

SOFTWARE

DelaySSA: stochastic simulation of biochemical systems and gene regulatory networks with or without time delays

Ziyan Jin¹, Xinyi Zhou², Zhaoyuan Fang^{1,3,4,5,6*}

1 Department of Colorectal Surgery and Oncology of the Second Affiliated Hospital, and Centre of Biomedical Systems and Informatics of Zhejiang University-University of Edinburgh Institute (ZJU-UoE Institute), Zhejiang University School of Medicine, Zhejiang University, Hangzhou, China, **2** State Key Laboratory of Bioreactor Engineering, East China University of Science and Technology, Shanghai, China, **3** Edinburgh Medical School, College of Medicine and Veterinary Medicine, The University of Edinburgh, Edinburgh, United Kingdom, **4** Key Laboratory of Cancer Prevention and Intervention, China National Ministry of Education, Hangzhou, China, **5** Biomedical and Health Translational Research Center of Zhejiang Province, Haining, China

* zhaoyuanfang@intl.zju.edu.cn



OPEN ACCESS

Citation: Jin Z, Zhou X, Fang Z (2025) DelaySSA: stochastic simulation of biochemical systems and gene regulatory networks with or without time delays. *PLoS Comput Biol* 21(4): e1012919. <https://doi.org/10.1371/journal.pcbi.1012919>

Editor: Chunhe Li, Fudan University, CHINA

Received: October 8, 2024

Accepted: February 26, 2025

Published: April 8, 2025

Copyright: © 2025 Jin 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: DelaySSA and tutorials with comprehensive examples and applications are available at <https://github.com/Zoey-JIN/DelaySSA> and <https://github.com/FangZY-Lab/DelaySSA-dev>.

Funding: This work was supported by the National Natural Science Foundation of China (grant number 82273093 and T2341024 to Z.F.), the Huadong Medicine Joint Funds of the Zhejiang Provincial Natural Science Foundation of China (No. LHDMZ23H160006 to Z.F.), and

Abstract

Stochastic Simulation Algorithm (SSA) is crucial for modeling biochemical reactions and gene regulatory networks. Traditional SSA is characterized by Markovian property and cannot naturally model systems with time delays. Several algorithms have already been designed to handle delayed reactions, yet few easy-to-use implementations exist. To address these challenges, we have developed DelaySSA, an R package that implements currently available algorithms for SSA with or without delays. Meanwhile, we also provided Matlab and Python versions to support wider applications. We demonstrated its accuracy and validity by simulating two classical models: the Bursty model and Refractory model. We then tested its capability to simulate the RNA Velocity model, where it successfully reproduced both the up- and down-regulation stages in the phase portrait. Finally, we extended its application to simulate a gene regulatory network of lung cancer adeno-to-squamous transition (AST) and qualitatively analyzed its bistability behavior by approximating the Waddington's landscape. Modeling the therapeutic intervention of a SOX2 degrader as a delayed degradation reaction, AST is effectively blocked and reprogrammed back to the adenocarcinoma state, providing a useful clue for targeting drug-resistant AST in the future. Taken together, DelaySSA is a powerful and easy-to-use software suite, facilitating accurate modeling of various kinds of biological systems and broadening the scope of stochastic simulations in systems biology.

Introduction

Stochastic Simulation Algorithm (SSA) has become a widely used tool in modeling biochemical reactions [1]. This algorithm directly simulates the temporal evolution of a spatially homogeneous system of molecular species undergoing reactions, providing exact time trajectories that reflect the stochastic and fluctuating nature of biochemical processes [2–5].

Medical Interdisciplinary Innovation Program 2024, Zhejiang University School of Medicine (to Z.F.). 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.

Importantly, SSA is capable of simulating many cellular and molecular models that are recently under extensive studies. For instance, one class of models is the molecular mechanisms driving gene expression state dynamics of individual cells, which have become a highly influential and fruitful field in single-cell studies of development and diseases [6–9]. Another example class of models includes the molecular driver networks underlying various biological processes, such as the lung cancer adeno-to-squamous (AST) lineage transition models explaining targeted-therapy-related drug resistance [10–12]. Simulating these models with SSA is expected to produce results that are consistent with those obtained using ordinary differential equation (ODE) methods, while offering a more natural examination of the fluctuations inherent to these complex biomedical systems.

One of the characteristics of SSA is the Markovian property, with the future state depending on the present state only and not on the past. Thus, the current state fully specifies the probability of a reaction occurring in the very next infinitesimal time interval [13,14]. This Markovian assumption is a simplified modeling of dynamical processes for efficient computation. However, such simplification also leads to its incapability of modeling processes where historical events influencing the timing of the future events. In biological systems, delays are widespread. For example, gene expression processes such as transcription and translation inherently do require time to complete [15–18].

To take into account those delay effects, several methods have been developed to extend the standard Gillespie algorithm by simulating delayed reactions [19–21]. These algorithms are effective in handling delays, though with additional complexity introduced and technically being more difficult to understand and apply, particularly for researchers without computational expertise [22]. On the other hand, currently available software packages mainly focus on simulate SSA reactions without delay effects [23,24] and very few efforts have been devoted to implementing delayed simulations [25]. In order to provide a user-friendly and open-source implementation of available SSA algorithms accounting for delay effects, we sought out to develop the software package DelaySSA in three data-science languages popular in bioinformatics (R, Python and Matlab). Researchers familiar with any of these three languages could now perform SSA simulations conveniently. DelaySSA and tutorials with comprehensive examples and applications are available at <https://github.com/Zoey-JIN/DelaySSA> and <https://github.com/FangZY-Lab/DelaySSA-dev>.

Design and implementation

In the stochastic simulation of biochemical reaction systems, Stochastic Simulation Algorithm (SSA)[1] is an important method that can describe the dynamic process of chemical reactions in the system with high accuracy. A detailed account of the theoretical background and algorithms could be found in S1 Appendix. Here, we implemented these algorithms with a common interface in the three most popular languages for systems biology, namely, R, Python, and Matlab (Fig A in S1 Appendix). Below we will illustrate the design considerations (Fig 1).

