HHS Public Access

Author manuscript

Cancer Res. Author manuscript; available in PMC 2026 February 11.

Published in final edited form as:

Cancer Res. 2025 December 15; 85(24): 4880–4889. doi:10.1158/0008-5472.CAN-25-0700.

A Roadmap for the Future of Systems Biology in Cancer Research

H. Steven Wiley[1,6], **Carlos F. Lopez**[2], **Andrei S. Rodin**[3], **Russell Rockne**[4], **Thomas E. Yankeelov**[5], **Herbert M. Sauro**[6], **Ghmkin Hassan**[7], **Sandhya Prabhakaran**[8], **Emek Demir**[9], **Mengzhou Hu**[10], **Juan Fuxman Bass**[11,22], **Kimberly Luddy**[12], **Hannah Newman**[12], **Jonathan P. Mochel**[13], **James C. Costello**[14], **Jianhua Xing**[15], **Said M Afify**[16], **Ahmet Acar**[17], **Melanie Haden Gephart**[18], **Song Feng**[19], **Orion Banks**[20], **Jineta Banerjee**[21], **Ashley Clayton**[20], **David E. Hill**[22], **Randy F. Sweis**[23], **Maxine C. Umeh-Garcia**[24], **Yapeng Su**[25], **Aaron S. Meyer**[26]

[1]Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, WA, 99352.

[2]Multiscale Modeling Group, Institute of Technology, Altos Labs, Redwood City, CA 94065.

[3]Department of Computational and Quantitative Medicine, Beckman Research Institute, City of Hope National Medical Center, Duarte CA, 91010

[4]Department of Computational and Quantitative Medicine, Beckman Research Institute, City of Hope National Medical Center, Duarte CA, 91010.

[5]Oden Institute for Computational Engineering and Sciences, Livestrong Cancer Institutes; Departments of Biomedical Engineering, Diagnostic Medicine, Oncology, The University of Texas at Austin; Department of Imaging Physics, MD Anderson Cancer Center

[6]Department of Bioengineering, University of Washington, Box 355061, Seattle, WA 98195

[7]Department of Surgery, Center for Bioengineering and Tissue Regeneration, University of California, San Francisco, San Francisco, CA 94143; Department of Cancer Stem Cell Engineering, Faculty of Interdisciplinary Science and Engineering in Health Systems, Institute of Academic & Research, Okayama University, Okayama 700-8530, Japan

[8]Integrated Mathematical Oncology department, Moffitt Cancer Center, Tampa, FL.

[9]Associate Professor of Molecular and Medical Genetics, OHSU Knight Cancer Institute, Oregon Health & Science University

[10]Department of Medicine, University of California San Diego, La Jolla, CA

[11]Biology Department, Boston University, Boston, MA

[12]Cancer Biology and Evolution Program, Moffitt Cancer Center, Tampa, FL

[13]Precision One Health Initiative, Department of Pathology, University of Georgia College of Veterinary Medicine, Athens, GA 30602

Conflict of interest disclosure statement: The authors declare no potential conflicts of interest.

[14]Department of Pharmacology, Department of Biomedical Informatics, University of Colorado Cancer Center, University of Colorado Anschutz Medical Campus, Aurora, CO, 80238

[15]Department of Computational and Systems Biology, Department of Physics and Astronomy, UPCI Hillman Cancer Center, University of Pittsburgh, Pittsburgh, PA 15260

[16]Department of Oncology, Lombardi Comprehensive Cancer Centre, Georgetown University Washington, DC, 20007

[17]Department of Biological Sciences, Middle East Technical University, Universiteler Mah. DuFLupınar Bulvarı 1, 06800 Cankaya, Ankara, Turkiye

[18]Department of Neurosurgery, Stanford Medical School, Stanford, CA

[19]Biological Sciences Division, Pacific Northwest National Laboratory, Richland, WA

[20]Sage Bionetworks, 2901 Third Ave., Suite 330, Seattle, WA 98121.

[21]Sage Bionetworks, 2901 Third Ave., Suite 330, Seattle, WA 98121.

[22]Center for Cancer Systems Biology (CCSB), Dana-Farber Cancer Institute, Boston, MA 02215

[23]Section of Hematology/Oncology, Department of Medicine, University of Chicago, 5841 S Maryland Ave, Chicago, IL 60637, USA

[24]Department of Neurosurgery, Stanford University, Stanford, CA 94305

[25]Herbold Computational Biology Program, Vaccine and Infectious Disease Division, Fred Hutchinson Cancer Research Center, Seattle, WA 98109; Program in Immunology, Translational Science and Therapeutics Division, Fred Hutchinson Cancer Research Center, Seattle, WA 98109

[26]Department of Bioengineering, University of California, Los Angeles, Los Angeles, CA 90024, USA

Abstract

Cancer systems biology seeks to understand how cancer arises as a system of interconnected molecules, cells, and tissues, with the goal of understanding, predicting, and controlling the disease. In the last decade, the field has rapidly grown as advances in experimental, computational, and analytical technologies have improved our ability to capture and recapitulate the complexities of cancer at multiple scales. However, the field’s promise to understand how specific molecular changes give rise to altered cancer outcomes remains incompletely fulfilled. Fortunately, an opportunity exists to accelerate progress by better coordinating modeling and data-gathering efforts across the cancer systems biology community. This will create the foundation for building accurate, multiscale cancer models that can better predict and identify improved therapeutic interventions. Here, we outline some of the current challenges in cancer systems biology research, how they can be addressed, and actions that the community can take to accelerate progress in the field.

Overview of a systems approach to cancer research

A system is any collection of interacting components that generates an output from a set of definable inputs. Systems can range from small scale, such as protein-protein interactions in a molecular pathway, to large scale, such as interacting populations of organisms. Biological systems are inherently complex, obscuring the role of individual components. Molecular-level perturbations (e.g., mutations or drug treatment) manifest as phenotypic changes (e.g., disease or health) across particularly vast spatiotemporal scales, from the atomic to the macroscale. The non-linear and recursive nature of many biological processes further complicates understanding these relationships, leading to non-intuitive emergent properties even in simple systems (1). Thus, one of the foundational goals in systems biology is to overcome these challenges so that we can predict how a system's behavior will respond to a perturbation based on the system's composition and context.

Cancer is a complex disease in which genetic and environmental perturbations combine to give rise to aberrant cell proliferation. Natural defenses that protect against inappropriate cell proliferation and migration are frequently compromised in cancer. The multisystem failures underlying most cancers would seem to make the disease an ideal target for systems biology. Indeed, multiple programs have been established in cancer systems biology and implemented across academic institutes, funding agencies, and commercial research.

Although considered an emerging field, cancer systems biology has already had substantial successes. The resulting systems-level studies have generated several significant tools and discoveries concerning the tumor microenvironment, mechanisms of drug resistance and sensitivity, and cancer metastasis. These include the development and implementation of predictive and adaptive therapies, to control tumor progression by acknowledging and modeling competition between cells in the tumor ecosystem (2,3). Computational models of cell component hierarchy have identified common network hubs where tumor mutations converge, leading to the identification of drug sensitivities in cancer cells (4). Within tumor immunity, the community has identified essential responses within tumor cells for effective response to immunotherapy (5). Each of these advancements has demonstrated the value of an integrative approach, as well as the synthesis of experimental and computational expertise. Indeed, cancer systems biology continues to explore how to integrate large datasets with computational models to understand a complex disease and predict successful treatments, which will continue to become ever more important given the expanded diversity and scale of data available.

