Integrating inverse reinforcement learning into data-driven mechanistic computational models: a novel paradigm to decode cancer cell heterogeneity

Patrick C. Kinnunen[1], Kenneth K. Y. Ho[2], Siddhartha Srivastava[3,4], Chengyang Huang[3], Wanggang Shen[3], Krishna Garikipati[3,4,5], Gary D. Luker[2,6], Nikola Banovic[7], Xun Huan[3,4], Jennifer J. Linderman[1,6]* and Kathryn E. Luker[2,8]*

[1]Departments of Chemical Engineering, University of Michigan, Ann Arbor, MI, United States, [2]Radiology, University of Michigan, Ann Arbor, MI, United States, [3]Mechanical Engineering, University of Michigan, Ann Arbor, MI, United States, [4]Michigan Institute for Computational Discovery and Engineering, University of Michigan, Ann Arbor, MI, United States, [5]Mathematics, University of Michigan, Ann Arbor, MI, United States, [6]Biomedical Engineering, University of Michigan, Ann Arbor, MI, United States, [7]Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI, United States, [8]Biointerfaces Institute, University of Michigan, Ann Arbor, MI, United States

Cellular heterogeneity is a ubiquitous aspect of biology and a major obstacle to successful cancer treatment. Several techniques have emerged to quantify heterogeneity in live cells along axes including cellular migration, morphology, growth, and signaling. Crucially, these studies reveal that cellular heterogeneity is not a result of randomness or a failure in cellular control systems, but instead is a predictable aspect of multicellular systems. We hypothesize that individual cells in complex tissues can behave as reward-maximizing agents and that differences in reward perception can explain heterogeneity. In this perspective, we introduce inverse reinforcement learning as a novel approach for analyzing cellular heterogeneity. We briefly detail experimental approaches for measuring cellular heterogeneity over time and how these experiments can generate datasets consisting of cellular states and actions. Next, we show how inverse reinforcement learning can be applied to these datasets to infer how individual cells choose different actions based on heterogeneous states. Finally, we introduce potential applications of inverse reinforcement learning to three cell biology problems. Overall, we expect inverse reinforcement learning to reveal why cells behave heterogeneously and enable identification of novel treatments based on this new understanding.

Introduction

There is an enigma at the heart of mammalian biology. Seemingly identical cells in a population exhibit distinct responses to the same environmental cues. *Consequences* of heterogeneity are readily apparent in normal biology and diseases such as cancer: specialized behaviors of cells, drug resistance, and fatal metastases. Mechanisms *causing* heterogeneity remain a mystery, impeding efforts to shift cell behaviors to prevent or cure disease.

The prevailing dogma is that heterogeneity among cancer cells arises randomly, generating "greedy individuals" that compete for growth factors and optimal environments. However, recent data suggest that cancer cells function cooperatively as a tissue-like entity, and work by our group and others demonstrate that single-cell differences in signaling and function among cancer cells can arise predictably with consistent variations across a population as a whole (Spencer et al., 2009; Overton et al., 2014; Spinosa et al., 2019; Spinosa et al., 2020; Zhan et al., 2020; Kinnunen et al., 2022). These observations imply that tumor progression benefits from or even requires interactions among distinct subgroups of cells (Marusyk et al., 2014). The idea that single, heterogeneous cancer cells work collectively within a constrained range of variability to drive population-level outputs in tumor progression is a concept that may revolutionize how we approach cancer biology and therapy.

To decipher mechanisms regulating single-cell heterogeneity and cooperative interactions among cells, we propose that the field adopt a conceptual approach that integrates: (1) high-dimensional single-cell data, (2) mechanistic modeling, and (3) inverse reinforcement learning (IRL). While typically used to imitate (Abbeel and Ng, 2004) or simulate (Banovic et al., 2016) human behavior, IRL is an artificial intelligence (AI) method that can interpret responses of single cells to multiple stimuli as a decision-making *policy* that is motivated by maximizing a *reward*. Key IRL terms, with application to cancer, are defined in Box 1. In the context of cancer, rewards exist at both single cancer cell and multicellular tumor-microenvironmental scales. For cancer cells positioned in nutrient-rich environments, a reward may be activation of signaling pathways that drive metabolic or cytoskeletal adaptations necessary for proliferation and invasion. Treatment with radiation or chemotherapy leads to rewards related to single and tumor-wide behaviors that promote survival (Shaffer et al., 2017). Single cancer cells may upregulate drug efflux transporters and DNA damage repair processes to resist therapy, while soluble and contact-mediated interactions among cancer and benign stromal cells promote survival of the tumor overall (Lim et al., 2011; Li et al., 2021; Xiao et al., 2021). Tumor-wide cellular and metabolic interactions generate immunosuppressive environments that restrain and exclude anti-cancer immune responses (DePeaux and Delgoffe, 2021; Luby and Alves-Guerra, 2021; Arner and Rathmell, 2023). These examples capture only a subset of the many possible reward-induced "decisions" cancer cells make that may support heterogeneity and drive tumor growth and metastasis.
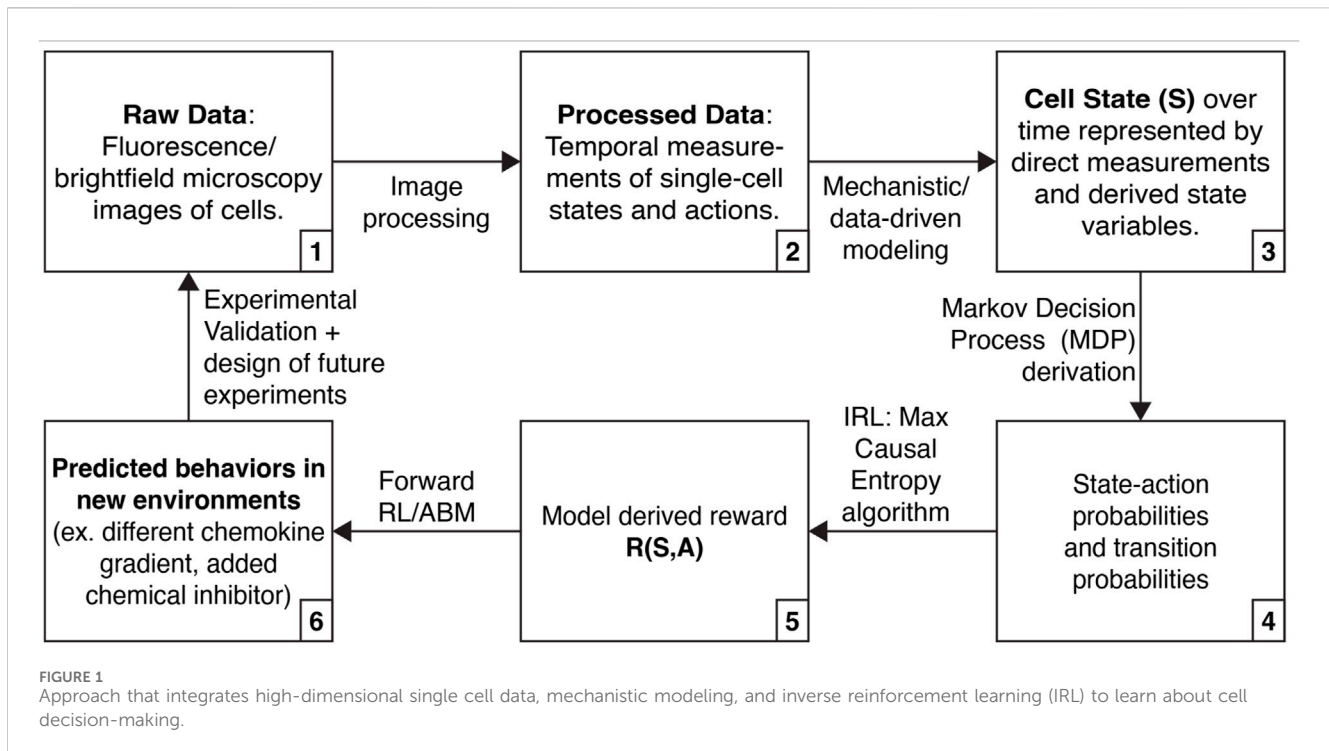
We describe below how high dimensional single-cell data, mechanistic modeling, and IRL might be integrated to discover molecular processes underlying "decision-making" by single cells and their "motivations" for acting competitively or collaboratively in cancer (Figure 1). By basing IRL findings on single cell data and mechanistic models, we can ensure that the approach yields biologically realistic hypotheses (for example, predicted behaviors in new environments, including reproducing heterogeneity in a population or evading drug treatment).