SSA can realistically reproduce the dynamic changes of species in the reaction system, including the intermediate processes of the system state. The core of the algorithm depends on the initial state of the system, including the types of species involved in the reaction and their initial quantities, as well as the reactions in the system. To prepare for simulation using the SSA, each reaction in the system must be described in detail, including reactants, products, and the rates of each reaction.

For non-Markovian processes with delays, we need to pay special attention to the characteristics of delayed reactions. The key to delayed reactions lies in two aspects, the changes in the quantities of reactants and products and the time points at which these changes occur.

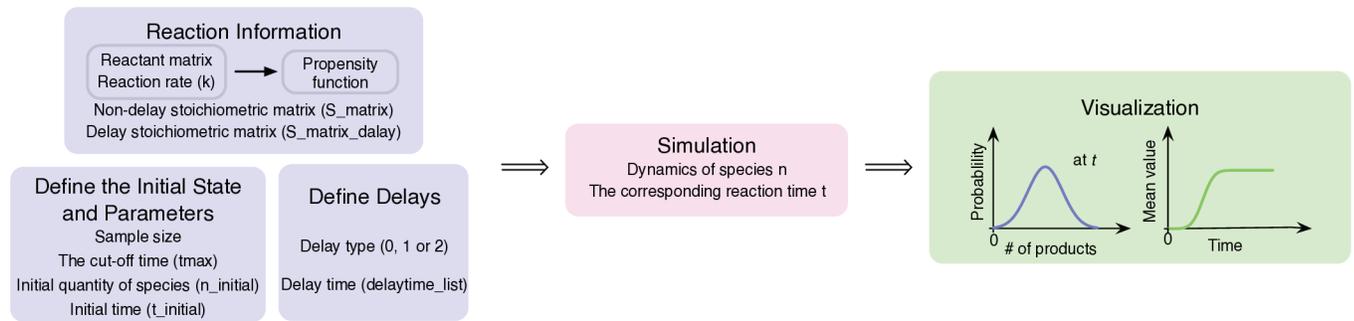


Fig 1. Design and implementation of DelaySSA. The input data for DelaySSA are reaction information, initial states, parameters and delays. Combining these into the simulation function, the temporal molecular states n at specific time t will be generated for further visualization.

<https://doi.org/10.1371/journal.pcbi.1012919.g001>

According to these criteria, reactions with delays can be divided into two categories: one with immediate changes in reactants at the start of the delay, and the other with latent changes in reactants at the end of the delay [21] (see S1 Appendix and Fig B in S1 Appendix for more comparison). For a reaction of the first type, the amount of reactants changes immediately at the beginning of the delay and generates products at the end of the delay when the reactions finish. Thus the amount of species is adjusted twice, both before and after the delay time τ . For a reaction of the other type, the amount of reactants does not change at the beginning of the delay when the reaction initiates, and it is not updated until the end of the delay. Thus the amount of reactants and of products are both updated at the end of the delay time τ . These two types of delayed reactions bring fundamental differences and diverge from the Markovian process which assumes an immediate change in response without any time delay.

In our implementation, reactant matrix represents the quantity of reactants (row of species and column of reactions). Reaction rate is the speed of chemical reaction. Two stoichiometric matrices are used to describe the non-delayed (S_{matrix}) and delayed ($S_{\text{matrix_delay}}$) reactions respectively. S_{matrix} represents the changes in molecular numbers for non-delayed reactions at the specific moment of reaction occurring. Each row corresponds to a species, while each column corresponds to a reaction. Each element of the matrix indicates the net changes of each species during each reaction. The quantity of species, the reaction rates, and the random numbers generated in the algorithm together specify when and which reaction is carried out. $S_{\text{matrix_delay}}$ is used to describe changes in material quantities for reactions that involve delays. Similar to the S_{matrix} , the rows represent species, the columns represent reactions, and the elements indicate the net change in each species after the delay time τ has elapsed.

The delay time τ plays a critical role in the dynamics of delayed reactions. It can be either a fixed value or generated by a function, enabling more complex and realistic modeling of delays based on reactions.

The propensity function f_r is fundamental for modeling reaction dynamics. We use the form of mass-action kinetics type [5] and $f_r(n)dt$ represents the probability of each reaction r occurring in the time interval dt , by taking into account the stochastic nature of the reaction process and the trend of the quantity of species. In combination with the random number generator, it determines which reaction to occur at each next moment and advances the simulation in discrete time steps. The system state is updated step by step iteratively until reaching a specific termination condition, either a maximum simulation time or a stable system state.

It is worth noting that each simulation results in one sequence or sample of the stochastic system. It is typically necessary to perform multiple repeated simulations to determine the average behavior of the system.

During the simulation, non-delayed reactions can influence the state of delayed reactions. For instance, if a non-delayed reaction causes a change in the quantity of a species that conflicts with the expected change from an ongoing delayed reaction, which must be resolved. A common method to handle such conflicts is to randomly remove one of the conflicting events from the delayed reactions. This approach maintains the consistency of the system state by preventing disruptions to the overall system dynamics caused by inconsistent species quantities.

Results

Evaluation on the Bursty and refractory models

To evaluate the validity of DelaySSA, we first applied it to two classical time-delayed models, the Bursty model and the refractory model.

The Bursty gene expression model describes the transcription bursts of mRNA [18,26] (Fig 2A). During short periods, multiple mRNA molecules are synthesized rapidly and then degraded after the time delay τ . Simulation results were consistent with the transcription rate as a function of burst frequency and burst size, and highlighted the phenomena that gene expression occurs in sporadic transcriptional bursts.

In the Refractory model [18,27,28], the gene state can switch between G_0 , G_1 and G_2 , but gene expression only occurs at G_2 (Fig 2B). The degradation of mRNA occurs after time delay τ . Simulation showed that mRNA expression has a bimodal distribution with a high probability of being zero, corresponding to refractory periods of remaining silence before switched to gene transcription, consistent with previous studies.