Despite several successful applications of systems biology to cancer research, the field has yet to realize its full potential, largely due to the problem of integrating systems across scales (6). Predicting how molecular perturbations propagate through molecular, cellular, tissue, organ, and whole-body levels to impact health necessitates models that can bridge these scales. Furthermore, cancer arises from diverse cell types and causes within the context of different people, meaning that a one-to-one correspondence between events at different scales and disease does not always exist. Because cancers evolve in response to tumor location and treatments, they are continuously moving targets. Current cancer systems biology models, however, tend to be focused on a single level of complexity and are difficult

to integrate (7). This restricts the ability of the systems biology community to build on the work of others and develop holistic models of the disease.

A roadmap to the future

Biological research has traditionally focused on a single scale, whether it be at the molecular or physiological level. This has been driven by the inherent complexity of biological systems and the limited experimental technologies available in the past. Recent advances in high-throughput multi-omics and imaging technologies have greatly expanded the availability of high-quality data; however, there has not been a corresponding improvement in the modeling frameworks needed to integrate and understand this data. Instead, most attempts to build multi-scale models are restricted to a small part of the continuum from molecular details to patient outcomes (8).

Integrating data across scales is essential for the future of cancer systems biology despite the immense challenge it represents. Other scientific fields have also faced similar challenges, especially in building digital representations of complex systems. How these fields have successfully addressed the issues of complexity and scale are instructive. For example, the current large scale earth system models started from basic weather models that evolved over time to include ever-increasing spatial and temporal details (9). Their successful evolution over the last 50 years required international collaborations, data and modeling standards, as well as rigorous evaluation and validation of model extensions. The highly accurate weather predictions that are currently generated by these models are a testament to the success of this approach.

The development of accurate models of cancer that could predict the impact of therapeutic interventions is likely to require a similar, long-term effort by the international cancer systems biology community. Fortunately, advances in computational technologies, such as artificial intelligence/machine learning (AI/ML) promise to accelerate this work. Nevertheless, the scope and scale of the effort will still require the development of appropriate computational frameworks, the participation of the cancer systems biology community, and the support from its sponsors.

Because biological processes at many different spatial-temporal scales are involved in cancer progression, linking models across scales is essential. This requires accepted community standards in data, models, and software. Data and modeling standards are critical for not only linking models across scales, but for integrating the contributions of the entire research community. To achieve this, we propose the following “roadmap” for cancer systems biology:

1. Establish multiscale frameworks for building cancer models
2. Create modeling standards compatible with the multiscale framework
3. Identify the data and metadata needed to build and parameterize multiscale models

4. Promote data-generating projects that are targeted at building and parameterizing community standard models
5. Identify specific cancer model systems that can serve as benchmarks for progress, especially with respect to clinical validation
6. Promote education for scientists to bridge the model building-data generation divide

This roadmap is not all-inclusive but addresses the major issues that we see as hindering progress in the field. The major challenges the community faces as well as our proposed actions are summarized in Table 1.

Establish multiscale frameworks for building cancer models

Models can help predict how events at one scale influence those at another. Multiscale linkage is especially critical in cancer systems biology because cancer is caused by disruption at the genetic level that negatively impacts cell behavior at the organismal level (10). Indeed, cancer outcomes depend not only on the aberrant behavior of the cancer cell itself, but on the response of the homeostatic systems in the body, which evolve over the course of the disease. Thus, predictive models of cancer need to include interactions across large temporal and spatial scales.

Unfortunately, most models of cellular processes involved in cancer focus on a single spatiotemporal scale because of technical limitations of current approaches. For example, molecular models simulate processes on the nanometer and nanosecond scale but would be impractical for investigating biochemical reactions at the network level. Similarly, reaction networks are usually represented as well-mixed dimensionless systems that represent an “average cell” behavior rather than the rich mosaic of cells that comprise a population.

Multiscale models that bridge the gap between molecular, cellular, and tissue-scale processes would be particularly useful for understanding how perturbations propagate across different spatiotemporal scales (11). Unfortunately, most current models were not designed to work with other model types. For example, multiple cancer systems biology studies have focused on understanding how the collective behavior of tumors arises from cells exhibiting distinct (epi)genetic and phenotypic variations (12). These studies typically use computational models together with multiple types of high-throughput data to correlate molecular state (“signatures”) and cell behavior. However, these models do not explicitly connect the two mechanistically. Recent work that links genetic changes to the assembly of protein complexes and their roles in regulatory networks and cell processes is an example of an approach to link processes at different scales, even in the absence of mechanistic details (13).

Successful multiscale models have been constructed in other fields, such as in cardiovascular research and earth systems science (14). The building of these complex models required a simulation framework that can link both mechanistic and data-driven models that are built around community standards. Several different frameworks have been proposed for linking multiple biological submodels in a modular fashion (11,15), but they need to address issues

of error propagation and interoperability of their subcomponents and data (14). Building data-driven and mechanistic hybrid models using current approaches is also challenging (11). A generally accepted standardized framework for linking the components of multiscale modeling would accelerate progress in the field by allowing independent groups to combine their work.

The need to include biological heterogeneity

Any framework for building multiscale biological models must adequately represent biological heterogeneity. Cancers exhibit heterogeneity at multiple scales—both within a tumor and between patients. This can arise from distinct genotypes or microenvironment-driven phenotypes and usually depends on tumor type. These heterogeneities introduce significant challenges in discovering cancer mechanisms and the development of effective treatments, because they greatly expand both the genotypic and phenotypic space that must be considered (16,17). Modeling the heterogeneity of cancer presents is both a challenge and an opportunity for system biology approaches to shine through tackling problems not otherwise solvable with traditional simplified experimental models.

One of the key questions in cancer systems biology is whether the complexity of genetic heterogeneity can be simplified by associating a range of genetic alterations with a common biological function. One approach used to simplify the relationship between heterogenous genotypes and specific phenotypes has been to map genetic changes to their underlying network (18–20). Recent studies have also shown that molecular-level interaction data can be integrated into a hierarchical proteomic framework, suggesting that tumor mutations converge on common assemblies across different scales (13). Dataset coupling approaches can statistically link datasets collected at different scales or with different levels of heterogeneity (21–24), but how to incorporate this information into hybrid models and how to map phenotypes to multiple genotypes is still unresolved.

One approach to reduce the complexity of cancer models is to build “coarse-grained” models that encapsulate molecular complexity by creating a level of abstraction within a general modeling framework (25). Coarse-grain modeling would also allow the inclusion of cell population heterogeneity within multiscale modeling (26), providing a more realistic representation of biological systems. However, integrating the diverse sets of data needed to build such complex models would be challenging and require active collaboration between modelers and experimentalists working at different scales. This, in turn, would require the establishment of standardized formats and ontologies that enable the exchange of different types of information and models.

Linking AI/ML with mechanistic modeling

The most common types of models used in cancer systems biology are data-driven because they do not require mechanistic insights to be useful. However, these types of models are only accurate for the conditions similar to those in which their underlying data were generated (14). In contrast, mechanism-based models can generate predictions outside of their training data by leveraging biophysics and known biological processes, but their construction requires detailed prior knowledge of the system.

