Journal of Theoretical Biology 454 (2018) 70–79

Contents lists available at ScienceDirect

Journal of Theoretical Biology

journal homepage: www.elsevier.com/locate/jtbi

# A Bayesian statistical analysis of stochastic phenotypic plasticity model of cancer cells

Da Zhou[a], Shanjun Mao[b], Jing Cheng[c], Kaiyi Chen[a], Xiaofang Cao[a], Jie Hu[a,*]

[a] School of Mathematical Sciences, Xiamen University, Xiamen 361005, PR China
[b] Department of Statistics, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong
[c] School of Statistics, Huaqiao University, Xiamen 361005, PR China

ARTICLE INFO

*Article history:*
Received 4 December 2017
Revised 25 May 2018
Accepted 28 May 2018
Available online 29 May 2018

*Keywords:*
Bayesian statistics
Model selection
Phenotypic plasticity
Cancer model

ABSTRACT

The phenotypic plasticity of cancer cells has received special attention in recent years. Even though related models have been widely studied in terms of mathematical properties, a thorough statistical analysis on parameter estimation and model selection is still very lacking. In this study, we present a Bayesian approach which is devised to deal with the data sets containing both mean and variance information of relative frequencies of cancer stem cells (CSCs). Both Gibbs sampling and Metropolis-Hastings (MH) algorithm are used to perform point and interval estimations of cell-state transition rates between CSCs and non-CSCs. Extensive simulations demonstrate the validity of our model and algorithm. By applying this method to a published data on SW620 colon cancer cell line, the model selection favors the phenotypic plasticity model, relative to conventional hierarchical model of cancer cells. Further quantitative analysis shows that, in the presence of phenotypic equilibrium, the variance data greatly influences the time-variant pattern of the parameters. Moreover, it is found that the occurrence of self-renewal of CSCs shows a strong negative correlation with de-differentiation rate from non-CSCs to CSCs, suggesting a balancing mechanism in the heterogenous population of cancer cells.



## 1. Introduction

The hypothesis of cancer stem cell theory (Jordan et al., 2006; Reya et al., 2001) postulates a hierarchical organization of cancer cells. A small number of tumorigenic cancer cells, also termed cancer stem cells (CSCs), reside at the apex of this cellular hierarchy (Dalerba et al., 2007). CSCs are capable of self-renewal and generating more differentiated cancer cells with lower tumorigenic potential. However, growing researches have extended the CSC model by proposing a *phenotypic plasticity* paradigm in which reversible transitions could happen between CSCs and non-CSCs (Marjanovic et al., 2013). That is, not only can CSCs give rise to non-CSCs, but a fraction of non-CSCs can reacquire CSC-like characteristics. This *de-differentiation* from non-CSCs to CSCs has been reported in quite a few types of cancers, such as breast cancer (Chaffer et al., 2013; Gupta et al., 2011; Meyer et al., 2009), melanoma (Quintana et al., 2010), colon cancer (Yang et al., 2012), and glioblastoma multiforme (Fessler et al., 2015).

Very recently special attention has been paid to reasonable mathematical models for quantifying the process of phenotypic plasticity. In particular, it was found that the phenotypic plasticity plays an important role in the stability of the quantitative models (Gupta et al., 2011; Niu et al., 2015; dos Santos and Silva, 2013a, 2013b; Wang et al., 2014; Zhou et al., 2014a, 2014b). That is, the phenotypic plasticity greatly contributes to stabilizing the phenotypic mixture of cancer cells (termed *phenotypic equilibrium*), thereby effectively maintaining the heterogeneity of cancer cell populations. Some other researches laid emphasis on the role of the phenotypic plasticity in transient dynamics. It was shown that an interesting overshoot phenomenon of CSCs observed in experiment can be well explained by de-differentiation from non-CSCs to CSCs (Chen et al., 2016; Sellerio et al., 2015). Besides, Leder et al. studied mathematical models of pdgf-driven glioblastoma and revealed that the effectiveness of radiotherapy is quite sensitive to the capability of de-differentiation from differentiated sensitive cells to stem-like resistant cells (Leder et al., 2014); Jilkine and Gutenkunst studied the effect of de-differentiation on time to mutation acquisition in cancers (Jilkine et al., 2014); Chen et al. studied transition model between endocrine therapy responsive and resistant states in breast cancer by Landscape Theory (Chen et al., 2014); Dhawan et al. showed with mathematical modeling that exposure to hypoxia enhanced the plasticity and heterogeneity of cancer cell populations (Dhawan et al., 2016); Tonekaboi et al. in-