Live cell imaging measures heterogeneous cell states and actions

Live-cell microscopy with advanced image processing methods can track and analyze single cells over space and time, measuring cellular phenotypes such as movement, division, proliferation, and death (Figure 1, steps 1 and 2). Stimuli can be applied to cells to measure the response of each cell, and multiple stimuli can be applied successively to determine how various inputs reinforce or counter outputs such as cell signaling and movement. Live-cell

BOX 1 Key terms in IRL with examples from cancer biology

Term	IRL Definition	Cancer biology example(s)
Agent	An autonomous entity that takes actions in a state-dependent manner to maximize some unknown reward	Cancer cell, stromal cell
State (S)	Variables defining the measurable or model-inferable properties of the agent	Size; location; level of activation of a signaling pathway; readiness to divide; cancer stemness
Action (A)	Performed by agents to transition between states	Moving up a chemical gradient; not moving; cell division; apoptosis; new activation of a signaling pathway
Reward R (S,A)	Benefit that the agent obtains from the environment by taking a particular action when in a particular state. In IRL, the reward is unknown *a priori* and is inferred from observed agent behaviors	High rewards under proliferation or invasion; survival during drug treatment; generation of immunosuppressive environment
Policy	A probabilistic or deterministic mapping from current state to immediate action. IRL assumes that the agent performs the optimal policy maximizing the expected cumulative (unknown) reward over some time horizon	A cancer cell may follow a policy of spending resources to send signals to neighboring stromal cells when it is surrounded by more stromal cells (state), causing them to alter their metabolism to support the cancer cell
Environment	Set of inputs sensed by the agent	Extracellular surroundings, including the presence and concentration of soluble signaling molecules and neighboring cells. Mechanical environment (stiffness) of surroundings

FIGURE 1
Approach that integrates high-dimensional single cell data, mechanistic modeling, and inverse reinforcement learning (IRL) to learn about cell decision-making.

microscopy has revealed previously unexamined dimensions of cellular heterogeneity, including morphology (Gordonov et al., 2016), engulfment (Chu et al., 2020), and migratory capacity (Ferreira et al., 2022). Combining live-cell microscopy with a growing array of optical imaging reporters vastly expands the number of measurable phenotypes per cell and dynamic responses of cells over time. As examples, investigators have used multiplexed fluorescent reporters of cell cycle phases, DNA damage, cell signaling pathways, or protein stability/degradation (Sakaue-Sawano et al., 2008; Regot et al., 2014; Spinosa et al., 2019; Suski et al., 2022; Abd El-Hafeez et al., 2023). Dynamic imaging studies generate large datasets by collecting information from thousands of cells over hours to days.

The application of live-cell fluorescence microscopy to cell biology has revealed two key principles. First, even genetically identical cells respond heterogeneously to identical stimuli. There are numerous examples of this heterogeneity in both continuous and discrete cell actions. For example, the Akt, ERK, and p38 kinase pathways display a continuum of signaling activities in response to chemokine stimulation (Kinnunen et al., 2022). Isogenic cells display a heterogeneous spectrum of chemotactic capacities under identical gradients (Ho et al., 2023). Heterogeneity is also present in cellular decisions relating to actions like cell death and cell-cycle progression. Imaging reporters for these processes have revealed intercellular variations in dynamics of cell division, inheritance of cell *states*, and responses to interventions such as chemotherapy drugs (Laughney et al., 2014; Kukhtevich et al., 2022; Arora et al., 2023). The second principle of cellular heterogeneity is that cellular behaviors are influenced by cell *state*, which is set by past stimuli. We and others have used imaging reporters to detect "memory" of past stimuli, responses to targeted therapy, and how oscillations in kinase activity can control single cell decisions regulating transcription,

chemotaxis, and apoptosis (Tomida et al., 2015; Hiratsuka et al., 2020; Wang et al., 2022; Heaton et al., 2023; Ho et al., 2023). Heterogeneity has been observed even in more complex environments, including in living tissues and organoids (Hiratsuka et al., 2015; de Witte et al., 2020; Ponsioen et al., 2021). Hence, heterogeneity in cell state appears to be a fundamental property of collections of cells.

Heterogeneity enables at least two emergent behaviors in cancer cells: cooperation and bet-hedging. Cooperation enables cells to specialize to create an overall more oncogenic environment. For example, cancer cells can exploit metabolic byproducts from the microenvironment (Richardson et al., 2018; Zhu et al., 2020), and chemokine-expressing metastatic cancer cells can create a favorable environment for non-expressing cells (Shahriari et al., 2017), which would otherwise die. We can think of the cells that rely on byproducts from other cells as selfish exploiters. Bet-hedging refers to the adoption of phenotypes that are suboptimal in the current environment but may be better suited to potential future environments, such as after the application of a cytotoxic drug (Sharma et al., 2010). Understanding how cancer cells collaborate, and when selfish cancer stem cells emerge, could enable the identification of novel cancer targets.

To work with IRL, we envision that live-cell, fluorescence microscopy combined with automated image processing will provide large data sets of cellular behaviors (Moen et al., 2019; Tian et al., 2020). These datasets can include multiple cell types, complex environments, and the addition of multiple exogenous stimuli (Zhang et al., 2019; Buschhaus et al., 2020; Ho et al., 2023). Such datasets can then be converted into sets of single-cell states and actions, a requirement for IRL. Our current microscopy datasets contain ~100,000 such data points (state-action pairs), and we can combine data from multiple experiments, providing ample data for

training IRL algorithms. Furthermore, innovations in live-cell microscopy and fluorescent reporter design will continue to expand the cell states and actions we can measure. IRL might also be combined with other sources of data that collect time-series data consisting of cell states and actions. However, IRL cannot be performed using only single-cell endpoint measurements, such as flow cytometry or single-cell (spatial) transcriptomics, because they do not provide time series data. Endpoint measurements that can be linked to the states and actions of a specific cell, such as cyclic immunofluorescence, could be used to associate specific behaviors with a wider range of endpoint measurements than can be measured in living cells.