There are additional benefits of bottom-up mechanistic modeling: First, mechanistic modeling can achieve higher predictive accuracy when the data are limited—and this is often the case with systems biology. For example, a simple mechanistic model of receptor-mediated endocytosis could recapitulate the results of thousands of measurements (27). It could also predict the mechanisms by which receptor-level mutational changes alter cell-level receptor behavior (28). A machine learning/deep learning (ML/DL) application could not be guaranteed to reach the same conclusions unless the pre-existing knowledge and appropriate well-defined features were incorporated into the framework. Second, ML/DL models might not provide full observability in the feature space for the system in question. Mechanistic understanding allows tailored feature selection and imposing biologically realistic, explicit constraints, leading to oftentimes more streamlined models that tend to show higher generalization capabilities. An instructive example can be found in the field of computational chemistry, where mechanistic modeling, such as molecular dynamic simulations, remains largely superior to the DL approaches for robust and stable dynamics predictions (29). Third, many stakeholders (e.g., research scientists, clinicians, public health administrators) prefer interpretable model representations to the proverbial “black box” of DL so that the basis of predictions can be reconciled with prior knowledge.

Even in its most mature and successful applications, such as AlphaFold (30), DL is mostly a predictive activity, with moderate ability to explain the answers and little to no interpretability. This is problematic for studying human diseases such as cancer. Without an explicit mechanistic understanding of how a complex network of molecular relationships gives rise to either health or disease, it will be difficult to engineer new therapies while avoiding predictions arising from spurious correlations. Emerging ML approaches seek to address interpretability, such as attention-based models with saliency maps, the recent trend towards graphical neural network (GNN)-probabilistic graphical model synthesis, or hierarchical models constrained by prior knowledge (4,31–35).

Regardless of their limitations, applications of AI/ML/DL approaches in cancer systems biology are likely to increase, if only because of a rapid growth in their availability. However, we advocate for a deeper integration of AI/ML/DL approaches with mechanism-driven modeling, perhaps by serving as surrogate models for well-defined mechanistic models that are computationally costly. Underlying biophysical processes should be incorporated into AI-based approaches to providing constraints to better leverage limited patient-specific data and potentially making patient-specific predictions that are more readily interpretable (36,37). Generally, successful applications of combined approaches have leveraged the compartmentalization of multiscale models, with different modeling paradigms handling different scales (38). Although the flexibility of an AI/ML model has been used to augment the rigid scaffolding of a mechanistic model, these models are typically developed with bespoke implementations that would be challenging to formalize into a sharable form (35).

Create modeling standards compatible with the multiscale frameworks

We posit that the most significant impediment in both the development and application of systems biology to cancer is the lack of community standards and resources that would

allow facile sharing of data and models between research groups, particularly for the construction of multiscale models. The lack of such standards and resources results largely from the complexity of the problem itself. Models have traditionally been used to address a specific research problem rather than build a scalable model and thus few research groups are working on more than just a small part of the overall problem. Efforts to promote multiscale modeling by the NIH, for example, have resulted in a large variety of different models that cross only a few scales without any restrictions on the use of different data or modeling standards (https://www.imagwiki.nibib.nih.gov). The result is a collection of models that can rarely be linked together. The effective solution to this problem will require the effort of the entire systems biology community to develop common standards and reusable software resources.

FAIR (i.e., Findable, Accessible, Interoperable, and Reusable) principles have brought attention to the need for preserving and making experimental data accessible. However, an equivalent for modeling does not currently exist. At the Interagency Modeling and Analysis (IMAG) meeting in 2023, a proposal was made to develop a similar set of principles for modeling. These principles focus on model credibility, understandability, reproducibility, and extensibility (CURE). A similar set of principles could also be developed for ML models. Although achieving understandability may be challenging for many model types, the community could develop general guidance for model evaluation, validation, and comparison.

There is currently no universal modeling framework optimal for all situations, but standards can be established for methods that are widely supported by the field. For example, the advent of SBML and associated software tools greatly enhanced mechanistic modeling at the cellular level. In fact, more than half of all models are currently published using SBML (Systems Biology Markup Language); models described with SBML also receive more citations than those distributed in other forms (39). Some areas of modeling remain difficult to reproduce and reuse, such as rule-based models; proposals exist for standardizing such models but have yet to be implemented (40). Another hurdle to model reuse is model annotation, which is metadata that allows a machine to unambiguously identify the components and processes in a model. This is particularly important as models become larger or if there is a need to carry out meta-analysis or use models as training data for ML algorithms. However, annotating the components of signaling networks prevalent in cancer studies is extremely difficult because of the plethora of components and modification states. Reproducibility of ML algorithms also remains a significant problem; there currently is no long-term, simple, or general solution. Standards that would promote connections between models at different scales still need to be developed (41–43).

Identify the data and metadata needed to build and parameterize multiscale models

Developing predictive, multiscale models of cancer will likely require vast amounts of complex, high-quality data available in easily translatable and adaptive formats. This realization has spurred increased interest in making data FAIR. The National Library

of Medicine (NLM) suggests that the first minimal requirement for datasets to be reusable across disciplines is a license that describes who can use the data and for what purpose (44). An open license is important for appropriate and ethical use of data in generating new and improved models (https://creativecommons.org/share-your-work/cclicenses/). Restrictive licenses can often limit the reuse of data, even if data is shared widely.

FAIR principles have generally focused on data sharing and "reuse" within the specific discipline or group that generated the data. More attention will need to be paid to the types of metadata and data conversion methods needed to make data reusable across disciplines and software platforms. At minimum, clear provenance should be provided to preserve the context of data collections even beyond the original discipline (44). Providing metadata that can be mapped to various available standards can also help preserve data collection context. Such provenance should include metadata that document all modifications to the cell line or model organism, in addition to all modifications that happen to the data after collection. These additional metadata must be in a machine-readable format to help preserve the provenance of both experiment and data, enabling re-use of a dataset outside the associated publication or discipline.

While processed and analyzed data can generally be reused in a qualitative fashion, for quantitative reuse, raw data that support re-processing will need to be shared along with metadata specifying the context of data collection. We recommend updating the concept of reusability to accommodate integration of shared data with newly generated data or allowing updated extraction of information using newly developed tools. This is key for facilitating data reuse in modeling. By enabling data to be re-processed, normalized, calibrated, and integrated, it can more easily be integrated into rapidly evolving modeling frameworks.

One key issue in data and model reuse is how to incentivize researchers to use best practices. A major reason why sharable, high-quality data is a secondary consideration in scientific publications is that the reward system in science largely focuses on the number of publications or journal citation factor rather than post-publication data use. Facile ways to track data reuse would allow the impact of a dataset to be determined.

Developing reusability standards for data and modeling will increase their value and impact, as well as help break down the current silos among different scientific disciplines involved in systems-level research. This will provide new opportunities for discoveries, collaborations, and applications. Systems biology has been seen as integrating modeling and experiments, but to accomplish this, we also need to focus on the cooperation between and among modelers and experimentalists.