* Corresponding author.
*E-mail address:* hujiechelsea@xmu.edu.cn (J. Hu).

https://doi.org/10.1016/j.jtbi.2018.05.031
0022-5193/

vestigated how cellular plasticity behaves differently in small and large cancer cell populations (Tonekaboni et al., 2017). A recent review by Jolly et al. (2017) focused on quantitative models of Epithelial-mesenchymal plasticity in cancer.

Even though the phenotypic plasticity has been extensively studied in terms of mathematical properties, the statistical analysis on parameter estimation and model selection is still very lacking. Actually, one of the crucial tasks in the research of phenotypic plasticity is to estimate the transition rates between different cell types. As a pioneering work, Gupta et al. (2011) established a discrete-time Markov state transition model and estimated the transition probabilities between different cell states by fitting the model to their FACS (Fluorescence-activated cell sorting) data on SUM159 and SUM149 breast cancer cell lines. Besides, continuous-time ordinary differential equations (ODEs) models were also developed (Wang et al., 2014; Zhou et al., 2014a), based on which de-differentiation rates were estimated by fitting to SW620 colon cancer cell line. However, the above mentioned works can only provide point estimations to the interested parameters, but not interval estimations. Comparatively, interval estimation is much more informative and frequently-used than point estimation in practice. For doing interval estimation, statistical modeling rather than deterministic modeling should be applied. Moreover, the time-variant pattern of the parameters for phenotypic plasticity is seldom investigated. In previous works, the parameters are normally treated as constants. However, it is more likely that the parameters are variables changing with environment, thus developing effective methods to explore the variability of the parameters should be interesting. Besides, an even more important issue is the model validation of phenotypic plasticity. Even though increasing evidence supports the paradigm of phenotypic plasticity, it is still questionable if this mechanism is a crucial improvement to the cellular hierarchy of cancer cells or just a minor extension to it. In other words, it is quite important to see if it is statistically significant when using different models with and without phenotypic plasticity to the validation data sets. Therefore, a thorough statistical analysis is of great value for further quantifying the biological process of phenotypic plasticity and exploring its biological significance.

In this research, a statistical framework is presented to analyze a two-phenotypic model of cancer cells. In this model, each cancer cell is either CSC phenotypic state or non-CSC phenotypic state. Both types of cells can divide symmetrically or asymmetrically with certain probabilities. A Bayesian approach (Hoff, 2009) is developed to deal with experimental data sets containing both mean and variance values of relative frequencies of cancer stem cells. Standard MCMC methods (such as Gibbs sampling (Geman and Geman, 1987) and MH algorithm (Hastings, 1970; Metropolis et al., 1953)) are used to perform statistical inference with Multivariate Potential Scale Reduction Factor (MPSRF) (Brooks and Gelman, 1998; Gelman and Rubin, 1992) checking the convergence of MCMC chains. Our simulation results demonstrate the precision and accuracy of our algorithm by both point estimation and interval estimation. By applying our approach to a published data on SW620 colon cancer cell line (Yang et al., 2012), we also perform model selection via deviance information criterion (DIC; Gelman et al., 2003). Our result shows that the phenotypic plasticity model with de-differentiation has superior quality relative to the hierarchical model without de-differentiation. Moreover, we provide a further quantitative analysis to the time-variant pattern of the parameters of the model. In the presence of the phenotypic equilibrium, i.e. the mean values of relative frequencies of CSCs tending to a steady value, the time-variant pattern is greatly influenced by the variance data. By using our model selection procedure, the favored model shows a strong negative correlation between symmetric division probability of CSCs and asymmetric division probability of non-CSCs. This result suggests an interesting balance mechanism between self-renewal of CSCs and de-differentiation of non-CSCs.