Evaluation on the RNA velocity model

RNA Velocity is a widely used model in computational methods inferring cell trajectories based on single-cell RNA-sequencing (scRNA-seq) data [7]. In this model, the unspliced mRNA (U) is initially produced at a rate denoted by α . The unspliced mRNA undergoes splicing at a rate β to generate mature spliced mRNA (S). After splicing, the mature mRNA undergo degradation at a rate γ . Simulation shows that before α becomes 0, the average values of U and S continue increasing until the steady state. In the down-regulation phase, α becomes 0, and the average value of U and S gradually decrease until returning to 0. The relationship between U and S can be seen from the U - S phase portrait (Fig 3A). This characteristic phase portrait is guaranteed by mRNA processing mechanism and have been widely accepted in single-cell data analysis to infer RNA-based cell differentiation. Since the gene transcriptional process (e.g. mRNA degradation) is inherent to time delays, and it still adheres to the maturation mechanism of mRNA, it would be reasonable to expect a similar U - S phase portrait after the introduction of time-delayed degradation. Indeed, when simulated with a time delay $\tau = 3$ in the mRNA degradation step, we again observed a similar U - S phase portrait (Fig 3B). Thus this simulation provides a further verification for DelaySSA.

Application to concentration-based gene regulatory network models

We further consider applying DelaySSA to gene regulatory network models. One thing to note here is that, a number of such networks are described in concentration-based mathematical models, whereas SSA simulations are molecule-based. To better approximate

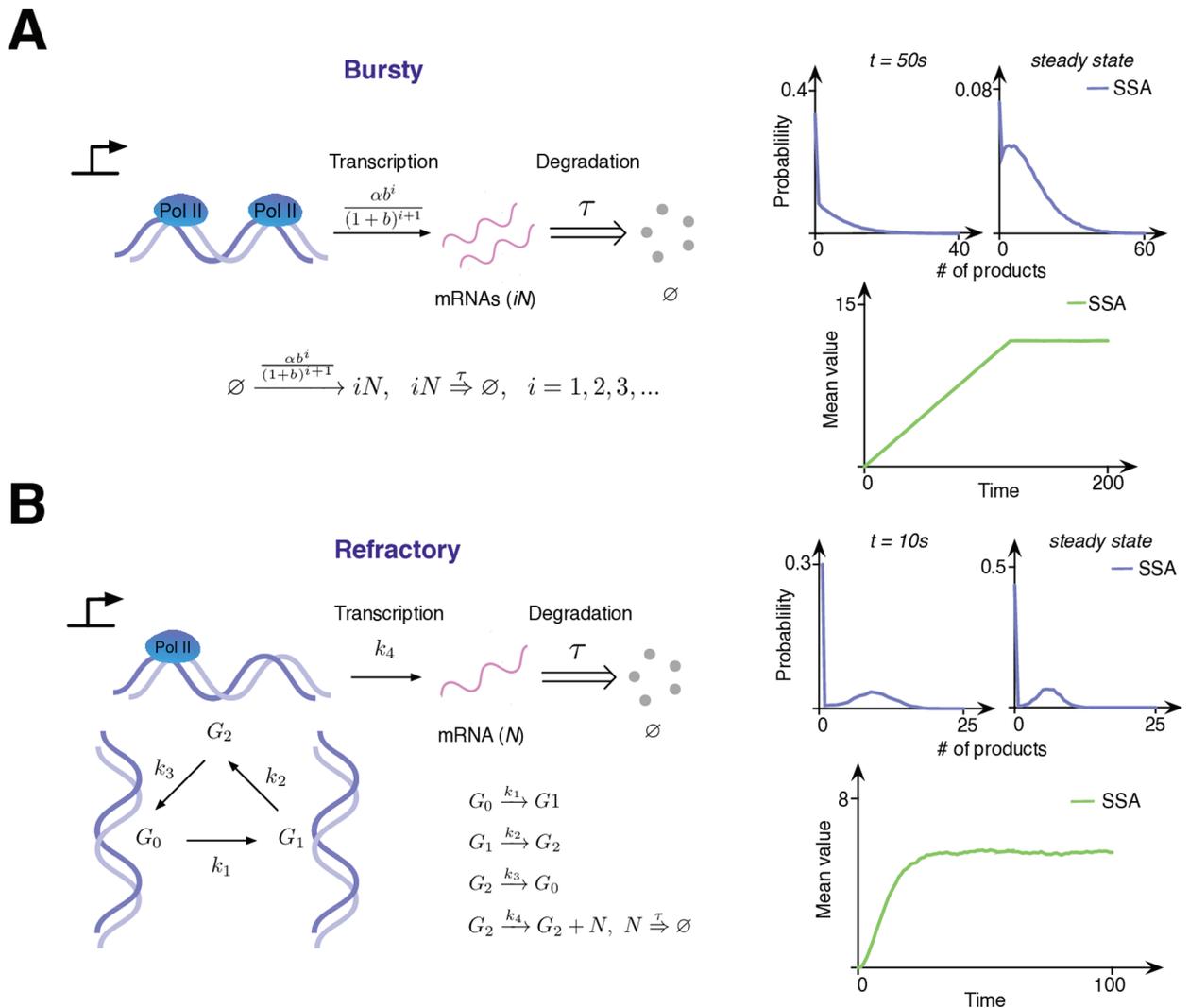


Fig 2. Bursty model and Refractory model simulation results. (A) Bursty model simulation results (10^5 samples). The production of nascent mRNA occurs in bursts and degrade after delay time τ . The burst frequency $\alpha = 0.0282$, the mean burst size $b = 3.46$ and $\tau = 120$. (B) Refractory model simulation results (10^4 samples). The gene has three states, G_0 , G_1 and G_2 , and can switch from one state to another. Moreover, nascent mRNA is produced only when the gene is in state G_2 . The parameters are $k_1 = 0.15$, $k_2 = 0.1$, $k_3 = 0.05$, $k_4 = 10$ and $\tau = 1$.

<https://doi.org/10.1371/journal.pcbi.1012919.g002>

concentration-based gene regulatory models, it would be useful to first perform a conversion from concentration-based models to molecule-based models. Here we demonstrate such a conversion in a lung cancer transition gene network model.

