that $\lambda_1$ is simple and positive. Conditioned on essential non-extinction, we have almost surely $(P_1^*(t), P_2^*(t), \ldots, P_\rho^*(t)) \to \bar{u} = (u_1, u_2, \ldots, u_n)$ as $t \to \infty$.

**Lemma 1.** Assume that $\lambda_1$ is simple. If for some $i \neq j$, $a_{i,j} > 0$ in $[\hat{T}]$, then $u_j > 0 \Rightarrow u_i > 0$. In the language of biology, if the phenotype $X_j$ does not die out, then the phenotype $X_i$ does not die out either.

**Proof.** Without loss of generality, let $i = 1, j = 2$. Assume $u_1 = 0$, $u_2 > 0$. Let $(p_1, p_2, \ldots, p_n) = \bar{u} = (u_1, u_2, \ldots, u_n)$ in the first equation of $[2]$. Since $a_{1,n} = a_{1,n} \sum_{k=2}^n u_k$, $a_{1,2} > 0$ and $u_2 > 0$, the equation becomes $dp_1/dt = \sum_{k=2}^n a_{1,k} u_k > 0$. However $\bar{u}$ is a fixed point of $[2]$ according to Theorem 4, thus we should have $dp_1/dt = 0$, which is a contradiction. \hfill $\square$

**Lemma 2.** Assume that $\lambda_1$ is simple. Then $u_i > 0 \iff X_i \in \bar{T}$.

**Proof.** Apply the Perron-Frobenius theorem to $A_T$, the restriction of $A$ on $\bar{T}$, and let $\bar{w}$ be one of its nonnegative Perron eigenvectors. $\bar{w}_T$, the restriction of $\bar{w}$ on $T$ is a Perron eigenvector of $A_T$, which implies that $\bar{w}_T$ is positive, since $A_T$ is irreducible. From Lemma 1 we know that $\bar{w}$ is positive. If necessary, normalize $\bar{w}$ to make that $||\bar{w}||_1 := \sum_{i \in \bar{T}} w_i = 1$. Set $u_i = w_i$ if $X_i \in \bar{T}$, and $u_j = 0$ if $X_j \notin \bar{T}$, then $\bar{u}$ is the normalized Perron eigenvector of $A$. Thus $u_i > 0 \iff X_i \in \bar{T}$.

Let $\bar{v} = (v_1, v_2, \ldots, v_n)$ be a left nonnegative eigenvector of $A$ corresponding to the eigenvalue $\lambda_1$, namely $\bar{v}A = \lambda_1 \bar{v}$ or $A'\bar{v} = \lambda_1 \bar{v}'$. Normalize $\bar{v}$ to make that $\bar{v} \cdot \bar{u} = 1$. From the previous lemma we have $v_i > 0 \iff X_i \in \bar{T}$.

**Lemma 3** (Lemma 9.7 (i) and Lemma 9.8 in [11]). Assume that $\lambda_1$ is simple and positive. Then we have almost surely $e^{-\lambda_1 t} \bar{v} : \bar{X}(t) \to W$ and $e^{-\lambda_1 t} \bar{X}(t) \to W \bar{u}$ as $t \to \infty$, where $W$ is a nonnegative random variable, and $\mathbb{P}(W > 0) > 0$.
Lemma 4 (Lemma 9.7 (ii) and (iii) in [1], originated from Theorem V.7.2 in [3]). Assume that $\lambda_k$ is simple and positive, and $T$ contains all phenotypes, then $W = 0$ if and only if the branching process becomes essentially extinct almost surely.

Remark 3. Janson’s paper [17] Section 2| has six fundamental assumptions (A1)-(A6). Assumptions (A1)-(A5) have been satisfied in this paper (regarding (A5) as “the process is not essentially extinct at time 0”). Assumption (A6) “$\hat{T}$ contains all phenotypes” is only used in Lemma 4. We will remove this assumption in Lemma 7.

The following lemma is a modification of the second Borel-Cantelli lemma. We base our proof on Theorem 2.3.6 in [6].

Lemma 5. Consider events $B_1, B_2, \ldots, B_n, \ldots$. If for any positive integers $m < n$, we have $\mathbb{P}(\cap_{i=m+1}^{n} B_i) \leq (1 - \epsilon)^{n-m}$, where $0 < \epsilon \leq 1$, then $\mathbb{P}(\limsup_{n \to \infty} B_n) = 1$. In other words, almost surely $\{ B_n : n \geq 1 \}$ will happen infinitely often.

