

RESEARCH ARTICLE

# Systems Mapping for Hematopoietic Progenitor Cell Heterogeneity

Linghua Zhou<sup>1</sup>, Yong Shen<sup>1</sup>, Libo Jiang<sup>1</sup>, Danni Yin<sup>1</sup>, Jingxin Guo<sup>1</sup>, Hui Zheng<sup>2</sup>, Hao Sun<sup>2</sup>, Rongling Wu<sup>1,3</sup>, Yunqian Guo<sup>1\*</sup>

**1** Center for Computational Biology, Beijing Forestry University, Beijing, People's Republic of China, **2** CAS Key Laboratory of Regenerative Biology, Guangdong Provincial Key Laboratory of Stem Cell and Regenerative Medicine, Guangzhou Institutes of Biomedicine and Health, Chinese Academy of Sciences, Guangzhou, Guangdong, People's Republic of China, **3** Center for Statistical Genetics, Pennsylvania State University, Hershey, Pennsylvania, United States of America

\* [guoyq04@gmail.com](mailto:guoyq04@gmail.com)



**OPEN ACCESS**

**Citation:** Zhou L, Shen Y, Jiang L, Yin D, Guo J, Zheng H, et al. (2015) Systems Mapping for Hematopoietic Progenitor Cell Heterogeneity. PLoS ONE 10(5): e0126937. doi:10.1371/journal.pone.0126937

**Academic Editor:** Song Wu, The State University of New York at Stony Brook, UNITED STATES

**Received:** February 4, 2015

**Accepted:** April 9, 2015

**Published:** May 13, 2015

**Copyright:** © 2015 Zhou et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the paper.

**Funding:** The National Natural Science Foundation of China (31370669, 30900854) and the Key Laboratory of Regenerative Biology, Guangzhou Institutes of Biomedicine and Health, Chinese Academy of Sciences (KLRB201308). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

## Abstract

Cells with the same genotype growing under the same conditions can show different phenotypes, which is known as “population heterogeneity”. The heterogeneity of hematopoietic progenitor cells has an effect on their differentiation potential and lineage choices. However, the genetic mechanisms governing population heterogeneity remain unclear. Here, we present a statistical model for mapping the quantitative trait locus (QTL) that affects hematopoietic cell heterogeneity. This strategy, termed systems mapping, integrates a system of differential equations into the framework for systems mapping, allowing hypotheses regarding the interplay between genetic actions and cell heterogeneity to be tested. A simulation approach based on cell heterogeneity dynamics has been designed to test the statistical properties of the model. This model not only considers the traditional QTLs, but also indicates the methylated QTLs that can illustrate non-genetic individual differences. It has significant implications for probing the molecular, genetic and epigenetic mechanisms of hematopoietic progenitor cell heterogeneity.

## Introduction

Cell fate decision is an important question during developmental processes, such as embryogenesis, neurogenesis, and hematopoiesis. During the hematopoiesis process, hematopoietic stem cells (HSCs) proliferate to self-renew or differentiate to progenitor cells, which generate mature blood cells. These progenitors, including common lymphoid progenitors (CLPs) and common myeloid progenitors (CMPs), can differentiate into more committed progenitors that give rise to blood cells. These progenitors can be used for bone marrow transplantation to treat diseases such as leukemia, sickle cell anemia, and thalassaemia [1–4].

Hematopoietic multipotential progenitors have two major differentiation choices: erythroid and myeloid lineages, which are regulated by the key transcription factors, Gata1 and PU.1. These two transcription factors positively regulate lineage-specific genes and repress each other

[5]. In addition to transcriptional networks, genetic and epigenetic mechanisms are critical in determining cell fate.

Genetically identical hematopoietic progenitor cells growing under the same conditions can show differences in phenotypic characteristics, which is known as “population heterogeneity”, and has attracted interest for many years. However, it remains unclear whether this non-genetic characteristic affects cell fate determination. A previous report published in Nature by Chang et al. [6] showed that gene expression of noise controls the lineage choices of hematopoietic progenitor cells. However, the genetic mechanisms that control this process have not been explored in that paper. Cell fate conversions are dynamic with changes in chromatin structure regulated by DNA and histone modifications, including DNA methylation at symmetrical CG dinucleotides (CpG) and histone methylation and acetylation. Epigenetic regulation has been studied in hematopoietic lineage specification based on coordinated changes in gene expression, chromatin state, and DNA methylation [5,7,8].

Genetic mapping can provide a view of network and gene actions, as well as interactions with quantitative trait loci (QTLs), which can demonstrate the effects of genetic variation. Functional mapping developed by Wu et al. differs from the traditional mapping strategies, and is a very useful method to analyze dynamic data, as well as mapping QTLs related to development processes including cell apoptosis, cancer stem cell proliferation [9–12], et al. Clonal population heterogeneity, which is known as “non-genetic cell individuality”, cannot be analyzed using traditional QTLs. Several studies have examined the link between DNA methylation and gene expression, as well as mapping the methylated QTLs (meQTLs) to interpret the mechanisms underlying genetic variants [13,14]. meQTLs may increase our understanding of population heterogeneity and lineage choice problems, which cannot be demonstrated by alterations in the DNA sequences.

The aim of this article was to explore the genetic mechanisms regulating cell population heterogeneity and hematopoietic progenitor cell lineage choices. Besides the traditional QTL analysis, we mapped the effects of genetic variation on DNA methylation, focusing on mapping meQTLs that determine population heterogeneity and lineage choices.

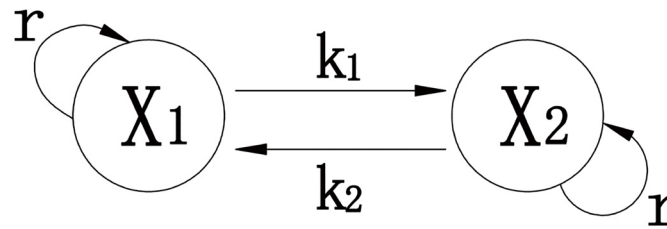
## Methods

### Mathematical modeling of the evolution of two hematopoietic progenitor cell subpopulations

Hematopoietic progenitor cells show heterogeneity in one clonal population. The expression level of stem-cell-surface marker Sca-1 showed an approximately 1000-fold range within one newly derived clonal cell population based on flow cytometric analysis. Cells with the highest, middle and lowest ~15% Sca-1 expression level were isolated from one clonal population as separate subpopulations using fluorescence-activated cell sorting (FACS). Within hours, all three subpopulations showed narrow Sca-1 histograms; however, the three fractions regenerated Sca-1 histograms similar to that of the parental (unsorted) population after 21-day culture. A two-Gaussian model that best fitted the observed histogram evolution and restoration of the parental distribution was predominantly driven by state transitions between the subpopulations. Linear and nonlinear ordinary differential equations (ODEs) are used to describe the transition of the two subpopulations, respectively.

Based on Fig 1, Chang et al. [6] proposed a linear model of equations for the size  $x_i$  of subpopulation  $i$ :

$$\begin{aligned}\frac{dx_1}{dt} &= rx_1 - k_1x_1 + k_2x_2 \\ \frac{dx_2}{dt} &= rx_2 + k_1x_1 - k_2x_2\end{aligned}\tag{1}$$



**Fig 1. Linear model of two interacting and growing progenitor cell subpopulations.**

doi:10.1371/journal.pone.0126937.g001

where  $x_1$  and  $x_2$  represent the sizes of subpopulations 1 and 2,  $r$  is the growth rate of both subpopulations,  $k_1$  is the transition rate from  $x_1$  to  $x_2$  and vice versa for  $k_2$ .

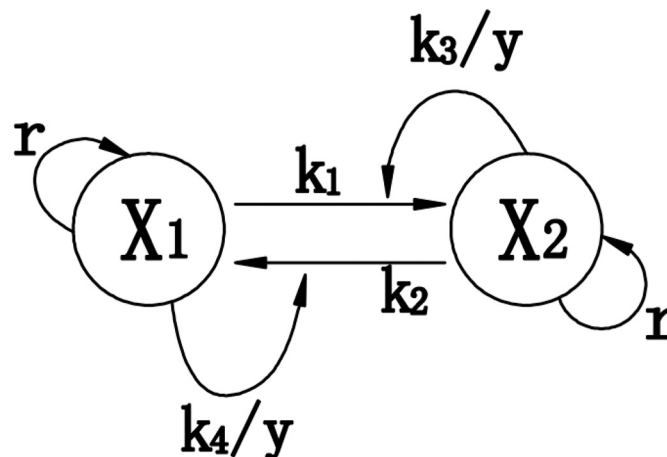
To increase our understanding of the asymptotic behavior and explain the sigmoidal increase of  $x_1$  and  $x_2$ , a nonlinear dynamic model of two interacting and growing progenitor cell populations was proposed based on Fig 2. The ordinary differential equations governing the growth of the two subpopulations will then be:

$$\begin{aligned} \frac{dx_1}{dt} &= [rx_1 - k_1x_1 + k_2x_2] - k_3w_2x_1 + k_4w_1x_2 \\ \frac{dx_2}{dt} &= [rx_2 + k_1x_1 - k_2x_2] + k_3w_2x_1 - k_4w_1x_2 \end{aligned} \tag{2}$$

where  $x_1$  and  $x_2$  represent the sizes of subpopulations 1 and 2,  $r$  is the growth rate of both subpopulations,  $k_1$  is the transition rate from  $x_1$  to  $x_2$  and vice versa for  $k_2$ , and  $k_3$  and  $k_4$  are parameters determining the feedback rate.

### Statistical modeling of systems mapping

Systems mapping is a statistical model that views a complex phenotype as a dynamic system, dissects it into its underlying components, coordinates different components in terms of biological laws through mathematical equations, and maps specific genes that mediate each component and its connection with other components [15]. As a bottom—top model, a systems approach can identify specific QTLs that govern the developmental interactions between various components, giving rise to the function and behavior of the system. By estimating and



**Fig 2. Nonlinear dynamic model of two interacting and growing progenitor cell subpopulations.**

doi:10.1371/journal.pone.0126937.g002

testing mathematical parameters that specify the system, systems mapping allows us to predict the critical genetic mechanisms governing alterations in the physiological status of a phenotype.