The network underlying lung cancer phenotypic transdifferentiation has been proposed and validated in recent studies of AST, which are closely related to drug-resistance during targeted therapies [10–12] (Fig 4A). The AST network has four core transcription factors: FOXA2 (F), NKX2-1 (N), P63 (P) and SOX2 (S). Among them, FOXA2 and NKX2-1 are in the adenocarcinoma lineage, whereas P63 and SOX2 are in the squamous carcinoma lineage.

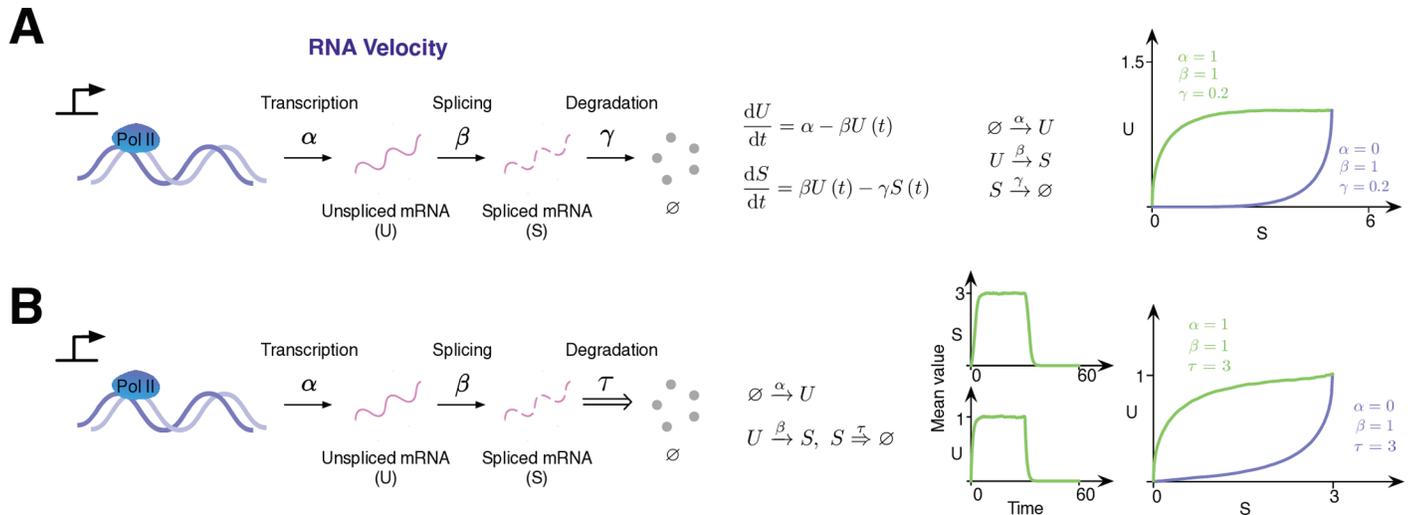


Fig 3. RNA Velocity model simulation results. (A) Simulation results for RNA Velocity model without delay (2×10^4 samples). The simulation is composed of an up-regulation phase (both production and degradation) and a down-regulation phase (degradation-only). In the first phase, $\alpha = 1, \beta = 1$ and $\gamma = 0.2$ for the reaction time $t = 50$. In the second phase, $\alpha = 0, \beta = 1$ and $\gamma = 0.2$. (B) Simulation results for RNA Velocity model with delay in degradation (2×10^4 samples). RNA degradation is specified as a delayed reaction. In the first (up-regulation) phase: $\alpha = 1, \beta = 1$ and $\tau = 3$. In the second (down-regulation) phase: $\alpha = 0, \beta = 1$ and $\tau = 3$.

<https://doi.org/10.1371/journal.pcbi.1012919.g003>

The set of ODEs describing this network is as following [11]:

$$\begin{cases} \frac{dF}{dt} = \frac{\alpha_F}{1 + a_1 P^2 + a_2 S^2} - d_F F \\ \frac{dN}{dt} = \alpha_N \frac{1}{1 + a_3 F^2 + a_4 S^2} - d_N N \\ \frac{dP}{dt} = \alpha_P \frac{1}{1 + a_5 S^2 + a_6 N^2} - d_P P \\ \frac{dS}{dt} = \beta_S \frac{S^2}{a_7^2 + S^2} + \frac{\alpha_S}{1 + a_8 F^2 + a_9 N^2} - d_S S \end{cases}$$

All these proteins are described in terms of concentrations [11], and thus it is a prerequisite to convert to a molecule-based model by rewriting these ODEs.

Given the concentration (x moles per liter, or mol/L) and the number of molecules (X), we have the relationship $X = xVN_A$, where $N_A \approx 6.023 \times 10^{23}$ is the Avogadro constant denoting the number of molecules in one mole, and V is the system volume. Whenever the system volume is not specifically modeled, we can define $B = VN_A$ and simplify the relationship as $X = Bx$. In the following simulations, we used $B = 6023$ that corresponds to a system volume of $V = 10^{-20}$ liters [29].

Taking the first equation of the above AST network model as an example, we will show how to deduce its molecule-based version below. In this equation, FOXA2 is inhibited by P63 and SOX2 (the first term on the right-hand side) and undergoes spontaneous degradation (the second term),

