CrossMark

Radiogenomics: What It Is and Why It Is Important

EC: Editor's Choice

SA-CME

Maciej A. Mazurowski, PhD

Abstract

In recent years, a new direction in cancer research has emerged that focuses on the relationship between imaging phenotypes and genomics. This direction is referred to as radiogenomics or imaging genomics. The question that subsequently arises is: What is the practical significance of elucidating this relationship in improving cancer patient outcomes. In this article, I address this question. Although I discuss some limitations of the radiogenomic approach, and describe scenarios in which radiogenomic analysis might not be the best choice, I also argue that radiogenomics will play a significant practical role in cancer research. Specifically, I argue that the significance of radiogenomics is largely related to practical limitations of currently available data that often lack complete characterization of the patients and poor integration of individual datasets. Radiogenomics offers a practical way to leverage limited and incomplete data to generate knowledge that might lead to improved decision making, and as a result, improved patient outcomes.

J Am Coll Radiol 2015;12:862-866. Copyright © 2015 American College of Radiology

INTRODUCTION TO RADIOGENOMICS

Increasingly, even casual readers of the scientific literature are encountering the terms "radiogenomics," "imaging genomics," and "radiomics." Because they have only recently been introduced, their usages and definitions are still in flux. The term "radiogenomics," in particular, has been inconsistently used to refer to a range of cancer-related endeavors and research topics.

Most often, "radiogenomics" refers to the relationship between the imaging characteristics of a disease (ie, the imaging phenotype or radiophenotype), and its gene expression patterns, gene mutations, and other genome-related characteristics [1,2]. As a simplification, I will refer to them collectively as "genomic characteristics" or simply "genomics." A particular focus of radiogenomic analysis has been on the relationship between imaging phenotypes and gene expression patterns which include expressions of individual genes as well as measures that summarize expressions of specific gene subsets (eg tumor molecular subtype, or Oncotype DX). 'Radiogenomics'

also refers to a research effort aimed at finding this relationship. Another term used to refer to this kind of research is imaging genomics.

Another, also very common, use of the term 'radiogenomics' is to refer to the analysis that looks for associations between patient genetics and his/her reaction to radiation therapy [3], with a focus on radiation toxicity. As opposed to an effort to match imaging phenotype and genomic characteristics, this genre of research focused on phenotypes representing radiation toxicity [3]. In 2009, the Radiogenomics Consortium was established in the United Kingdom [4] in relation to this research area.

Finally, 'radiogenomics' has been equated with another approach called 'radiomics' [5-7]. However, rather than describing a particular relationship of interest, radiomics focuses on the methodology used in the analysis. Specifically, radiomics involves extraction of many quantitative features from images, using computer algorithms. The extracted features can be evaluated in relation to other data of interest, including patient outcomes. These features can also be related to genomic characteristics and such a pursuit could be referred to as the 'radiomics approach to radiogenomics.'

STATE OF THE ART IN BRIEF

The literature on radiogenomics is limited, but a rapidly increasing number of articles are appearing in relation to

Department of Radiology, Duke Cancer Institute, Duke Medical Physics Program, Duke University School of Medicine, Durham, North Carolina.
Corresponding author and reprints: Department of Radiology, Duke Cancer Institute, Duke Medical Physics Program, Duke University School of Medicine, Duke University, 2424 Erwin Rd, Suite 302, Durham, NC 27705; e-mail: maciej.mazurowski@duke.edu.

brain cancer [8-11], particularly glioblastoma, breast cancer [12-16], lung cancer [17], and other cancers. Because the objective of this article is to discuss the significance of radiogenomics research, below I provide an overview of the general level of advancement, rather than an exhaustive review of specific studies in the field.

In glioblastoma (GBM), Zinn et al [8] showed that an upregulated PERIOSITIN gene is associated with a high tumor volume in FLAIR MRI exams. Jamshidi et al [9] showed that specific molecular phenotypes correlate with some imaging traits in GBM. Further evidence of the potential association between molecular phenotypes and imaging can be found in [10] and [11].

In breast cancer, Yamamoto et al [12] showed the potential for an association between imaging and genomics with a small sample of 10 patients. This was followed up by discovering a relationship between semi-automatically extracted imaging features describing MRI enhancement dynamics with Luminal A and Luminal B subtypes [13], [14] and Oncotype DX [15], [16]. Semi-automatic feature extraction involves both a human reader and a computer algorithm.