### Sampling strategy

Suppose there is a natural population drawn at random. From a genetic perspective, we assume the original population follows Hardy-Weinberg equilibrium. The  $n$  sampled subjects are genotyped for a set of molecular markers, such as single nucleotide polymorphisms (SNPs). Assume that there exists a particular gene that controls the dynamics of cell interactions and growth. This gene has three alleles; namely,  $Q$ ,  $q$ , and  $q^+$ , whose frequencies are  $q_1$ ,  $q_2$ , and  $1-q_1-q_2$  in the population, and  $q^+$  represents the methylated state of  $q$ . Let  $\mu_j$  denote the genotypic value of size for  $x_1$  and  $x_2$  with  $j = 0$  to 5 for  $QQ$ ,  $Qq$ ,  $Qq^+$ ,  $qq$ ,  $qq^+$ , and  $q^+q^+$ , respectively.

The actual gene for cell dynamics, called the quantitative trait locus (QTL), cannot be observed directly, but it can be inferred from associated markers. Consider a SNP with alleles  $M$  vs.  $m$ , with respective frequencies  $p_1$  vs.  $1-p_1$  in the population. The SNP and QTL form six haplotypes; namely,  $MQ$ ,  $Mq$ ,  $Mq^+$ ,  $mQ$ ,  $mq$  and  $mq^+$ , with respective frequencies  $P_{MQ} = P_M P_Q + D_1$ ,  $P_{Mq} = P_M P_q + D_2$ ,  $P_{Mq^+} = P_M P_{q^+} - D_1 - D_2$ ,  $P_{mQ} = P_m P_Q - D_1$ ,  $P_{mq} = P_m P_q - D_2$ , and  $P_{mq^+} = P_m P_{q^+} + D_1 + D_2$  in the population, where  $D_1$  and  $D_2$  are a linkage disequilibrium between the marker and QTL.

In the parental progeny of the population, the haplotypes unite randomly to generate 21 diplotypes and 18 genotypes, whose frequencies can be expressed in terms of haplotype frequencies. We express the genotype frequencies in matrix notation, with rows representing marker genotypes and columns representing QTL genotypes (Table 1). Thus, the conditional probabilities of QTL genotypes, conditional upon marker genotypes, can be derived according to Bayes' theorem.

**Likelihood.** Systems mapping embeds a system of differential equations into a genetic mapping setting constructed using a segregating population. Genetic mapping uses a mixture model-based likelihood to estimate genotype-specific parameters by assuming  $j$  QTL genotypes. For a progenitor cell population  $i$ , we obtain the size,  $x_1$  and  $x_2$ , for the two following cell subpopulations at a finite set of times,  $1, \dots, T$ . A linear model for describing the phenotypic values of population  $i$  controlled by a putative gene is expressed as:

$$\begin{aligned}
 x_{1i}(t) &= \sum_{j=0}^5 \xi_i \mu_j^{(x_1)}(t) + e_i^{(x_1)} \\
 x_{2i}(t) &= \sum_{j=0}^5 \xi_i \mu_j^{(x_2)}(t) + e_i^{(x_2)}
 \end{aligned}
 \tag{3}$$

where  $\xi_i$  is an indicator variable defined as 1 if this population belongs to a specific genotype and 0 otherwise; time  $t$ ,  $\mu_j^{(x_1)}(t)$ , and  $\mu_j^{(x_2)}(t)$  are the genotypic values of type  $k$  for the cell size at time  $t$ , respectively; and  $e_i^{(x_1)}$  and  $e_i^{(x_2)}$  are the residual errors of subject  $i$  for cell  $x_1$  and  $x_2$  at time  $t$ , respectively.

**Table 1. Conditional probabilities of QTL genotypes given marker genotypes in a natural population.**

	QQ	Qq	Qq <sup>+</sup>	qq	qq <sup>+</sup>	q <sup>+</sup> q <sup>+</sup>
<b>MM</b>	$P_{MQ}^2$	$2P_{MQ}P_{Mq}$	$2P_{MQ}P_{Mq^+}$	$P_{Mq}^2$	$2P_{Mq}P_{Mq^+}$	$P_{Mq^+}^2$
<b>Mm</b>	$2P_{MQ}P_{mQ}$	$2P_{MQ}P_{mq} + 2P_{Mq}P_{mQ}$	$2P_{MQ}P_{mq^+} + 2P_{Mq}P_{mQ}$	$2P_{Mq}P_{mq}$	$2P_{Mq}P_{mq^+} + 2P_{Mq^+}P_{mQ}$	$2P_{Mq^+}P_{mq^+}$
<b>mm</b>	$P_{mQ}^2$	$2P_{mQ}P_{mq}$	$2P_{mQ}P_{mq^+}$	$P_{mq}^2$	$2P_{mq}P_{mq^+}$	$P_{mq^+}^2$

doi:10.1371/journal.pone.0126937.t001

For the mapping population of  $n$  members with marker information ( $M$ ) and phenotypic data ( $Y_i$ ) for the two subpopulations, we formulate the likelihood as:

$$L(M, Y) = \prod_{i=1}^n \left[ \sum_{j=0}^5 \pi_{ji} f_j(Y_i; \Theta_j, \Psi) \right] \tag{4}$$

where  $Y_i = (x_{1i(1)}, \dots, x_{1i(T)}, x_{2i(1)}, \dots, x_{2i(T)})$  represents phenotypic data for the subpopulation weights,  $\pi_{ji}$  is the mixture proportion representing the conditional probability of QTL genotype  $k$  given the marker genotype of progenitor cell population  $i$ ; and  $f_j(Y_i; \Theta_j, \Psi)$  is a multivariate normal distribution with an expected mean vector,  $\mu_j = (\mu_{x_{1j(1)}}; \dots; \mu_{x_{1j(T)}}; \mu_{x_{2j(1)}}; \dots; \mu_{x_{2j(T)}})$ , for population  $i$  that belongs to QTL genotype  $j$ , and covariance matrix

$$\Sigma = \begin{pmatrix} \Sigma_1 & \Sigma_{1 \times 2} \\ \Sigma_{2 \times 1} & \Sigma_2 \end{pmatrix} \tag{5}$$

with  $\Sigma_1$  and  $\Sigma_2$  being the time-dependent covariance matrix for  $x_1$  and  $x_2$ , respectively, and  $\Sigma_{1 \times 2} = \Sigma_{2 \times 1}^T$  being the time-dependent covariance matrix between these two variables.

**Modeling mean vectors.** The biological merit of systems mapping is to model genotype-specific differences in the time-dependent mean vector using a biologically meaningful mathematical equation [9]. When modeling the dynamic behavior of progenitor cell transition, Chang et al. [6] developed a set of ordinary differential equations (ODEs) as shown in (1) and (2). The overall behavior of progenitor cell heterogeneity can be described and quantified by ODE solution's parameters ( $r, k_1, k_2$ ) for the linear model and parameters ( $r, k_1, k_2, k_3, k_4$ ) for the nonlinear model. Differences in any one or more of these parameters will influence cell transition dynamics.

For a particular QTL genotype  $j$ , the dynamic behavior of ODE can be specified by a set of parameters  $\Theta_{uj} = (r_j, k_{1j}, k_{2j})$  (for the linear differential equation) or  $\Theta_{uj} = (r_j, k_{1j}, k_{2j}, k_{3j}, k_{4j})$  (for nonlinear differential equations). In other words, by changing any one or more parameters in  $\Theta_{uj}$ , the trajectory of cell transition dynamics may change. Thus, by incorporating ODE into functional mapping, a statistical framework originally derived to estimate QTL genotype-specific ODE parameters,  $\Theta_{uj}$ , can be used to assess the genetic effects of QTLs on progenitor cell transition by comparing the genotypic differences of the parameters.

**Modeling covariance structure.** The statistical power of systems mapping is partly the result of structured modeling of the covariance matrix. There are many approaches available to model the covariance structure, including autoregressive [9], antedependent [16], autoregressive moving average [17], nonparametric [18], and semiparametric [19]. These approaches have their own advantages and disadvantages in terms of efficiency, flexibility, and parsimony. Of these approaches, the antedependent model shows a good balance between these properties, and Zhao et al. [20] extended it to model the covariance structure of multiple variables. In this article, we use a bivariate antedependence approach for modeling the structure of the covariance matrix (5).

The residual errors of the linear model (3) at time  $t$  depend on the previous residuals with the degree of dependence decaying with time lag. If the current residuals only depend on their first preceding ones, the model is called a first-order antedependence model [SAD(1)]. The SAD(1) model specifies time-dependent residual errors for population  $i$  as:

$$\begin{aligned} e_i^{(1)}(t) &= \varphi_1 e_i^{(1)}(t-1) + \phi_1 e_i^{(1)}(t-1) + \varepsilon_i^{(1)}(t) \\ e_i^{(2)}(t) &= \varphi_2 e_i^{(2)}(t-1) + \phi_2 e_i^{(2)}(t-1) + \varepsilon_i^{(2)}(t) \end{aligned} \tag{6}$$

where  $\phi_1$  (or  $\phi_2$ ) and  $\varphi_1$  (or  $\varphi_2$ ) are the unrestricted antedependence parameters induced by variable 1 (or 2) itself and by the other variable 2 (or 1), respectively;  $\varepsilon_i^1(t)$  and  $\varepsilon_i^2(t)$  are the “innovation” errors assumed to be bivariate-normally distributed with mean zero and variance matrix:

$$\Sigma_\varepsilon(t) = \begin{pmatrix} v_1^2(t) & v_1(t)v_2(t)\rho(t) \\ v_1(t)v_2(t)\rho(t) & v_2^2(t) \end{pmatrix} \tag{7}$$

For simplicity, we assume that  $\Sigma_\varepsilon(t)$  is time-independent; i.e.,  $v_1$ ,  $v_2$ , and  $\rho$  are not dependent on time. Based on the above assumptions,  $\Sigma_\varepsilon$  becomes a block matrix with the diagonal repeating the following unit  $T$  times,

$$\Sigma_\varepsilon = \begin{pmatrix} v_1^2 & v_1v_2\rho \\ v_1v_2\rho & v_2^2 \end{pmatrix} \tag{8}$$

All parameters modeling the covariance matrix  $\Sigma_\varepsilon$  are arrayed in vector  $\Psi_\varepsilon = (\phi_1, \phi_2, \varphi_1, \varphi_2, v_1, v_2, \rho)$ . With model (6), closed forms for the determinant and inverse of the structured matrix, which enhances computing efficiency, were derived [20].