Promote data-generating projects that are targeted at building and parameterizing community standard models

When large data-generating consortia are assembled, the systems biology community should be asked to provide input on which data types are most useful and what metadata are needed to support systems biology models. Currently, most data in the cancer field gathered

by large scale consortia are descriptive and intended to populate online data resources (45,46). Unfortunately, most traditional experimental approaches do not generate data that can be directly used in modeling. Most biological data are qualitative, such as the absence or presence of a marker (e.g., cellular staining) or the relative change in the abundance of cell components (e.g., western blots, RNA-Seq, proteomics). Mechanistic models in systems biology are usually based on approaches from chemistry and physics that require quantitative measurements of the number of molecules or reaction rates. Recognizing this data gap, several methods have been proposed to integrate quantitative and non-quantitative datasets (47–50). Specifically, these approaches recognize that all data, whether quantitative or not, contain intrinsic information about cellular processes that can be extracted using various techniques. A better approach, however, is to encourage practices that generate data that can be directly used in models. This requires bringing experimentalists, modelers, data scientists, and clinicians together to better understand what data is needed, what samples are available and how they can be collected and processed for optimal results.

Quantitative data at multiple scales are needed to build mechanistic cancer models. Gathering data for building models should be an explicit aim of at least a subset of large-scale data gathering enterprises. At the lowest level, whole genome/exome data are needed to quantify DNA mutations, structural alterations, copy number variations, and chromosomal rearrangements in cancer patients (51,52). At the next level, transcriptomics data (e.g., RNA-seq) quantify the expression of specific genes under various conditions and infer gene dependencies and complex regulatory mechanisms, but these data need to be normalized and calibrated. Single-cell sequencing provides data at the level of individual cells, allowing the measurement of cancer heterogeneity both at the cellular level and across spatial and temporal dimensions. These data can also reveal heterogeneous non-cancer cell types within the tumor microenvironments that interact with cancers and impact tumor behavior following different treatments. Tumors comprise many cell types in addition to cancer cells, including fibroblast, T cells, and myeloid cells. Ideally, data should also be collected on both the abundance and localization of these diverse cell types in the so-called tumor microenvironment (5). Spatial transcriptomics data could then allow researchers to map gene expression within the spatial architecture of tissues, which would enable modeling of cancer cell behavior within the context of its immediate microenvironment.

Because cell functions are primarily executed by their proteins, the quantitative abundance profile of proteins and their post-translational modifications are particularly useful for modeling the signaling and metabolic pathways critical in cancer development (53). In addition, technologies such as affinity-purification mass spectrometry (AP-MS), size-exclusion chromatography (SEC), and cross-linking mass spectrometry (XL-MS) can provide information on the physical and structural interactions between proteins. Protein-protein interactions, measured by both direct and indirect means, provide critical information on how mutations disrupt cellular circuits and macromolecular interaction networks. While more difficult to capture experimentally, several other types of measurements, such as extracellular matrix composition and metabolic state (54,55), could directly inform modeling across scales, as these processes influence cell-to-cell communication.

The potential for data collection is endless, while the capacity to make new measurements is limited, and not all types of measurements are possible or ethical. Mathematical models can be used to identify specific measurements of high value, but most studies are currently designed around a technology of interest and restricted by the available budget. Model-driven data collection efforts should focus limited resources on the highest value measurements. This approach is likely to be especially important for data integration across scales.

Each of the diverse data types described above can provide important insights into cancer mechanisms from different perspectives and scales. However, because the underlying technologies were developed by different disciplines with different scientific goals, each has unique metadata standards. To build integrative models from these types of data, common data representation standards must be defined, codified, and supported.

Identify specific cancer model systems that can serve as benchmarks for progress, especially with respect to clinical validation

The current paradigm to model the dynamics of a biological process typically involves observation and measurement of the processes, typically through experiments (e.g., single-cell sequencing, physiological response) at a single scale. This is followed by formulation of a mathematical model to explain the process, calibration of the model to experimental data, and validation of model predictions. The calibration and validation of the model are done within the range of the original experimental conditions. This is also the case with data-driven models where data are gathered over a narrow set of conditions and used to predict system performance over similar conditions. When multiple mechanistic models are linked or combined with data-driven models to create hybrid multiscale models, validation and identifying sources of error become much more difficult. Thus, developing robust approaches to identify and localize modeling errors and how they propagate through multiple scales will be essential.

An additional problem is that quantitative measurements of biological processes are sparse, usually generated by indirect quantification techniques such as antibody- or fluorophore-based methods, ordinal data such as western blots, or nominal data derived from staining or clustering methods. These measurements are typically a proxy for the data of interest and thus must be normalized and calibrated before they can be used in modeling. One approach is statistical, where both quantitative and non-quantitative data are used to estimate model parameters and their uncertainties (47,48). Another is to use analytical methods to directly estimate model parameters, such as protein distribution (56) and modification stoichiometry (57).

Once data are generated, they must be converted into a useful form for modeling. For example, ion intensities from peptides measured by mass spectrometry need to be converted to protein concentration. All such conversions require a set of underlying assumptions, which can vary among research groups and over time (58). In addition, models often have parameters that cannot be directly measured and thus must be inferred from the primary data; for example, the fraction of proteins in a specific cell type that are activated. This is

almost always done on an ad hoc basis, from various on-hand data, adding to the difficulty of reproducing results. Unfortunately, there is currently no consensus in the field as to the best way to estimate different types of parameters for different types of models.

For multiscale models to be useful for clinical use, they will need to be rigorously validated. Unfortunately, there are no accepted standards for validating multi-scale cancer models. There are commonly used cell types that are used in cancer research (59), but none have been proposed by the cancer systems biology community as a multi-scale modeling standard. A major problem is that these cell lines evolve over time, and their functional state is very context dependent. Building and validating multiscale models from components built under different contexts is a major challenge but must be addressed by the cancer systems biology community. We suggest that the community adopt a limited number of cancer model systems that operate at different scales to serve as testbeds for developing and validating multiscale models of cancer. An ideal model system would be one that can be transplanted into syngeneic hosts for replicable *in vivo* studies, while also allowing propagation *in vitro* as organoids. Given the importance of genetic heterogeneity, it would be important for the community to either identify genetically engineered models in which a variety of driver mutations can be introduced, or a panel of patient-derived models that represent the range of the disease. Given limited resources, the community might focus on a particular cancer as a focus, preferably one that has shown poor responses to existing therapies and therefore has the potential for transformative advancements in treatment.

The widespread acceptance of predictive, multiscale models of cancer will require a rigorous assessment of their accuracy and an understanding of how to best improve them. The scale of the challenge is such that it will take the cooperative work of the entire systems biology community to succeed. But what this “cooperation” looks like is currently unclear. Competitions, such as Critical Assessment of Structure Prediction (CASP), played an important role in accelerating protein structure determination, and a similar strategy could be applied to cancer systems biology. While DREAM Challenges (Dialogue for Reverse Engineering Assessment and Methods) have made some progress in this direction, more targeted and varied competitions for cancer systems biology could be organized, such as predicting the effect of a given set of genetic mutations on cell proliferation or migration (60–63). Predicting such functional outcomes would demonstrate the potential utility of the competing models, determine the most effective modeling frameworks and data types that can advance the field, and encourage standards for data and model evaluation and sharing.