Physically-based mechanistic models ground IRL findings in reality

Predictions about the drivers of cell behavior need to be placed in a readily understandable, real-world framework for cell signaling and function: physically-based mechanistic models. Systems biologists have already created a broad corpus of knowledge about heterogeneity in single- and multicellular behavior and regulation. For example, mechanistic models in cancer have been developed for multiple signaling pathways, tissue formation, cell migration, and drug treatment, primarily by combining knowledge of biology with principles from biochemistry, biophysics, and engineering, e.g., diffusion and convection, mechanics, and biochemical reaction networks (Spinosa et al., 2020; Kinnunen et al., 2022; Menezes et al., 2022). Such models may include both deterministic and stochastic elements. An emerging data-driven approach for modeling is system inference. For example, Variational System Identification (VSI) techniques allow estimation of the parametric form of the governing partial differential equations–such as reaction-diffusion and phase field models–that may underlie cell migration and signaling, directly from experimental data (Wang et al., 2019; Wang et al., 2021; Ho et al., 2023; Kinnunen et al., 2023).

An important class of models for our discussion is agent-based models (ABMs). In the current context, the *agents* in the models are individual cells, and they behave and interact in their environment according to probabilistic rules. In particular, and relevant to our IRL discussion, we describe the behavior of agents in an ABM through a Markov Decision Process (MDP), a mathematical framework where cell-agents decide their actions from their current states motivated by gaining higher rewards. ABMs model cellular heterogeneity by explicitly representing cell state, placing heterogeneous cells in a varied environment, and following the state changes and actions taken by individual cells over time as the simulation proceeds. There is now a fairly long history of ABMs in biology with rules informed by our knowledge of biology and also, more recently, by machine learning (Norton et al., 2017; Rikard et al., 2019; Hult et al., 2021; Sivakumar et al., 2022). Yet deducing a rule, for example, that cells are likely to move in a certain way in a certain gradient, does not tell us if or why this *action* supports heterogeneity and ultimately cancer survival. This is a difficult problem because the final result (cancer survival, say) is likely many steps removed from any individual cell's action. For this,

we can turn to IRL to determine the *rewards* that drive the *policies* the cells follow.

We envision using mechanistic modeling to improve the interpretability of IRL inference in three ways (Figure 1, steps 3, 4, and 6). First, mechanistic modeling can expand the number of cell states we can use for IRL. Many cell states do not have associated live-cell reporters, and there are limitations on the number of fluorescence reporters that can be simultaneously measured. However, we can fit data to mechanistic models, elucidating additional states (Yao et al., 2016; Spinosa et al., 2020). Second, mechanistic modeling can identify physical limits in cellular actions or state transitions. For instance, previous work has derived physical limits on a cell's ability to sense a chemical gradient (Mugler et al., 2016). By incorporating these limits into measured state-action transitions, we can prevent IRL from needlessly exploring solutions that are physically inadmissible. Finally, we can use IRL in combination with ABMs to simulate cells following the inferred rewards with controlled perturbations, yielding actionable hypotheses and guiding the design of future experiments (Huan and Marzouk, 2013; Shen and Huan, 2023).

IRL uncovers cell- and tumor-level "motivations" from observed cell states and actions

Uncovering the underlying incentive mechanism in a complex decision-making system is a formidable task, especially when the system is stochastic and its constituent agents possess substantial heterogeneity. IRL is a powerful tool that harnesses agent-scale data to infer the unknown incentive mechanisms governing the behavior of individual agents. IRL differs from the more commonly used reinforcement learning (RL): in RL (or forward RL) an agent learns a good policy for taking actions from trial and error based on a given (known) reward function; in IRL one tries to discover a reward function based on the behavior of an agent that follows an optimal policy in its environment.

In the IRL framework, we model a cancer cell as a decision-making agent under the mathematical formalism of an MDP (Bellman, 1957). This approach is rooted in the assumption that the agent is a rational actor, and the observed data reflect the agent choosing the optimal state-dependent action to maximize its expected cumulative reward while navigating the constraints of its environment. In other words, the agent is assumed to adhere to an optimal policy for some unknown, underlying reward mechanism. For example, we know that only a small subpopulation of cancer cells in a tumor are metastatic (Luzzi et al., 1998). Using IRL, and assuming that these cells are maximizing an unknown reward, might reveal that metastatic cells undergo a set of specific states prior to metastasis, where migration is highly rewarded. Meanwhile, other nonmetastatic cells do not pass through these states (Marusyk et al., 2014). Furthermore, by comparing the magnitude of the rewards accumulated at each step on the path to metastasis, we could identify the steps taken by metastatic cells that are most important to target therapeutically. IRL provides the mathematical and computational tools to systematically identify

other cases where individual cells may adopt seemingly suboptimal phenotypes in order to optimize tumor growth.

In the IRL framework, cells and their surrounding environment (e.g., a neighborhood consisting of various other cells, soluble factors, and mechanical properties) are represented by a set of states. The framework also specifies a set of actions that a cell can take in each of those states (e.g., movement, division, secretion). The cell transitions from state to state appear stochastic for two reasons. First, cell actions can change the environment; for example, the secretion of a cytokine will change the local concentration. Second, cells do not have full control over their environment, and some changes in the environment happen irrespective of cell actions. For instance, a moving cell may intend to move to a region of lower cell density, but since other cells are also moving, it may end up in a region of similar or even higher density. Cells perform actions according to a policy that maximizes a reward function the cell receives after reaching a new state for each action. IRL models cell behavior as a sequence of actions the cell performs as it moves from state to state until reaching some final goal state, such as continuing to proliferate after exposure to a chemotherapeutic drug.

IRL is a method for estimating the rewards of an MDP (Figure 1, steps 3–5). To perform IRL, state-action probabilities are calculated. Here, we envision state-action probabilities being determined both from measured data and augmentation of measurements using mechanistic and data-driven modeling. Next, we parameterize the reward function and use the MaxCausalEntropy algorithm to identify the most likely rewards for each state and action (Ziebart et al., 2010). MaxCausalEntropy is particularly well suited for modeling cellular behaviors because it explicitly models the connection between cell state and cell action, which we assume are connected by (currently incompletely understood) physical and chemical laws.

With an MDP and rewards in hand, we can formulate and test critical hypotheses in cancer biology. We can test whether individual cells in new conditions are behaving consistently with the model, or if they represent outliers displaying new behavior; in other words, how heterogeneous, and in what ways, is the new population? We can calculate the probabilities that cells will exist in particular states, or take particular sequences of actions, to better understand the scope of cell behavior. We can simulate populations of cells under different situations, i.e., make predictions that can then be tested in experiments (Figure 1, step 6). Finally, we can identify a final state of interest (for instance, metastatic or drug resistant cells) and identify the states and actions most likely to lead to that state. These latter examples highlight the ability of the model to enable us to develop targeted interventions to control the behavior of cells.