**Computational algorithms.** Three algorithms were integrated to estimate marker-QTL haplotype frequencies using the EM algorithm; the genotype-specific ODE parameters for progenitor cell transition dynamics by the fourth-order Runge-Kutta algorithm [21,22] and the SAD (1) parameters for the covariance structure using simplex algorithm. We derived a closed-form solution for the EM algorithm to estimate haplotype frequencies. In the E step, we calculated the posterior probability with which population  $i$  carries QTL mating type genotype  $j_m$  using:

$$\Pi_{ji} = \frac{\pi_{ji}f_j(Y_i)}{\sum_{j=0}^5 \pi_{ji}f_j(Y_i)} \tag{9}$$

In the M step, the calculated posterior probability was used to estimate marker-QTL haplotype frequencies by:

$$\begin{aligned} \hat{P}_{MQ} &= \frac{1}{2n} (2\Pi_{5|2i} + \Pi_{4|2i} + \Pi_{3|2i} + \Pi_{5|1i} + \theta_1\Pi_{4|1i} + \theta_2\Pi_{3|1i}) \\ \hat{P}_{Mq} &= \frac{1}{2n} (2\Pi_{2|2i} + \Pi_{4|2i} + \Pi_{1|2i} + \Pi_{2|1i} + (1 - \theta_1)\Pi_{4|1i} + \theta_3\Pi_{1|1i}) \\ \hat{P}_{Mq^+} &= \frac{1}{2n} (2\Pi_{0|2i} + \Pi_{3|2i} + \Pi_{1|2i} + \Pi_{0|1i} + (1 - \theta_2)\Pi_{3|1i} + (1 - \theta_3)\Pi_{1|1i}) \\ \hat{P}_{mQ} &= \frac{1}{2n} (2\Pi_{5|0i} + \Pi_{5|1i} + \Pi_{4|0i} + \Pi_{3|0i} + (1 - \theta_1)\Pi_{4|1i} + (1 - \theta_2)\Pi_{3|1i}) \\ \hat{P}_{mq} &= \frac{1}{2n} (2\Pi_{2|0i} + \Pi_{2|1i} + \Pi_{4|0i} + \Pi_{1|0i} + \theta_1\Pi_{4|1i} + (1 - \theta_3)\Pi_{1|1i}) \\ \hat{P}_{mq^+} &= \frac{1}{2n} (2\Pi_{0|0i} + \Pi_{0|1i} + \Pi_{3|0i} + \Pi_{1|0i} + \theta_2\Pi_{3|1i} + \theta_3\Pi_{1|1i}) \end{aligned}$$

$$\begin{aligned} \theta_1 &= P_{MQ}P_{mq} / (P_{MQ}P_{mq} + P_{Mq}P_{mQ}) \\ \theta_2 &= P_{MQ}P_{mq^+} / (P_{MQ}P_{mq^+} + P_{Mq^+}P_{mQ}) \\ \theta_3 &= P_{Mq}P_{mq^+} / (P_{Mq}P_{mq^+} + P_{Mq^+}P_{mq}) \end{aligned} \tag{10}$$

In each iteration of the EM step, we estimated ODE parameters,  $\Theta_{u_j} = (r_j, k_{1j}, k_{2j})$  (for the linear differential equation) or  $\Theta_{u_j} = (r_j, k_{1j}, k_{2j}, k_{3j}, k_{4j})$  (for nonlinear differential equations) using the Runge-Kutta algorithm, and covariance-structuring parameters, and  $\Psi_e = (\phi_1, \phi_2, \phi_1, \phi_2, v_1, v_2, \rho)$ , using the simplex algorithm. The Runge—Kutta and simplex algorithms are integrated within the EM framework to generate each iterative procedure, which is repeated until the parameters converge to stable values.

### Hypothesis tests

Whether there are specific QTLs for progenitor cell transition dynamics can be tested by formulating the following hypotheses:

$$H_0 : \Psi_{\mu_{jm}} \equiv \Psi_{\mu^j} = 5, 4, 3, 2, 1, 0; m = 2, 1, 0 \tag{11}$$

$$H_1 : \text{At least one of the equalities above does not hold,}$$

where the  $H_0$  corresponds to the reduced model, in which only one single curve of transition dynamics exists; and  $H_1$  corresponds to the full model, in which there exists six such different curves to fit the data. We calculate this hypothesis to reduce the full model using the log-likelihood ratio (LR). The critical threshold is determined empirically from permutation tests [23].

Two variables,  $x_1$  and  $x_2$ , may or may not be controlled by the same gene. The pleiotropic control of the gene over these two variables can be tested by formulating the null hypotheses:  $H_0: r_{jm} \equiv r$  and  $k_{ijm} \equiv k_i j = 5, 4, 3, 2, 1, 0; m = 2, 1, 0; i = 1, 2$  for the linear model and 1, 2, 3, 4 for the nonlinear model

$$H_1 : \text{At least one of the equalities above does not hold} \tag{12}$$

If the null hypothesis is rejected, this indicates that the gene detected exerts a pleiotropic effect on progenitor cell heterogeneity.

### Results

We performed simulation experiments to examine the statistical properties of the model built for genetic mapping of the two subpopulations of hematopoietic progenitor cell transitions. Fu et al. provided much more detailed information regarding the test and validation of systems mapping [24]. The transition process was described by linear and nonlinear ODEs, and the simulation assumed different heritabilities (0.05, 0.1 and 0.2) under different human population sizes (200 and 400) at Hardy-Weinberg equilibrium, considering a SNP marker that is associated with a putative QTL (with six genotypes  $QQ, Qq, Qq^+, qq, qq^+,$  and  $q^+q^+$ ) through linkage disequilibrium (LD). The transition parameters  $\Theta_{u_j} = (r_j, k_{1j}, k_{2j})$  and  $\Theta_{u_j} = (r_j, k_{1j}, k_{2j}, k_{3j}, k_{4j})$  for the six genotypes were determined in the ranges of empirical estimates of these parameters. Note that for computational simplicity,  $r, k_1, k_2, k_3$  and  $k_4$  are provided as shown in Tables 2, 3, 4 and 5. Based on allele frequencies of the marker, QTL, and their LD, joint marker and phenotypic data were simulated. The genetic parameters ( $P, Q_1, Q_2, D_1,$  and  $D_2$ ) of the QTL can be estimated with high precision using the EM algorithm. The genotype-specific mean vectors were modeled based on ODE's solution (3) where the covariance matrix was structured by the first-order antedependence model [SAD(1)] with correlation and variance parameters,  $\Psi_e = (\phi_1, \phi_2, \phi_1, \phi_2, v_1, v_2, \rho)$ , for  $w_1$  and  $w_2$ .

The simulation results of cell transition parameters, genetic parameters, and covariance-structural parameters with the linear and nonlinear model are shown in Tables 2 and 3 and Tables 4 and 5, respectively. The cell transition parameters can be estimated with high precision



**Table 2. MLEs of linear ODE parameters that define the dynamics of cell transition for six different QTL genotypes and the association between the marker and QTL in a natural population of 200 based on different assumptions of heritability of the simulated QTL.**

Cell Transition Parameters											
QQ		Qq		Qq+		qq		qq+		q+q+	
Given	MLE(Std.)	Given	MLE(Std.)	Given	MLE(Std.)	Given	MLE(Std.)	Given	MLE(Std.)	Given	MLE(Std.)
Heritability of the simulated QTL is $H^2 = 0.05$											
$r$	0.12 0.1184(0.0053)	0.15 0.1502(0.0023)	0.24	0.2396 (0.0027)	0.2	0.1993 (0.0019)	0.27	0.2707(0.0020)	0.3	0.3045(0.0163)	
$k_1$	0.003 0.0045(0.0011)	0.027 0.0307(0.0017)	0.05	0.0637 (0.0093)	0.015	0.0167 (0.0013)	0.018	0.0197(0.0016)	0.011	0.0183(0.0027)	
$k_2$	0.006 0.0075(0.0028)	0.0011 0.0014(0.0012)	0.0007	0.0009 (0.0002)	0.0009	0.0010 (0.0002)	0.007	0.0072(0.0018)	0.003	0.0035(0.0011)	
Heritability of the simulated QTL is $H^2 = 0.1$											
$r$	0.12 0.1165(0.0039)	0.15 0.1504(0.0016)	0.24	0.2396 (0.0024)	0.2	0.1995 (0.0017)	0.27	0.2706(0.0014)	0.3	0.3042(0.0129)	
$k_1$	0.003 0.0046(0.0010)	0.027 0.0307(0.0016)	0.05	0.0607 (0.0067)	0.015	0.0164 (0.0001)	0.018	0.0194(0.0011)	0.011	0.0127(0.0030)	
$k_2$	0.006 0.0075(0.0020)	0.0011 0.0014(0.0008)	0.0007	0.0009 (0.0002)	0.0009	0.0010 (0.0002)	0.007	0.0072(0.0015)	0.003	0.0035(0.0009)	
Heritability of the simulated QTL is $H^2 = 0.2$											
$r$	0.12 0.1175(0.0031)	0.15 0.1504(0.0015)	0.24	0.2397 (0.0023)	0.2	0.1996 (0.0011)	0.27	0.2703(0.0012)	0.3	0.3018(0.0070)	
$k_1$	0.003 0.0044(0.0010)	0.027 0.0299(0.0016)	0.05	0.0572 (0.0074)	0.015	0.0162 (0.0008)	0.018	0.0194(0.0009)	0.011	0.013(0.0019)	
$k_2$	0.006 0.0073(0.0019)	0.0011 0.0014(0.0007)	0.0007	0.0008 (0.0002)	0.0009	0.0010 (0.0001)	0.007	0.0071(0.0013)	0.003	0.0033(0.0009)	
Genetic Parameters											
$H^2 = 0.05$											
Given	MLE(Std.)	Given	MLE(Std.)	Given	MLE(Std.)	Given	MLE(Std.)	Given	MLE(Std.)	Given	MLE(Std.)
$P$	0.6 0.6062(0.0216)	0.6	0.5962(0.0295)	0.6	0.5976 (0.0243)	0.6	0.5976 (0.0243)	0.6	0.5976 (0.0243)	0.6	0.5976 (0.0243)
$Q_1$	0.4 0.3972(0.0211)	0.4	0.4026(0.0218)	0.4	0.4008 (0.0251)	0.4	0.4008 (0.0251)	0.4	0.4008 (0.0251)	0.4	0.4008 (0.0251)
$Q_2$	0.5 0.5036(0.0223)	0.5	0.4963(0.0231)	0.5	0.4973 (0.0272)	0.5	0.4973 (0.0272)	0.5	0.4973 (0.0272)	0.5	0.4973 (0.0272)
$D_1$	0.01 0.0152(0.0191)	0.01	0.0104(0.0177)	0.01	0.0114 (0.0182)	0.01	0.0114 (0.0182)	0.01	0.0114 (0.0182)	0.01	0.0114 (0.0182)
$D_2$	0.02 0.0146(0.0213)	0.02	0.0172(0.0175)	0.02	0.0213 (0.0203)	0.02	0.0213 (0.0203)	0.02	0.0213 (0.0203)	0.02	0.0213 (0.0203)
$H^2 = 0.1$											
Given	MLE(Std.)	Given	MLE(Std.)	Given	MLE(Std.)	Given	MLE(Std.)	Given	MLE(Std.)	Given	MLE(Std.)
$\phi_1$	0.76 0.7644(0.0387)	0.7	0.7204(0.0424)	0.6	0.6004 (0.0362)	0.6	0.6004 (0.0362)	0.6	0.6004 (0.0362)	0.6	0.6004 (0.0362)
$\phi_2$	0.62 0.6238(0.0321)	0.564	0.5566(0.0392)	0.47	0.4694 (0.0362)	0.47	0.4694 (0.0362)	0.47	0.4694 (0.0362)	0.47	0.4694 (0.0362)
$\theta_1$	1.22 1.2194(0.0351)	1.24	1.2412(0.0288)	1.25	1.2569 (0.0329)	1.25	1.2569 (0.0329)	1.25	1.2569 (0.0329)	1.25	1.2569 (0.0329)
$\theta_2$	1.23 1.2279(0.0249)	1.21	1.2134(0.0298)	1.2	1.1992 (0.0245)	1.2	1.1992 (0.0245)	1.2	1.1992 (0.0245)	1.2	1.1992 (0.0245)
$v_1$	0.01 0.0112(0.0031)	0.01	0.0111(0.0018)	0.01	0.0108 (0.0017)	0.01	0.0108 (0.0017)	0.01	0.0108 (0.0017)	0.01	0.0108 (0.0017)
$v_2$	0.01 0.0106(0.0052)	0.01	0.0101(0.0027)	0.01	0.0103 (0.0024)	0.01	0.0103 (0.0024)	0.01	0.0103 (0.0024)	0.01	0.0103 (0.0024)
$\rho$	0.4824 0.4849(0.0681)	0.4824	0.4809(0.0642)	0.4824	0.4891 (0.4525)	0.4824	0.4891 (0.4525)	0.4824	0.4891 (0.4525)	0.4824	0.4891 (0.4525)
$H^2 = 0.2$											
Given	MLE(Std.)	Given	MLE(Std.)	Given	MLE(Std.)	Given	MLE(Std.)	Given	MLE(Std.)	Given	MLE(Std.)
$\phi_1$	0.76 0.7644(0.0387)	0.7	0.7204(0.0424)	0.6	0.6004 (0.0362)	0.6	0.6004 (0.0362)	0.6	0.6004 (0.0362)	0.6	0.6004 (0.0362)
$\phi_2$	0.62 0.6238(0.0321)	0.564	0.5566(0.0392)	0.47	0.4694 (0.0362)	0.47	0.4694 (0.0362)	0.47	0.4694 (0.0362)	0.47	0.4694 (0.0362)
$\theta_1$	1.22 1.2194(0.0351)	1.24	1.2412(0.0288)	1.25	1.2569 (0.0329)	1.25	1.2569 (0.0329)	1.25	1.2569 (0.0329)	1.25	1.2569 (0.0329)
$\theta_2$	1.23 1.2279(0.0249)	1.21	1.2134(0.0298)	1.2	1.1992 (0.0245)	1.2	1.1992 (0.0245)	1.2	1.1992 (0.0245)	1.2	1.1992 (0.0245)
$v_1$	0.01 0.0112(0.0031)	0.01	0.0111(0.0018)	0.01	0.0108 (0.0017)	0.01	0.0108 (0.0017)	0.01	0.0108 (0.0017)	0.01	0.0108 (0.0017)
$v_2$	0.01 0.0106(0.0052)	0.01	0.0101(0.0027)	0.01	0.0103 (0.0024)	0.01	0.0103 (0.0024)	0.01	0.0103 (0.0024)	0.01	0.0103 (0.0024)
$\rho$	0.4824 0.4849(0.0681)	0.4824	0.4809(0.0642)	0.4824	0.4891 (0.4525)	0.4824	0.4891 (0.4525)	0.4824	0.4891 (0.4525)	0.4824	0.4891 (0.4525)