Promote education for scientists to bridge the model building-data generation divide

Systems biology research requires the engagement of a wide range of modelers alongside clinicians, biological researchers, and technologists. This cooperation is even more essential in the framework of cancer research, given the nonlinearity resulting from heterogeneous cell populations and the tumor microenvironment. Because of the enormous complexity of cancer itself, a wide range of different technologies and modeling approaches will be needed. The field will need modelers focused on both targeted small-scale and systems-scale

representations, as well as a wide variety of data generators supported by funding agencies that understand the value of making data useful outside of their own focus area. It will also require the engagement of early career scientists, who traditionally perform the experiments and drive innovations.

Train and incentivize early career scientists

The training of experimental cancer biologists has traditionally focused on reductionist approaches with minimal exposure to the importance of bridging scales. Modules in graduate education that focus on systems biology would be beneficial and likely thought-provoking to cancer biology students. Additionally, students require training in modern experimental techniques that emphasize complex, systems-level data analysis and how to test computationally driven hypotheses. The next generation of experimentalists will likely have access to sophisticated bioengineering and nanotechnology tools, as well as enhanced approaches for (epi)genetic and spatial transcriptomic profiling and novel animal models. Their early training needs to include a focus on collecting and formatting data to test systems-level models and allow reusability across disciplines. It should also include a multiscale perspective on how processes at each level give rise to higher-level processes.

Even if students are exposed to systems biology at an early stage, they often struggle with obtaining usable data sources. To address this issue, multidisciplinary conferences and workshops should be supported to allow early career scientists from different backgrounds to network and interact with potential mentors and collaborators. Events hosted by NCI, including the annual junior investigator meeting, are excellent examples of such events (64). In these meetings, early career physicists, biologists, computer scientists, clinicians, and patient advocates meet annually to discuss progress in oncology research. Direct interaction between trainees and mentors are facilitated through small group discussions. Increased funding and support for additional conferences and workshops would empower the next generation of scientists with the necessary tools to navigate the ever-evolving landscape of oncology research and contribute to more effective and personalized cancer care for patients.

Diversity of scientific talent and perspective as well as a diversity of quantitative data are crucial to drive innovation in cancer systems biology. Early career scientists also face unique challenges in building a career in cancer systems biology. In the current academic publishing environment, there is a tendency to prioritize reductionist approaches over systems-level models and insights. There is also the issue of establishing scientific contributions on large multi-author manuscripts. Promotion committees will need to value collaborations between clinical, experimental, and computational researchers. As collaborative studies with several lead and senior authors become more common, they need to be accepted as a central and valuable output of young scientists' careers. Transition grants that understand the demands of contemporary research will be imperative for identifying and providing incentives to early career scientists to establish and excel in team science.

Next steps and conclusions

Despite the idea of cancer systems biology being a multi-institutional, community-based effort to understand and effectively treat cancer, the reality falls short of this vision.

The reasons include the complexity of the problem itself, its multiscale nature, and institutional incentives that remain largely aligned with single investigator, reductionist science. Currently, advancements occur as isolated components of a larger problem. Realizing the full impact of the field will require assembling the puzzle pieces of technical solutions, investigator talent, and mechanisms of collaboration into a coordinated effort (Fig. 1). Solutions are also likely to be complex but are not insurmountable. The urgency of the issue requires a concerted effort to find a solution.

At its heart, cancer systems biology is community-based team science. This must be recognized by individual investigators, research institutions, and funders. Programs such as the NCI Cancer Systems Biology Consortium primarily served to provide financial support, but they also organized meetings that supported trainee development, the exchange of ideas, and collaboration. The only long-term sustainable model will be to continue these activities through an independent organization. The cancer systems biology community comprises a wide range of disciplines developing unique experimental and computational approaches and technologies. Efforts to engage the community must necessarily engage all these disparate groups and disciplines. If sustainable, the Systems Approaches to Cancer Biology meetings, organized by the Association of Cancer Systems Biologists, may be a home for these community-building and maintaining efforts (65).

The field needs to come together and define specific, testable, currently unachievable grand challenges that would represent major advancements in cancer systems biology. These undertakings do not require closed datasets; in some cases, solving them might necessitate the collection of new experimental data and knowledge. For instance, these challenges may include predicting how perturbations in one cell type would affect the others within the tumor microenvironment or forecasting how a tumor evolves based on information at the initiation of treatment. Focusing the field on common goals would facilitate iterative advancements and incentivize researchers to develop improved standards for data inter-operability and reuse.

The community should work together to develop blinded benchmarks for testing to drive success in solving these challenges. The AI community has shown the value of open-source benchmarks of increasing difficulty for advancing the frontier of models. These benchmarks have proven their value time and again by showing that researchers can even fool themselves without objective assessment. In contrast to the approach taken by competitions such as DREAM (66), these benchmarks would be designed with the intention that they are currently not solvable with existing technologies or knowledge. Thus, either new modeling advancements, experimental characterization, or both, might be necessary to fulfill the tasks.

In addition to focusing the scientific community, such challenges could clarify the role of funding agencies in supporting cancer systems biology. Funding should support the development of modeling and data standards and benchmarks that would objectively and efficiently demonstrate progress in the field. Competitive granting mechanisms could be developed that bring together teams centered around well-defined problems rather than a diffusely defined field. Finally, funding agencies could motivate progress and external investment toward major advancements through large prizes to incentivize teams to pass

certain thresholds in performance. Similar programs, such as the Xprize competitions (https://www.xprize.org/competitions), have repeatedly served to focus community efforts and inspire major advancements.

The protein structure prediction field and AlphaFold together serve as an excellent example of the value provided by a grand challenge (30). The field provided a standardized task—predicting protein structure from sequence—with clear, quantifiable metrics for success. This undertaking remained effectively unsolved for decades and required both new experimental data and mechanistic understanding of protein folding. Success also depended on well accepted data and modeling standards around which the field could focus its efforts and benchmark progress (67). At its core, a challenge would serve as a common reference by which to define the field and allow members to build on advancements from one another. Without this common reference frame, and objective measures of progress, the field risks languishing in disparate ideas, techniques, and goals.

Progress will still take time. Despite computational advances in AI and modeling, success will depend on an enormous amount of good data, much of which is likely to require technologies that do not yet exist amidst a competitive funding landscape. Models thus need to be adaptable to adjust to the ever-changing data landscape and the evolvability of cancer itself. Despite these challenges, we feel that embracing a systems-level perspective of cancer is the best way to exploit the accelerating pace of scientific technologies to find cures.

Acknowledgments

This work is supported by the National Cancer Institute of the National Institutes of Health (NIH) under award numbers U01CA227544, U01CA250044, U01CA232161, U01CA231978, U24CA274494 [Multi-Consortia Coordinating (MC2) Center], U01CA250046, OT2OD032742, U54CA274502, 1U01CA253540, R01GM148525, [JSPS Grant-in-Aid for Early-Career Scientists No.23K15972], R01LM013138, U01CA232216, U01CA243075. This work is partially supported by an Mark Foundation for Cancer Research Emerging Leader Award to A.S.M. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

We are grateful to Lily Dell for creating the artwork.