IRL has had remarkable success in various fields, including human behavior modeling (Antar et al., 2022) and robotics (Finn et al., 2016), but has only recently been applied to biology. IRL was used to understand the clonal evolution of tumors (Kalantari et al., 2020) and mimic the behavior of physicians making cancer treatment decisions (Imani and Braga-Neto, 2019). Two papers apply IRL to study the migration behavior of roundworms (Yamaguchi et al., 2018) and mice (Ashwood et al., 2022), which are particularly relevant for our application. Yamaguchi et al. used IRL to study thermotaxis in roundworms (Yamaguchi et al., 2018). They tracked roundworm migration in a thermal gradient using recordings and automated video analysis, which generated hundreds
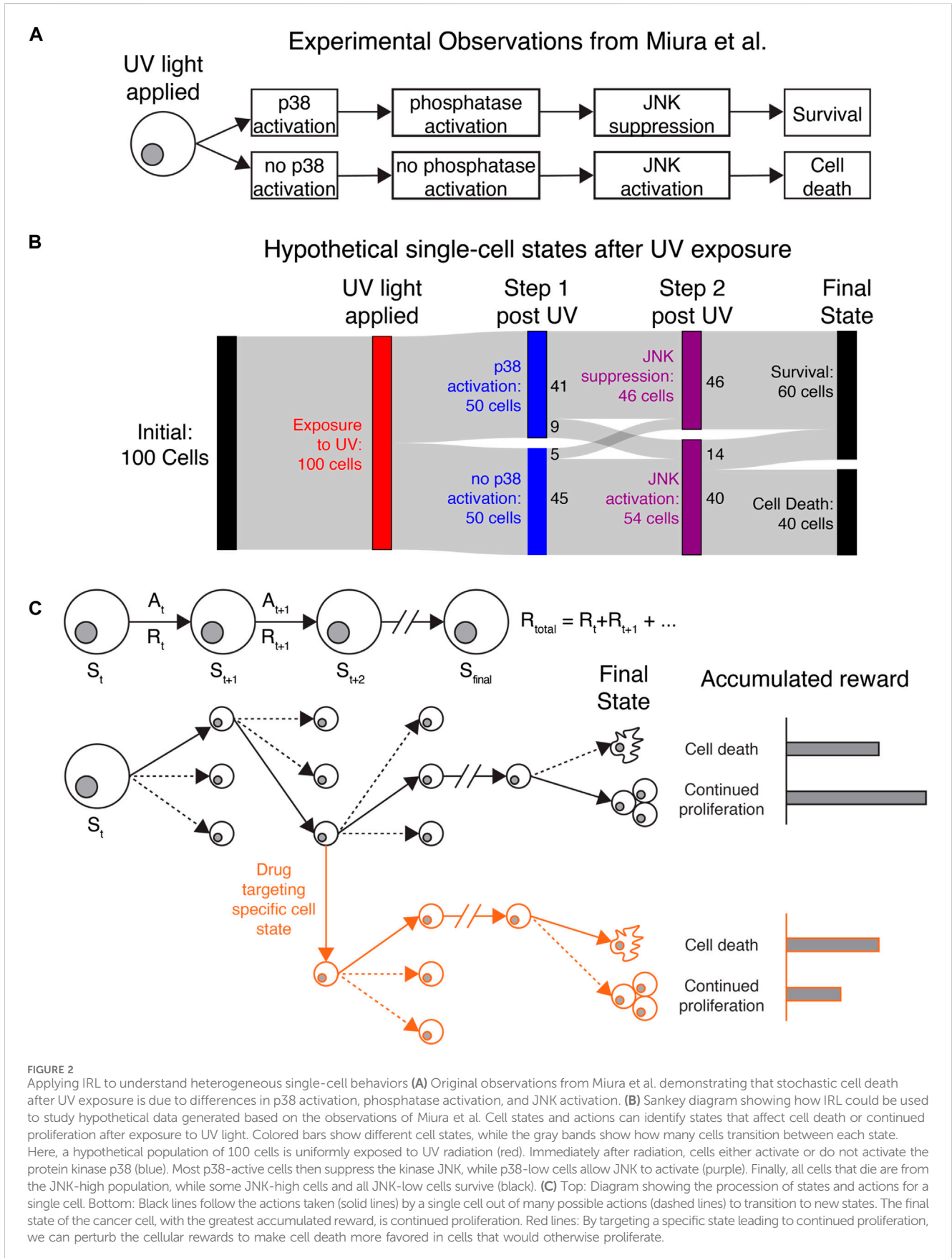
of single-worm trajectories. They modeled the worm state based on the current temperature and the current temperature gradient. IRL revealed different migration strategies for worms grown in different conditions, which recapitulated prior knowledge about worm thermotaxis. Ashwood et al. applied IRL to mice navigating a maze (Rosenberg et al., 2021; Ashwood et al., 2022). They also used video recordings as a data source and were able to identify different time-varying rewards for water-restricted and -unrestricted mice. The data used in these studies are structurally very similar to the data collected from live-cell microscopy, which suggests that similar techniques may be effective.

Challenges in applying IRL to cellular behaviors

Despite the effectiveness of IRL in various fields, it comes with significant challenges and limitations. First, IRL is inherently ill-posed since many reward functions exist that can explain the demonstrated trajectories equally well, which can lead to overfitting. Moreover, the ill-posedness can be exacerbated by incomplete or imperfect knowledge about the environmental dynamics and where an explicit, analytical form of the state transition function is unavailable, as in many biological scenarios. These challenges emphasize the need to embed IRL within an experimental framework where inferred rewards can be tested using new experiments incorporating genetic, chemical, or environmental perturbations. Second, IRL may infer rewards that do not make physical sense–for instance, predicting cell division more quickly than cells could possibly divide–or are difficult to interpret. Thirdly, IRL faces challenges related to computational complexity and sample size requirements, both of which usually increase with the dimensionality of the state-action space. Meanwhile, as the problem size increases, more diverse examples of behavior are needed to maintain sufficient coverage in the training data. This need highlights another challenge: generalizability. The difficulty lies in accurately extrapolating to unobserved spaces using data that often covers only a fraction of the total space. Relying solely on observations to generalize to state and action regions beyond training samples becomes especially difficult in high-dimensional settings, compounded when training data are limited and noisy. To help resolve these problems, we emphasize that combining IRL with more traditional biochemical and biophysical modeling will ensure that the learned rewards are physically meaningful and interpretable. An example of this approach recently developed by our team is Fokker-Planck-based IRL (FP-IRL), which we will elaborate below.

Toward integrating IRL, mechanistic models, and single-cell biology: three potential applications

We present three potential applications where IRL may help uncover key insights for understanding cancer cell heterogeneity. The first two fall into the category of *single-agent IRL*. Here we consider a population of cells observed in our microscopy experiments (e.g., all cancer cells in the field of view) as a

FIGURE 2
Applying IRL to understand heterogeneous single-cell behaviors **(A)** Original observations from Miura et al. demonstrating that stochastic cell death after UV exposure is due to differences in p38 activation, phosphatase activation, and JNK activation. **(B)** Sankey diagram showing how IRL could be used to study hypothetical data generated based on the observations of Miura et al. Cell states and actions can identify states that affect cell death or continued proliferation after exposure to UV light. Colored bars show different cell states, while the gray bands show how many cells transition between each state. Here, a hypothetical population of 100 cells is uniformly exposed to UV radiation (red). Immediately after radiation, cells either activate or do not activate the protein kinase p38 (blue). Most p38-active cells then suppress the kinase JNK, while p38-low cells allow JNK to activate (purple). Finally, all cells that die are from the JNK-high population, while some JNK-high cells and all JNK-low cells survive (black). **(C)** Top: Diagram showing the procession of states and actions for a single cell. Bottom: Black lines follow the actions taken (solid lines) by a single cell out of many possible actions (dashed lines) to transition to new states. The final state of the cancer cell, with the greatest accumulated reward, is continued proliferation. Red lines: By targeting a specific state leading to continued proliferation, we can perturb the cellular rewards to make cell death more favored in cells that would otherwise proliferate.

collection of single agents operating independently and with no awareness of each other's actions but obeying the same policy. For a concrete example, consider the behavior of individual cells collected by Miura et al. after exposure to UV-C radiation (Miura et al., 2018). The study identified a molecular determinant of UV-induced cell death by tracking cell motion, kinase activity, and cell survival over time (Figure 2A). Radiation activates JNK kinase after several hours, which induces cell death. Cells that survive radiation first activate p38 kinase, which induces transcription of a regulatory phosphatase that inhibits JNK and prevents cell death.