In lung cancer Gevaert et al [17] showed a correlation between molecular phenotypes and some imaging traits in lung computed tomography (CT). Radiogenomic analysis has also been applied to hepatocellular carcinoma [18] and clear cell renal cell carcinoma [19].

A typical research study in radiogenomics involves manual or semiautomatic assessment of imaging features and their correlation with individual gene expressions, combined gene expression patterns, such as previously defined genomic subtypes, and other molecular phenotypes. The currently available studies are typically characterized by smaller sample sizes (<100), which limit the conclusions that can be drawn.

LIMITATIONS AND SIGNIFICANCE

Radiogenomics attempts to establish and examine the relationship between tumor genomic characteristics and their radiologic appearance. Although there is certainly a lot to learn about these relationships, one could ask: what is the practical significance of radiogenomic discoveries? From the perspective of the patients, cancer patients in this case, it is their outcomes that are of greatest interest, such as survival, time to recurrence, or response to a particular treatment. A question appears: If imaging data and particularly specific features extracted from the images are available along with the outcome of interest, why not simply build a model that relates the imaging features to the outcomes directly? Relation between some imaging

features and outcomes is already established and utilized in treatment planning. What is the benefit of including genomics in the mix?

One could argue that using genomics as an intermediate step in the analysis could damage the potential of imaging to predict patient outcomes. Specifically, current models that show associations of molecular phenotypes to outcomes and usage of different therapeutic regimens are highly imperfect. These models often show only minor differences in outcome for various molecular phenotypes (eg, for different molecular subtypes) and therefore provide limited prognostic/predictive values. Radiogenomic models relating imaging data to genomics, especially now, in their early days, are also capturing fairly weak or noisy relationships. When the tenuous imaging to genomics and genomics to outcomes relationships are combined to establish an imaging to outcomes relationship, the resulting link might be very weak or nonexistent.

Another reason for relating imaging features directly to outcomes is that imaging phenotypes potentially contain information that is not available in genomics data. For example, gene expression patterns are typically assessed based on a relatively small tumor tissue sample, or "averaged" from tissue samples from multiple tumor regions, and therefore may not reflect the usual heterogeneity of cancerous tumors [20]. Imaging on the other hand can potentially capture this heterogeneity [21]. Constructing a radiogenomic model first and then applying it to predict outcomes without incorporating the imaging-outcomes data in the model limits the information from imaging available to predict outcomes to what is already contained within tumor genomics. The imaging information that complements genomics is not used in such a scenario. To utilize such complementary information, a model directly relating imaging to outcomes is needed.

These are limitations of radiogenomics. However, this does not at all mean that radiogenomic analysis is without use. I will argue that the significance of discoveries in radiogenomics is largely related to a very practical aspect of science: availability of data and availability of knowledge.

As a result of prior and current data collection efforts, various data sets, both private and public [22,23], are available containing different combinations of imaging, genomics, and outcomes data (often just one or two components). The quality of the data components may differ dramatically among data sets.

Specifically, well organized molecular data repositories are publically available. To develop the field of radiogenomics, recent efforts have been undertaken to assemble large cancer imaging data sets (eg, The Cancer Imaging

Archive [22]). On the other hand, routine imaging data are readily available in large quantities in patient records and can be relatively easily and cheaply collected retrospectively by investigators working at large clinical institutions (although sharing such data poses difficulties). Outcomes data may be equally easy to collect if available, however, outcomes of interest may occur after an extended time period, limiting their availability.

Radiogenomics allows imperfect data sets such as these to be leveraged in conjunction with prior knowledge of the relationship between outcomes and imaging genomics to draw new conclusions. For example, let's assume that some research has shown that a specific genomic characteristic related to poorer survival. A research group has a dataset available that contains genome-wide gene expression data and imaging data but no outcomes data. The research group builds a model that represents specific genomic characteristics in terms of imaging features, and conducts an analysis that establishes the relationship between some imaging features and genomic data. This analysis also determines which imaging features are the most predictive of the genomic characteristic. This is significant because it allows for identifying imaging features (or even defining new features) correlated with the genomic characteristic previously shown to be related to outcomes. These individual imaging features are likely to be correlated with outcomes too, and they are good candidates for further analysis. Moreover, the constructed model that predicts the genomic characteristic using multiple imaging features can be directly applied to predict outcomes. Conversely, if a relationship between a particular imaging feature and an outcome is known, then finding a correlation between that feature and specific genomic characteristics may identify the relationship of those characteristics with that outcome.

