

# Harnessing Big Data with Machine Learning in Precision Oncology

Nirmish Singla, MD, MSCS,<sup>1#\*</sup> Shyamli Singla, MD<sup>2#</sup>

<sup>1</sup> Urology Service, Department of Surgery, Memorial Sloan Kettering Cancer Center, New York, NY

<sup>2</sup> Division of Hematology/Oncology, Department of Pediatrics, Mount Sinai Hospital, New York, NY

## ABSTRACT

**While multi-level molecular “omic” analyses have undoubtedly increased the sophistication and depth with which we can understand cancer biology, the challenge is to make this overwhelming wealth of data relevant to the clinician and the individual patient. Bridging this gap serves as the cornerstone of precision medicine, yet the expense and difficulty of executing and interpreting these molecular studies make it impractical to routinely implement them in the clinical setting. Herein, we propose that machine learning may hold the key to guiding the future of precision oncology accurately and efficiently. Training deep learning models to interpret the histopathologic or radiographic appearance of tumors and their microenvironment—a phenotypic microcosm of their inherent molecular biology—has the potential to output relevant diagnostic, prognostic, and therapeutic patient-level data. This type of artificial intelligence framework may effectively shape the future of precision oncology by fostering multidisciplinary collaboration.**

## INTRODUCTION

Undoubtedly, we have become deeply immersed in an oncologic era defined by omics. In our quest to achieve precision medicine, we have attempted to unpeel several omic layers in cancer, including the genome, epigenome, transcriptome, proteome, lipidome, glycome, metabolome, and microbiome. These approaches have yielded an unparalleled wealth of data and breakthrough discoveries that have enabled us to better deconvolute the biology and aggressiveness of tumors, decipher the process of metastatic dissemination and tropism, characterize patterns of inheritance, identify candidate biomarkers, and understand mechanisms of therapeutic response and resistance. For the clinician and the individual patient, however, this information becomes relevant only if it can be translated into informing prognosis, guiding therapeutic decisions, and improving outcomes overall. As more sophisticated omic levels are introduced, the integration of these

data becomes increasingly complex, and the clinical interpretation is made even more challenging. Furthermore, the necessary technical and bioinformatics expertise coupled with the financial expense of executing omic studies makes it impractical to routinely implement such studies in the clinical setting.

As we delve into deeper layers of tumor omics, it is worth taking a step back and revisiting these tumors on a more phenotypic level—both pathologically and radiographically—which is easily overlooked with this newfound molecular knowledge. In particular, a closer assessment of the histopathology and morphologic architecture of tumors and their microenvironment reveals that, indeed, the appearance of tumor cells and their surroundings under the microscope can serve as a microcosm of the molecular milieu that defines these tumors. That is, precise histologic features conceivably appear the way they do as a consequence of omics. Likewise, with improvements in anatomical and

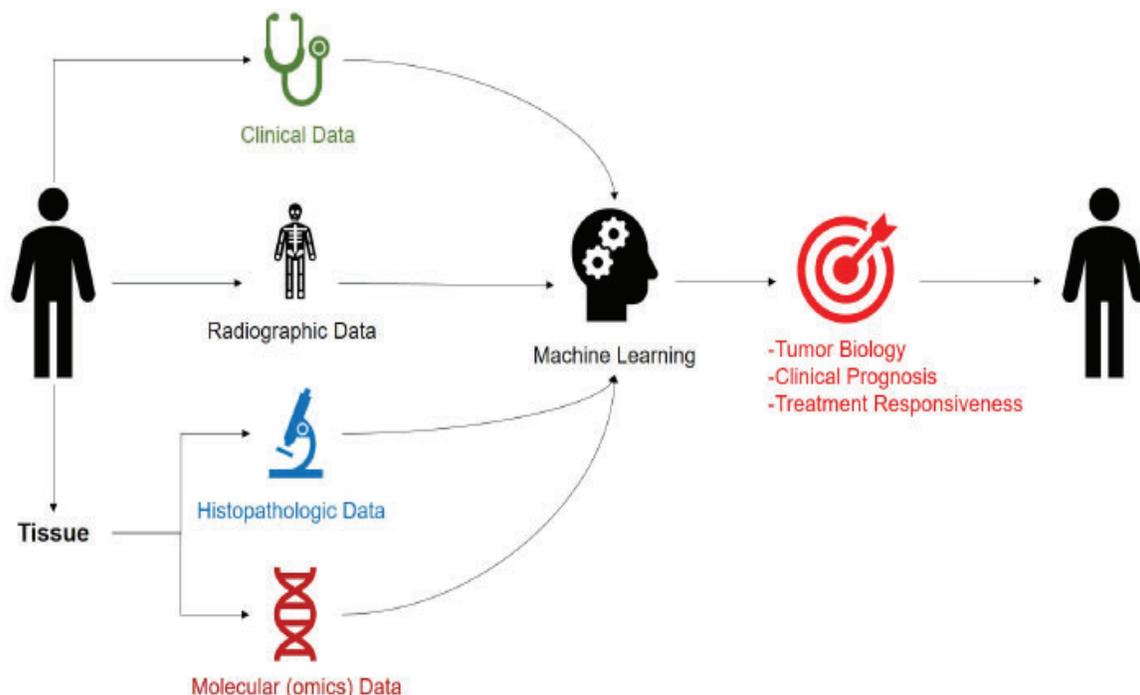
functional imaging techniques, correlating the radiographic appearance of tumors with omics—an emerging field termed radiomics—may similarly reveal heretofore uncaptured information about a tumor’s biology from radiography alone. Novel approaches that integrate the pathologic and radiographic phenotypes of cancers with their inherent omics may thus serve as a powerful means by which to glean information about their biological behavior and potentially inform clinical outcomes and therapeutic responsiveness in a more cost-sensitive and practical manner.