$$\frac{dF}{dt} = \frac{\alpha_F}{1 + a_1 P^2 + a_2 S^2} - d_F F \tag{1}$$

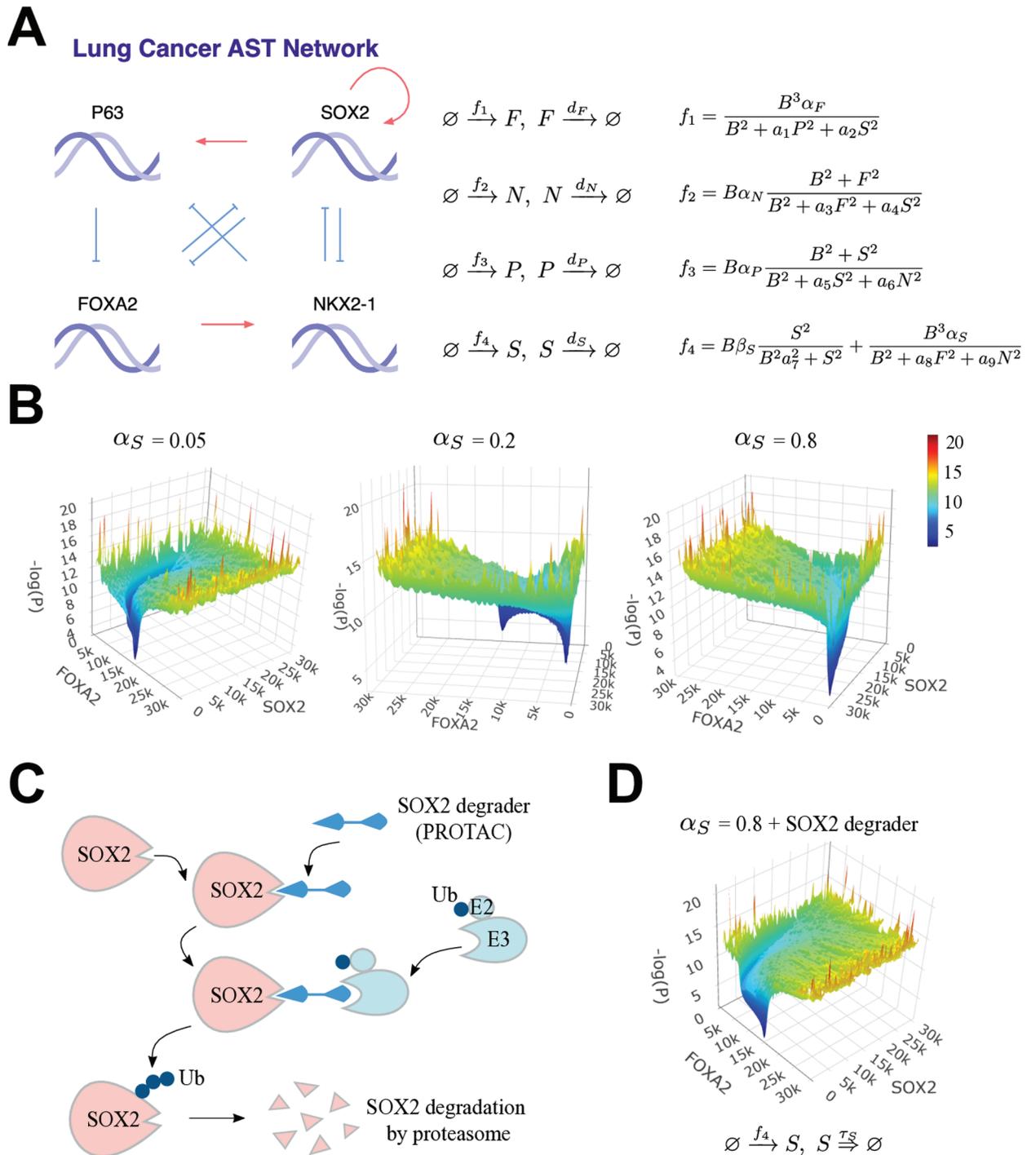


Fig 4. Modeling lung cancer AST network and its perturbation with a SOX2 degrader. (A) Lung Cancer AST Network and the corresponding stochastic equations. Parameters are $\alpha_N = \alpha_P = 1$, $d_F = d_N = d_P = 1$, $\alpha_1 = \alpha_2 = \alpha_6 = \alpha_8 = \alpha_9 = 1$, $\alpha_F = 2.9$, $\beta_S = 1.3$, $\alpha_3 = 0.8$, $\alpha_4 = \alpha_7 = 2$, $\alpha_5 = 0.8$, $d_S = 0.4$ with different α_S . The control parameter is α_S . (B) The potential landscape is approximated by simulating a large number of molecules corresponding to concentrations (3×10^3 samples) with parameter values $\alpha_S = 0.05$, $\alpha_S = 0.2$, and $\alpha_S = 0.8$, respectively representing the adenocarcinoma state, the intermediate state, and the squamous carcinoma state. FOXA2 and SOX2 amounts are chosen to visualize the landscape. (C) Principle of SOX2-targeted degradation by the PROTAC technique. The PROTAC molecule here is designed to recognize SOX2 proteins and to recruit E3 ligases to induce an addition of ubiquitin (Ub) proteins onto SOX2, which causes degradation of SOX2 subsequently. (D) Application of a SOX2 degrader to the squamous carcinoma state ($\alpha_S = 0.8$) reprogram the AST network back to the adenocarcinoma state. Here the treatment effect of the SOX2 degrader is modeled as a delayed reaction ($\tau_S = 1$).

<https://doi.org/10.1371/journal.pcbi.1012919.g004>

Multiplying B to both sides of Eq. 1, we have

$$\frac{dF}{dt} B = B \frac{\alpha_F}{1 + a_1 P^2 + a_2 S^2} - B d_F F,$$

$$\frac{dF}{dt} B = B \frac{B^2 \alpha_F}{B^2 (1 + a_1 P^2 + a_2 S^2)} - B d_F F,$$

$$\frac{d(FB)}{dt} = \frac{B^3 \alpha_F}{B^2 + a_1 (PB)^2 + a_2 (SB)^2} - d_F (FB).$$

Here, FB , PB , and SB respectively represent the numbers of molecules corresponding to the concentrations F , P , and S . Thus, scaling the four molecules by a factor of B , we obtain the following molecule-based version of equation,

$$\frac{dF}{dt} = \frac{B^3 \alpha_F}{B^2 + a_1 P^2 + a_2 S^2} - d_F F. \tag{2}$$

Similarly, We can obtain the other three equations. Together, the four molecule-based version of equations of the system are as following:

$$\begin{cases} \frac{dF}{dt} = \frac{B^3 \alpha_F}{B^2 + a_1 P^2 + a_2 S^2} - d_F F \\ \frac{dN}{dt} = B \alpha_N \frac{B^2 + F^2}{B^2 + F^2} - d_N N \\ \frac{dP}{dt} = B \alpha_P \frac{B^2 + a_3 F^2 + a_4 S^2}{B^2 + S^2} - d_P P \\ \frac{dS}{dt} = B \beta_S \frac{S^2}{B^2 a_7^2 + S^2} + \frac{B^3 \alpha_S}{B^2 + a_8 F^2 + a_9 N^2} - d_S S \end{cases}$$

Modeling the lung cancer cell transition network with therapeutic interventions

We stochastically simulated the molecule-based equations to illustrate how the system state transits from the adenocarcinoma lineage to the squamous carcinoma lineage (Fig 4A). We sought out to portray the quasi-potential landscape of the network, which is also an evaluation of the capability of our package in simulating Waddington’s landscapes.