One example of this approach is in study [19] where the authors apply radiogenomic analysis to clear renal cell carcinoma. In this paper, the authors identified several imaging features which they correlated to mutations in VHS (von Hippel-Lindau tumor suppressor), PBRM1, SETD2 (SET domain containing 2), KDM5C, and BAP1 genes that were previously indicated in relation to clinically significant factors of advanced grade, stage, and diminished survival prognosis. The authors found correlations between some imaging features and the gene mutations and through this discovery identified imaging features that are potentially predictive of outcomes.

Another example is our own studies with colleagues [13,14], where we found a correlation between computer-extracted imaging features and intrinsic tumor subtypes in breast cancer. Tumor molecular subtypes are defined using a set of tumor gene expression data: subtypes include luminal A, luminal B, basal, and HER2 type. We demonstrated that computer-assessed features that describe tumor enhancement dynamics can distinguish luminal B from other subtypes. The practical application of this result is to identify luminal B patients and adapt therapy accordingly. Furthermore, this result suggests that tumor enhancement dynamics in MRI might be predictive of outcomes.

Another approach to using radiogenomic analysis to address the issue of limited data is the topic of a recent article by Gevaert et al [17]. In this study, the authors had 2 sets of data on non—small-cell lung cancer. The first contained imaging and gene expression data (no outcomes); the second contained genomic and outcomes data (no imaging). The authors used specific genomic characteristics as input to a model that maps them into imaging features. Multiple models were constructed for the imaging features. Using the second data set, the authors tested how well the imaging features, based on specific genomic characteristics, predicted patient survival. The goal of the analysis, as stated by the authors, was to identify imaging features that are predictive of survival by evaluating their gene-based surrogates. They showed that this approach can identify imaging features that are related to survival outcomes.

Finally, beyond filling the gaps in knowledge, radiogenomics discoveries have a more basic significance of building a better understanding of the imaging representations of various molecular phenotypes, uncovering biological processes that are underlying phenotypes seen in imaging [2] (ie, causal relationship between the two), which could drive future discoveries in cancer research.

EXTRACTION OF IMAGING FEATURES

To conduct radiogenomic analysis, specific features have to be extracted from images. Extraction can be done by radiologists (or other qualified individuals), or with the minor assistance of computer programs [11,19]. This is typical in this early stage of research.

Lexicons

Reproducibility of results depends on use of a well-defined lexicon to guide extraction. The lexicon must contain a set of terms (features) that describe the tumor and its surroundings, along with definitions of those terms. An established example of a radiology lexicon is the ACR BI-RADS®, which, for example, contains descriptions of mammographic masses, such as "mass margin," with its 5

possible values: circumscribed, microlobulated, obscured, indistinct, and spiculated. Limiting readers to choosing 1 of the previously specified answers allows for better comparison of features across readers and cases. Another example of a radiology lexicon, developed specifically in the context of radiogenomics and imaging biomarker development is the VASARI [24] lexicon for the annotation of brain tumors. Specific terms include: major axis length, proportion of the enhancing tumor, and deep white matter invasion. The development of lexicons in radiology has recently become an area of substantial work and interest in ontology [25]. The goal is to develop lexicons such as RadLex [26] that could lend themselves naturally to radiogenomic analysis.

Although radiology lexicons facilitate the extraction of easily usable features for radiogenomic analysis, manual analysis of images has some disadvantages, most significantly, interobserver variability. Although the lexicon terms can be well defined, there are still often notable differences between the assessments of individual features by different radiologists. This introduces noise to radiogenomic analysis or simply means that a particular correlation of imaging features with genomics does not hold when different radiologists annotate images. Related to interobserver variability is the lack of precision when human observers conduct quantitative assessment of features such as mass volume, or volume of a specific part of a tumor. Another disadvantage is the significant time commitment required to assess individual cases. For example, the VASARI lexicon contains more than 20 features, many of which require inspection of more than one MRI sequence. This means that a skilled individual will have to expand a good deal of time analyzing an imaging examination.

