Learning to Fight Against Cell Stimuli: A Game Theoretic Perspective

Seyed Hamid Hosseini, and Mahdi Imani

2023 IEEE Conference on Artificial Intelligence (CAI) | 979-8-3503-3984-0/23/$31.00 ©2023 IEEE | DOI: 10.1109/CAI54212.2023.00127

***Abstract*— Current genomics interventions have limitations in accounting for cell stimuli and the dynamic response to intervention. Although genomic sequencing and analysis have led to significant advances in personalized medicine, the complexity of cellular interactions and the dynamic nature of the cellular response to stimuli pose significant challenges. These limitations can lead to chronic disease recurrence and inefficient genomic interventions. Therefore, it is necessary to capture the full range of cellular responses to develop effective interventions. This paper presents a game-theoretic model of the fight between the cell and intervention, demonstrating analytically and numerically why current interventions become ineffective over time. The performance is analyzed using melanoma regulatory networks, and the role of artificial intelligence in deriving effective solutions is described.**

I. Introduction

Advancements in genomics have enabled significant breakthroughs in the field of personalized medicine [1]–[3]. However, there are limitations in the current genomic interventions that fail to account for the dynamic cellular responses to intervention and stimuli [4]–[11]. The complexities of cellular interactions and responses often lead to a recurrence of chronic diseases upon genomic interventions. These limitations require effective therapeutic policies that can learn, understand, and effectively react to the full range of cellular responses.

Unhealthy or cancerous cells constantly fight against interventions/therapies. The cell aims to keep the system in a cancerous condition and perceives its actions as essential in keeping the cell alive. Modeling the cell's response is key to deriving effective interventions or designing drugs to fight against diseases. Most existing interventions assume that the cell is non-responsive and design the intervention under this naive assumption. This often leads to the short-term success of these therapies before cells find new ways to fight against them and partially or fully return the system to undesirable conditions.

This paper models the fight between the intervention and the cell as a two-player zero-sum game and highlights the potential of this model, combined with artificial intelligence, in deriving effective and adaptable genomics interventions. We demonstrate the reasons for the success of existing interventions by representing the cell as an adaptable reinforcement learning (RL) player in the cell-intervention fight. Our model and results provide insights and mathematical justifications for the early-stage success and long-term failure of existing interventions and provide the valuable potential for learning-based and game-theoretic approaches for treating chronic diseases and designing new drugs.

S. H. Hosseini and M. Imani are with the Department of Electrical and Computer Engineering at Northeastern University. Emails: hosseini.ha@northeastern.edu, m.imani@northeastern.edu

II. Proposed Game-Theoretic Model

This paper employs a general form of Boolean network with perturbation (BNp) model [12]–[17] for capturing the dynamics of gene regulatory networks. This model properly captures the stochasticity in GRNs, coming from intrinsic uncertainty or unmodeled parts of systems. The battle between the cell and interventionist is modeled in this paper as a two-player zero-sum game [18]–[21]. This can be characterized by a tuple $\langle \mathcal{X}, \mathcal{A}, \mathcal{U}, R^a, T \rangle$, where $\mathcal{X} = \{0,1\}^d$ is the *state space*, d is the number of genes, $\mathcal{A}$ is the *action (e.g., intervention) space*, $\mathcal{U}$ is the *internal cell control (i.e., internal stimuli) space*, $T : \mathcal{X} \times \mathcal{A} \times \mathcal{U} \times \mathcal{X}$ is the *state transition probability function* such that $p(\mathbf{x}' \mid \mathbf{x}, \mathbf{a}, \mathbf{u})$ represents the probability of moving to state $\mathbf{x}'$ according to the external and internal inputs $\mathbf{a}$ and $\mathbf{u}$ in state $\mathbf{x}$. $R^a : \mathcal{X} \times \mathcal{A} \times \mathcal{U} \times \mathcal{X}$ denotes the reward functions for an interventionist, where $R^a(\mathbf{x}, \mathbf{a}, \mathbf{u}, \mathbf{x}')$ denotes the immediate shift from the cancerous states (i.e., reduction in cell proliferation) if the system moves from state $\mathbf{x}$ to state $\mathbf{x}'$ after the intervention $\mathbf{a}$ and the internal cell response $\mathbf{u}$. The cell aims to increase cell proliferation in cancerous cells, while the interventionist aims to reduce cell proliferation. Thus, the the reward for the cell R^u takes negative of the interventionist reward function, i.e., $R^u(\mathbf{x}, \mathbf{a}, \mathbf{u}, \mathbf{x}') = -R^a(\mathbf{x}, \mathbf{a}, \mathbf{u}, \mathbf{x}')$.

