Imaging-Genomic Study of Head and Neck Squamous Cell Carcinoma: Associations Between Radiomic Phenotypes and Genomic Mechanisms via Integration of The Cancer Genome Atlas and The Cancer Imaging Archive

Yitan Zhu, PhD¹; Abdallah S.R. Mohamed, MD, MSc^{2,3}; Stephen Y. Lai, MD, PhD²; Shengjie Yang, PhD¹; Aasheesh Kanwar, MD²; Lin Wei, PhD¹; Mona Kamal, MD, PhD²; Subhajt Sengupta, PhD¹; Hesham Elhalawani, MD, MSc²; Heath Skinner, MD, PhD²; Dennis S. Mackin, PhD²; Jay Shiao, MD, MPH²; Jay Messer, MD²; Andrew Wong, MD²; Yao Ding, PhD²; Lifei Zhang, PhD²; Laurence Court, PhD²; Yuan Ji, PhD^{1,4}; and Clifton D. Fuller, MD, PhD²

abstract

PURPOSE Recent data suggest that imaging radiomic features of a tumor could be indicative of important genomic biomarkers. Understanding the relationship between radiomic and genomic features is important for basic cancer research and future patient care. We performed a comprehensive study to discover the imaging-genomic associations in head and neck squamous cell carcinoma (HNSCC) and explore the potential of predicting tumor genomic alternations using radiomic features.

METHODS Our retrospective study integrated whole-genome multiomics data from The Cancer Genome Atlas with matched computed tomography imaging data from The Cancer Imaging Archive for the same set of 126 patients with HNSCC. Linear regression and gene set enrichment analysis were used to identify statistically significant associations between radiomic imaging and genomic features. Random forest classifier was used to predict the status of two key HNSCC molecular biomarkers, human papillomavirus and disruptive *TP53* mutation, on the basis of radiomic features.

RESULTS Widespread and statistically significant associations were discovered between genomic features (including microRNA expression, somatic mutations, and transcriptional activity, copy number variations, and promoter region DNA methylation changes of pathways) and radiomic features characterizing the size, shape, and texture of tumor. Prediction of human papillomavirus and *TP53* mutation status using radiomic features achieved areas under the receiver operating characteristic curve of 0.71 and 0.641, respectively.

CONCLUSION Our exploratory study suggests that radiomic features are associated with genomic characteristics at multiple molecular layers in HNSCC and provides justification for continued development of radiomics as biomarkers for relevant genomic alterations in HNSCC.

Clin Cancer Inform. © 2019 by American Society of Clinical Oncology

INTRODUCTION

Head and neck squamous cell carcinomas (HNSCCs) prevail as the sixth most common cancer worldwide, with > 500,000 expected newly diagnosed cases reported annually.¹ In the United States, 40,000 new HNSCC cases are reported, with approximately 7,890 deaths per year.² HNSCCs encompass a diverse array of cancers that can originate from subsites within the oral cavity (44%), larynx (31%), or pharynx (25%).³ Viral infections, specifically human papillomavirus (HPV), primarily type 16, and Epstein-Barr virus, are associated with higher risk of oropharynx and nasopharynx cancers, respectively.^{4,5} Protracted tobacco and alcohol use and ultraviolet light exposure are

among the traditional risk factors for development of HNSCC.⁶ There has been a dramatic change in the affected patient cohort as risk factors have changed, represented by a decrease in tobacco use and concomitant increase in HPV-associated disease. This was reflected as a substantial rise in the incidence of HPV-associated oropharynx cancers as compared with a decline in cancers of the larynx and hypopharynx.⁷ Given the high morbidity and mortality associated with HNSCC, this type of cancer represents a major health burden.

The refinement in head and neck irradiation techniques, specifically introduction of intensity-modulated radiotherapy approximately 15 years ago, marked

ASSOCIATED CONTENT

Data Supplement

Author affiliations and support information (if applicable) appear at the end of this article.