Stochastic simulation produces large numbers of molecules for each protein that better correspond to concentrations. We collected dynamical samples with random initiation and appropriate noise, estimated the frequency P in two-dimensional bins (FOXA2 and SOX2 as variables), and the quasi-potential U can be approximated through $U \approx -\ln P$ (Fig 4B). When α_S is low (0.05), the system converges to a single stable point of FOXA2-high and SOX2-low, which corresponds to the lung adenocarcinoma phenotype (Fig 4B). When α_S is at an intermediate level (0.2), the system has two stable points simultaneously, suggesting a mixture of lung adenocarcinoma and squamous carcinoma phenotypes (Fig 4B). When α_S is large (0.8), the system again converges to a single stable point but of FOXA2-low and SOX2-high, which corresponds to the lung squamous carcinoma phenotype (Fig 4B). Taken together, stochastic simulation of this system clearly demonstrates the bistability of two phenotypic states of lung cancer, as well as the qualitative transition from FOXA2-high to SOX2-high when α_S increases dramatically.

From the therapeutic perspective, blocking lung cancer AST might reprogram squamous-state cells back to the original adenocarcinoma state and re-sensitize them to anticancer drugs. We can see that SOX2 plays a critical role in the AST network, based on above simulations, making it a seemingly candidate target for blocking AST. The technical challenge, however, is that transcription factors (TFs) like SOX2 are often considered “undruggable” with conventional inhibitors [30]. Recent advances in the proteolysis-targeting chimera (PROTAC) technology might alleviate this issue with rationally designed TF degraders [31]. In principle, SOX2-targeting PROTACs could be developed to degrade SOX2 (Fig 4C). Although research in this direction is still in its early stage [30] and it remains largely unclear whether SOX2 PROTACs would effectively block AST, we could test this rationale through model-based stochastic simulations. To model the effect of SOX2 degraders, we represent it as a delayed SOX2 protein degradation reaction (Fig 4D). Interestingly, whereas the system was stabilized in the squamous state given a large α_S (0.8), *in silico* addition of SOX2 degraders successfully produces a ‘reprogramming’ effect that pushes the system back to its original adenocarcinoma state (Fig 4D).

In conclusion, DelaySSA not only captures the potential landscapes underlying gene regulatory networks, but also proves to be highly valuable in simulating therapeutic interventions of network nodes.

Computational efficiency

We have taken advantages of the Rcpp package to accelerate the simulation processes. Rcpp allows the incorporation of fast C++ routines within the R framework. In addition, we also employed a dynamical memory allocation mechanism to reduce the need of frequent memory allocation at each iteration during simulation, which reveals significant improvements for prolonged sampling.

We first compared the three delay algorithms in Rcpp-accelerated DelaySSA on the classical Bursty and Refractory models. DelayDirect and DelayRejection are faster on the Bursty model, whereas DelayMNR is faster on the Refractory model (Fig C in S1 Appendix). We further evaluated the Genetic Toggle Switch model with or without delays (see S1 Appendix for model details). There are slight differences among the three non-delay algorithms, whereas all three delay algorithms perform similarly (Fig D in S1 Appendix).

To compare the Rcpp-accelerated version of algorithms with the non-accelerated pure R version, we benchmarked the running time by simulation on all the three above models. As expected, optimization with Rcpp has significantly improved the computational efficiency for all algorithms, almost 10 times faster than their corresponding unoptimized versions on the Refractory and Toggle Switch models (Fig E in S1 Appendix). In comparison with the respective algorithms developed in Julia, a programming language featured by efficiency, Rcpp-optimized version of algorithms show a closely comparable performance (Fig E in S1 Appendix), indicating a high efficiency of the current DelaySSA implementation.

Availability and future directions

DelaySSA (<https://github.com/Zoey-JIN/DelaySSA> and <https://github.com/FangZY-Lab/DelaySSA-dev>) currently implements a range of stochastic simulation algorithms with or without delays. For both simple and more complex biochemical gene networks, it already reaches a fast execution time. For extremely large-scale networks, further reducing the computation time remains a valuable direction in the future. Interesting explorations include integration of multi-scale approaches, hybrid deterministic-stochastic algorithms, and dimensionality-reduction methods.

In biological research, stochastic simulations offer several advantages compared to actual biological experiments. Experiments typically require significant expenditure whereas simulations primarily depend on computational resources. Biomedical experiments are often time-consuming whereas simulations can be completed in a shorter time. In certain cases, experimental techniques might not even yet be fully developed (such as in the above SOX2-targeting example). In all these scenarios, stochastic simulations offer a unique and important value in rationale development or evaluation.

Given the importance of SSA in simulation of biochemical and molecular networks, it is surprising that few user-friendly packages have been developed in popular programming languages such as R, Python and Matlab, especially those non-Markovian algorithms capable of handling time delays inherent in more realistic biological systems. This new implementation DelaySSA fills the current gap in three programming languages altogether, thus providing a powerful platform for enormous researchers from different backgrounds and for applications in diverse fields. By integrating both traditional SSA and delayed SSA, DelaySSA provides a unified and intuitive interface for a wide range of stochastic simulations, allowing users to describe and simulate reaction networks with minimal coding effort. In addition, DelaySSA does not require additional dependencies, simplifying the installation and setup process. A series of concrete tutorials are also accompanied in the Github websites of DelaySSA.