Proof. Let $0 < M < N < \infty$. $\mathbb{P}(\cap_{i=m+1}^{N} B_i) \leq (1 - \epsilon)^{N-M}$ as $N \to \infty$. So $\mathbb{P}(\cup_{i=M+1}^{N} B_i) = 1$ for all $M$, and since $\cup_{i=M+1}^{\infty} B_i \downarrow \limsup_{n \to \infty} B_n$, it follows that $\mathbb{P}(\limsup_{n \to \infty} B_n) = 1$. 

Lemma 6. For almost every essentially non-extinct trajectory (according to Lemma 4, the set of such trajectories has positive probability), we can find an essentially non-extinct cell with phenotype in $T$ within finite time. If we can find such cell at time $t$, then we can find such cell at any time $\tau > t$.

Proof. If at some time $t$ all cells with phenotypes in $\hat{T} \setminus T$ die out, then at least one of the remaining cells with phenotypes in $T$ is not essentially extinct.

Otherwise, at each time $t = k$ ($k \in \mathbb{Z}^+$), there exists one cell $E_k$ with phenotype in $\hat{T} \setminus T$. (For different $k$, $E_k$ may be the same cell.) Let $B_k$ ($k \in \mathbb{Z}^+$) be the event that during the time interval $[k, k+1)$, the cell $E_k$ produces (directly or indirectly) at least one cell with phenotype in $T$.

If $B_k$ happens, choose one such cell with phenotype in $T$ and put it in a special set $S$. Consider any two cells $F$ and $G$ in $S$, and assume $F$ is produced in the time interval $[i, i+1)$, $G$ is produced in the time interval $[j, j+1)$, and $i < j$, where $i, j \in \mathbb{Z}^+$. Then $E_j$ is the ancestor of $G$. Since $E_j$ has phenotype in $\hat{T} \setminus T$, and $F$ has phenotype in $T$, $F$ cannot be the ancestor of $E_j$. Since $E_j$ is still alive at time $t = j$, when $F$ has been produced, $E_j$ cannot be the ancestor of $F$. Thus $F$ cannot be the ancestor of $G$. Since $G$ is produced after $F$, $G$ cannot be the ancestor of $F$. In sum, one cell in $S$ cannot be the ancestor of another cell in $S$. Thus all cells in $S$ are independent.

Consider two phenotypes $X_i$ and $X_j$, and assume a cell with phenotype $X_i$ can produce a cell with phenotype $X_j$ directly, namely $\mathbb{P}(d_{ij} > 0) > 0$. Because of Markov property, within a time span of $1/n$, the probability for a cell with phenotype $X_i$ to produce a cell with phenotype $X_j$ directly is $n_{ij} = [1 - \exp(-\alpha_i/n)]\mathbb{P}(d_{ij} > 0) > 0$. Let $\eta = \min_{i,j} \{n_{ij} : \mathbb{P}(d_{ij} > 0) > 0 \}$. For a cell with phenotype in $\hat{T} \setminus T$, it can produce a cell with phenotype in
within $n$ steps. Thus the probability of $B_k$ is no less than $\eta^n$, regardless of what happens before time $t = k$.

Now we can use Lemma 5 with $\epsilon = \eta^n$, and there will be an infinite number of cells in $S$, except for a zero-measure set of trajectories. According to Lemma 3, the probability for one cell in $S$ to become essentially extinct is less than 1, thus the probability for all cells in $S$ to become essentially extinct is 0, and at least one cell in $S$ is not essentially extinct, except for a zero-measure set of trajectories.

**Lemma 7.** Assume that $\lambda_1$ is simple and positive, then $W = 0$ if and only if the branching process becomes essentially extinct almost surely.

**Proof.** $\Leftarrow$: For a trajectory $X^*()$ outside the zero-measure exclusion set of Lemma 3, assume that at some time $t \geq 0$ (dependent on the trajectory), $X_i(t) = 0$ for all $i \in \hat{T}$. Since $v_i > 0 \iff X_i \in \hat{T}$, we have $e^{-\lambda_1 \tau \mathbf{v} \cdot \mathbf{X}^*(\tau)} = 0$ for all $\tau > t$. Thus $W = \lim_{\tau \to \infty} e^{-\lambda_1 \tau \mathbf{v} \cdot \mathbf{X}^*(\tau)} = 0$ almost surely.

$\Rightarrow$: Assume that $P(W = 0 \&$ the trajectory is not essentially extinct) = $P_0 > 0$. According to Lemma 6, we can find time $t_0 > 0$ large enough such that $P(W = 0 \&$ the trajectory is not essentially extinct & there exists an essentially non-extinct cell with phenotype in $T$ at time $t_0) \geq P_0/2 > 0$. On this set, only consider this essentially non-extinct cell and its descendants from time $t \geq t_0$, then the population is restricted on $\bar{T}$ and we can use Lemma 4. Now we have $W > 0$ except for a zero-measure set of trajectories, which is a contradiction.