Let $\pi^a : \mathcal{X} \to \mathcal{A}$ denote the intervention strategy, prescribing an intervention to any given state $\mathbf{x} \in \mathcal{X}$. Let also $\pi^u : \mathcal{X} \to \mathcal{U}$ be the cell policy, representing cell actions/simuli at all system states. For the system with no intervention (i.e., $\mathcal{A} = \{\mathbf{0}\}$), the cell can aggressively push the system to the undesirable states. The cell policy in this case can be expressed as:

$$\pi^u_{\text{No}}(\mathbf{x}) = \underset{\pi^u \in \Pi}{\operatorname{argmin}}\, \mathbb{E}\Bigg[\sum_{t \geq 0} \gamma^t R^a(\mathbf{x}_t, \mathbf{a}_t = \mathbf{0}, \mathbf{u}_t, \mathbf{x}_{t+1}) \mid \mathbf{a}_{0:\infty} = \mathbf{0}, \mathbf{u}_{0:\infty} \sim \pi^u, \mathbf{x}_0 = \mathbf{x} \Bigg], \text{ for } \mathbf{x} \in \mathcal{X}, \tag{1}$$

where $0 < \gamma < 1$ is a discount factor that indicates the importance of early-stage rewards compared to future ones. This models the behavior of cell in chronic diseases such as cancer, leading to uncontrolled proliferation of cancer cells and often death.

Most existing intervention strategies are deterministic, meaning they assume the cell has no defense mechanism against the interventions [5], [22]. Under this simplistic

979-8-3503-3984-0/23/$31.00 ©2023 IEEE
DOI 10.1109/CAI54212.2023.00127

assumption, the Markov game is equivalent of the Markov decision process with a single agent/player, where the interventionist deals with an unresponsive cell, i.e., $\mathcal{U} = \{\mathbf{0}\}$. The conventional intervention policy can be obtained as a solution of the following optimization process:

$$\pi^a_{\text{naive}}(\mathbf{x}) = \underset{\pi^a \in \Pi}{\operatorname{argmax}} \mathbb{E}\Bigg[\sum_{t\geq 0} \gamma^t R^a(\mathbf{x}_t, \mathbf{a}_t, \mathbf{u}_t = \mathbf{0}, \mathbf{x}_{t+1}) \mid \mathbf{a}_{0:\infty} \sim \pi^a, \mathbf{u}_{0:\infty} = \mathbf{0}, \mathbf{x}_0 = \mathbf{x} \Bigg], \text{ for } \mathbf{x} \in \mathcal{X}. \tag{2}$$

The naive intervention policy often has a positive short-term impact on the cell, but these impacts fade out as the cell understands the impact of therapies and finds new ways to fight back against intervention. Given our model of the fight between the cell and intervention, the best defense policy for the cell against the naive intervention policy π^a_{naive} in (2) can be formulated as:

$$\pi^u(\mathbf{x}) = \underset{\pi^u \in \Pi}{\operatorname{argmin}} \mathbb{E}\Bigg[\sum_{t\geq 0} \gamma^t R^a(\mathbf{x}_t, \mathbf{a}_t = \pi^a_{\text{naive}}(\mathbf{x}_t), \mathbf{u}_t, \mathbf{x}_{t+1}) \mid \mathbf{a}_{0:\infty} \sim \pi^a_{\text{naive}}, \mathbf{u}_{0:\infty} \sim \pi^u, \mathbf{x}_0 = \mathbf{x} \Bigg], \text{ for } \mathbf{x} \in \mathcal{X}. \tag{3}$$

The cell figures out the intervention policy over time and through dynamic stimuli, it recurs the disease. The solutions to (1)-(3) can be obtained using dynamic programming or reinforcement learning, depending on the size of the regulatory network and action spaces [21], [23].

The analytical performance analysis can be achieved in terms of the state value function and the steady state probability. In particular, the difference in the expected discounted rewards under the naive intervention and no intervention can be expressed for any $\mathbf{x} \in \mathcal{X}$ as:

$$e(\mathbf{x}) = V_{\pi^a_{\text{naive}}, \pi^u}(\mathbf{x}) - V_{\mathbf{a}=\mathbf{0}, \pi^u_{\text{No}}}(\mathbf{x}). \tag{4}$$

In this case, $e(\mathbf{x}) \approx 0$ for all $\mathbf{x} \in \mathcal{X}$ means that the cell response ultimately has returned the system close to the original space through new internal stimuli (e.g., increase cancer cell proliferation).