References

1. Jensen HJ. Complexity science: the study of emergence. Cambridge, United Kingdom; New York, NY: Cambridge University Press; 2023.
2. Gatenby RA, Silva AS, Gillies RJ, Frieden BR. Adaptive Therapy. Cancer Res. American Association for Cancer Research (AACR); 2009;69:4894–903. [PubMed: 19487300]
3. Liu Z, Zhang J, Hong L, Nie Q, Sun X. Multiscale mathematical model-informed reinforcement learning optimizes combination treatment scheduling in glioblastoma evolution. Sci Adv. 2025;11:eadv3316. [PubMed: 40779623]
4. Zhao X, Singhal A, Park S, Kong J, Bachelder R, Ideker T. Cancer Mutations Converge on a Collection of Protein Assemblies to Predict Resistance to Replication Stress. Cancer Discov. American Association for Cancer Research (AACR); 2024;14:508–23. [PubMed: 38236062]
5. Hickey JW, Agmon E, Horowitz N, Tan T-K, Lamore M, Sunwoo JB, et al. Integrating multiplexed imaging and multiscale modeling identifies tumor phenotype conversion as a critical component of therapeutic T cell efficacy. Cell Syst. 2024;15:322–338.e5. [PubMed: 38636457]
6. Ford Versypt AN. Multiscale modeling in disease. Curr Opin Syst Biol. 2021;27:100340. [PubMed: 35310906]

7. Cristini V, Lowengrub J. Multiscale modeling of cancer: an integrated experimental and mathematical modeling approach. Cambridge; New York: Cambridge University Press; 2010.
8. Qu Z, Garfinkel A, Weiss JN, Nivala M. Multi-scale modeling in biology: how to bridge the gaps between scales? Prog Biophys Mol Biol. 2011;107:21–31. [PubMed: 21704063]
9. Prinn RG. Development and application of earth system models. Proc Natl Acad Sci. Proceedings of the National Academy of Sciences; 2013;110:3673–80.
10. Deisboeck TS, Wang Z, Macklin P, Cristini V. Multiscale Cancer Modeling. Annu Rev Biomed Eng. 2011;13:127–55. [PubMed: 21529163]
11. Agmon E, Spangler RK, Skalnik CJ, Poole W, Peirce SM, Morrison JH, et al. Vivarium: an interface and engine for integrative multiscale modeling in computational biology. Valencia A, editor. Bioinformatics. 2022;38:1972–9. [PubMed: 35134830]
12. Meyer AS, Heiser LM. Systems biology approaches to measure and model phenotypic heterogeneity in cancer. Curr Opin Syst Biol. 2019;17:35–40. [PubMed: 32864511]
13. Zheng F, Kelly MR, Ramms DJ, Heintschel ML, Tao K, Tutuncuoglu B, et al. Interpretation of cancer mutations using a multiscale map of protein systems. Science. 2021;374:eabf3067. [PubMed: 34591613]
14. Committee on Foundational Research Gaps and Future Directions for Digital Twins, Board on Mathematical Sciences and Analytics, Committee on Applied and Theoretical Statistics, Computer Science and Telecommunications Board, Board on Life Sciences, Board on Atmospheric Sciences and Climate, et al. Foundational Research Gaps and Future Directions for Digital Twins [Internet]. Washington, D.C.: National Academies Press; 2024 [cited 2025 July 19]. Available from: https://www.nap.edu/catalog/26894
15. Masison J, Beezley J, Mei Y, Ribeiro H, Knapp AC, Sordo Vieira L, et al. A modular computational framework for medical digital twins. Proc Natl Acad Sci U S A. 2021;118:e2024287118. [PubMed: 33972437]
16. Burrell RA, McGranahan N, Bartek J, Swanton C. The causes and consequences of genetic heterogeneity in cancer evolution. Nature. 2013;501:338–45. [PubMed: 24048066]
17. Dagogo-Jack I, Shaw AT. Tumour heterogeneity and resistance to cancer therapies. Nat Rev Clin Oncol. 2018;15:81–94. [PubMed: 29115304]
18. Kim Y-A, Cho D-Y, Przytycka TM. Understanding Genotype-Phenotype Effects in Cancer via Network Approaches. Karchin R, editor. PLOS Comput Biol. 2016;12:e1004747. [PubMed: 26963104]
19. Chuang H, Lee E, Liu Y, Lee D, Ideker T. Network-based classification of breast cancer metastasis. Mol Syst Biol. 2007;3:140. [PubMed: 17940530]
20. Hofree M, Shen JP, Carter H, Gross A, Ideker T. Network-based stratification of tumor mutations. Nat Methods. 2013;10:1108–15. [PubMed: 24037242]
21. Tan ZC, Meyer AS. The structure is the message: Preserving experimental context through tensor decomposition. Cell Syst. 2024;15:679–93. [PubMed: 39173584]
22. Kelley DZ, Flam EL, Izumchenko E, Danilova LV, Wulf HA, Guo T, et al. Integrated Analysis of Whole-Genome ChIP-Seq and RNA-Seq Data of Primary Head and Neck Tumor Samples Associates HPV Integration Sites with Open Chromatin Marks. Cancer Res. 2017;77:6538–50. [PubMed: 28947419]
23. Acar E, Papalexakis EE, Gürdeniz G, Rasmussen MA, Lawaetz AJ, Nilsson M, et al. Structure-revealing data fusion. BMC Bioinformatics. 2014;15:239. [PubMed: 25015427]
24. Stein-O'Brien GL, Arora R, Culhane AC, Favorov AV, Garmire LX, Greene CS, et al. Enter the Matrix: Factorization Uncovers Knowledge from Omics. Trends Genet. 2018;34:790–805. [PubMed: 30143323]
25. Andrews SS, Wiley HS, Sauro HM. Design patterns of biological cells. BioEssays. 2024;46:e2300188. [PubMed: 38247191]
26. Ruiz-Martinez A, Gong C, Wang H, Sové RJ, Mi H, Kimko H, et al. Simulations of tumor growth and response to immunotherapy by coupling a spatial agent-based model with a whole-patient quantitative systems pharmacology model. PLoS Comput Biol. 2022;18:e1010254. [PubMed: 35867773]