If we were to use IRL to understand the observations of Miura et al., we could consider JNK, p38, and cell survival/death as key state variables. IRL would first reveal the most common series of events (change in p38, followed by change in JNK, possibly followed by cell death) based on the transitions between states that are present in the data. Identifying the most common state transitions may be trivial in this application, but if more reporters were used or more states were identified from the data, it could be more difficult to identify common series of events. IRL would also show the dominant state-action transitions leading to cell death or survival, where most cells that survive first activate p38. A Sankey diagram (Antar et al., 2022) showing the behavior of 100 hypothetical cells is shown in Figure 2B. Most cells follow the series of events shown in Figure 2A, while a minority do not because of unknown sources of regulation affecting p38 and JNK activity. Miura et al. used separate experiments inspired by biological knowledge to reveal the phosphatase dynamics underlying JNK suppression. Since the phosphatase was not captured in the live-cell imaging experiments, IRL would not be able to identify it. However, IRL would demonstrate that most cells that first activate p38 do not activate JNK, which could generate hypotheses about how these two molecules are connected. Furthermore, after IRL reward inference, we could use the observed rewards to simulate realistic cellular behaviors in different environments or in the presence of different perturbations (Figure 2C). The inferred reward and measured state-action transitions could be used to identify states most likely to lead to cell survival. Identifying these states and targeting them could reveal novel, experimentally testable perturbations to prevent cell survival. In this example, IRL provides a unique, data-driven lens for identifying granular cellular activities that drive specific phenotypes.

As another application, we developed a novel IRL algorithm, called Fokker-Planck IRL (Garikipati et al., 2023), to better understand how chemokine gradients affect cell migration decisions (Ho et al., 2023). FP-IRL infers the transition and reward function simultaneously in a physics-constrained manner by leveraging a mathematical conjecture on a structural isomorphism (i.e., equivalence mapping) between the FP equation, which governs particle motion affected by diffusive and advective forces, and MDP, which is the mathematical basis for IRL. We found that the injection of physical principles mitigates some of the aforementioned challenges, including ill-posedness, physical interpretability, and computational efficiency. We first validated FP-IRL on a synthetic problem that mimics cell migration under a chemotactic gradient. Computational convergence studies showed that FP-IRL can accurately estimate the reward and transition functions we defined in the simulation. To test the method, we then applied FP-IRL to an experimental dataset (1,332 cells over 361 total timesteps) of MDA-MB-231 breast cancer cells expressing

fluorescent reporters for Akt and ERK kinases in a chemotaxis assay (Ho et al., 2023). We applied a chemical gradient of chemoattractant CXCL12, which induced cells to move up the gradient. We modeled the cancer cells as decision-making agents under the mathematical formalism of an MDP. We defined x- and y-velocity as state variables and changes in Akt and ERK signaling as actions. FP-IRL identified that cells have a higher reward for migrating up the gradient with relatively high speed, in agreement with our understanding of chemotaxis. Going forward, this method can be applied to understand cell migration strategies in new environments.

Our third potential application employs *multi-agent* IRL to understand competitive and cooperative cellular interactions that support overall tumor progression. Here, we can model each agent (in an overall multi-agent setting) to have its own individual reward function, for instance that might correspond to short-term and long-term goals, or local (agent-level) and global (population-level) goals. Using this approach, we could understand how multiple cancer cells adopt a range of phenotypes (following cell-level rewards) to support the overall proliferation of a tumor (a population-level reward). Experimentally, we could monitor cell proliferation from a small, sparse population of cells to a monolayer, and expose them to sequential doses of different cytotoxic drugs. In this case, we could track the emergence of heterogeneity and the eventual death of part of the population in response to different stressors. By assuming that the dead cells provided some benefit to the living cells and that cell death was state dependent, we could apply multi-agent IRL to understand what state-action pairs had high rewards for the individual cell and which had high rewards for survival of the population as a whole. Multi-agent IRL is much more computationally demanding than traditional, single-agent IRL since it must track and capture the interplay of actions by different agents. New algorithms and methodology are currently under development to overcome the computational challenges.

Discussion

The framework described in this paper—using IRL together with physically-based mechanistic models to interpret high-dimensional live-cell imaging datasets—has potentially game-changing implications for how we understand and treat cancer. First, it provides a rigorous framework for testing if the hypothesis that cells pursue rewards is relevant to cancer. It is likely true that in some cases, clear rewards can be inferred from heterogeneous cellular behaviors (e.g., cooperation or bet-hedging). However, since cellular regulation is imperfect and generally mediated by local signals, it is also likely that some heterogeneity is random, unregulated, or not driven by cellular cooperation. For behaviors that are reward-driven, we will also learn some of the molecular drivers of cell behavior and potential interventions. Analyzing the reward function will further enable us to develop targeted interventions to control the behavior of cells. By inferring decision-making policies for single-cell and population-scale outputs, we may be able to design therapies to pre-emptively shift cells from aggressive

behaviors and disrupt collaborative interactions among subpopulations of cells in a tumor, rather than reacting to these processes after they occur. Combining IRL with physically-based mechanistic models means that we will be able to identify specific, and potentially targetable, drivers of collaborative behaviors.

Although IRL is an emerging technique and questions remain about the application of IRL to single-cell behavioral data, we emphasize that techniques for measuring cell states and actions and achieving granular control over single cells are expanding rapidly. IRL will serve as a powerful method for modeling these new data streams. Specifically, novel reporters have multiplexed up to seven separate fluorescent channels (Qian et al., 2023) and demonstrated the ability to extract single-cell biological information from novel frequency-based fluorescent reporters (Rajasekaran et al., 2024). Such capabilities dramatically expand the range of single-cell states and actions that can be measured. Another emerging approach, where individual cells record a specific physiological variable, such as promoter activity or chemical exposure, onto a protein- (Ravindran et al., 2022; Linghu et al., 2023) or DNA-based (Park et al., 2021) recorder analyzed using endpoint methods, could serve as a novel source for cell state-action data. Finally, recent work has recapitulated fully synthetic kinase networks in mammalian cells (Yang et al., 2023), and optogenetics enables the activation of signaling molecules in cells (Wilson et al., 2017). These tools offer finely tuned control over specific cell behaviors in experimental formats that are compatible with long-term single-cell measurements.