III. Numerical Experiments

The performance of the proposed model is investigated using the melanoma regulatory network [5], [24]. This network is essential for studying and understanding the molecular mechanisms that drive the development and progression of melanoma, a deadly form of skin cancer. The regulatory relationships for this network are presented in Fig. 1, where the system consists of 10 genes. This regulatory network contains 10 genes, which lead to $2^{10} = 1,024$ possible states. The activation of WNT5A has been directly connected to the metastatic condition and reducing the activation of this gene using antibodies has been shown to be highly effective in preventing melanoma from metastasizing and achieving a desirable outcome [5]. Thus, we consider $R^a(\mathbf{x}, \mathbf{a}, \mathbf{u}, \mathbf{x}') = -5\mathbf{x}'(1)$ as an intervention reward function, where the reward of -5 is assigned for each activation of WNT5A. The internal

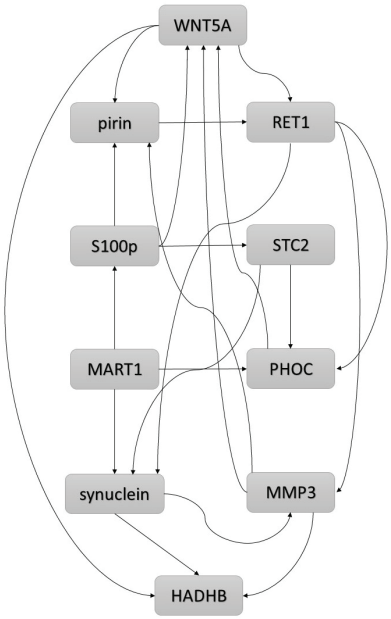

Fig. 1: The pathway diagram for the melanoma regulatory network.

cell stimuli space is $\mathcal{U} = \{\mathbf{u}^1, \mathbf{u}^2, \mathbf{u}^3\}$, where $\mathbf{u}^1$ corresponds to no stimuli, and $\mathbf{u}^2$ and $\mathbf{u}^3$ represent stimuli over the S100P and MMP3 genes, respectively. For the intervention space, we consider $\mathcal{A} = \{\mathbf{a}^1, \mathbf{a}^2\}$, where $\mathbf{a}^1$ represents no control and $\mathbf{a}^2$ represents intervention over the PHOC gene.

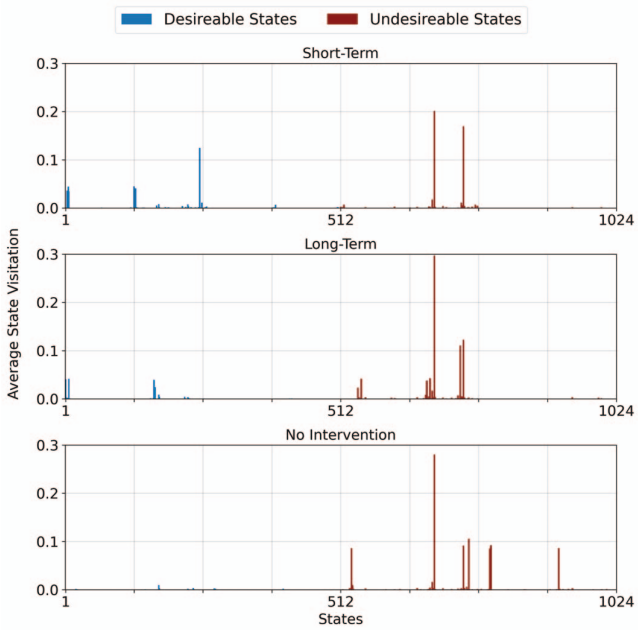

Fig. 2: The short-term and long-term desirable and undesirable state visitations under intervention and under no intervention.

Fig. 2 represents the early-stage and long-term state visitations of desirable and undesirable states, along with

the state visitation under the no-intervention case. One can see a significant increase in the short-term visitation of desirable states under the intervention policy compared to no intervention. However, the shift toward desirable states has faded out in the long term, and the condition has recurred to the case with no intervention. Our future work will study the use of reinforcement learning to understand the cell stimuli and policy and adaptively fight against cells as more gene-expression data becomes available.

IV. Conclusion

This paper introduces a game-theoretic model for deriving genomic interventions that can capture the full range of cellular responses. Current genomic interventions have limitations in accounting for the dynamic responses of cells, which pose challenges that can lead to chronic disease recurrence and inefficient interventions. This paper models the fight between the intervention and the cell as a two-player zero-sum game and highlights the potential of this model, combined with artificial intelligence, in deriving genomic interventions with long-term effectiveness.

Acknowledgment

The authors acknowledge the support of the National Institute of Health award 1R21EB032480-01, National Science Foundation award IIS-2202395, ARMY Research Office award W911NF2110299, and Oracle for Research program.

References

[1] Q. Liu, Y. He, and J. Wang, "Optimal control for probabilistic Boolean networks using discrete-time markov decision processes," *Physica A: Statistical Mechanics and its Applications*, vol. 503, pp. 1297–1307, 2018.