Automatic and Semi-Automatic Feature Extraction

Automatic or semi-automatic feature extraction alleviates some of the issues of interobserver variability, limited precision, and time commitment associated with manual feature extraction. In automatic feature extraction, computer vision algorithms are used to automatically segment the abnormality and extract a variety of features. In semi-automatic segmentation, a large part of the segmentation and feature extraction is still carried out by a computer, but a human reader is involved in the process. An example is when a reader indicates roughly the boundaries of the tumor and a computer algorithm completes segmentation and feature extraction. Involvement of computer algorithms in the process allows for a more precise and more consistent assessment of quantitative features and extraction of a larger number of features, some of which might not be easily perceptible by a human eye. Although embraced in the field of radiomics, this approach to extraction of imaging features for radiogenomic analysis has not reached its full potential.

CONCLUSIONS AND DISCUSSION

The main limitation of radiogenomics is related to the fact that if strong imaging and outcomes data are available, and the prediction of outcomes is the primary goal, the radiogenomic analysis may not bring an additional contribution to the analysis. In such cases, imaging features can be directly correlated with the outcome of interest. However, in the most common situation, where the data is limited, radiogenomic analysis can be used, along with previously generated knowledge to identify imaging features that might be related to outcomes. Furthermore, elucidating the relationship between imaging and genomics will lead to a better understanding of cancer in general and may lead to improved treatment protocols. Finally, radiogenomic findings could guide the collection of future datasets by identifying imaging or genomics with a high potential to predict outcomes.

In this article, I focused on radiogenomic analysis directed at finding a relationship between genomics and imaging. However, a broader understanding of the term is sometimes applied to include any analysis that involves genomics and imaging. Various research directions within this broader theme and related to it are emerging and gaining popularity, some of which with very high significance. A very promising research direction is using computer- and radiologist-extracted features to directly predict patient outcomes [11], [27], [28]. Another involves combining imaging and genomic markers. Still another is traditional imaging-based surveillance of patients with higher disease risk determined by their genotype, for example higher breast cancer risk determined by BRCA1 or BRCA2 mutations.

A significant opportunity related to radiogenomics lies in the development of new statistical methodologies for the construction and evaluation of radiogenomic models. New methodology may instruct us on how to better leverage radiogenomic models in the development of clinically useful tools. As the discipline of radiogenomics matures, the significance of radiogenomic analysis will become clearer and new applications will undoubtedly emerge.

TAKE-HOME POINTS

- Radiogenomics investigates the relationship between disease genomic characteristics and its radiology phenotypes.
- In some scenarios, direct prediction of outcomes using imaging features might be preferable to radiogenomic analysis.
- Radiogenomics will play a significant role in cancer research as it creates a new avenue of generating important knowledge from limited data.