IRL is a general framework that can be adopted for other biological contexts where agent-based perspectives are appropriate. For example, bacteria function as integrated communities, generating interconnected biofilms under stressful conditions. Inflammation in cancer, infections, and other diseases represents a delicate balance between pro-inflammatory and anti-inflammatory cells and molecules. Inferring the cellular reward structure for sustaining or ending inflammation may reveal decision points controlling immunosuppression in tumors and persistent immune responses in autoimmune disorders. We believe IRL will help us understand the underlying causes of cellular heterogeneity by quantifying state-dependent rewards and ultimately contribute to a novel biological paradigm in which the individual roles of heterogeneous cells are considered as the basis of physiological processes.

Data availability statement

The raw data supporting the conclusion of this article will be made available by the authors, without undue reservation.

Ethics statement

Ethical approval was not required for the studies on animals in accordance with the local legislation and institutional requirements because only commercially available established cell lines were used.

Author contributions

Funding

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Abbeel, P., and Ng, A. Y. (2004). "Apprenticeship learning via inverse reinforcement learning," in Twenty-first international conference on Machine learning - ICML '04 1, Banff, Alberta, Canada, July, 2004. doi:10.1145/1015330.1015430

Abd El-Hafeez, A. A., Sun, N., Chakraborty, A., Ear, J., Roy, S., Chamarthi, P., et al. (2023). Regulation of DNA damage response by trimeric G-proteins. *iScience* 26, 105973. doi:10.1016/j.isci.2023.105973

Antar, A. D., Kratz, A., and Banovic, N. (2022). Behavior modeling approach for forecasting physical functioning of people with multiple sclerosis. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 7, 1–29. doi:10.1145/3580887

Arner, E. N., and Rathmell, J. C. (2023). Metabolic programming and immune suppression in the tumor microenvironment. *Cancer Cell* 41, 421–433. doi:10.1016/j.ccell.2023.01.009

Arora, M., Moser, J., Hoffman, T. E., Watts, L. P., Min, M., Musteanu, M., et al. (2023). Rapid adaptation to CDK2 inhibition exposes intrinsic cell-cycle plasticity. *Cell* 186, 2628–2643.e21. doi:10.1016/j.cell.2023.05.013

Ashwood, Z., Jha, A., and Pillow, J. W. (2022). "Dynamic inverse reinforcement learning for characterizing animal behavior," in *Advances in neural information processing systems*. Editor S. Koyejo, et al. (Red Hook, New York, United States: Curran Associates, Inc.), 29663–29676.

Banovic, N., Buzali, T., Chevalier, F., Mankoff, J., and Dey, A. K. (2016). "Modeling and understanding human routine behavior," in Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, San Jose, California, USA, May, 2016. doi:10.1145/2858036.2858557

Bellman, R. A. (1957). A markovian decision process. *J. Math. Mech.* 6, 679–684. doi:10.1512/iumj.1957.6.56038

Buschhaus, J. M., Humphries, B. A., Eckley, S. S., Robison, T. H., Cutter, A. C., Rajendran, S., et al. (2020). Targeting disseminated estrogen-receptor-positive breast cancer cells in bone marrow. *Oncogene* 39, 5649–5662. doi:10.1038/s41388-020-01391-z

Chu, C. C., Pinney, J. J., Whitehead, H. E., Rivera-Escalera, F., VanDerMeid, K. R., Zent, C. S., et al. (2020). High-resolution quantification of discrete phagocytic events by live cell time-lapse high-content microscopy imaging. *J. Cell Sci.* 133, jcs237883. doi:10.1242/jcs.237883

DePeaux, K., and Delgoffe, G. M. (2021). Metabolic barriers to cancer immunotherapy. *Nat. Rev. Immunol.* 21, 785–797. doi:10.1038/s41577-021-00541-y

de Witte, C. J., Espejo Valle-Inclan, J., Hami, N., Lõhmussaar, K., Kopper, O., Vreuls, C. P. H., et al. (2020). Patient-derived ovarian cancer organoids mimic clinical response and exhibit heterogeneous inter- and intrapatient drug responses. *Cell Rep.* 31, 107762. doi:10.1016/j.celrep.2020.107762

Ferreira, A., Bressan, C., Hardy, S. V., and Saghatelyan, A. (2022). Deciphering heterogeneous populations of migrating cells based on the computational assessment of their dynamic properties. *Stem Cell Rep.* 17, 911–923. doi:10.1016/j.stemcr.2022.02.011

Finn, C., Levine, S., and Abbeel, P. (2016). "Guided cost learning: deep inverse optimal control via policy optimization," in Proceedings of The 33rd International Conference on Machine Learning, New York NY USA, June, 2016, 49–58.

Garikipati, K., Huang, C., Srivastava, S., and Huan, X. (2023). FP-IRL Fokker-Planck-based inverse reinforcement learning -- A physics-constrained approach to Markov decision processes. https://arxiv.org/abs/2306.10407.

Gordonov, S., Hwang, M. K., Wells, A., Gertler, F. B., Lauffenburger, D. A., and Bathe, M. (2016). Time series modeling of live-cell shape dynamics for image-based phenotypic profiling. *Integr. Biol.* 8, 73–90. doi:10.1039/c5ib00283d

Heaton, A. R., Rehani, P. R., Hoefges, A., Lopez, A. F., Erbe, A. K., Sondel, P. M., et al. (2023). Single cell metabolic imaging of tumor and immune cells *in vivo* in melanoma bearing mice. *Front. Oncol.* 13, 1110503. doi:10.3389/fonc.2023.1110503

Hiratsuka, T., Bordeu, I., Pruessner, G., and Watt, F. M. (2020). Regulation of ERK basal and pulsatile activity control proliferation and exit from the stem cell compartment in mammalian epidermis. *Proc. Natl. Acad. Sci.* 117, 17796–17807. doi:10.1073/pnas.2006965117

Hiratsuka, T., Fujita, Y., Naoki, H., Aoki, K., Kamioka, Y., and Matsuda, M. (2015). Intercellular propagation of extracellular signal-regulated kinase activation revealed by *in vivo* imaging of mouse skin. *eLife* 4, e05178. doi:10.7554/eLife.05178

Ho, K. K. Y., Srivastava, S., Kinnunen, P. C., Garikipati, K., Luker, G. D., and Luker, K. E. (2023). Oscillatory ERK signaling and morphology determine heterogeneity of breast cancer cell chemotaxis via MEK-ERK and p38-MAPK signaling pathways. *Bioeng. Basel Switz.* 10, 269. doi:10.3390/bioengineering10020269

Huan, X., and Marzouk, Y. M. (2013). Simulation-based optimal Bayesian experimental design for nonlinear systems. *J. Comput. Phys.* 232, 288–317. doi:10.1016/j.jcp.2012.08.013

Hult, C., Mattila, J. T., Gideon, H. P., Linderman, J. J., and Kirschner, D. E. (2021). Neutrophil dynamics affect *Mycobacterium tuberculosis* granuloma outcomes and dissemination. *Front. Immunol.* 12, 712457. doi:10.3389/fimmu.2021.712457

Imani, M., and Braga-Neto, U. M. (2019). Control of gene regulatory networks using bayesian inverse reinforcement learning. *IEEE/ACM Trans. Comput. Biol. Bioinform.* 16, 1250–1261. doi:10.1109/TCBB.2018.2830357

Kalantari, J., Nelson, H., and Chia, N. (2020). The unreasonable effectiveness of inverse reinforcement learning in advancing cancer research. *Proc. AAAI Conf. Artif. Intell.* 34, 437–445. doi:10.1609/aaai.v34i01.5380

Kinnunen, P. C., Srivastava, S., Wang, Z., Ho, K. K. Y., Humphries, B. A., Chen, S., et al. (2023). Partial differential equation-based inference of migration and proliferation mechanisms in cancer cell populations. Preprint at http://arxiv.org/abs/2302.09445.