Std. represents the standard deviation of the estimates obtained from 100 simulation replicates.

doi:10.1371/journal.pone.0126937.t002



**Table 3. MLEs of linear ODE parameters that define the dynamics of cell transition for six different QTL genotypes and the association between the marker and QTL in a natural population of 400 based on different assumptions of heritability of the simulated QTL.**

Cell Transition Parameters												
	QQ		Qq		Qq+		qq		qq+		q+q+	
	Given	MLE(Std.)	Given	MLE(Std.)	Given	MLE(Std.)	Given	MLE(Std.)	Given	MLE(Std.)	Given	MLE(Std.)
Heritability of the simulated QTL is $H^2 = 0.05$												
$r$	0.12	0.1197(0.0054)	0.15	0.1483(0.002)	0.24	0.2401(0.0032)	0.2	0.1993(0.0021)	0.27	0.2713(0.0019)	0.3	0.3024(0.0049)
$k_1$	0.003	0.0047(0.0010)	0.027	0.0308(0.0013)	0.05	0.065(0.0101)	0.015	0.0163(0.0012)	0.018	0.0197(0.0013)	0.011	0.0135(0.0019)
$k_2$	0.006	0.0087(0.0019)	0.0011	0.0014(0.0002)	0.0007	0.0010(0.0002)	0.0009	0.0010(0.0002)	0.007	0.0073(0.0017)	0.003	0.0035(0.0010)
Heritability of the simulated QTL is $H^2 = 0.1$												
$r$	0.12	0.1165(0.0042)	0.15	0.1507(0.0016)	0.24	0.2399(0.0027)	0.2	0.1994(0.0013)	0.27	0.2708(0.0012)	0.3	0.3016(0.0060)
$k_1$	0.003	0.0047(0.0010)	0.027	0.0308(0.0014)	0.05	0.063(0.0067)	0.015	0.0168(0.0009)	0.018	0.0196(0.0011)	0.011	0.013(0.002)
$k_2$	0.006	0.0076(0.0018)	0.0011	0.0015(0.0002)	0.0007	0.0009(0.0002)	0.0009	0.0010(0.0002)	0.007	0.0072(0.0013)	0.003	0.0035(0.0009)
Heritability of the simulated QTL is $H^2 = 0.2$												
$r$	0.12	0.1167(0.0027)	0.15	0.1504(0.0013)	0.24	0.2397(0.0022)	0.2	0.1995(0.0011)	0.27	0.2707(0.0011)	0.3	0.3013(0.0036)
$k_1$	0.003	0.0046(0.0008)	0.027	0.0305(0.0013)	0.05	0.060(0.0062)	0.015	0.0165(0.0007)	0.018	0.0194(0.0008)	0.011	0.0135(0.0017)
$k_2$	0.006	0.0069(0.0016)	0.0011	0.0015(0.0001)	0.0007	0.0009(0.0001)	0.0009	0.0010(0.0001)	0.007	0.0073(0.0009)	0.003	0.0035(0.0008)
Genetic Parameters												
$H^2 = 0.1$												
Given	MLE(Std.)		Given	MLE(Std.)		Given	MLE(Std.)		Given	MLE(Std.)		
$P$	0.6	0.6009(0.0172)	0.6	0.6006(0.0196)	0.6	0.5992(0.0134)	0.6	0.5992(0.0134)	0.6	0.5992(0.0134)	0.6	0.5853(0.0418)
$Q_1$	0.4	0.3953(0.021)	0.4	0.3979(0.0171)	0.4	0.3979(0.0171)	0.4	0.402(0.0173)	0.4	0.402(0.0173)	0.4	0.4532(0.0341)
$Q_2$	0.5	0.5048(0.0192)	0.5	0.503(0.0169)	0.5	0.503(0.0169)	0.5	0.5011(0.0147)	0.5	0.5011(0.0147)	0.5	1.2605(0.0293)
$D_1$	0.01	0.0111(0.0097)	0.01	0.0111(0.0122)	0.01	0.0111(0.0122)	0.01	0.0109(0.0109)	0.01	0.0109(0.0109)	0.01	1.2216(0.0337)
$D_2$	0.02	0.0198(0.0103)	0.02	0.0171(0.0117)	0.02	0.0171(0.0117)	0.02	0.0188(0.0112)	0.02	0.0188(0.0112)	0.02	0.0105(0.0013)
$H^2 = 0.2$												
Given	MLE(Std.)		Given	MLE(Std.)		Given	MLE(Std.)		Given	MLE(Std.)		
$\phi_1$	0.76	0.7566(0.0513)	0.7	0.6947(0.0478)	0.6	0.6947(0.0478)	0.6	0.5853(0.0418)	0.6	0.5853(0.0418)	0.6	0.4532(0.0341)
$\phi_2$	0.62	0.6102(0.0501)	0.564	0.5375(0.0386)	0.47	0.5375(0.0386)	0.47	0.4532(0.0341)	0.47	0.4532(0.0341)	0.47	1.2605(0.0293)
$\rho_1$	1.22	1.2338(0.0371)	1.24	1.2561(0.0336)	1.25	1.2561(0.0336)	1.25	1.2605(0.0293)	1.25	1.2605(0.0293)	1.25	1.2216(0.0337)
$\rho_2$	1.23	1.2395(0.0310)	1.21	1.2385(0.0356)	1.2	1.2385(0.0356)	1.2	1.2216(0.0337)	1.2	1.2216(0.0337)	1.2	0.0105(0.0013)
$v_1$	0.01	0.0120(0.002)	0.01	0.0111(0.0017)	0.01	0.0111(0.0017)	0.01	0.0105(0.0013)	0.01	0.0105(0.0013)	0.01	0.0101(0.0011)
$v_2$	0.01	0.0114(0.0043)	0.01	0.0103(0.0014)	0.01	0.0103(0.0014)	0.01	0.0101(0.0011)	0.01	0.0101(0.0011)	0.01	0.4503(0.0377)
$\rho$	0.4824	0.4135(0.0477)	0.4824	0.4304(0.0393)	0.4824	0.4304(0.0393)	0.4824	0.4503(0.0377)	0.4824	0.4503(0.0377)	0.4824	0.4503(0.0377)

Std. represents the standard deviation of the estimates obtained from 100 simulation replicates.

doi:10.1371/journal.pone.0126937.t003

**Table 4. MLEs of nonlinear ODE parameters that define the dynamics of cell transition for six different QTL genotypes and the association between the marker and QTL in a natural population of 200 based on different assumptions of heritability of the simulated QTL.**