Accepted on August 8, 2018 and published at ascopubs.org/journal/cci on February 7, 2019; DOI <https://doi.org/10.1200/CCI.18.00073>

a paradigm shift in HNSCC management that resulted in improvement of treatment outcomes.⁸ Continued efforts have been made to investigate potential prognostic and predictive biomarkers to establish the conceptual framework for precision medicine in management of HNSCC.⁹ One example is the exploration of the correlation between disruptive alteration of the gene encoding the tumor-suppressor protein p53 (*TP53*) and treatment failure, with subsequent decreased survival in patients with HNSCC.¹⁰

Radiographic images, such as computed tomography (CT), have been routinely used for diagnosis and treatment of HNSCC. However, the relationship between tumor imaging phenotypes and underlying tumor genomic mechanisms remains underexplored. Precise and effective treatment of cancer requires the integration of disease information from multiple sources. Imaging-genomic research combines radiographic image analysis with genomic research to improve disease diagnosis and prognosis, discover novel biomarkers, and identify genomic mechanisms associated with phenotype formation.¹¹⁻¹⁵ Such imaging-genomic studies have been performed for multiple cancer types, including breast invasive carcinoma,¹¹⁻¹⁵ lung cancer,^{16,17} glioblastoma multiforme,¹⁸ and clear cell renal cell carcinoma.¹⁹

To our knowledge, there are few existing imaging-genomic studies in HNSCC. One of the earliest studies, reported by Yang et al²⁰ in 2003, investigated the correlation between temporal changes in T1- and T2-weighted contrast-enhanced magnetic resonance imaging and genomic analysis using oligonucleotide microarrays in murine squamous cell carcinoma tumor models. Aerts et al¹⁷ developed a multifeature radiomic signature capturing intratumoural heterogeneity that was linked to gene expression patterns, validated in three independent data sets of patients with lung or head and neck cancer. Recently, Pickering et al²¹ correlated radiologist-selected CT imaging features of 27 oral cavity squamous cell carcinomas with the expression of cyclin D1, angiogenesis-related genes, and epidermal growth factor receptors.

In our study, we innovatively investigated the comprehensive relationship between the multilayer tumor genomic system and multiple aspects of tumor imaging phenotype for HNSCC. We integrated multiomics whole-genome data from The Cancer Genome Atlas (TCGA)²² with radiomic data derived on CT images from The Cancer Imaging Archive (TCIA)²³ for matched patients and identified statistically significant associations. We also explored the potential of using CT imaging as a noninvasive marker predicting tumor molecular status in HNSCC.

METHODS

Clinical, radiologic, and genomic data (Data Supplement) for 126 patients with HNSCC from TCGA and TCIA were integrated and analyzed. CT images of the patients were

downloaded from TCIA and processed using Imaging Biomarker Explorer,²⁴ a public medical image analysis software pipeline that can automatically generate tumor radiomic features. The radiomic features were grouped into five categories: gray-level cooccurrence matrix, gray-level run length matrix, neighbor intensity difference, intensity direct, and size/shape.¹⁷ The Data Supplement introduces how the radiomic features were generated. Multiomic genomic data and patient clinical information were acquired from TCGA using the open-source software tool TCGA-Assembler.²⁵ The collection and processing of genomic data are also described in the Data Supplement. Genomic, clinical, and radiomic data were integrated to form the imaging-genomic data (Table 1) for subsequent analysis. A total of 126 patient samples were used for analysis, representing all matched HNSCC cases in TCGA and TCIA databases.

A multistep informatic and statistical pipeline was built to perform integrative data processing and analysis (Fig 1). First, linear regression was used to identify statistically significant associations between radiomic features and gene-level genomic features, including expressions of microRNAs (miRNAs) and somatic mutations summarized at the gene level, adjusting for patient age, tumor grade, tumor subsite, and patient smoking status (Data Supplement). Second, for whole-genome measurements, including gene expressions, copy number variations (CNVs), and promoter region DNA methylation, we investigated associations with tumor radiomic features at the pathway level using a modified gene set enrichment analysis²⁶ scheme that was also adjusted for the same confounding factors (Data Supplement). The genetic pathways in consideration were from the Kyoto Encyclopedia of Genes and Genomes (KEGG)²⁷ database characterizing various aspects of the biomolecular system. Third, on the basis of radiomic features, random forest classifiers²⁸ were used to predict patient HPV status and *TP53* mutation status in HNSCC (Data Supplement).

RESULTS

A total of 5,347 statistically significant associations (adjusted $P \leq .05$) were identified between various radiomic and genomic features. The Data Supplement includes a graphical presentation of the identified associations. Table 2 lists the numbers of identified associations between different categories of genomic and radiomic features, on the basis of which Fisher's exact test^{29,30} indicated that the frequency of statistically significant associations depended on the feature category ($P < .001$), meaning some feature categories had more associations than others. The identified associations were statistically significantly enriched among pathway transcriptional activity and all five categories of radiomic features, with adjusted $P < .001$ (Data Supplement). This implies that transcriptional activity of genetic pathways modulates various aspects of tumor imaging phenotype.