From Lemma 5 and Lemma 7, we can obtain Theorem 5.

**Remark 4.** The assumption of $\lambda_1 > 0$ is necessary. If $\lambda_1 < 0$, then from Theorem 4, it follows that the expected populations $(x_1(t), x_2(t), \ldots, x_n(t))$ decays to 0. (As shown in Section 2, $(x_1(t), x_2(t), \ldots, x_n(t)) = \mathbb{E}(X_1^*(t), X_2^*(t), \ldots, X_n^*(t))$ is the solution to (1).) Therefore this process will become extinct almost surely. For $\lambda_1 = 0$, consider a special case that phenotype $X_1$ can only transform to phenotype $X_2$ and vice versa. Starting from one cell with phenotype $X_1$, $(P_1^*, P_2^*)$ will jump between $(1,0)$ and $(0,1)$. This process will not become essentially extinct, and the proportions will not tend to constants $[?]$.

The probability of essential extinction in real environment is very small since the initial cell population is very large. Thus the proportions will almost always tend to the same constants. This gives a satisfactory stochastic explanation of the phenotypic equilibrium phenomenon reported in Gupta et al.’s paper [1].

5 When will one proportion tend to 0 or 1?

In population dynamics, we are also concerned about when one phenotype dies out or dominates. In terms of the notations in this paper, we need to consider when $P_i^*(t) \to 0$ or
\( P_i^*(t) \to 1 \) as \( t \to \infty \).

In this section, we will still assume that the Perron eigenvalue \( \lambda_1 \) of \( A \) is simple and positive. Then from Theorem \[9\] we have \((P_1^*(t), P_2^*(t), \cdots, P_n^*(t)) \to \vec{u} = (u_1, u_2, \cdots, u_n)\) almost surely in the stochastic model. Thus we can get the following theorems from Lemma \[2\]

**Theorem 6.** \( P_i^*(t) \to 0 \iff X_i \notin \bar{T} \).

**Theorem 7.** \( P_i^*(t) \to 1 \iff \bar{T} = T = \{X_i\} \).

**Remark 5.** If we find that \( P_i^*(t) \to 0, P_j^*(t) \not\rightarrow 0 \) in an experiment, then we know that the phenotype \( X_j \) will never transform to \( X_i \) in any way. If we find that \( P_i^*(t) \to 1 \), then we know that the phenotype \( X_i \) will never transform to any other phenotypes.

**Remark 6.** We can substitute \( P_i^*(t) \) by \( p_i(t) \) and apply Theorem \[4\]. Then the conclusions in this section also work in the deterministic model, except for the initial values in the zero-measure exclusion set.

## 6 Conclusion

We have presented a unified stochastic model for the population dynamics with cellular phenotypic conversions. We have given the sufficient and necessary condition under which the dynamical behavior of our model can be described by an \( n \)-state Markov chain. In general case, we have proved that the proportions of different phenotypes will tend to constants regardless of their initial values, and we have investigated the sufficient and necessary conditions under which one phenotype will die out or dominate. In this way we have rigorously explained the experimental phenomenon in Gupta et al.’s paper \[9\].

Since the phenotypic conversions have been reported in various cellular systems, such as *E. coli* \[18\] and cancer cells \[7, 22\], we hope that our model here could be applied as a general framework in the study of multi-phenotypic populations of cells.

## Acknowledgements

We would like to thank Professor Min-Ping Qian, Da-Yue Chen and Svante Janson for helpful advice and discussions. D.-Q. J. acknowledges the support of NSFC 11271029 and NSFC 11171024. D. Z. acknowledges the support of NSFC 11401499 and Fundamental Research Funds for the Central Universities in China.

## References

[1] Altschuler, S.J. and Wu, L.F. (2010). Cellular heterogeneity: do differences make a difference? *Cell* **141**(4), 559–563.
[2] Athreya, K. B. (1968) Some results on multitype continuous time Markov branching processes. *Ann. Math. Statist.* **39**(2), 347–357.

[3] Athreya, K.B. and Ney, P.E. (1972) *Branching Processes*, Springer-Verlag, Berlin, pp.199–206.

[4] Clayton, E., Doup, D. P., Klein, A. M., Winton, D. J., Simons, B. D., and Jones, P. H. (2007). A single type of progenitor cell maintains normal epidermis. *Nature* **446**(7132), 185–189.