	Cell Transition Parameters													
	QQ		Qq		Qq+		qq		qq+		qq++			
	Given	MLE(Std.)	Given	MLE(Std.)	Given	MLE(Std.)	Given	MLE(Std.)	Given	MLE(Std.)	Given	MLE(Std.)		
$r$	0.12	0.1329 (0.0429)	0.22	0.2292 (0.0148)	0.24	0.2611(0.0137)	0.2	0.2257 (0.0084)	0.27	0.3077(0.0134)	0.3	0.3081(0.0108)		
$k_1$	0.003	0.0035 (0.0035)	0.027	0.0272 (0.0010)	0.09	0.083(0.0245)	0.015	0.0097 (0.0028)	0.0018	0.0019(0.0002)	0.011	0.0109(0.0015)		
$k_2$	0.0006	0.0003 (0.0009)	0.0011	0.0003 (0.0006)	0.0007	0.0015(0.0017)	0.0009	0.0030 (0.0008)	0.001	0.0018(0.0017)	0.0009	0.0007(0.0008)		
$k_3$	0.66	0.6547 (0.0096)	0.69	0.6921 (0.0104)	0.6	0.597(0.0083)	0.55	0.5354 (0.0105)	0.5	0.5054(0.0108)	0.65	0.6472(0.0105)		
$k_4$	0.55	0.5469 (0.0079)	0.5	0.4938 (0.0083)	0.6	0.6009(0.0081)	0.65	0.6579 (0.0111)	0.6	0.5989(0.0087)	0.5	0.4973(0.0084)		
Heritability of the simulated QTL is $H^2 = 0.1$														
$r$	0.12	0.1378 (0.0258)	0.22	0.2271 (0.0118)	0.24	0.2517(0.0095)	0.2	0.2057 (0.0179)	0.27	0.2969(0.0116)	0.3	0.3044(0.0093)		
$k_1$	0.003	0.0035 (0.0020)	0.027	0.0267 (0.0021)	0.09	0.0931(0.0122)	0.015	0.0106 (0.0019)	0.0018	0.0020(0.0002)	0.011	0.0109(0.0012)		
$k_2$	0.0006	0.0004 (0.0008)	0.0011	0.0005 (0.0006)	0.0007	0.0013(0.0022)	0.0009	0.0025 (0.0008)	0.001	0.0019(0.0016)	0.0009	0.0009(0.0010)		
$k_3$	0.66	0.6628 (0.0074)	0.69	0.6884 (0.0200)	0.6	0.5992(0.0062)	0.55	0.5387 (0.0090)	0.5	0.5063(0.0143)	0.65	0.6502(0.0078)		
$k_4$	0.55	0.5361 (0.0157)	0.5	0.4982 (0.0077)	0.6	0.5994(0.0118)	0.65	0.6619 (0.0070)	0.6	0.5969(0.0138)	0.5	0.4965(0.0183)		
Heritability of the simulated QTL is $H^2 = 0.2$														
$r$	0.12	0.1055 (0.0092)	0.22	0.2253 (0.0039)	0.24	0.2629(0.0216)	0.2	0.2059 (0.0179)	0.27	0.2894(0.0069)	0.3	0.2939(0.0073)		
$k_1$	0.003	0.0035 (0.0004)	0.027	0.0271 (0.0012)	0.09	0.0901(0.0153)	0.015	0.0108 (0.0032)	0.0018	0.0019(0.0001)	0.011	0.0105(0.0008)		
$k_2$	0.0006	0.0005 (0.0007)	0.0011	0.0009 (0.0005)	0.0007	0.0010(0.0009)	0.0009	0.0024 (0.0004)	0.001	0.0023(0.0012)	0.0009	0.0009(0.0007)		
$k_3$	0.66	0.6591 (0.0052)	0.69	0.6899 (0.0059)	0.6	0.6056(0.0086)	0.55	0.5521 (0.0083)	0.5	0.5081(0.0128)	0.65	0.6477(0.0056)		
$k_4$	0.55	0.5464 (0.0054)	0.5	0.4993 (0.0057)	0.6	0.6017(0.0091)	0.65	0.6632 (0.0052)	0.6	0.6020(0.0083)	0.5	0.5003(0.0062)		
Genetic Parameters														
$H^2 = 0.05$														
Given	MLE(Std.)		Given	MLE(Std.)		Given	MLE(Std.)		Given	MLE(Std.)		Given	MLE(Std.)	
$P$	0.6	0.6062(0.0216)	0.6	0.5962(0.0295)	0.6	0.5976 (0.0243)	0.6	0.5976 (0.0243)	0.6	0.5976 (0.0243)	0.6	0.5976 (0.0243)	0.6	0.5976 (0.0243)
$Q_1$	0.4	0.3972(0.0211)	0.4	0.4026(0.0218)	0.4	0.4008 (0.0251)	0.4	0.4008 (0.0251)	0.4	0.4008 (0.0251)	0.4	0.4008 (0.0251)	0.4	0.4008 (0.0251)
$Q_2$	0.5	0.5036(0.0223)	0.5	0.4963(0.0231)	0.5	0.4973 (0.0272)	0.5	0.4973 (0.0272)	0.5	0.4973 (0.0272)	0.5	0.4973 (0.0272)	0.5	0.4973 (0.0272)
$D_1$	0.01	0.0152(0.0191)	0.01	0.0104(0.0177)	0.01	0.0114 (0.0182)	0.01	0.0114 (0.0182)	0.01	0.0114 (0.0182)	0.01	0.0114 (0.0182)	0.01	0.0114 (0.0182)
$D_2$	0.02	0.0146(0.0213)	0.02	0.0172(0.0175)	0.02	0.0213 (0.0203)	0.02	0.0213 (0.0203)	0.02	0.0213 (0.0203)	0.02	0.0213 (0.0203)	0.02	0.0213 (0.0203)
Matrix Structuring Parameters														
$H^2 = 0.05$														
Given	MLE(Std.)		Given	MLE(Std.)		Given	MLE(Std.)		Given	MLE(Std.)		Given	MLE(Std.)	
$\phi_1$	0.76	0.7400(0.0119)	0.7	0.6927(0.0089)	0.6	0.6927(0.0089)	0.6	0.6927(0.0089)	0.6	0.6927(0.0089)	0.6	0.6927(0.0089)	0.6	0.6927(0.0089)
$\phi_2$	0.62	0.6161(0.0089)	0.564	0.5581(0.0099)	0.47	0.5581(0.0099)	0.47	0.5581(0.0099)	0.47	0.5581(0.0099)	0.47	0.5581(0.0099)	0.47	0.5581(0.0099)
$\theta_1$	1.06	1.0521(0.0089)	1.0824	1.0782(0.0121)	1.11	1.0782(0.0121)	1.11	1.0782(0.0121)	1.11	1.0782(0.0121)	1.11	1.0782(0.0121)	1.11	1.0782(0.0121)
$\theta_2$	1.19	1.1829(0.0086)	1.1757	1.1735(0.0094)	1.16	1.1735(0.0094)	1.16	1.1735(0.0094)	1.16	1.1735(0.0094)	1.16	1.1735(0.0094)	1.16	1.1735(0.0094)

(Continued)

**Table 4. (Continued)**

Cell Transition Parameters											
QQ		Qq		Qq+		qq		qq+		qq+q+	
Given	MLE(Std.)	Given	MLE(Std.)	Given	MLE(Std.)	Given	MLE(Std.)	Given	MLE(Std.)	Given	MLE(Std.)
$v_1$	0.01	0.0108(0.0041)	0.01	0.0108(0.0022)	0.01	0.0107(0.0018)	0.01	0.0101(0.0012)	0.0107(0.0018)	Given	MLE(Std.)
$v_2$	0.01	0.0103(0.0027)	0.01	0.0102(0.0018)	0.01	0.0101(0.0012)	0.01	0.0101(0.0012)	0.0101(0.0012)	Given	MLE(Std.)
$\rho$	0.6624	0.6721(0.0183)	0.6624	0.6619(0.0072)	0.6624	0.6631(0.0084)	0.6624	0.6631(0.0084)	0.6631(0.0084)	Given	MLE(Std.)

Std. represents the standard deviation of the estimates obtained from 100 simulation replicates.

doi:10.1371/journal.pone.0126937.t004

**Table 5. MLEs of nonlinear ODE parameters that define the dynamics of cell transition for six different QTL genotypes and the association between the marker and QTL in a natural population of 400 based on different assumptions of heritability of the simulated QTL.**