The paper is organized as follows. The model assumptions and Bayesian framework are presented in Section 2. Main results including simulations and real data analysis are shown in Section 3. Conclusions are presented in Section 4.

## 2. Methods

### 2.1. Model assumptions

In this section we describe the model assumptions. Note that the salient feature of the phenotypic plasticity model is the reversibility between CSCs and non-CSCs, i.e., not only can CSCs differentiate into non-CSCs, but non-CSCs are also capable of de-differentiating into CSCs. Consider a population of cancer cells comprising two phenotypes: CSC represents cancer stem cell phenotypic state, non-CSC represents non-stem cancer cell phenotypic state. Even though this two-phenotypic assumption simplifies the biological complexity of highly diverse phenotypes in cancer, the two-phenotypic setting has been proved as an effective and reasonable simplification for highlighting the minimal process of phenotypic plasticity (Leder et al., 2014; dos Santos and Silva, 2013a, 2013b; Wang et al., 2014). Similar bidirectional transition cascade models were also studied in bacterial community (Mao et al., 2015; Pei et al., 2015).

We now present the cellular process of the two-phenotypic model. From probabilistic point of view, this model can be seen as a discrete-time two-type branching process (Haccou et al., 2005). Each cell lives for a fixed time (suppose one unit of time). At the moment of death it gives birth to two daughter cells. More specifically, for each CSC, it gives birth to two identical CSC daughter cells with probability $\alpha$ (symmetric division), otherwise (with probability $1-\alpha$) it gives birth to one CSC daughter cell and one non-CSC daughter cell (asymmetric division). For each non-CSC, it divides symmetrically into two non-CSC daughter cells with probability $1-\beta$, whereas it divides asymmetrically into one non-CSC daughter cell and one CSC daughter cell with probability $\beta$ (de-differentiation). The model will reduce to conventional hierarchical model if letting $\beta = 0$, i.e. de-differentiation is not allowed to happen. Hence the model selection with respect to $\beta$ provides an efficient way to evaluate the significance of phenotypic plasticity. It should be pointed out that, $\alpha$ and $\beta$ may not be constant, so it is interesting to see whether the parameters are time-varying and how to quantify the time-variant pattern of them.

The statistical inference of branching processes has been studied for a long time (Guttorp, 1991). The usage of statistical methods strongly depends on the data types available. Normally, the observation of the whole genealogy tree generated from underlying process is quite difficult to obtain (except in very limited experiments (Hu et al., 2015)). More often, only the absolute numbers or relative frequencies of distinct cell types are recorded at given moment, and it is even easier to collect relative frequencies than absolute numbers of given cell types (Yakovlev and Yanev, 2009). Thus developing statistical approaches for proportion data has a wider range of application. In this work our proposed method is used for the time-series data on relative frequencies of CSC phenotypic state.

Let $x_A(t)$ be the frequency of CSC state at time $t$, $\mu_t$ be the expectation of $x_A(t)$, i.e. $\mu_t = \mathbf{E}(x_A(t))$, and $\sigma_t^2$ be the variance of $x_A(t)$, i.e. $\sigma_t^2 = \mathbf{Var}(x_A(t))$. Then we can obtain two important recurrence formulas as follows (see Appendix A for more details):

$$\mu_{t+1} = \frac{1+\alpha-\beta}{2}\mu_t + \frac{\beta}{2}, \tag{1}$$