[2] N. S. Taou, D. W. Corne, and M. A. Lones, "Investigating the use of Boolean networks for the control of gene regulatory networks," *Journal of computational science*, vol. 26, pp. 147–156, 2018.

[3] G. Papagiannis and S. Moschoyiannis, "Deep reinforcement learning for control of probabilistic Boolean networks," *arXiv preprint arXiv:1909.03331*, 2019.

[4] I. Shmulevich and E. R. Dougherty, *Probabilistic Boolean networks: the modeling and control of gene regulatory networks*. SIAM, 2010.

[5] X. Qian and E. R. Dougherty, "Intervention in gene regulatory networks via phenotypically constrained control policies based on long-run behavior," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 9, no. 1, pp. 123–136, 2011.

[6] M. Imani and U. M. Braga-Neto, "Finite-horizon LQR controller for partially-observed Boolean dynamical systems," *Automatica*, vol. 95, pp. 172–179, 2018.

[7] M. Imani and U. M. Braga-Neto, "Point-based methodology to monitor and control gene regulatory networks via noisy measurements," *IEEE Transactions on Control Systems Technology*, 2018.

[8] M. Imani and U. M. Braga-Neto, "Control of gene regulatory networks with noisy measurements and uncertain inputs," *IEEE Transactions on Control of Network Systems*, vol. 5, no. 2, pp. 760–769, 2018.

[9] M. Imani and U. Braga-Neto, "Optimal control of gene regulatory networks with unknown cost function," in *Proceedings of the 2018 American Control Conference (ACC)*, IEEE, 2018.

[10] M. Imani and U. Braga-Neto, "Control of gene regulatory networks using Bayesian inverse reinforcement learning," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 16, no. 4, pp. 1250–1261, 2019.

[11] M. Imani, R. Dehghannasiri, U. M. Braga-Neto, and E. R. Dougherty, "Sequential experimental design for optimal structural intervention in gene regulatory networks based on the mean objective cost of uncertainty," *Cancer informatics*, vol. 17, p. 1176935118790247, 2018.

[12] L. E. Chai, S. K. Loh, S. T. Low, M. S. Mohamad, S. Deris, and Z. Zakaria, "A review on the computational approaches for gene regulatory network construction," *Computers in biology and medicine*, vol. 48, pp. 55–65, 2014.

[13] M. Alali and M. Imani, "Reinforcement learning data-acquiring for causal inference of regulatory networks," in *American Control Conference (ACC)*, IEEE, 2023.

[14] M. Alali and M. Imani, "Inference of regulatory networks through temporally sparse data," *Frontiers in control engineering*, vol. 3, 2022.

[15] A. Ravari, S. F. Ghoreishi, and M. Imani, "Optimal recursive expert-enabled inference in regulatory networks," *IEEE Control Systems Letters*, vol. 7, pp. 1027–1032, 2022.

[16] M. Imani and U. Braga-Neto, "Gene regulatory network state estimation from arbitrary correlated measurements," *EURASIP Journal on Advances in Signal Processing*, vol. 2018, no. 1, pp. 1–10, 2018.

[17] L. D. McClenny, M. Imani, and U. Braga-Neto, *BoolFilter package vignette*, 2017.

[18] L. S. Shapley, "Stochastic games," *Proceedings of the national academy of sciences*, vol. 39, no. 10, pp. 1095–1100, 1953.

[19] K. Zhang, Z. Yang, and T. Bacsar, "Multi-agent reinforcement learning: A selective overview of theories and algorithms," *Handbook of reinforcement learning and control*, pp. 321–384, 2021.

[20] K. Zhang, S. Kakade, T. Basar, and L. Yang, "Model-based multi-agent RL in zero-sum Markov games with near-optimal sample complexity," *Advances in Neural Information Processing Systems*, vol. 33, pp. 1166–1178, 2020.

[21] K. Zhang, Z. Yang, and T. Basar, "Policy optimization provably converges to nash equilibria in zero-sum linear quadratic games," *Advances in Neural Information Processing Systems*, vol. 32, 2019.

[22] R. Pal, A. Datta, and E. R. Dougherty, "Optimal infinite-horizon control for probabilistic Boolean networks," *IEEE Transactions on Signal Processing*, vol. 54, no. 6, pp. 2375–2387, 2006.

[23] S. Gronauer and K. Diepold, "Multi-agent deep reinforcement learning: a survey," *Artificial Intelligence Review*, pp. 1–49, 2022.

[24] M. Imani and S. F. Ghoreishi, "Optimal finite-horizon perturbation policy for inference of gene regulatory networks," *IEEE Intelligent Systems*, vol. 36, no. 1, pp. 54–63, 2020.