TABLE 1. Summary of the Integrative Imaging-Genomic Data Used in the Analysis

Variable	
Data platform	No. of Features/Tumors
Radiomics	
Features	187
Tumors	126
miRNA expression	
miRNAs	1,046
Tumors	125
Mutated genes	
Genes	16,573
Tumors	122
Gene expression	
Genes	20,531
Pathways	179
Tumors	125
CNVs	
Genes	19,921
Pathways	179
Tumors	126
Promoter region DNA methylation	
Genes	19,325
Pathways	179
Tumors	126
Patient characteristic	No. of Tumors/Patients
AJCC tumor stage	
I	4
II	14
III	22
IV	86
Tumor site	
Oral cavity	69
Larynx	36
Oropharynx	21
Smoking behavior	
Current	52
Former	45
Never	29
Sex	
Male	97
Female	29
HPV status	
Positive	29
Negative	96
Not measured	1

(Continued in next column)

TABLE 1. Summary of the Integrative Imaging-Genomic Data Used in the Analysis (Continued)

Variable	
Disruptive <i>TP53</i> mutation	
Yes	33
No	89
Not measured	4
CT parameter	No. of Tumors/Patients
CT scanner manufacturer	
GE	61
Siemens	33
Philips	26
Toshiba	5
Picker	1
Slice thickness, mm	
Median	2.5
IQR	1.25-3
Contrast	All were contrast enhanced
Tube voltage, kVp	
120	98
130	7
135	1
140	20
Tube current range, mA	59-600
Data collection diameter, mm	
Median	500
IQR	320-500
Reconstruction diameter, mm	
Median	250
IQR	242-268
Other	
Age, years	
Mean	59.81
SD	11.28

NOTE. The No. of tumors for radiomics (ie, 126) is the No. of tumor cases with radiomic features. For genomic data platforms, the No. of tumors is the No. of tumor cases with both radiomic features and genomic data of the specific platform used in our study.

Abbreviations: AJCC, American Joint Committee on Cancer; CNV, copy number variation; CT, computed tomography; HPV, human papillomavirus; IQR, interquartile range; miRNA, microRNA; SD, standard deviation.

Associations Between Radiomic Features and Genetic Pathways

The Data Supplement describes all identified associations involving transcriptional activity, gene CNVs, and promoter region DNA methylation changes of all KEGG pathways.

FIG 1. Flowchart of processing The Cancer Genome Atlas (TCGA) and The Cancer Imaging Archive (TCIA) data and conducting the imaging-genomic analyses. CT, computed tomography; HPV, human papilloma-virus; IBEX, Imaging Biomarker Explorer; miRNA, microRNA.

Figure 2 specifically shows that radiomic features were associated with cancer-related KEGG pathways²⁷ covering multiple aspects of the cancer molecular system, such as signal transduction, cell growth and death, immune system, and cellular interactions and community. Figures 2A, 2B, and 2C show the associations of transcriptional activity, gene CNV, and promoter region DNA methylation change of cancer-related KEGG pathways, respectively. There are many interesting findings shown in Figure 2A indicating pathway transcriptional activity is correlated with and modulates multiple aspects of tumor imaging phenotype; these are described in the following sections.

Cell growth and death. Multiple associations related to cell growth and death were identified in our analysis. Transcriptional activity of ribosome genes was correlated with multiple aspects of tumor imaging phenotype, including tumor texture heterogeneity characterized by positive association with entropy and negative associations with energy 1, homogeneity, and homogeneity 2; tumor size features, including convex hull volume, convex hull volume three dimensional (3D), mass, maximum 3D diameter,

mean breadth, number of voxel, and surface area; and tumor shape irregularity, characterized by negative associations with roundness, sphericity, and convex and positive association with spherical disproportion. Ribosome genes support protein synthesis and are important for various cellular processes, such as cell proliferation and growth. Our result shows that they were more transcriptionally active in larger, more irregular, heterogeneous tumors. The apoptosis pathway played a tumor suppressive role by eliminating damaged or redundant cells through activation of caspases. Disruption or evasion of apoptosis can lead to tumor initiation, progression, or metastasis.³¹ Consistently, we found that transcriptional activity of the apoptosis pathway was negatively associated with tumor size (characterized by convex hull volume, convex hull volume 3D, maximum 3D diameter, mean breadth, and surface area) and tumor shape irregularity (characterized by its positive associations with convex and sphericity and negative association with spherical disproportion).