27. Lund KA, Opresko LK, Starbuck C, Walsh BJ, Wiley HS. Quantitative analysis of the endocytic system involved in hormone-induced receptor internalization. J Biol Chem. 1990;265:15713–23. [PubMed: 1975591]
28. Chang CP, Lazar CS, Walsh BJ, Komuro M, Collawn JF, Kuhn LA, et al. Ligand-induced internalization of the epidermal growth factor receptor is mediated by multiple endocytic codes analogous to the tyrosine motif found in constitutively internalized receptors. J Biol Chem. 1993;268:19312–20. [PubMed: 8396132]
29. Stocker S, Gasteiger J, Becker F, Günnemann S, Margraf JT. How robust are modern graph neural network potentials in long and hot molecular dynamics simulations? Mach Learn Sci Technol. 2022;3:045010.
30. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. Nature. 2021;596:583–9. [PubMed: 34265844]
31. Yu Y, Chen J, Gao T, Yu M. DAG-GNN: DAG Structure Learning with Graph Neural Networks [Internet]. arXiv; 2019 [cited 2023 Nov 3]. Available from: http://arxiv.org/abs/1904.10098
32. Vu MN, Thai MT. PGM-Explainer: Probabilistic Graphical Model Explanations for Graph Neural Networks [Internet]. arXiv; 2020 [cited 2023 Nov 3]. Available from: http://arxiv.org/abs/2010.05788
33. Hua C, Luan S, Zhang Q, Fu J. Graph Neural Networks Intersect Probabilistic Graphical Models: A Survey [Internet]. arXiv; 2023 [cited 2023 Nov 3]. Available from: http://arxiv.org/abs/2206.06089
34. Kuenzi BM, Park J, Fong SH, Sanchez KS, Lee J, Kreisberg JF, et al. Predicting Drug Response and Synergy Using a Deep Learning Model of Human Cancer Cells. Cancer Cell. 2020;38:672–684.e6. [PubMed: 33096023]
35. Ma J, Yu MK, Fong S, Ono K, Sage E, Demchak B, et al. Using deep learning to model the hierarchical structure and function of a cell. Nat Methods. 2018;15:290–8. [PubMed: 29505029]
36. West J, Robertson-Tessi M, Anderson ARA. Agent-based methods facilitate integrative science in cancer. Trends Cell Biol. 2023;33:300–11. [PubMed: 36404257]
37. Yankeelov TE, Quaranta V, Evans KJ, Rericha EC. Toward a science of tumor forecasting for clinical oncology. Cancer Res. 2015;75:918–23. [PubMed: 25592148]
38. Zangooei MH, Margolis R, Hoyt K. Multiscale computational modeling of cancer growth using features derived from microCT images. Sci Rep 2021;11: 18524. [PubMed: 34535748]
39. Höpfl S, Pleiss J, Radde NE. Bayesian estimation reveals that reproducible models in Systems Biology get more citations. Sci Rep. 2023;13:2695. [PubMed: 36792648]
40. Zhang F, Meier-Schellersheim M. SBML Level 3 package: Multistate, Multicomponent and Multicompartment Species, Version 1, Release 1. J Integr Bioinforma [Internet]. 2018 [cited 2024 Apr 5];15. Available from: https://www.degruyter.com/document/doi/10.1515/jib-2017-0077/html
41. Hucka M, Finney A, Sauro HM, Bolouri H, Doyle JC, Kitano H, et al. The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. Bioinformatics. 2003;19:524–31. [PubMed: 12611808]
42. Porubsky VL, Goldberg AP, Rampadarath AK, Nickerson DP, Karr JR, Sauro HM. Best Practices for Making Reproducible Biochemical Models. Cell Syst. 2020;11:109–20. [PubMed: 32853539]
43. Keating SM, Waltemath D, König M, Zhang F, Dräger A, Chaouiya C, et al. SBML Level 3: an extensible format for the exchange and reuse of biological models. Mol Syst Biol. 2020;16:e9110. [PubMed: 32845085]
44. Wilkinson MD, Dumontier M, Aalbersberg IjJ, Appleton G, Axton M, Baak A, et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data. 2016;3:160018. [PubMed: 26978244]
45. Petralia F, Tignor N, Reva B, Koptyra M, Chowdhury S, Rykunov D, et al. Integrated Proteogenomic Characterization across Major Histological Types of Pediatric Brain Cancer. Cell. 2020;183:1962–1985.e31. [PubMed: 33242424]
46. Li Y, Lih T-SM, Dhanasekaran SM, Mannan R, Chen L, Cieslik M, et al. Histopathologic and proteogenomic heterogeneity reveals features of clear cell renal cell carcinoma aggressiveness. Cancer Cell. 2023;41:139–163.e17. [PubMed: 36563681]

47. Irvin MW, Ramanathan A, Lopez CF. Model certainty in cellular network-driven processes with missing data. Meier-Schellersheim M, editor. PLOS Comput Biol. 2023;19:e1011004. [PubMed: 37099625]
48. Schmiester L, Schälte Y, Fröhlich F, Hasenauer J, Weindl D. Efficient parameterization of large-scale dynamic models based on relative measurements. Schwartz R, editor. Bioinformatics. 2020;36:594–602. [PubMed: 31347657]
49. Mitra ED, Suderman R, Colvin J, Ionkov A, Hu A, Sauro HM, et al. PyBioNetFit and the Biological Property Specification Language. iScience. 2019;19:1012–36. [PubMed: 31522114]
50. Mitra ED, Hlavacek WS. Bayesian inference using qualitative observations of underlying continuous variables. Wren J, editor. Bioinformatics. 2020;36:3177–84. [PubMed: 32049328]
51. Berger MF, Mardis ER. The emerging clinical relevance of genomics in cancer medicine. Nat Rev Clin Oncol. 2018;15:353–65. [PubMed: 29599476]
52. Wheeler DA, Wang L. From human genome to cancer genome: The first decade. Genome Res. 2013;23:1054–62. [PubMed: 23817046]
53. Shi T, Niepel M, McDermott JE, Gao Y, Nicora CD, Chrisler WB, et al. Conservation of protein abundance patterns reveals the regulatory architecture of the EGFR-MAPK pathway. Sci Signal. 2016;9:rs6. [PubMed: 27405981]
54. Chen P-H, Cai L, Huffman K, Yang C, Kim J, Faubert B, et al. Metabolic Diversity in Human Non-Small Cell Lung Cancer Cells. Mol Cell. 2019;76:838–851.e5. [PubMed: 31564558]
55. Naba A, Clauser KR, Hoersch S, Liu H, Carr SA, Hynes RO. The matrisome: in silico definition and in vivo characterization by proteomics of normal and tumor extracellular matrices. Mol Cell Proteomics MCP. 2012;11:M111.014647.
56. Orre LM, Vesterlund M, Pan Y, Arslan T, Zhu Y, Fernandez Woodbridge A, et al. SubCellBarCode: Proteome-wide Mapping of Protein Localization and Relocalization. Mol Cell. 2019;73:166–182.e7. [PubMed: 30609389]
57. Lim MY, O'Brien J, Paulo JA, Gygi SP. Improved Method for Determining Absolute Phosphorylation Stoichiometry Using Bayesian Statistics and Isobaric Labeling. J Proteome Res. 2017;16:4217–26. [PubMed: 28985074]
58. Ning K, Fermin D, Nesvizhskii AI. Comparative analysis of different label-free mass spectrometry based protein abundance estimates and their correlation with RNA-Seq gene expression data. J Proteome Res. 2012;11:2261–71. [PubMed: 22329341]
59. Goodspeed A, Heiser LM, Gray JW, Costello JC. Tumor-Derived Cell Lines as Molecular Models of Cancer Pharmacogenomics. Mol Cancer Res MCR. 2016;14:3–13. [PubMed: 26248648]
60. Gabor A, Tognetti M, Driessen A, Tanevski J, Guo B, Cao W, et al. Cell-to-cell and type-to-type heterogeneity of signaling networks: insights from the crowd. Mol Syst Biol. 2021;17:e10402. [PubMed: 34661974]
61. Karr JR, Williams AH, Zucker JD, Raue A, Steiert B, Timmer J, et al. Summary of the DREAM8 Parameter Estimation Challenge: Toward Parameter Identification for Whole-Cell Models. PLoS Comput Biol. 2015;11:e1004096. [PubMed: 26020786]
62. Choobdar S, Ahsen ME, Crawford J, Tomasoni M, Fang T, Lamparter D, et al. Assessment of network module identification across complex diseases. Nat Methods. 2019;16:843–52. [PubMed: 31471613]
63. Hill SM, Heiser LM, Cokelaer T, Unger M, Nesser NK, Carlin DE, et al. Inferring causal molecular networks: empirical assessment through a community-based effort. Nat Methods. 2016;13:310–8. [PubMed: 26901648]
64. Walker SA, Pham A, Nizzero S, Kim M, Riter B, Bletz J, et al. Education and Outreach in Physical Sciences in Oncology. Trends Cancer. 2021;7:3–9. [PubMed: 33168416]
65. Archer TC, Fertig EJ, Gosline SJC, Hafner M, Hughes SK, Joughin BA, et al. Systems Approaches to Cancer Biology. Cancer Res. 2016;76:6774–7. [PubMed: 27864348]
66. Saez-Rodriguez J, Costello JC, Friend SH, Kellen MR, Mangravite L, Meyer P, et al. Crowdsourcing biomedical research: leveraging communities as innovation engines. Nat Rev Genet. 2016;17:470–86. [PubMed: 27418159]
67. Moult J A decade of CASP: progress, bottlenecks and prognosis in protein structure prediction. Curr Opin Struct Biol. 2005;15:285–9. [PubMed: 15939584]