Cell Transition Parameters											
Q0		Q1		Q2		Q3		Q4		Q5	
Given	MLE(Std.)	Given	MLE(Std.)	Given	MLE(Std.)	Given	MLE(Std.)	Given	MLE(Std.)	Given	MLE(Std.)
Heritability of the simulated QTL is $H^2 = 0.05$											
$r$	0.12 0.1321(0.0315)	0.22 0.2277(0.0122)	0.24 0.2711(0.0120)	0.2	0.2184 (0.0075)	0.27	0.3049(0.0126)	0.3	0.3052(0.0096)	0.3	0.3052(0.0096)
$k_1$	0.003 0.0039(0.0028)	0.027 0.0270(0.0019)	0.09 0.0852(0.0123)	0.015	0.0103 (0.0028)	0.0018	0.0020(0.0003)	0.011	0.0109(0.0018)	0.011	0.0109(0.0018)
$k_2$	0.0006 0.0005(0.0007)	0.0011 0.0013(0.0006)	0.0007 0.0014(0.0037)	0.0009	0.0028 (0.0008)	0.001	0.0023(0.0016)	0.009	0.0074(0.0096)	0.009	0.0074(0.0096)
$k_3$	0.66 0.6578(0.0088)	0.69 0.6901(0.0204)	0.6 0.5985(0.0086)	0.55	0.5554 (0.0096)	0.5	0.5094(0.0128)	0.65	0.6578(0.0122)	0.65	0.6578(0.0122)
$k_4$	0.55 0.5389(0.0085)	0.5 0.4963(0.0085)	0.6 0.6008(0.0114)	0.65	0.6658 (0.0112)	0.6	0.5962(0.0097)	0.5	0.4957(0.0165)	0.5	0.4957(0.0165)
Heritability of the simulated QTL is $H^2 = 0.1$											
$r$	0.12 0.1338(0.0210)	0.22 0.2275(0.0112)	0.24 0.2516(0.0083)	0.2	0.2064 (0.0151)	0.27	0.2949(0.0077)	0.3	0.3044(0.0076)	0.3	0.3044(0.0076)
$k_1$	0.003 0.0038(0.0015)	0.027 0.0269(0.0017)	0.09 0.0931(0.0087)	0.015	0.0104 (0.0017)	0.0018	0.0020(0.0002)	0.011	0.0110(0.0013)	0.011	0.0110(0.0013)
$k_2$	0.0006 0.0006(0.0007)	0.0011 0.0015(0.0005)	0.0007 0.0014(0.0024)	0.0009	0.0026 (0.0005)	0.001	0.0026(0.0015)	0.009	0.0062(0.0082)	0.009	0.0062(0.0082)
$k_3$	0.66 0.6630(0.0078)	0.69 0.6888(0.0198)	0.6 0.6014(0.0074)	0.55	0.5594 (0.0082)	0.5	0.5082(0.0149)	0.65	0.6497(0.0137)	0.65	0.6497(0.0137)
$k_4$	0.55 0.5346(0.0208)	0.5 0.4985(0.0061)	0.6 0.5983(0.0131)	0.65	0.6626 (0.0067)	0.6	0.5977(0.0125)	0.5	0.4957(0.0159)	0.5	0.4957(0.0159)
Heritability of the simulated QTL is $H^2 = 0.2$											
$r$	0.12 0.1247(0.0075)	0.22 0.2241(0.0031)	0.24 0.2735(0.0230)	0.2	0.2084 (0.0146)	0.27	0.2873(0.0047)	0.3	0.2952(0.0062)	0.3	0.2952(0.0062)
$k_1$	0.003 0.0038(0.0002)	0.027 0.0270(0.0010)	0.09 0.105(0.0163)	0.015	0.0107 (0.0314)	0.0018	0.0019(0.00001)	0.011	0.0107(0.0008)	0.011	0.0107(0.0008)
$k_2$	0.0006 0.0006(0.0005)	0.0011 0.0011(0.0005)	0.0007 0.0008(0.0013)	0.0009	0.0022 (0.0003)	0.001	0.0023(0.0008)	0.009	0.0110(0.0073)	0.009	0.0110(0.0073)
$k_3$	0.66 0.6583(0.0063)	0.69 0.6878(0.0063)	0.6 0.6089(0.0110)	0.55	0.5531 (0.0068)	0.5	0.5065(0.0072)	0.65	0.6484(0.0055)	0.65	0.6484(0.0055)
$k_4$	0.55 0.5477(0.0047)	0.5 0.5003(0.0054)	0.6 0.6053(0.0075)	0.65	0.6631 (0.0036)	0.6	0.6010(0.0087)	0.5	0.5009(0.0062)	0.5	0.5009(0.0062)
Genetic Parameters											
$H^2 = 0.1$											
$H^2 = 0.2$											
Matrix Structuring Parameters											
$H^2 = 0.1$											
$H^2 = 0.2$											
Heritability of the simulated QTL is $H^2 = 0.05$											
$P$	0.6 0.5991(0.0178)	0.6 0.6025(0.5991)	0.6 0.6025(0.5991)	0.6	0.6 0.0178(0.6025)	0.6	0.6 0.0178(0.6025)	0.6	0.6 0.0178(0.6025)	0.6	0.6 0.0178(0.6025)
$Q_1$	0.4 0.4005(0.0181)	0.4 0.3995(0.4005)	0.4 0.3995(0.4005)	0.4	0.4 0.0181(0.3995)	0.4	0.4 0.0181(0.3995)	0.4	0.4 0.0181(0.3995)	0.4	0.4 0.0181(0.3995)
$Q_2$	0.5 0.5006(0.0187)	0.5 0.5007(0.5006)	0.5 0.5007(0.5006)	0.5	0.5 0.0187(0.5007)	0.5	0.5 0.0187(0.5007)	0.5	0.5 0.0187(0.5007)	0.5	0.5 0.0187(0.5007)
$D_1$	0.01 0.09(0.0121)	0.01 0.0093(0.009)	0.01 0.0093(0.009)	0.01	0.01 0.0121(0.0093)	0.01	0.01 0.0121(0.0093)	0.01	0.01 0.0121(0.0093)	0.01	0.01 0.0121(0.0093)
$D_2$	0.02 0.0212(0.0126)	0.02 0.0202(0.0212)	0.02 0.0202(0.0212)	0.02	0.02 0.0126(0.0202)	0.02	0.02 0.0126(0.0202)	0.02	0.02 0.0126(0.0202)	0.02	0.02 0.0126(0.0202)
Heritability of the simulated QTL is $H^2 = 0.1$											
$\phi_1$	0.76 0.7504(0.0132)	0.7 0.6906(0.0068)	0.7 0.6906(0.0068)	0.6	0.6 0.5975(0.0157)	0.6	0.6 0.5975(0.0157)	0.6	0.6 0.5975(0.0157)	0.6	0.6 0.5975(0.0157)
$\phi_2$	0.62 0.6206(0.0072)	0.564 0.5589(0.0068)	0.564 0.5589(0.0068)	0.47	0.47 0.4657(0.0064)	0.47	0.47 0.4657(0.0064)	0.47	0.47 0.4657(0.0064)	0.47	0.47 0.4657(0.0064)
$\psi_1$	1.06 1.0598(0.0069)	1.0824 1.0801(0.0069)	1.0824 1.0801(0.0069)	1.11	1.11 1.1061(0.0044)	1.11	1.11 1.1061(0.0044)	1.11	1.11 1.1061(0.0044)	1.11	1.11 1.1061(0.0044)
$\psi_2$	1.19 1.1879(0.0056)	1.1757 1.1721(0.0106)	1.1757 1.1721(0.0106)	1.16	1.16 1.1551(0.0058)	1.16	1.16 1.1551(0.0058)	1.16	1.16 1.1551(0.0058)	1.16	1.16 1.1551(0.0058)
$v_1$	0.01 0.0107(0.0033)	0.01 0.0106(0.0021)	0.01 0.0106(0.0021)	0.01	0.01 0.0106(0.0021)	0.01	0.01 0.0106(0.0021)	0.01	0.01 0.0106(0.0021)	0.01	0.01 0.0106(0.0021)
$v_2$	0.01 0.0104(0.0024)	0.01 0.0104(0.0012)	0.01 0.0104(0.0012)	0.01	0.01 0.0104(0.0012)	0.01	0.01 0.0104(0.0012)	0.01	0.01 0.0104(0.0012)	0.01	0.01 0.0104(0.0012)
$p$	0.6624 0.6668(0.0152)	0.6624 0.6598(0.0130)	0.6624 0.6598(0.0130)	0.6624	0.6624 0.6616(0.0108)	0.6624	0.6624 0.6616(0.0108)	0.6624	0.6624 0.6616(0.0108)	0.6624	0.6624 0.6616(0.0108)

Std. represents the standard deviation of the estimates obtained from 100 simulation replicates.

doi:10.1371/journal.pone.0126937.t005

in both linear and nonlinear models. The precision of estimation of marker allele frequency is not affected by differences in heritability, but estimates of transition parameters, QTL allele frequency, and marker-QTL linkage disequilibrium are more precise for a higher than a lower heritability and for a bigger than a smaller population. The covariance-structural parameters can also be estimated, partly because of their simple structure for covariance modeling.

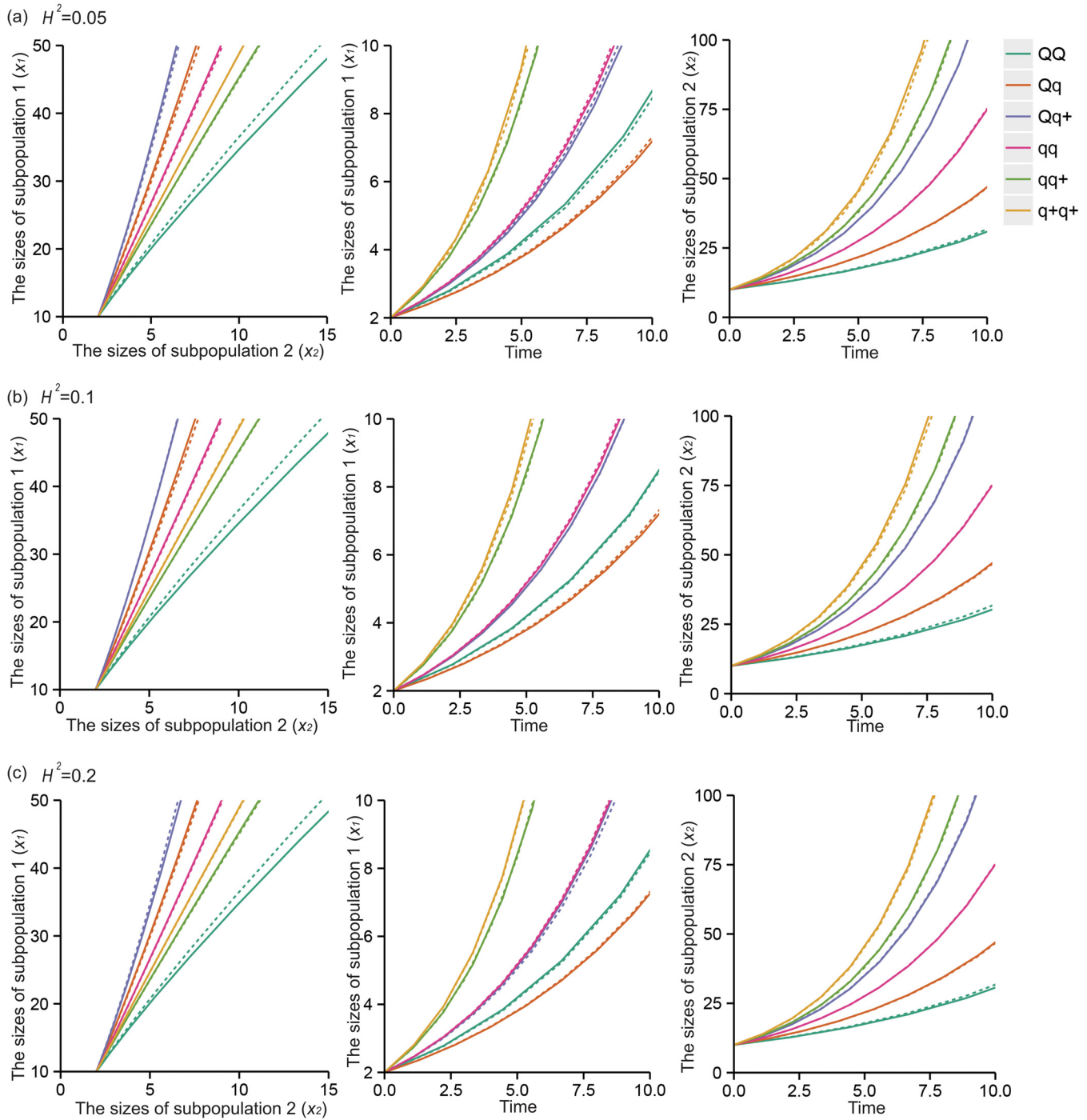
The six QTL genotypes  $QQ$ ,  $Qq$ ,  $Qq^+$ ,  $qq$ ,  $qq^+$ , and  $q^+q^+$  are each hypothesized to have different response curves for different human populations. Figs 3 and 4 illustrate different forms of the cell population transition for six QTL genotypes,  $QQ$ ,  $Qq$ ,  $Qq^+$ ,  $qq$ ,  $qq^+$ , and  $q^+q^+$ , under different heritabilities (0.05, 0.1 and 0.2) with size 400 populations, with the transitional values given in Table 3 for the linear model and Table 5 for the nonlinear model. Pronounced differences among the genotypes suggest that the QTL may affect cell transitions, resulting in different cellular phenotypes. Meanwhile, we considered the methylated QTL status ( $Qq^+$ ,  $qq^+$ , and  $q^+q^+$ ), which could illustrate phenotypic differences among cells with the same DNA sequences. The cell transition values can be estimated from the model. The model displays great power in detecting a QTL responsible for cell transitions using the associated marker. The trajectories of additive and dominant effects of subpopulations  $x_1$  and  $x_2$  are shown as Figs 5 and 6 for the linear model and nonlinear model, respectively. We could observe the genetic architecture of the transition dynamics of the two subpopulation cells.