Immune system. Pathways related to immune regulation, including pathways of natural killer cell-mediated cytotoxicity,

TABLE 2. No. of Statistically Significant Associations Between Genomic Features of Different Platforms and Radiomic Features of Different Categories

Feature Category	Gray-Level Cooccurrence Matrix	Gray-Level Run Length Matrix	Neighbor Intensity Difference	Intensity Direct	Size/Shape
Transcriptional activity of pathway	1709	304	131	1413	884
CNV of pathway	196	37	28	161	62
Promoter region DNA methylation change of pathway	156	30	15	119	29
miRNA expression	6	0	0	10	0
Mutated gene	16	5	0	36	0

Abbreviations: CNV, copy number variation; miRNA, microRNA.

FIG 2. Statistically significant associations between radiomic features and (A) transcriptional activity of cancer-related genetic pathways, (B) gene copy number variations (CNVs) of cancer-related genetic pathways, and (C) gene promoter region DNA methylation changes of cancer-related genetic pathways. In each heat map, only genetic pathways and radiomic features with statistically significant associations are shown. Each of the gray-level cooccurrence matrix features can be calculated using different offset parameter values (ie, 1, 2, 3, 4, and 5), which results in five different instances of a feature. Because the five instances of a feature were usually correlated, the directions (ie, positive or negative) of the associations between a cancer-related pathway and the different instances of a radiomic feature were always the same. Thus, in the heat maps, associations between different instances of a radiomic feature and a pathway could be collapsed into one association. If a pathway had an association with at least one instance of a radiomic feature, the association between the pathway and the radiomic feature was included in the heat map. Percentile and quantile radiomic features from the intensity direct category were not included in the heat maps for simplicity, because they had many instances with different percentile or quantile values. 3D, three dimensional; ECM, extracellular matrix; JAK-STAT, Janus kinase–signal transducers and activators of transcription; MAPK, mitogen-activated protein kinase; mTOR, mammalian target of rapamycin; TGF- β , transforming growth factor β ; VEGF, vascular endothelial growth factor.

T-cell receptor signaling, B-cell receptor signaling, antigen processing and presentation, and chemokine signaling, were negatively associated with tumor size features. One possible explanation is that patients with larger tumors have a less active immune system and therefore are unable to effectively destroy tumor cells and curb tumor growth. Similarly, we found a correlation between immune system activity and tumor shape regularity, because the pathway activity was positively associated with sphericity and convex and negatively associated with spherical disproportion.

Cellular interactions and community. Pathways related to cell adhesion molecules, cytokine-cytokine receptor interaction, ECM-receptor interaction, adherens junction, gap junction, and focal adhesion regulate cell-cell interaction and signaling, acting as intercellular regulators and mobilizers of cells, and maintain cell and tissue architecture that limits cell movement and proliferation, which are two important factors in cancer progression. Aberrant activity of these pathways can lead to the development and metastasis of many types of cancer,

including HNSCC.³² We found that their activity was negatively associated with multiple tumor size features, indicating smaller tumors tend to have stronger activity in these pathways than larger tumors. The activity of all these pathways, except gap junction, was also correlated with tumor shape regularity, characterized by positive association with sphericity and negative association with spherical disproportion.

Signal transduction. Transcriptional activity of several molecular signaling pathways, including mitogen-activated protein kinase signaling, transforming growth factor β (TGF- β) signaling, JAK-STAT signaling, vascular endothelial growth factor signaling, WNT signaling, and ERBB signaling pathways, was negatively associated with tumor size features, indicating that they are more active in smaller tumors than larger tumors. A previous report³³ has suggested TGF- β signaling as a potent tumor suppressor in HNSCC, which is supported by its negative association with tumor size identified in our study. The activity of mitogen-activated protein kinase, TGF- β , JAK-STAT, and vascular endothelial growth factor signaling pathways was positively associated with tumor shape regularity.

