Published in final edited form as:
Nat Med. 2021 December ; 27(12): 2065–2066. doi:10.1038/s41591-021-01558-5.

Digital Twins for Predictive Oncology will be a Paradigm Shift for Precision Cancer Care

Tina Hernandez-Boussard[*,+,1], **Paul Macklin**[+,2], **Emily J. Greenspan**[+,3], **Amy L. Gryshuk**[4], **Eric Stahlberg**[5], **Tanveer Syeda-Mahmood**[6], **Ilya Shmulevich**[7]

[1]Department of Medicine, Stanford University, Stanford CA, USA

[2]Department of Medicine, Indiana University, Stanford CA, USA

[3]Center for Biomedical Informatics and Information Technology, National Cancer Institute, Rockville, MD, USA

[4]Physical and Life Sciences Directorate, Lawrence Livermore National Laboratory, Livermore, CA, USA

[5]Biomedical Informatics and Data Science Directorate, Frederick National Laboratory for Cancer Research, Frederick, MD, USA

[6]IBM Almaden Research Center, San Jose CA, USA

[7]Institute for Systems Biology, Seattle, WA, USA

To the Editor - In medicine, digital twin models use real-time data to adjust treatment, monitor response, and track lifestyle modifications. Similarly, cancer patient digital twins (CPDTs) use emerging computing and biotechnologies to build *in silico* individual representations that dynamically reflect molecular, physiological and lifestyle status across different treatments and time. We propose a CPDT framework with a continuous life cycle for shared decision-making (Figure 1).

The proposed CPDT framework integrates individual-level data, such as proteome and clinical characteristics, with other factors, like clinical trials and population studies, to create a multiscale and multimodal data set for model training. To ensure rapid and comprehensive data integration, data must be captured under FAIR (Findability, Accessibility, Interoperability, Reusability) principles[1,2] and across diverse populations to ensure all patients equally benefit.[3]

A revolutionary concept of the proposed CPDT will be its ability to bridge size and time scales of biological organization to address changes that span the full patient experience,

*Corresponding Author: Tina Hernandez-Boussard, PhD., Department of Medicine, 1265 Welch Road, #245, Stanford University, Stanford, CA. boussard@stanford.edu.
+Shared First Authors

Author Contributions
THB, PM, EG, AG and IS drafted the manuscript; All authors were responsible for the concept, design, and critical revision of the manuscript.

Competing Interests
The authors declare no competing interests.

from the molecular level over nanoseconds to the population levels across decades. As the patient's physical state evolves, their CPDT must incorporate observational data to represent the patient's current state and reliably forecast future state transitions. A range of multiscale models exist for various cancer-related processes. The envisioned CPDT will need to connect scales and processes by adapting existing techniques for simulation, model inference, data assimilation, and high performance computing to build and test real-time, dynamic models at scale.[4] Throughout development and once complete, technical validation and rigorous software engineering best practices will be critical to ensuring that the future CPDT system is trustworthy.[5]

Extending current, focused pilot studies that use mathematical models to predict and plan therapy,[6] it is envisioned that clinical teams will use future CPDTs to perform virtual experiments, by simulating the model forward without treatment, under the current standard of care, and under treatment variations. Each simulation will predict a trajectory for the patient's cancer under one of the treatment options. At each clinical encounter, the previous forecast for the chosen treatment will be compared to the patient's newest measurements to assess performance of the digital twin. The new measurements will then be assimilated to update the patient's CPDT, and the process will begin anew. CPDTs must be seamlessly integrated into medical workflows to achieve clinical utility by helping the doctor and patient to explore treatment options with intuitive visualizations. Dashboards need to be optimized to not burden the clinician or interfere with patients' care experiences.

When fully realized, CPDTs will usher in a new age in medicine by increasing the probability that the optimal treatment is chosen each time. The optimality criteria will be chosen to include the patient's care goals as well as objective clinical endpoints. Equal and equitable CPDT performance across diverse populations is crucial to their successful integration into clinical practice. CPDTs are susceptible to biases when learning from potentially biased data, reflecting existing healthcare systems that are rife with inequalities.[7] Tight controls and rigorous standards are necessary to ensure CPDTs do not reinforce pre-existing biases.[8,9]

CPDTs start with a patient model template that is based on retrospective data and a continuous learning process. Continuous learning maximizes predictive capacity while accounting for uncertainty and variability in measurements, missing data, and incomplete mechanistic knowledge. The systematic accumulation of CPDTs from real world deployment will enable cohorts of hundreds or thousands of CPDTs that may be used for *in silico* clinical trials and population studies. Although key technologies and data are rapidly evolving, significant hurdles remain.