This concept forms the fundamental basis of machine learning in precision oncology. Iterative artificial intelligence strategies in medicine are designed to distill big data into practically useful means by streamlining the analysis, decreasing cost, increasing accuracy via automation, and yielding clinically relevant information at the patient level. Conceptually, a deep learning model could be trained to harness and integrate clinical data, radiographic data, histopathologic data, and molecular (omic) data to yield information that could be used by the clinician in counseling patients and in guiding treatment strategies (Figure 1). Following training and validation of the initial model, the machine learning algorithm would then be able to output similar information with a high degree of accuracy, but with the need for less inputs.

As a case study, clear cell renal cell carcinoma (ccRCC) nicely exemplifies these concepts. By virtue of its considerable histologic and molecular intratumoral heterogeneity, broad spectrum of biological and clinical behavior, and recent groundbreaking discoveries, ccRCC serves as a robust platform to illustrate the utility of machine learning in guiding precision oncology. Arguably the most comprehensive molecular characterization of ccRCC to date, Clark et al.

\* Correspondence: Nirmish Singla, MD, MSCS. Departments of Urology and Oncology, The James Buchanan Brady Urological Institute, The Johns Hopkins University School of Medicine, 600 North Wolfe Street, Park 213, Baltimore, MD 21287, Phone: (410) 502-3692; Fax: (410) 955-0833. Email: nsingla2@jhmi.edu

Disclosures: None



**Fig. 1 | Simplified schematic conceptually depicting large-scale integration of big data into a machine learning algorithm for precision oncology. Inputs from a single patient, including clinical, radiographic, histopathologic, and/or molecular data, can be used to train a machine learning model to accurately and efficiently predict personalized, clinically relevant data including information about the biological behavior of the patient's tumor(s), clinical prognosis, and treatment responsiveness.**

recently conducted a multi-level omics analysis of ccRCC tumors and matched normal tissue by combining genomics, epigenomics, transcriptomics, proteomics, and phosphoproteomics.<sup>1</sup> Through proteogenomic integration, they discerned the functional impact of genomic alterations in ccRCC and further characterized novel immune signatures in the tumor microenvironment. Their findings provide evidence for rational selection of personalized therapies for patients based on ccRCC pathobiology, which is urgently needed in an era in which multiple frontline therapeutic regimens are available for metastatic ccRCC without a clear algorithmic approach.<sup>2</sup> On a biological and prognostic level, the study by Clark *et al.* extends the recent findings of the TRACERx Renal Consortium, who, in a series of three elegant studies, used multiregional targeted genomic analyses to describe clonal evolutionary patterns and explain metastatic competence of ccRCC tumors.<sup>3-5</sup>

Given the cost and difficulty of conducting analyses of such depth, Cai *et al.* employed an alternative approach based entirely on histopathology to develop a systematic ontology of ccRCC phenotypic variability across tumor

architecture, cytology, and the micro-environment.<sup>6</sup> Remarkably, the authors reveal that a meticulous analysis of the histopathology alone may recapitulate clonal evolutionary trajectories, portend patient outcomes, and even inform differential response to therapies, in much the same way that Clark *et al.* and the TRACERx Renal Consortium implicated using molecular data.<sup>1,3-5</sup> They suggest that the morphologic appearance of tumors captures their molecular environment on a phenotypic level. Logically, this rather traditional approach of analyzing tumors pathologically, when integrated with multi-level omics, may serve as the basis to train future deep learning models that may then require less data input to yield the same prognostic and therapeutic patient-level information relevant to personalized medicine.

Indeed, a similar conceptual framework for artificial intelligence can be applied across other cancer types as well. As we traverse an age of big data, the challenge we will increasingly face is not so much how to generate more information, but rather how to use the information we gather. Through multidisciplinary integration of radiology, pathology, bioinformatics, and bench science, machine learning will likely hold

the key to harnessing this information efficiently and help translate precision oncology into a clinical reality.

## REFERENCES

1. Clark DJ, Dhanasekaran SM, Petralia F, *et al.* Integrated Proteogenomic Characterization of Clear Cell Renal Cell Carcinoma. *Cell*. 2019;179(4):964-983 e931.
2. Singla N. Progress Toward Precision Medicine in Frontline Treatment of Metastatic Renal Cell Carcinoma. *JAMA Oncol*. 2019.
3. Mitchell TJ, Turajlic S, Rowan A, *et al.* Timing the Landmark Events in the Evolution of Clear Cell Renal Cell Cancer: TRACERx Renal. *Cell*. 2018;173(3):611-623 e617.
4. Turajlic S, Xu H, Litchfield K, *et al.* Tracking Cancer Evolution Reveals Constrained Routes to Metastases: TRACERx Renal. *Cell*. 2018;173(3):581-594 e512.
5. Turajlic S, Xu H, Litchfield K, *et al.* Deterministic Evolutionary Trajectories Influence Primary Tumor Growth: TRACERx Renal. *Cell*. 2018;173(3):595-610 e511.
6. Cai Q, Christie A, Rajaram S, *et al.* Ontological analyses reveal clinically-significant clear cell renal cell carcinoma subtypes with convergent evolutionary trajectories into an aggressive type. *EBioMedicine*. 2019.