REFERENCES

1. Rutman AM, Kuo MD. Radiogenomics: creating a link between molecular diagnostics and diagnostic imaging. Eur J Radiol 2009;70:232-41.
2. Kuo MD, Jamshidi N. Behind the numbers: decoding molecular phenotypes with radiogenomics—guiding principles and technical considerations. Radiology 2014;270:320-5.
3. Rosenstein BS, West CM, Bentzen SM, et al. Radiogenomics: Radiobiology enters the era of big data and team science. Int J Radiat Oncol Biol Phys 2014;89:709.
4. West C, Rosenstein BS. Establishment of a radiogenomics consortium. Radiother Oncol 2010;94:117-8.
5. Lambin P, Rios-Velazquez E, Leijenaar R, et al. Radiomics: extracting more information from medical images using advanced feature analysis. Eur J Cancer 2012;48:441-6.
6. Kumar V, Gu Y, Basu S, Berglund A, et al. Radiomics: the process and the challenges. Magn Reson Imaging 2012;30:1234-48.
7. Aerts HJWL, Velazquez ER, Leijenaar RTH, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. Nat Commun 2014;5:4006.
8. Zinn PO, Majadan B, Sathyan P, et al. Radiogenomic mapping of edema/cellular invasion MRI-phenotypes in glioblastoma multiforme. PLoS One 2011;6:e25451.
9. Jamshidi N, Diehn M, Bredel M, Kuo MD. Illuminating radiogenomic characteristics of glioblastoma multiforme through integration of MR imaging, messenger RNA expression, and DNA copy number variation. Radiology 2013;270:1-2.
10. Gevaert O, Mitchell LA, Achrol AS, et al. Glioblastoma multiforme: exploratory radiogenomic analysis by using quantitative image features. Radiology 2014;273:168-74.
11. Gutman DA, Cooper LA, Hwang SN, et al. MRI predictors of molecular profile and survival: multi-institutional study of the TCGA glioblastoma data set. Radiology 2013;267:560-9.
12. Yamamoto S, Maki DD, Korn RL, Kuo MD. Radiogenomic analysis of breast cancer using MRI: a preliminary study to define the landscape. AJR Am J Roentgenol 2012;199:654-63.
13. Mazurowski MA, Zhang J, Grimm LJ, Yoon SC, Silber JI. Radiogenomic analysis of breast cancer: Luminal B molecular subtype is associated with enhancement dynamics at MR imaging. Radiology 2014;273:365-72.
14. Grimm LJ, Zhang J, Mazurowski MA. A computational approach to radiogenomics of breast cancer: Luminal A and luminal B molecular subtypes are associated with imaging features on routine breast MRI extracted using computer vision algorithms. J Magn Reson Imaging. 2015 Mar 17. http://dx.doi.org/10.1002/jmri.24879. [Epub ahead of print]
15. Ashraf AB, Gavenonis SC, Daye D, Mies C, Rosen MA, Kontos D. A multichannel Markov random field framework for tumor segmentation with an application to classification of gene expression-based breast cancer recurrence risk. IEEE Trans Med Imaging 2013;32:637-48.
16. Ashraf AB, Daye D, Gavenonis S, et al. Identification of intrinsic imaging phenotypes for breast cancer tumors: preliminary associations with gene expression profiles. Radiology 2014;272:374-84.
17. Gevaert O, Xu J, Hoang CD, Leung AN, Xu Y, Quon A, et al. Non–small cell lung cancer: identifying prognostic imaging biomarkers by leveraging public gene expression microarray data—methods and preliminary results. Radiology 2012;264(2):387-96.
18. Kuo MD, Gollub J, Sirlin CB, Ooi C, Chen X. Radiogenomic analysis to identify imaging phenotypes associated with drug response gene expression programs in hepatocellular carcinoma. J Vasc Interv Radiol 2007;18:821-30.
19. Karlo CA, Di Paolo PL, Chaim J, et al. Radiogenomics of clear cell renal cell carcinoma: associations between CT imaging features and mutations. Radiology 2014;270:464-71.
20. Lawrence MS, Stojanov P, Polak P, et al. Mutational heterogeneity in cancer and the search for new cancer-associated genes. Nature 2013;499:214-8.
21. O'Connor JPB, Rose CJ, Waterton JC, Carano RAD, Parker GJM, Jackson A. Imaging intratumor heterogeneity: role in therapy response, resistance, and clinical outcome. Clin Cancer Res 2014;21:249-57.
22. Cancer Imaging Archive. Available at: http://www.cancerimagingarchive.net/. Accessed May 29, 2015
23. National Cancer Institute. The Cancer Genome Atlas. Available at: http://cancergenome.nih.gov/. Accessed May 29, 2015.
24. Visually Accessable REMBRANDT [Repository for Molecular Brain Neoplasia Data] Images, Available at: http://cabig.cancer.gov/action/collaborations/vasari. Accessed May 29, 2015.
25. Rubin DL. Creating and curating a terminology for radiology: ontology modeling and analysis. J Digit Imaging 2008;21:355-62.
26. Langlotz CP. RadLex: a new method for indexing online educational materials. 1. Radiographics 2006;26:1595-7.
27. Mazurowski MA, Desjardins A, Malof JM. Imaging descriptors improve the predictive power of survival models for glioblastoma patients. Neuro Oncol 2013;15:1389-94.
28. Mazurowski MA, Zhang J, Peters KB, Hobbs H. Computer-extracted MR imaging features are associated with survival in glioblastoma patients. J Neurooncol 2014;120:483-8.