An example of a CPDT could be for an acute myeloid leukemia (AML) patient who received a hematopoietic stem cell transplantation from an unmatched donor. For patients whose disease relapses, the best treatment plan may involve combinations of drugs and immunotherapies at multiple time points. The patient's host and tumor genomic and other multi-omic measurements from the bone marrow and peripheral blood collected can be used to create updated predictions for various clinical scenarios including drug combinations, doses and durations, or a decision for no action, which are then intuitively presented to the patient and doctor. The CPDT continuously accounts for the evolving cancer state and the donor (graft) immune system to reduce the uncertainty inherent in clinical decision making, thereby improving outcomes and patient-clinician interactions.

In 2019, the National Cancer Institute, the Department of Energy, several government national laboratories, and a consortium of academic and industrial partners formed the Envisioning Computational Innovations for Cancer Challenges (ECCIC) community at the intersection of cancer research and advanced computing to frame forward-looking approaches to accelerate predictive oncology – and the CPDT idea began to grow.[10] However, the full realization of CPDTs can only succeed with contributions by the experimental, computational, and clinical communities.

Developing CPDTs is a grand challenge for the convergence of advanced computing technologies and oncology. Using a CPDT for individualized patient care decision making has enormous potential for advancing predictive oncology. With further development, refinement, and eventual implementation into clinical practice, CPDTs are poised to revolutionize how cancer and a host of other complex diseases are treated and managed.

CPDTs offer far more than individual patient predictions. The accumulated patient trajectories, decision making, outcomes, and match or discordance between predictions and reality will provide invaluable evidence for research investment, enabling policymakers to channel resources into therapies that show the most effectiveness. CPDTs could help to structure existing healthcare systems to better respond to real-time public health situations, addressing healthcare needs and health disparities as they occur.

ACKNOWLEDGEMENTS

We thank Jeannette Yusko [LLNL] and Janelle Cataldo [LLNL] for their contributions to the development of Figure 1. This work was supported in part by Lawrence Livermore National Laboratory (LLNL). Lawrence Livermore National Laboratory is operated by Lawrence Livermore National Security, LLC, for the U.S. Department of Energy, National Nuclear Security Administration under Contract DE-AC52-07NA27344. LLNL-JRNL-823517. This work has been supported in part by the Joint Design of Advanced Computing Solutions for Cancer (JDACS4C) program established by the U.S. Department of Energy (DOE) and the National Cancer Institute (NCI) of the National Institutes of Health, Leidos Biomedical Research Contract #75N91019D00024. Drs. Hernandez-Boussard, Macklin, Shmulevich, and Syeda-Mahood were supported in part by Cancer Moonshot^{SM} funds from the National Cancer Institute, Leidos Biomedical Research Subcontract 21X126F. Dr. Hernandez-Boussard was supported in part by the National Cancer Institute of the National Institutes of Health under Award Number R01CA183962.

The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. This document was prepared as an account of work sponsored by an agency of the United States government. Neither the United States government nor Lawrence Livermore National Security, LLC, nor any of their employees makes any warranty, expressed or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States government or Lawrence Livermore National Security, LLC. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States government or Lawrence Livermore National Security, LLC, and shall not be used for advertising or product endorsement purposes.

