

Special Section Guest Editorial: Radiogenomics in Prognosis and Treatment

Karen Drukker,^{a,*} Despina Kontos,^b and Hui Li^a

^aUniversity of Chicago, Department of Radiology, United States

^bUniversity of Pennsylvania, Department of Radiology, United States

1 What is Radiogenomics?

In recent years, in the scientific literature there has been a steady increase in the use of the terms “radiomics” and “radiogenomics” (or “imaging genomics,” which is the sense in which we use the term here). The term “radiomics” was coined about five years ago and refers to the high-throughput extraction of large amounts of quantifiable features from medical imaging data. The foundation of the field of radiomics lies in the widely known field of computer-aided diagnosis and image processing/analysis techniques incepted decades ago when interest sparked in the use of imaging markers derived from routine clinical images to provide insight into diseases in a non-invasive manner. In fact, features often evaluated in computer-aided diagnosis and image processing for medical imaging were originally introduced in the field of computer vision decades ago, including the popular Haralick texture features first introduced in the 1970s. The increase in both size and availability of medical imaging datasets, and especially the digital format of the contemporary medical imaging modalities, the advances in computer power, and the emergence of deep learning, have made it possible to develop increasingly complex models combining multiple data sources for data mining and discovery. Many technical and clinical challenges remain, however, especially since more complex tasks are being investigated, such as patient prognosis, prediction of response to therapy, and patient survival outcomes, which are closely related to the genetic and/or genomic make-up of the host and the disease. “Radiogenomics” denotes the relationship between the imaging features of a disease and various genetic or genomic/molecular features; in other words, it is the “marriage” between radiomics and genetics/genomics research and explores the relationship between the imaging traits—the imaging phenotype—and the genotype of a disease—its gene expression patterns, gene mutations, and other genome-related characteristics. Radiogenomics, therefore, provides an avenue to correlate imaging characteristics to genetic or molecular markers of disease processes (such as cancer or Alzheimer’s) to help develop precision medicine approaches and guide targeted therapy for patients.

2 Why is Radiogenomics Important?

What is the practical significance of radiogenomic research, and why would one benefit from combining imaging phenotypes and genetics/genomics in discovery? In some instances, the relationship between imaging phenotypes and outcomes is already rather well understood and used in treatment planning; e.g., tumor enhancement on dynamic contrast-enhanced MRI can serve as a surrogate marker for angiogenesis and cancer aggressiveness. Another advantage of relating imaging phenotypes directly to outcome is that imaging phenotypes capture the entire region of disease, and therefore its heterogeneity, rather than a small tissue sample as analyzed for gene expression patterns or tumor subtypes. In that sense, imaging represents a unique opportunity to characterize disease heterogeneity *in vivo* and potentially augment the information obtained via selected tissue sampling with biopsy. On the other hand, genetic mapping and genomic/molecular profiling helps understand disease since genomic instability and mutations are a hallmark of cancer, and mutations—either germline or somatic—play a role in many other diseases. In cancer,

*Address all correspondence to Karen Drukker, kdrukker@uchicago.edu

genomic profiling is often used to predict survival—as a prognostic biomarker—or response to treatment—as a predictive biomarker—helping inform clinical decisions and treatment planning. Radiogenomics exploits the strengths of both medical image analysis and genomics, in that it not only aims to relate the imaging phenotypes to information already captured in genotypes, but also aims to maximize the synergy between the two, by identifying potentially clinically relevant orthogonal information. It is an exciting emerging field of research that to date has generated important knowledge. However, conclusions that can be drawn from currently available studies are typically constrained by their limited sample sizes and generalizability remains an issue. With the goal to better understand disease and improve patient outcomes, the success of radiogenomics will ultimately greatly depend on the availability of large high quality curated datasets with imaging, genomics, and outcome data, in publicly accessible repositories. While there are extensive repositories for genomic data alone—such as the Genomic Data Commons—and for imaging data alone—such as the Imaging Data Commons—there are much fewer publicly available datasets combining imaging, genomics, and outcome data—such as the Cancer Genome Atlas in combination with the Cancer Imaging Archive. Equally important are on-going efforts to standardize and harmonize such multi-omic and imaging data across different acquisition equipment and protocols, clinical sites, and analysis platforms. Nevertheless, as evident in this issue, increasingly accurate and robust radiogenomic models are being presented and, with future validation of these models, many of the challenges standing between radiogenomics and clinical implementation may be overcome.

3 In this Issue

This special section of *Journal of Medical Imaging, Volume 8 Issue 3*, includes two review papers, one by Coates et al. and one by Singh et al.:

The review by Coates et al., “[Radiomic and radiogenomic modeling for radiotherapy: strategies, pitfalls, and challenges](#),” offers a comprehensive review on the topic of predictive modeling for radiotherapy outcomes using two complementary but distinct approaches, radiomics and radiogenomics. Many technical aspects are covered, from data collection, processing, harmonization, analysis, model development and validation. The authors also include a useful discussion on outcome modeling strategies using the integration of heterogeneous and high-dimensional multiomics datasets (panomics). Many challenges and pitfalls as well as strategies to overcome them are also discussed in detail.