## Discussion

As more investigators use stem/progenitor cells to study the mechanisms of pluripotency and differentiation, they pay greater attention to cell heterogeneity and variability, which affects the use of pluripotent cells in regenerative medicine, disease modeling, and studying development processes. New methodologies, including single-cell and single-molecule analysis, as well as mathematical and computational modeling have been developed to explore pluripotent cell population heterogeneity. Emerging evidence has found differences in gene expression profiles at the mRNA level and functional differences in the differentiation within a heterogeneous cell population. However, the genetic status such as genetic backgrounds and epigenetic profiles including methylation of CpG islands and histone modifications remain unclear. Previous reports by Chang et al. [6] showed that transcriptome noise leads to clonal heterogeneity and controls lineage choices, but the genetic mechanisms of the heterogeneity and differentiation potential were not described.

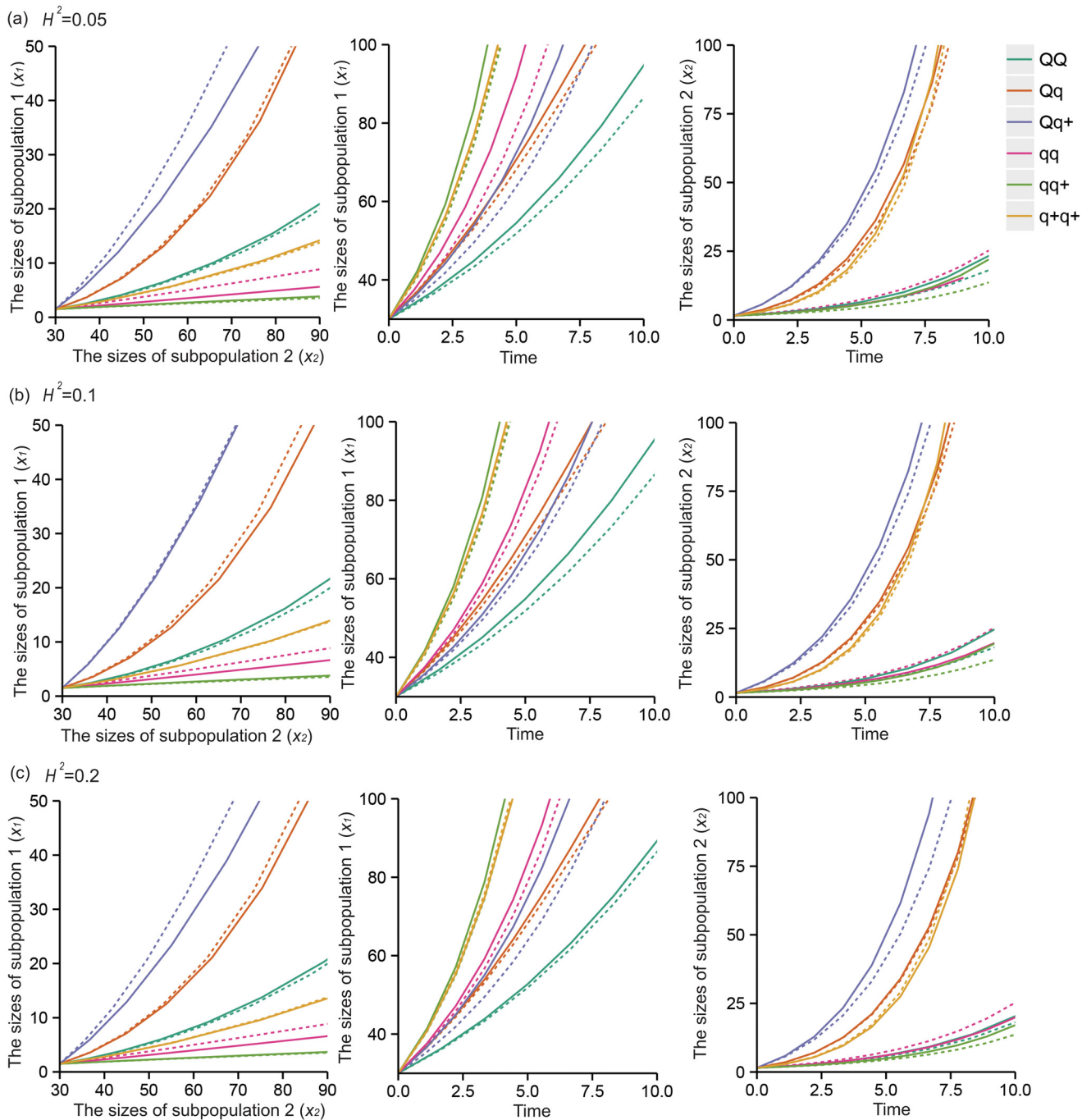
In this article, we developed a statistical model that combined the mathematical concepts regarding cellular mechanisms and a general framework for mapping dynamic traits [15]. Both linear and nonlinear ODEs use specific parameters to describe how cells transit from one state to another during cell culture and test the magnitude and patterns of genetic effects in the process. The estimates of response curves are more precise for the linear model (Fig 3). However, the nonlinear model includes more parameters and information regarding complex cell proliferation and transition processes, so it may be more close to the true conditions. This model first considers methylated DNA as an allele in this “heterogeneity” based question, which indicates that different methylation states affect the phenotypes of the cells with the same genotype. Therefore, the model provides a tool to generate biologically meaningful hypotheses for understanding the genetic control of population cell heterogeneity.

The model proposed in this article considered only six genotypes; namely,  $QQ$ ,  $Qq$ ,  $Qq^+$ ,  $qq$ ,  $qq^+$ , and  $q^+q^+$ , but biologically there should be more methylated states of the genotypes such as  $Q^+Q^+$ ,  $Q^+q$ , and  $Q^+q^+$ . The allele frequencies of the marker, QTL and their LD, joint marker and phenotypic data, and the statistical model will be more complex in this situation. In addition, further cellular and molecular experiments are required to confirm the exact DNA



**Fig 3. Estimated and true curves of systems mapping for the linear progenitor cell transition dynamics model.** The curves show a putative QTL having six genotypes, QQ, Qq, Qq<sup>+</sup>, qq, qq<sup>+</sup>, and q<sup>+</sup>q<sup>+</sup>, as indicated by colors in a natural population of 400 assuming heritability  $H^2 = 0.05$  (a),  $H^2 = 0.1$  (b), and  $H^2 = 0.2$  (c), respectively. The broad consistency between the estimated (solid) and true curves (broken) suggests that the model provides a good estimate of the dynamic system.

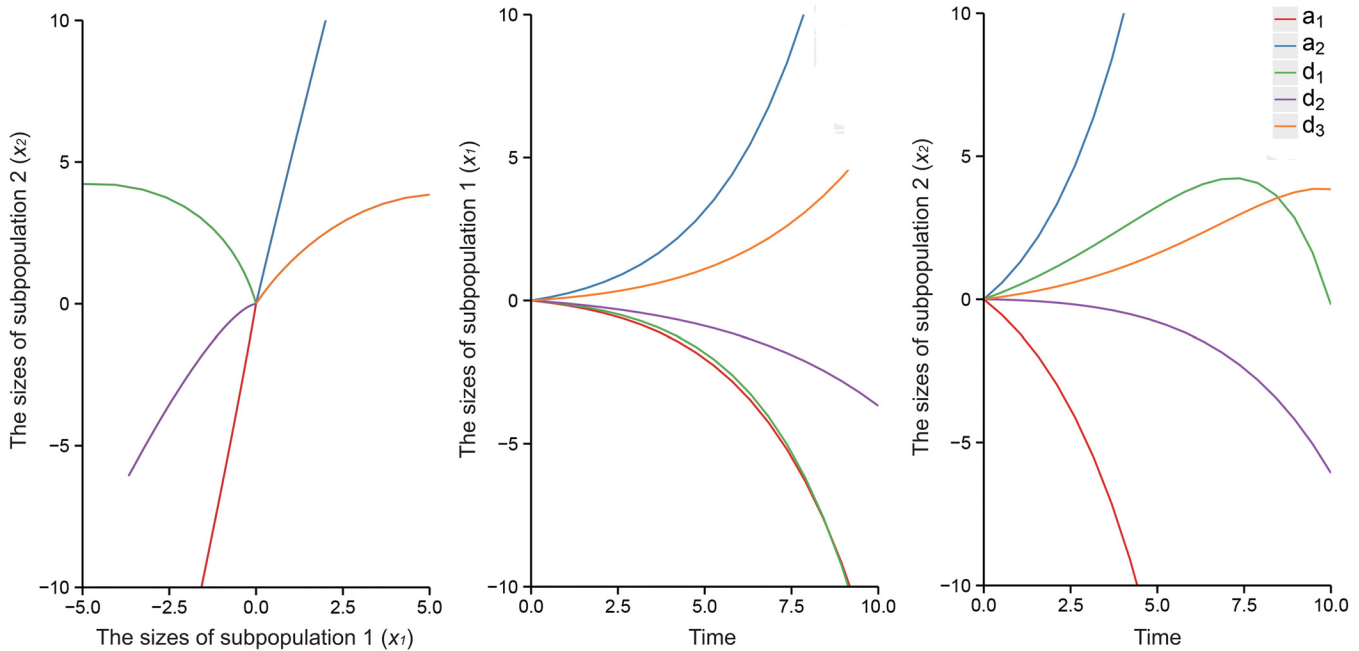
doi:10.1371/journal.pone.0126937.g003



**Fig 4. Estimated and true curves of systems mapping for the nonlinear progenitor cell transition dynamics model.** The curves show a putative QTL having six genotypes,  $QQ$ ,  $Qq$ ,  $Qq^+$ ,  $qq$ ,  $qq^+$ , and  $q^+q^+$ , as indicated by colors in a natural population of 400 assuming heritability  $H^2 = 0.05$  (a),  $H^2 = 0.1$  (b), and  $H^2 = 0.2$  (c), respectively. The broad consistency between the estimated (solid) and true curves (broken) suggests that the model provides a good estimate of the dynamic system.