Kinnunen, P. C., Luker, G. D., Luker, K. E., and Linderman, J. J. (2022). Computational modeling implicates protein scaffolding in p38 regulation of Akt. *J. Theor. Biol.* 555, 111294. doi:10.1016/j.jtbi.2022.111294

Kukhtevich, I. V., Rivero-Romano, M., Rakesh, N., Bheda, P., Chadha, Y., Rosales-Becerra, P., et al. (2022). Quantitative RNA imaging in single live cells reveals age-dependent asymmetric inheritance. *Cell Rep.* 41, 111656. doi:10.1016/j.celrep.2022.111656

Laughney, A. M., Kim, E., Sprachman, M. M., Miller, M. A., Kohler, R. H., Yang, K. S., et al. (2014). Single-cell pharmacokinetic imaging reveals a therapeutic strategy to overcome drug resistance to the microtubule inhibitor eribulin. *Sci. Transl. Med.* 6, 261ra152. doi:10.1126/scitranslmed.3009318

Li, L., Guan, Y., Chen, X., Yang, J., and Cheng, Y. (2021). DNA repair pathways in cancer therapy and resistance. *Front. Pharmacol.* 11, 629266. doi:10.3389/fphar.2020.629266

Lim, P. K., Bliss, S. A., Patel, S. A., Taborga, M., Dave, M. A., Gregory, L. A., et al. (2011). Gap junction–mediated import of MicroRNA from bone marrow stromal cells can elicit cell cycle quiescence in breast cancer cells. *Cancer Res.* 71, 1550–1560. doi:10.1158/0008-5472.CAN-10-2372

Linghu, C., An, B., Shpokayte, M., Celiker, O. T., Shmoel, N., Zhang, R., et al. (2023). Recording of cellular physiological histories along optically readable self-assembling protein chains. *Nat. Biotechnol.* 41, 640–651. doi:10.1038/s41587-022-01586-7

Luby, A., and Alves-Guerra, M.-C. (2021). Targeting metabolism to control immune responses in cancer and improve checkpoint blockade immunotherapy. *Cancers* 13, 5912. doi:10.3390/cancers13235912

Luzzi, K. J., MacDonald, I. C., Schmidt, E. E., Kerkvliet, N., Morris, V. L., Chambers, A. F., et al. (1998). Multistep nature of metastatic inefficiency: dormancy of solitary cells after successful extravasation and limited survival of early micrometastases. *Am. J. Pathol.* 153, 865–873. doi:10.1016/S0002-9440(10)65628-3

Marusyk, A., Tabassum, D. P., Altrock, P. M., Almendro, V., Michor, F., and Polyak, K. (2014). Non-cell-autonomous driving of tumour growth supports sub-clonal heterogeneity. *Nature* 514, 54–58. doi:10.1038/nature13556

Menezes, B., Linderman, J. J., and Thurber, G. M. (2022). Simulating the selection of resistant cells with bystander killing and antibody coadministration in heterogeneous human epidermal growth factor receptor 2–positive tumors. *Drug Metab. Dispos.* 50, 8–16. doi:10.1124/dmd.121.000503

Miura, H., Kondo, Y., Matsuda, M., and Aoki, K. (2018). Cell-to-Cell heterogeneity in p38-mediated cross-inhibition of JNK causes stochastic cell death. *Cell Rep.* 24, 2658–2668. doi:10.1016/j.celrep.2018.08.020

Moen, E., Bannon, D., Kudo, T., Graf, W., Covert, M., and Van Valen, D. (2019). Deep learning for cellular image analysis. *Nat. Methods* 16, 1233–1246. doi:10.1038/s41592-019-0403-1

Mugler, A., Levchenko, A., and Nemenman, I. (2016). Limits to the precision of gradient sensing with spatial communication and temporal integration. *Proc. Natl. Acad. Sci. U. S. A.* 113, E689–E695. doi:10.1073/pnas.1509597112

Norton, K.-A., Wallace, T., Pandey, N. B., and Popel, A. S. (2017). An agent-based model of triple-negative breast cancer: the interplay between chemokine receptor CCR5 expression, cancer stem cells, and hypoxia. *BMC Syst. Biol.* 11, 68. doi:10.1186/s12918-017-0445-x

Overton, K. W., Spencer, S. L., Noderer, W. L., Meyer, T., and Wang, C. L. (2014). Basal p21 controls population heterogeneity in cycling and quiescent cell cycle states. *Proc. Natl. Acad. Sci.* 111, E4386–E4393. doi:10.1073/pnas.1409797111

Park, J., Lim, J. M., Jung, I., Heo, S. J., Park, J., Chang, Y., et al. (2021). Recording of elapsed time and temporal information about biological events using Cas9. *Cell* 184, 1047–1063.e23. doi:10.1016/j.cell.2021.01.014

Ponsioen, B., Post, J. B., Buissant des Amorie, J. R., Laskaris, D., van Ineveld, R. L., Kersten, S., et al. (2021). Quantifying single-cell ERK dynamics in colorectal cancer organoids reveals EGFR as an amplifier of oncogenic MAPK pathway signalling. *Nat. Cell Biol.* 23, 377–390. doi:10.1038/s41556-021-00654-5

Qian, Y., Celiker, O. T., Wang, Z., Guner-Ataman, B., and Boyden, E. S. (2023). Temporally multiplexed imaging of dynamic signaling networks in living cells. *Cell* 186, 5656–5672.e21. doi:10.1016/j.cell.2023.11.010

Rajasekaran, R., Chang, C.-C., Weix, E. W. Z., Galateo, T. M., and Coyle, S. M. (2024). A programmable reaction-diffusion system for spatiotemporal cell signaling circuit design. *Cell* 187, 345–359.e16. doi:10.1016/j.cell.2023.12.007

Ravindran, P. T., McFann, S., Thornton, R. H., and Toettcher, J. E. (2022). A synthetic gene circuit for imaging-free detection of signaling pulses. *Cell Syst.* 13, 131–142.e13. doi:10.1016/j.cels.2021.10.002

Regot, S., Hughey, J. J., Bajar, B. T., Carrasco, S., and Covert, M. W. (2014). High-sensitivity measurements of multiple kinase activities in live single cells. *Cell* 157, 1724–1734. doi:10.1016/j.cell.2014.04.039

Richardson, A. M., Havel, L. S., Koyen, A. E., Konen, J. M., Shupe, J., Wiles, W. G., et al. (2018). Vimentin is required for lung adenocarcinoma metastasis via heterotypic tumor cell–cancer-associated fibroblast interactions during collective invasion. *Clin. Cancer Res.* 24, 420–432. doi:10.1158/1078-0432.CCR-17-1776