References

1. Wilkinson MD, Dumontier M, Aalbersberg IJ, et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data Mar 2016;3:160018. doi:10.1038/sdata.2016.18 [PubMed: 26978244]
2. Fagnan K, Nashed Y, Perdue G, Ratner D, Shankar A, Yoo S. Data and models: a framework for advancing AI in science 2019.
3. Bozkurt S, Cahan EM, Seneviratne MG, et al. Reporting of demographic data and representativeness in machine learning models using electronic health records. J Am Med Inform Assoc. Dec 2020;27(12):1878–1884. doi:10.1093/jamia/ocaa164 [PubMed: 32935131]
4. Kapteyn MG, Pretorius JV, Willcox KE. A probabilistic graphical model foundation for enabling predictive digital twins at scale. Nature Computational Science. 2021;1(5):337–347.
5. Taschuk M, Wilson G. Ten simple rules for making research software more robust. PLoS Comput Biol 04 2017;13(4):e1005412. doi:10.1371/journal.pcbi.1005412 [PubMed: 28407023]
6. Zhang J, Cunningham JJ, Brown JS, Gatenby RA. Integrating evolutionary dynamics into treatment of metastatic castrate-resistant prostate cancer. Nat Commun. 2017;8(1):1816. [PubMed: 29180633]
7. Zou J, Schiebinger L. AI can be sexist and racist - it's time to make it fair. Nature. 07 2018;559(7714):324–326. doi:10.1038/d41586-018-05707-8 [PubMed: 30018439]
8. Corbett-Davies S, Goel S. The measure and mismeasure of fairness: A critical review of fair machine learning. arXiv preprint arXiv:180800023. 2018;
9. Hernandez-Boussard T, Bozkurt S, Ioannidis JPA, Shah NH. MINIMAR (MINimum Information for Medical AI Reporting): Developing reporting standards for artificial intelligence in health care. J Am Med Inform Assoc Jun 2020;doi:10.1093/jamia/ocaa088
10. Greenspan E, Lauzon C, Gryshuk A, et al. CAFCW 113 Digital Twins for Predictive Cancer Care: an HPC-Enabled Community Initiative. 2019; https://ncihub.org/resources/2296

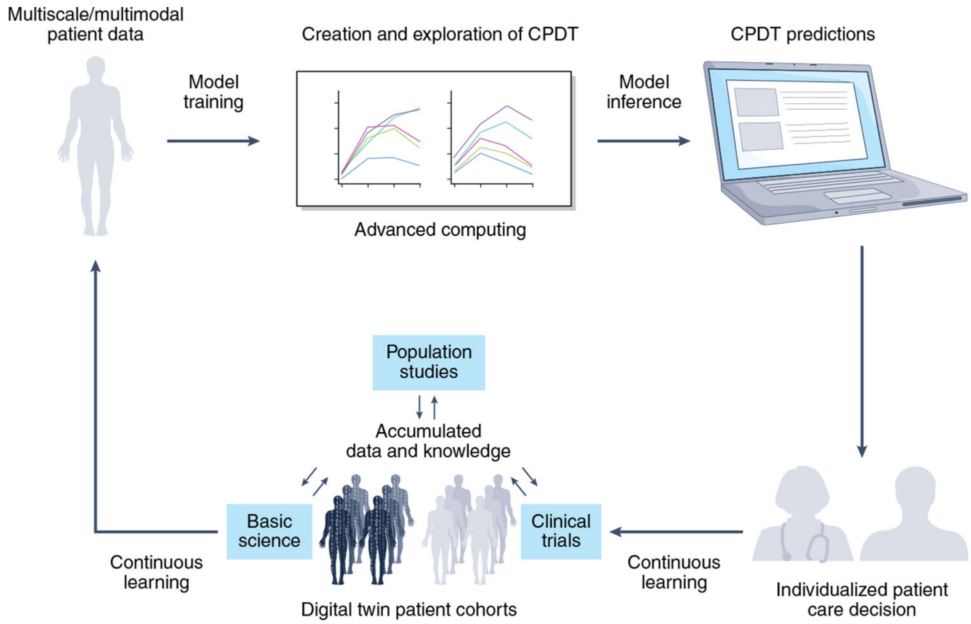

Figure 1. The Cancer Patient Digital Twin Life Cycle.

The CPDT is envisioned to have a real-time, dynamic life cycle with multiscale/multimodal data harmonization and integrated model training and inference. Advanced computing will be used to create and explore mathematical, statistical, mechanistic and AI models. CPDT predictions based on virtual experiments will be integrated into medical workflows for patient decision making and continuous learning.

Author Manuscript

Author Manuscript

Author Manuscript

Table 1.

Key challenges facing the development of CPDTs

Data challenges
Generating and acquiring high-volume, high-quality, multiscale data
Ensuring multiscale data represent both healthy and diseased states
Ensuring data are derived from diverse populations
Modeling and Integration challenges
Harmonizing and aggregating new and existing data
Developing multimodal data fusion methods that align and combine information to more accurately characterize disease states
Seamlessly integrating data-driven and mechanistic modeling
Reducing model uncertainty via standardized training and validation
Improving access to data, workflows, and HPC across the workforce
Ethical and community challenges
Ensuring all stakeholders have a voice in CPDT development
Addressing ethical biases and privacy concerns
Ensuring built-in compliance through CPDT development
Establishing and adhering to regulatory standards for data goverance and usage