The comprehensive 4-part review by Singh et al., “[Radiogenomics in brain, breast, and lung cancer: opportunities and challenges](#),” highlights the current state of the art and the role of radiogenomics in cancer research. This review discusses (1) the biologic basis of radiomic signatures using gene expression and molecular profiling information, (2) the non-invasive prediction of molecular subtypes of tumors through radiomic signatures, (3) the potential to augment the performance of established prognostic signatures by combining complementary information encoded by radiomic and genomic signatures, and (4) the biological significance of radiomic phenotypes. The authors conclude their paper by discussing current challenges and opportunities in the field.

The remainder of this special section contains original research papers:

Budzikowski et al., in “[Radiomics-based assessment of idiopathic pulmonary fibrosis is associated with genetic mutations and patient survival](#),” demonstrate that radiomics features extracted from CT scans of patients with idiopathic pulmonary fibrosis (IPF) combined with logistic regression modeling can be used to identify genetic variations and patient survival. Specifically, the authors show that the developed radiomic signatures correlate with the TOLLIP-1 (rs4963062) and TOLLIP-2 (rs5743905) mutations, and that particularly first-order and fractal features demonstrate the greatest discrimination between Kaplan–Meyer survival curves.

Oh et al., in “[Reproducibility of radiomic features using network analysis and its application in Wasserstein \$k\$ -means clustering](#),” investigate innovative methods for identifying radiomic features that are reproducible over varying image acquisition settings and propose a k -means clustering algorithm coupled with the optimal mass transport theory. It was demonstrated that,

when applied to a set of computed tomography (CT) scans from patients with head and neck squamous cell carcinoma, the resulting clusters separate tumor subsites as well as HPV status. Importantly, this was also validated on an independent dataset and the authors showed that their network-based analysis enables identifying reproducible radiomic features and that the use of those features can enhance clustering results.

Santinha et al., in [“Improving performance and generalizability in radiogenomics: a pilot study for prediction of IDH1/2 mutation status in gliomas with multicentric data,”](#) address radiogenomics models’ generalizability issue by applying the feature selection as a potential mean to overcome this issue. Specifically, in this paper the authors apply the proposed feature selection method, where “environments” are known variances in data, which considers causal structures in the development of a radiogenomics model to predict the IDH1/2 mutation status with publicly available gliomas and glioblastomas multicenter data in TCGA and TCIA. The proposed feature selection method achieves good performances, both in terms of robustness and generalizability by comparing with traditional method.

Smedley et al., in [“Using deep neural networks and interpretability methods to identify gene expression patterns that predict radiomic features and histology in non-small cell lung cancer,”](#) leverage deep learning architectures as a means to predict quantitative image (radiomic) features and histology from gene expression in non-small cell lung cancer (NSCLC). Using publicly available datasets, and a technique called “gene masking,” the authors train deep feedforward neural networks and show that they can predict the patterns of hundreds of radiomics features from CT and tumor histology from specific subsets of gene expression data, including specific gene signatures related to hypoxia, and the immune and cardiac systems.

Disclosures

The authors declare no conflicts of interest.

Karen Drukker is a research associate professor of radiology at the University of Chicago where she has been involved in medical imaging research for 20+ years. She received her PhD in physics from the University of Amsterdam. Her research interests include machine learning applications in the detection, diagnosis, and prognosis of breast cancer and, more recently, of COVID-19 patients, focusing on rigorous training/testing protocols, generalizability, and performance evaluation of machine learning algorithms.

Despina Kontos is an associate professor of radiology at the Radiology Department of the University of Pennsylvania. She received her C.Eng. diploma in computer engineering and informatics from the University of Patras in Greece and her MSc and PhD degrees in computer science from Temple University in Philadelphia. Her current research interests focus on investigating the role of quantitative imaging as a predictive biomarker for guiding personalized clinical decisions in cancer screening, prognosis and treatment. She currently co-leads the Radiomics Working Group of the ECOG-ACRIN and leads several on-going studies, funded both by the NIH/NCI and private foundations, to incorporate novel quantitative multi-modality imaging measures of tumor and normal tissue composition into cancer risk prediction models.

Hui Li is a research associate professor of radiology at the University of Chicago, and has been involved in quantitative imaging analysis on medical images for over a decade. His research interests include breast cancer risk assessment, diagnosis, prognosis, response to therapy, understanding the relationship between radiomics and genomics, and their future roles in precision medicine with both conventional and deep learning approaches.