Rikard, S. M., Athey, T. L., Nelson, A. R., Christiansen, S. L. M., Lee, J. J., Holmes, J. W., et al. (2019). Multiscale coupling of an agent-based model of tissue fibrosis and a logic-based model of intracellular signaling. *Front. Physiol.* 10, 1481. doi:10.3389/fphys.2019.01481

Rosenberg, M., Zhang, T., Perona, P., and Meister, M. (2021). Mice in a labyrinth show rapid learning, sudden insight, and efficient exploration. *eLife* 10, e66175. doi:10.7554/eLife.66175

Sakaue-Sawano, A., Kurokawa, H., Morimura, T., Hanyu, A., Hama, H., Osawa, H., et al. (2008). Visualizing spatiotemporal dynamics of multicellular cell-cycle progression. *Cell* 132, 487–498. doi:10.1016/j.cell.2007.12.033

Shaffer, S. M., Dunagin, M. C., Torborg, S. R., Torre, E. A., Emert, B., Krepler, C., et al. (2017). Rare cell variability and drug-induced reprogramming as a mode of cancer drug resistance. *Nature* 546, 431–435. doi:10.1038/nature22794

Shahriari, K., Shen, F., Worrede-Mahdi, A., Liu, Q., Gong, Y., Garcia, F. U., et al. (2017). Cooperation among heterogeneous prostate cancer cells in the bone metastatic niche. *Oncogene* 36, 2846–2856. doi:10.1038/onc.2016.436

Sharma, S. V., Lee, D. Y., Li, B., Quinlan, M. P., Takahashi, F., Maheswaran, S., et al. (2010). A chromatin-mediated reversible drug-tolerant state in cancer cell subpopulations. *Cell* 141, 69–80. doi:10.1016/j.cell.2010.02.027

Shen, W., and Huan, X. (2023). Bayesian sequential optimal experimental design for nonlinear models using policy gradient reinforcement learning. *Comput. Methods Appl. Mech. Eng.* 416, 116304. doi:10.1016/j.cma.2023.116304

Sivakumar, N., Mura, C., and Peirce, S. M. (2022). Innovations in integrating machine learning and agent-based modeling of biomedical systems. *Front. Syst. Biol.* 2, 959665. doi:10.3389/fsysb.2022.959665

Spencer, S. L., Gaudet, S., Albeck, J. G., Burke, J. M., and Sorger, P. K. (2009). Non-genetic origins of cell-to-cell variability in TRAIL-induced apoptosis. *Nature* 459, 428–432. doi:10.1038/nature08012

Spinosa, P. C., Humphries, B. A., Lewin Mejia, D., Buschhaus, J. M., Linderman, J. J., Luker, G. D., et al. (2019). Short-term cellular memory tunes the signaling responses of the chemokine receptor CXCR4. *Sci. Signal* 12, eaaw4204. doi:10.1126/scisignal.aaw4204

Spinosa, P. C., Kinnunen, P. C., Humphries, B. A., Luker, G. D., Luker, K. E., and Linderman, J. J. (2020). Pre-existing cell states control heterogeneity of both EGFR and CXCR4 signaling. *Cell. Mol. Bioeng.* 14, 49–64. doi:10.1007/s12195-020-00640-1

Suski, J. M., Ratnayeke, N., Braun, M., Zhang, T., Strmiska, V., Michowski, W., et al. (2022). CDC7-independent G1/S transition revealed by targeted protein degradation. *Nature* 605, 357–365. doi:10.1038/s41586-022-04698-x

Tian, C., Yang, C., and Spencer, S. L. (2020). EllipTrack: a global-local cell-tracking pipeline for 2D fluorescence time-lapse microscopy. *Cell Rep.* 32, 107984. doi:10.1016/j.celrep.2020.107984

Tomida, T., Takekawa, M., and Saito, H. (2015). Oscillation of p38 activity controls efficient pro-inflammatory gene expression. *Nat. Commun.* 6, 8350. doi:10.1038/ncomms9350

Wang, A. G., Son, M., Kenna, E., Thom, N., and Tay, S. (2022). NF-κB memory coordinates transcriptional responses to dynamic inflammatory stimuli. *Cell Rep.* 40, 111159. doi:10.1016/j.celrep.2022.111159

Wang, Z., Huan, X., and Garikipati, K. (2019). Variational system identification of the partial differential equations governing the physics of pattern-formation: inference under varying fidelity and noise. *Comput. Methods Appl. Mech. Eng.* 356, 44–74. doi:10.1016/j.cma.2019.07.007

Wang, Z., Huan, X., and Garikipati, K. (2021). Variational system identification of the partial differential equations governing microstructure evolution in materials: inference over sparse and spatially unrelated data. *Comput. Methods Appl. Mech. Eng.* 377, 113706. doi:10.1016/j.cma.2021.113706

Wilson, M. Z., Ravindran, P. T., Lim, W. A., and Toettcher, J. E. (2017). Tracing information flow from erk to target gene induction reveals mechanisms of dynamic and combinatorial control. *Mol. Cell* 67, 757–769. doi:10.1016/j.molcel.2017.07.016

Xiao, H., Zheng, Y., Ma, L., Tian, L., and Sun, Q. (2021). Clinically-relevant ABC transporter for anti-cancer drug resistance. *Front. Pharmacol.* 12, 648407. doi:10.3389/fphar.2021.648407

Yamaguchi, S., Naoki, H., Ikeda, M., Tsukada, Y., Nakano, S., Mori, I., et al. (2018). Identification of animal behavioral strategies by inverse reinforcement learning. *PLOS Comput. Biol.* 14, e1006122. doi:10.1371/journal.pcbi.1006122

Yang, X., Rocks, J. W., Jiang, K., Walters, A. J., Rai, K., Liu, J., et al. (2023). Engineering synthetic phosphorylation signaling networks in human cells. http://biorxiv.org/lookup/doi/10.1101/2023.09.11.557100.

Yao, J., Pilko, A., and Wollman, R. (2016). Distinct cellular states determine calcium signaling response. *Mol. Syst. Biol.* 12, 894. doi:10.15252/msb.20167137

Zhan, H., Bhattacharya, S., Cai, H., Iglesias, P. A., Huang, C. H., and Devreotes, P. N. (2020). An excitable ras/PI3K/ERK signaling network controls migration and oncogenic transformation in epithelial cells. *Dev. Cell* 54, 608–623. doi:10.1016/j.devcel.2020.08.001

Zhang, C., Tu, H. L., Jia, G., Mukhtar, T., Taylor, V., Rzhetsky, A., et al. (2019). Ultra-multiplexed analysis of single-cell dynamics reveals logic rules in differentiation. *Sci. Adv.* 5, eaav7959. doi:10.1126/sciadv.aav7959

Zhu, Z., Achreja, A., Meurs, N., Animasahun, O., Owen, S., Mittal, A., et al. (2020). Tumour-reprogrammed stromal BCAT1 fuels branched-chain ketoacid dependency in stromal-rich PDAC tumours. *Nat. Metab.* 2, 775–792. doi:10.1038/s42255-020-0226-5

Ziebart, B. D., Bagnell, J. A., and Dey, A. K. (2010). "Modeling interaction via the principle of maximum causal entropy," in Proceedings of the 27th International Conference on International Conference on Machine Learning, Madison, WI, USA, June, 2010.