doi:10.1371/journal.pone.0126937.g004

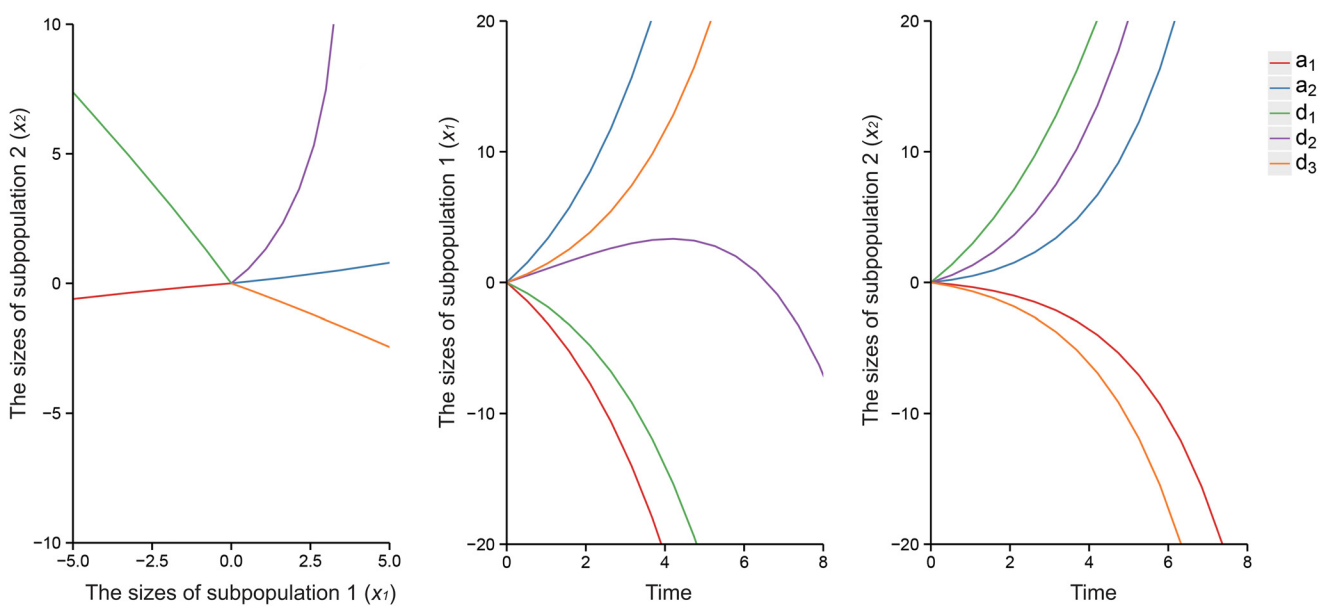




**Fig 5. Trajectories of genetic effects on the linear progenitor cell transition dynamics model, including additive and dominant effects from  $x_1$  and  $x_2$ .**

doi:10.1371/journal.pone.0126937.g005

modification states of these cells, and a model incorporating multiple QTL and their interactive networks should be derived. However, the model will be useful in elucidating the genetic architecture, especially epigenetic mechanisms of cell heterogeneity, and could increase our understanding of the driving forces behind cell heterogeneity and transitions.



**Fig 6. Trajectories of genetic effects on nonlinear progenitor cell transition dynamics model, including additive and dominant effects from  $x_1$  and  $x_2$ .**

doi:10.1371/journal.pone.0126937.g006

## Acknowledgments

We thank Dr. Wang Z and Ye M for beneficial revisions. We thank all members from Dr. Wu's Lab for their support during the study.

## Author Contributions

Conceived and designed the experiments: YG RW. Performed the experiments: LZ RW. Analyzed the data: LZ DY JG HZ HS. Contributed reagents/materials/analysis tools: LZ YS. Wrote the paper: YG LZ YS LJ.

## References

1. Akashi K, Traver D, Miyamoto T, Weissman IL (2000) A clonogenic common myeloid progenitor that gives rise to all myeloid lineages. *Nature* 404: 193–197. PMID: [10724173](#)
2. Kondo M, Weissman IL, Akashi K (1997) Identification of clonogenic common lymphoid progenitors in mouse bone marrow. *Cell* 91: 661–672. PMID: [9393859](#)
3. Suh HC, Gooya J, Renn K, Friedman AD, Johnson PF, Keller JR (2006) C/EBPalpha determines hematopoietic cell fate in multipotential progenitor cells by inhibiting erythroid differentiation and inducing myeloid differentiation. *Blood* 107: 4308–4316. PMID: [16469877](#)
4. Clements WK, Traver D (2013) Signalling pathways that control vertebrate haematopoietic stem cell specification. *Nat Rev Immunol* 13: 336–348. doi: [10.1038/nri3443](#) PMID: [23618830](#)
5. Wolff L, Humeniuk R (2013) Concise review: erythroid versus myeloid lineage commitment: regulating the master regulators. *Stem Cells* 31: 1237–1244. doi: [10.1002/stem.1379](#) PMID: [23559316](#)
6. Chang HH, Hemberg M, Barahona M, Ingber DE, Huang S (2008) Transcriptome-wide noise controls lineage choice in mammalian progenitor cells. *Nature* 453: 544–547. doi: [10.1038/nature06965](#) PMID: [18497826](#)
7. Mikkelsen TS, Ku M, Jaffe DB, Issac B, Lieberman E, Giannoukos G, et al. (2007) Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* 448: 553–560. PMID: [17603471](#)
8. Ji H, Ehrlich LI, Seita J, Murakami P, Doi A, Lindau P, et al. (2010) Comprehensive methylome map of lineage commitment from haematopoietic progenitors. *Nature* 467: 338–342. doi: [10.1038/nature09367](#) PMID: [20720541](#)
9. Ma CX, Casella G, Wu R (2002) Functional mapping of quantitative trait loci underlying the character process: a theoretical framework. *Genetics* 161: 1751–1762. PMID: [12196415](#)
10. Cui Y, Zhu J, Wu R (2006) Functional mapping for genetic control of programmed cell death. *Physiol Genomics* 25: 458–469. PMID: [16464975](#)
11. Liu T, Liu X, Chen Y, Wu R (2007) A computational model for functional mapping of genes that regulate intra-cellular circadian rhythms. *Theor Biol Med Model* 4: 5. PMID: [17261199](#)
12. Wang Z, Liu J, Wang J, Wang Y, Wang N, Li Y, et al. (2012) Dynamic modeling of genes controlling cancer stem cell proliferation. *Front Genet* 3: 84. doi: [10.3389/fgene.2012.00084](#) PMID: [22661984](#)
13. Gibbs JR, van der Brug MP, Hernandez DG, Traynor BJ, Nalls MA, Lai SL, et al. (2010) Abundant quantitative trait loci exist for DNA methylation and gene expression in human brain. *PLoS Genet* 6: e1000952. doi: [10.1371/journal.pgen.1000952](#) PMID: [20485568](#)
14. Bell JT, Pai AA, Pickrell JK, Gaffney DJ, Pique-Regi R, Degner JF, et al. (2011) DNA methylation patterns associate with genetic and gene expression variation in HapMap cell lines. *Genome Biol* 12: R10. doi: [10.1186/gb-2011-12-1-r10](#) PMID: [21251332](#)
15. Wu R, Cao J, Huang Z, Wang Z, Gai J, Vallejos E (2011) Systems mapping: how to improve the genetic mapping of complex traits through design principles of biological systems. *BMC Syst Biol* 5: 84. doi: [10.1186/1752-0509-5-84](#) PMID: [21615967](#)
16. Zhao W, Hou W, Littell RC, Wu R (2005) Structured antedependence models for functional mapping of multiple longitudinal traits. *Stat Appl Genet Mol Biol* 4: Article33. PMID: [16646852](#)
17. Li N, McMurry T, Berg A, Wang Z, Berceli SA, Wu R (2010) Functional clustering of periodic transcriptional profiles through ARMA(p,q). *PLoS One* 5: e9894. doi: [10.1371/journal.pone.0009894](#) PMID: [20419127](#)
18. Yap JS, Fan J, Wu R (2009) Nonparametric modeling of longitudinal covariance structure in functional mapping of quantitative trait loci. *Biometrics* 65: 1068–1077. doi: [10.1111/j.1541-0420.2009.01222.x](#) PMID: [19302406](#)

19. Das K, Li J, Fu G, Wang Z, Li R, Wu R (2013) Dynamic semiparametric Bayesian models for genetic mapping of complex trait with irregular longitudinal data. *Stat Med* 32: 509–523. doi: [10.1002/sim.5535](https://doi.org/10.1002/sim.5535) PMID: [22903809](https://pubmed.ncbi.nlm.nih.gov/22903809/)
20. Zhao W, Li H, Hou W, Wu R (2007) Wavelet-based parametric functional mapping of developmental trajectories with high-dimensional data. *Genetics* 176: 1879–1892. PMID: [17435222](https://pubmed.ncbi.nlm.nih.gov/17435222/)
21. Runge C (1895) Ueber die numerische Auflösung von Differentialgleichungen. *Math Ann* 46: 167–178.
22. Kutta W (1901) Beitrag zur näherungsweise Integration totaler Differentialgleichungen. *Z Math Phys* 46: 435–453.
23. Churchill GA, Doerge RW (1994) Empirical threshold values for quantitative trait mapping. *Genetics* 138: 963–971. PMID: [7851788](https://pubmed.ncbi.nlm.nih.gov/7851788/)
24. Fu G, Luo J, Berg A, Wang Z, Li J, Das K, et al. (2011) A dynamic model for functional mapping of biological rhythms. *J Biol Dyn* 5: 84–101. doi: [10.1080/17513758.2010.491558](https://doi.org/10.1080/17513758.2010.491558) PMID: [21278847](https://pubmed.ncbi.nlm.nih.gov/21278847/)