[5] Dos Santos, R.V. and da Silva, L.M. (2013). The noise and the KISS in the cancer stem cells niche. *J. Theoret Biol.* **335**, 79–87.

[6] Durrett, Rick. (2010) *Probability: Theory and Examples*, 4th Edition, Cambridge University Press, Cambridge, pp.58–59.

[7] Fidler, I.J. and Kripke, M.L. (1977) Metastasis results from preexisting variant cells within a malignant tumor. *Science* **197**, 893–895.

[8] Fillmore, C.M. and Kuperwasser, C. (2008) Human breast cancer cell lines contain stem-like cells that self-renew, give rise to phenotypically diverse progeny and survive chemotherapy. *Breast Cancer Res.* **10**(2), R25.

[9] Gupta, P.B., Fillmore, C.M., Jiang, G., Shapira, S.D., Tao, K., Kuperwasser, C. and Lander, E.S. (2011). Stochastic state transitions give rise to phenotypic equilibrium in populations of cancer cells. *Cell* **146**(4), 633–644.

[10] Jagers, P. (1969) The proportions of individuals of different kinds in two-type populations. A branching problem arising in biology. *J. Appl. Prob.* **6**, 249–260.

[11] Janson, S. (2004) Functional limit theorems for multitype branching processes and generalized Pólya urns. *Stoch. Proc. Appl.* **110**(2), 177–245.

[12] Karlin, S., and Taylor, H. M. (1975) *A First Course in Stochastic Processes*, 2nd Edition, Academic Press, New York, pp.547–551.

[13] Kesten, H. and Stigum, B.P. (1967) Limit theorems for decomposable multi-dimensional Galton-Watson processes. *J. Math. Anal. Appl.* **17**, 309–338.

[14] Kimmel, M., and Axelrod, D. E. (2002) *Branching Processes in Biology*, Springer, New York, pp.103–140.

[15] Kussell, E. and Leibler, S. (2005). Phenotypic diversity, population growth, and information in fluctuating environments. *Science* **309**, 2075–2078.

[16] Murray, J.D. (2001) *Mathematical Biology I: An Introduction*, Springer-Verlag, Berlin, Heidelberg, pp.1–6, 101–105.

[17] Norris, J.R. (1997) *Markov Chains*, Cambridge University Press, Cambridge, p.122.

[18] Ozbudak, E.M., Thattai, M., Lim, H.N., Shraiman, B.I. and van Oudenaarden, A. (2004) Multistability in the lactose utilization network of *Escherichia coli*. *Nature* **427**(6976), 737–740.

[19] Patrawala, L., Calhoun, T., Schneider-Broussard, R., Zhou, J., Claypool, K., and Tang, D. G. (2005) Side population is enriched in tumorigenic, stem-like cancer cells, whereas ABCG2+ and ABCG2− cancer cells are similarly tumorigenic. *Cancer Res.* **65**(14), 6207–6219.

[20] Seneta, E. (1981) *Non-negative Matrices and Markov Chains*, 2nd Edition, Springer, New York, p.22.
[21] Smythe, R. T. (1996) Central limit theorems for urn models. *Stoch. Proc. Appl.* 65(1), 115–137.

[22] Yang, G., Quan, Y., Wang, W., Fu, Q., Wu, J., Mei, T., Li, J., Tang, Y., Luo, C., Ouyang, Q., Chen, S., Wu, L., Hei, T.K. and Wang, Y. (2012). Dynamic equilibrium between cancer stem cells and non-stem cancer cells in human SW620 and MCF-7 cancer cell populations. *Br. J. Cancer* 106(9), 1512–1519.

[23] Yakovlev, A. Y., Mayer-Proschel, M., and Noble, M. (1998) A stochastic model of brain cell differentiation in tissue culture. *J. Math. Biol.* 37(1), 49–60.

[24] Yakovlev, A. Y., and Yanev, N. M. (2009) Relative frequencies in multitype branching processes. *Ann. Appl. Prob.* 19(1), 1–14.

[25] Yakovlev, A. Y., and Yanev, N. M. (2010) Limiting distributions for multitype branching processes. *Stoch. Anal. Appl.* 28(6), 1040–1060.

[26] Zhou, D., Wang, Y., and Wu, B. (2014) A multi-phenotypic cancer model with cell plasticity. *J. Theoret. Biol.* 357, 35–45.

[27] Zhou, D., Wu, D.M., Li, Z., Quan, M.P. and Zhang, M.Q. (2013). Population dynamics of cancer cells with cell state conversions. *Quant. Biol.* 1(3), 201–208.