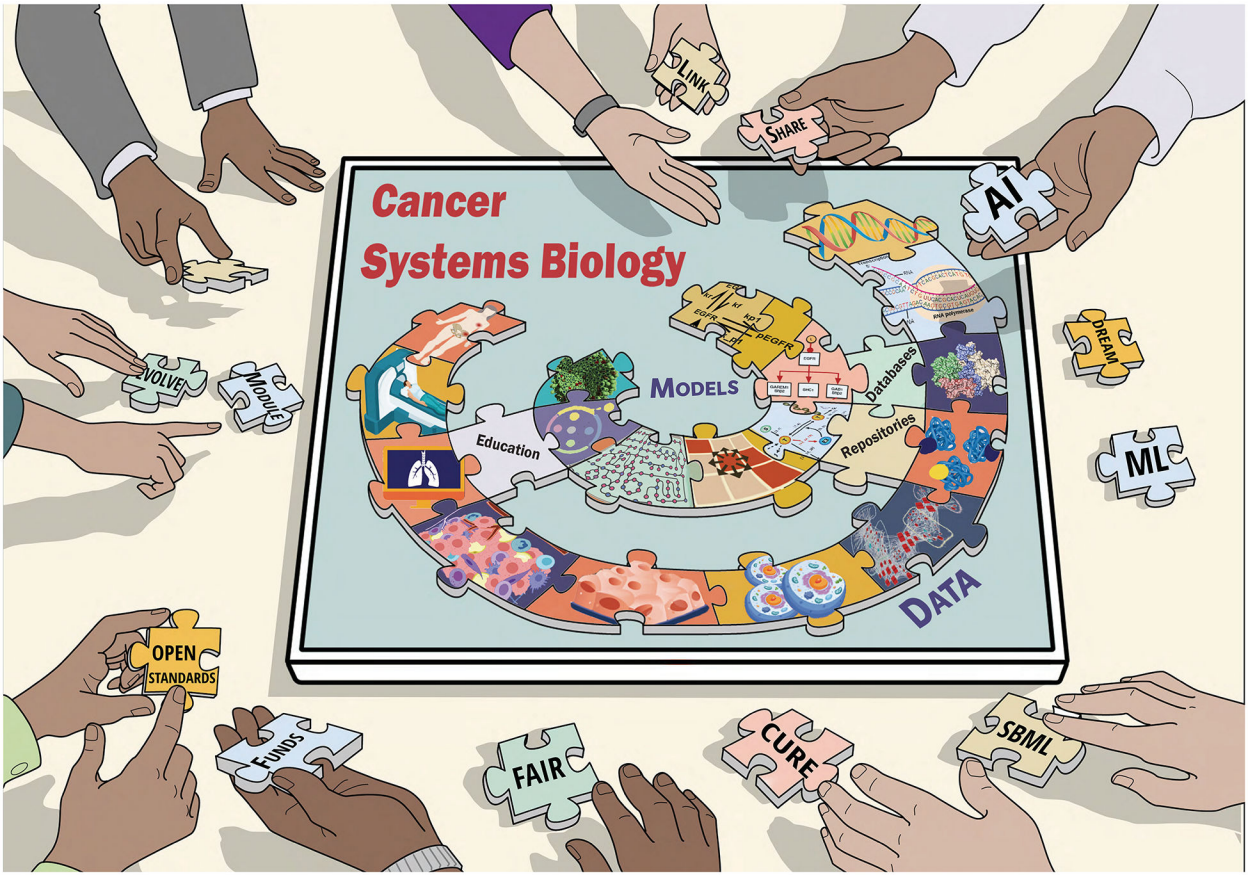

Figure 1.

Progress in understanding cancer from a systems perspective will require cooperation and collaboration across the entire systems biology community. Although progress has been made in collecting multi-scale data and building complex models, there are many gaps in our collective understanding of how molecular perturbations impact cancer progression and disease outcomes. There is an urgent need for community data and modeling standards to support multi-scale modeling which in turn will allow the entire community to work together to build more realistic models of cancer.

Table 1.

Challenges and needed actions for cancer systems biology.

Challenge	Proposed Action	Primary Stakeholders
1. Lack of multiscale modeling frameworks	Develop and adopt standardized frameworks for building and integrating models across spatial and temporal scales	Modelers, Computational scientists, Research consortia
2. Inconsistent data and model standards	Create and promote community-wide standards (e.g., FAIR for data; CURE for models); expand use of SBML and metadata annotations	Research community, Sponsors, Journals
3. Limited integration of AI/ML with mechanistic models	Combine mechanistic and data-driven approaches to improve interpretability, generalizability, and utility in clinical settings	Modelers, Computational scientists, AI/ML experts
4. Lack of community standards for data, models, and software	Create exchangeable modeling frameworks compatible with a multiscale modeling; develop common formats and ontologies	Modelers, Systems biology community
5. Biological heterogeneity across tumors and patients complicates modeling	Develop coarse-grained models to encapsulate heterogeneity; integrate diverse datasets using standardized formats	Modelers, Experimentalists, Data analysts
6. Limited mechanistic interpretability of AI/ML models	Deeper integration of AI/ML approaches with mechanism-driven modeling; use hybrid models	Computational scientists, AI/ML experts, Mechanistic modelers
7. Sparse and inconsistent quantitative data for multiscale modeling	Identify and generate data and metadata needed to build and parameterize multiscale models	Experimentalists, Data scientists, Modelers, Research consortia
8. Absence of community-defined grand challenges	Define specific, currently unachievable challenges (e.g., predicting tumor evolution or perturbation effects) and create blinded benchmarks to track progress	Research consortia, Funders, Systems biology community
9. Lack of standard cancer model systems for benchmarking and clinical validation	Identify specific cancer model systems to serve as benchmarks for multiscale model development and validation	Cancer systems biology community
10. Gaps in education and training	Promote education for scientists to bridge the model building–data generation divide; support multidisciplinary conferences and workshops	Institutions, Graduate programs, Sponsors, Mentors
11. Academic incentives undervalue collaborative team science	Adjust promotion and grant review criteria to value multi-author, interdisciplinary collaborations. Create reward systems for data reuse and sharing; track post-publication data use	Institutions, Promotion committees, Sponsors