One of the significant benefits of DelaySSA is its ability to generate accurate simulation data. This feature is particularly valuable for researchers who may lack precise experimental data [32–35], enabling them to test and improve biological models, and to optimize or fine tune computational methods according to simulated data. As we have demonstrated in the lung cancer AST intervention example, although targeting SOX2 with PROTAC is practical in principle, in reality it is not yet fully established. Nonetheless, with DelaySSA we could simulate a therapeutic intervention model with the SOX2-targeting PROTAC modeled as a delayed degradation reaction. Our simulation data allow a portrait of the ‘reprogrammed’ system state landscape after SOX2-degrader intervention, theoretically predicting the efficacy of SOX2-based intervention for blocking lung cancer AST. This provides a new valuable clue for future research on blocking drug-resistant AST in lung cancer [12,36,37]. Following our application case, researchers could further perform valuable applications to reveal novel biomedical insights in various studies of systems biology.

Supporting information

S1 Appendix. Mathematical foundations and supplementary data. The S1 Appendix file contains Mathematical Foundations, Algorithmic Details, and Supplementary Data which includes five supplementary figures. (PDF).

Acknowledgments

We thank helpful support from Prof. Jingyi Lu’s team. We also thank Profs. Luonan Chen and Ruiqi Wang for helpful discussion.

Author contributions

Conceptualization: Ziyang Jin, Xinyi Zhou, Zhaoyuan Fang.

Formal analysis: Ziyang Jin, Xinyi Zhou.

Funding acquisition: Zhaoyuan Fang.

Investigation: Ziyang Jin, Xinyi Zhou.

Methodology: Ziyang Jin, Xinyi Zhou, Zhaoyuan Fang.

Project administration: Zhaoyuan Fang.

Resources: Zhaoyuan Fang.

Software: Ziyang Jin, Xinyi Zhou, Zhaoyuan Fang.

Supervision: Zhaoyuan Fang.

Validation: Ziyang Jin, Xinyi Zhou, Zhaoyuan Fang.

Visualization: Ziyang Jin, Xinyi Zhou, Zhaoyuan Fang.

Writing – original draft: Ziyang Jin, Xinyi Zhou, Zhaoyuan Fang.

Writing – review & editing: Ziyang Jin, Xinyi Zhou, Zhaoyuan Fang.

References

1. Gillespie DT. Exact stochastic simulation of coupled chemical reactions. *J Phys Chem.* 1977;81(25):2340–61. <https://doi.org/10.1021/j100540a008>
2. Bernstein D. Simulating mesoscopic reaction-diffusion systems using the Gillespie algorithm. *Phys Rev E.* 2005;71(4):041103. <http://dx.doi.org/10.1103/PhysRevE.71.041103>
3. Cao Z, Grima R. Linear mapping approximation of gene regulatory networks with stochastic dynamics. *Nat Commun.* 2018;9(1):3305. <https://doi.org/10.1038/s41467-018-05822-0> PMID: 30120244
4. Rao CV, Arkin AP. Stochastic chemical kinetics and the quasi-steady-state assumption: application to the Gillespie algorithm. *J Chem Phys.* 2003;118(11):4999–5010. <https://doi.org/10.1063/1.2971036> PMID: 19044893
5. Van Kampen NG. *Stochastic processes in physics and chemistry.* Elsevier; 1992.
6. Cao Z, Wang Y, Grima R. Universal deterministic patterns in stochastic count data. *arXiv preprint.* arXiv:2405.17530. 2024.
7. La Manno G, Soldatov R, Zeisel A, Braun E, Hochgerner H, Petukhov V, et al. RNA velocity of single cells. *Nature.* 2018;560(7719):494–8. <https://doi.org/10.1038/s41586-018-0414-6> PMID: 30089906
8. Li S, Zhang P, Chen W, Ye L, Brannan KW, Le NT, et al. A relay velocity model infers cell-dependent RNA velocity. *Nat Biotechnol.* 2024;42(1):99–108. <https://doi.org/10.1038/s41587-023-01728-5> PMID: 37012448
9. Wang K, Hou L, Wang X, Zhai X, Lu Z, Zi Z, et al. PhyloVelo enhances transcriptomic velocity field mapping using monotonically expressed genes. *Nat Biotechnol.* 2024;42(5):778–89. <https://doi.org/10.1038/s41587-023-01887-5> PMID: 37524958
10. Fang Z, Han X, Chen Y, Tong X, Xue Y, Yao S, et al. Oxidative stress-triggered Wnt signaling perturbation characterizes the tipping point of lung adeno-to-squamous transdifferentiation. *Signal Transduct Target Ther.* 2023;8(1):16. <https://doi.org/10.1038/s41392-022-01227-0> PMID: 36627278
11. Tang S, Xue Y, Qin Z, Fang Z, Sun Y, Yuan C, et al. Counteracting lineage-specific transcription factor network finely tunes lung adeno-to-squamous transdifferentiation through remodeling tumor immune microenvironment. *Natl Sci Rev.* 2023;10(4):nwad028. <https://doi.org/10.1093/nsr/nwad028> PMID: 37051524
12. Tong X, Patel AS, Kim E, Li H, Chen Y, Li S, et al. Adeno-to-squamous transition drives resistance to KRAS inhibition in LKB1 mutant lung cancer. *Cancer Cell.* 2024;42(3):413–28. <https://doi.org/10.1016/j.ccell.2024.01.012> PMID: 38402609
13. Gillespie DT. *Markov processes: an introduction for physical scientists.* Elsevier; 1991.
14. Peccoud J, Ycart B. Markovian modeling of gene-product synthesis. *Theor Popul Biol.* 1995;48(2):222–34. <https://doi.org/10.1006/tpbi.1995.1027>

15. Fu X, Patel HP, Coppola S, Xu L, Cao Z, Lenstra TL, et al. Quantifying how post-transcriptional noise and gene copy number variation bias transcriptional parameter inference from mRNA distributions. *Elife*. 2022;11:e82493. <https://doi.org/10.7554/elifesciences.82493> PMID: 36250630
16. Gibson MA, Bruck J. Efficient exact stochastic simulation of chemical systems with many species and many channels. *J Phys Chem A*. 2000;104(9):1876–89. <https://doi.org/10.1021/jp993732q>
17. Vilk O, Metzler R, Assaf M. Non-Markovian gene expression. *Phys Rev Res*. 2024;6(2):L022026.
18. Suter DM, Molina N, Gattfield D, Schneider K, Schibler U, Naef F. Mammalian genes are transcribed with widely different bursting kinetics. *Science*. 2011;332(6028):472–4. <https://doi.org/10.1126/science.1198817> PMID: 21415320
19. Anderson DF. A modified next reaction method for simulating chemical systems with time dependent propensities and delays. *J Chem Phys*. 2007;127(21):214107. <https://doi.org/10.1063/1.2799998> PMID: 18067349
20. Barrio M, Burrage K, Leier A, Tian T. Oscillatory regulation of Hes1: discrete stochastic delay modelling and simulation. *PLoS Comput Biol*. 2006;2(9):e117. <https://doi.org/10.1371/journal.pcbi.0020117> PMID: 16965175
21. Cai X. Exact stochastic simulation of coupled chemical reactions with delays. *J Chem Phys*. 2007;126(12). <https://doi.org/10.1063/1.2710253> PMID: 17411109
22. Slepoy A, Thompson AP, Plimpton SJ. A constant-time kinetic Monte Carlo algorithm for simulation of large biochemical reaction networks. *J Chem Phys*. 2008;128(20):205101. <https://doi.org/10.1063/1.2919546> PMID: 18513044
23. Abel JH, Drawert B, Hellander A, Petzold LR. GillesPy: a python package for stochastic model building and simulation. *IEEE life Sci Lett*. 2016;2(3):35–38.
24. Pineda-Krch M. GillespieSSA: implementing the Gillespie stochastic simulation algorithm in R. *J. Stat. Softw*. 2008;25:1–18.
25. Fu X, Zhou X, Gu D, Cao Z, Grima R. DelaySSAToolkit.jl: stochastic simulation of reaction systems with time delays in Julia. *Bioinformatics*. 2022;38(17):4243–5. <https://doi.org/10.1093/bioinformatics/btac472> PMID: 35799359
26. Jiang Q, Fu X, Yan S, Li R, Du W, Cao Z, et al. Neural network aided approximation and parameter inference of non-Markovian models of gene expression. *Nat Commun*. 2021;12(1):2618. <https://doi.org/10.1038/s41467-021-22919-1> PMID: 33976195
27. Cao Z, Filatova T, Oyarzún DA, Grima R. A stochastic model of gene expression with polymerase recruitment and pause release. *Biophys J*. 2020;119(5):1002–14. <https://doi.org/10.1016/j.bpj.2020.07.020> PMID: 32814062
28. Zhang J, Zhou T. Stationary moments, distribution conjugation and phenotypic regions in stochastic gene transcription. *Math Biosci Eng*. 2019;16(5):6134–6166. <https://doi.org/10.3934/mbe.2019307> PMID: 31499756
29. Zhang X, Chong KH, Zhu L, Zheng J. A Monte Carlo method for in silico modeling and visualization of Waddington's epigenetic landscape with intermediate details. *Biosystems*. 2020;198:104275. <https://doi.org/10.1016/j.biosystems.2020.104275> PMID: 33080349
30. Choi J, Lim S, Bai X, Song H, Oh Y, Yong Y, Lee S. Abstract 3256: Novel SOX2-targeting bioPROTACs for the treatment of cancers. *Cancer Res*. 2024;84:3256–6. <http://dx.doi.org/10.1158/1538-7445.AM2024-3256>
31. Békés M, Langley DR, Crews CM. PROTAC targeted protein degraders: the past is prologue. *Nat Rev Drug Discov*. 2022;21:181–200. <https://doi.org/10.1038/s41573-021-00371-6> PMID: 35042991
32. Bergen V, Lange M, Peidli S, Wolf FA, Theis FJ. Generalizing RNA velocity to transient cell states through dynamical modeling. *Nat Biotechnol*. 2020;38(12):1408–14. <https://doi.org/10.1038/s41587-020-0591-3> PMID: 32747759
33. Chen S, Mar JC. Evaluating methods of inferring gene regulatory networks highlights their lack of performance for single cell gene expression data. *BMC Bioinformatics*. 2018;19:1–21. <https://doi.org/10.1186/s12859-018-2217-z> PMID: 29914350
34. Maturana MI, Meisel C, Dell K, Karoly PJ, D'Souza W, Grayden DB, et al. Critical slowing down as a biomarker for seizure susceptibility. *Nat Commun*. 2020;11(1):2172. <https://doi.org/10.1038/s41467-020-15908-3> PMID: 32358560
35. Shu J, Wu C, Wu Y, Li Z, Shao S, Zhao W, et al. Induction of pluripotency in mouse somatic cells with lineage specifiers. *Cell*. 2013;153(5):963–75. <https://doi.org/10.1016/j.cell.2015.05.020> PMID: 28910643

36. Leonetti A, Verzè M, Minari R, Perrone F, Gnetti L, Bordi P, et al. Resistance to osimertinib in advanced EGFR-mutated NSCLC: a prospective study of molecular genotyping on tissue and liquid biopsies. *Br J Cancer*. 2024;130(1):135–42. <https://doi.org/10.1038/s41416-023-02475-9> PMID: [37938348](https://pubmed.ncbi.nlm.nih.gov/37938348/)
37. Quintanal-Villalonga A, Taniguchi H, Zhan YA, Hasan MM, Chavan SS, Meng F, et al. Comprehensive molecular characterization of lung tumors implicates AKT and MYC signaling in adenocarcinoma to squamous cell transdifferentiation. *J Hematol Oncol*. 2021;14(1):170. <https://doi.org/10.1186/s13045-021-01186-z> PMID: [34656143](https://pubmed.ncbi.nlm.nih.gov/34656143/)