Compared with pathway transcriptional activity, CNVs of cancer-related pathways had fewer statistically significant associations with radiomic features (Fig 2B). CNVs of JAK-STAT signaling pathway, cytokine-cytokine receptor interaction, natural killer cell-mediated cytotoxicity, and antigen processing and presentation genes were correlated with tumor shape regularity, characterized by their positive associations with convex and sphericity and negative association with spherical disproportion. CNVs of apoptosis genes were positively associated with tumor texture homogeneity, characterized by homogeneity and homogeneity 2, indicating tumors with heterogeneous texture may have fewer copies of apoptosis genes than tumors with homogeneous texture.

Figure 2C shows the statistically significant associations between radiomic features and promoter region DNA methylation changes of cancer-related pathways. DNA methylation changes of ribosome genes had the largest number of associations with radiomic features (first row in Fig 2C), including negative associations with two tumor size features (maximum 3D diameter and surface area) and positive association with tumor shape regularity (characterized by positive association with sphericity and negative association with spherical disproportion). The directions of these associations were opposite of those of the transcriptional activity of ribosome genes, which is expected, because methylation at promoter region usually negatively affects gene expression. In addition, we found that DNA methylation changes of three immune-related pathways (ie, natural killer cell-mediated cytotoxicity, T-cell receptor signaling pathway, and chemokine signaling pathway) were negatively associated with tumor shape regularity (Fig 2C). These are new results that may shed light on

the connection between immune pathways and radiomic phenotypes.

Details of the analysis scheme and other findings are described in the Data Supplement.

Associations Between Radiomic Features and miRNA Expression and Mutated Genes

MiRNAs. The Data Supplement describes statistically significant associations between miRNA expression and radiomic features. *Mir-320a* has been reported as a negative regulator of tumor invasion and metastasis.³⁴ Its expression correlated with tumor texture homogeneity, characterized by positive associations with homogeneity and homogeneity 2 and negative associations with entropy and global entropy. The radiomic feature global uniformity measures the overall homogeneity of tumor pixel intensity¹⁷ and was positively associated with the expression of eight miRNAs, including both antitumorigenic/antimetastatic and oncogenic miRNAs. The antitumorigenic/antimetastatic miRNAs include *mir-101*,³⁵ *mir-15b*,³⁶ and *mir-320a*; the oncogenic miRNAs include *mir-106b* and *mir-25*,³⁷ *mir-155*,³⁸ and *mir-378*³⁹; the last miRNA, *mir-7*, is involved in multiple cancer-related signaling pathways and has been reported to have both oncogenic and antitumorigenic roles.³⁷

Somatic mutations. The Data Supplement describes statistically significant associations between radiomic features and genes with somatic mutations in at least 10 patients. *EP300* encodes the E1A binding protein p300, a histone acetyltransferase regulating the transcription of genes involved in cell proliferation and differentiation. Mutations in *EP300* have been reported for HNSCC and may contribute to disease initiation and progression.⁴⁰ Our analysis shows somatic mutations in *EP300* were negatively associated with inverse variance and positively associated with median absolute deviation. *COL11A1* encodes one of the two alpha chains of type XI collagen, which is an essential component of the interstitial extracellular matrix. *COL11A1* may contribute to HNSCC tumorigenesis and be a potential therapeutic target.⁴¹ We found mutations in *COL11A1* were negatively associated with inverse variance.

We report the analysis schemes and other details regarding the identified associations involving miRNAs and somatic mutations in the Data Supplement.

Predictions of Patient HPV Status and Disruptive *TP53* Mutation Using Radiomic Features

We applied the random forest classifier²⁸ to predict patient HPV status on the basis of tumor radiomic features. Two-tier five-fold cross-validation was used to tune the classifier parameters and evaluate the generalization prediction performance. Predictive radiomic features were selected through a recursive feature elimination scheme. Table 3 lists the means and standard deviations of the areas under the receiver operating characteristic curve (AUCs) across 30 cross-validation trials, measuring prediction accuracy.

There was no significant difference between the average AUCs obtained using different numbers of features for prediction. The highest average AUC achieved was 0.71; however, the average AUC using only five features in each cross-validation trial still reached 0.706. Using the same classification and feature selection scheme, we predicted whether a tumor possessed any disruptive *TP53* mutation, a biomarker in HNSCC development and treatment.¹⁰ Loss-of-function alterations were dominant among the *TP53* mutations in the cancer cases. All disruptive *TP53* mutations were loss-of-function alterations; only one of the nondisruptive *TP53* mutations was a gain-of-function alteration. Table 3 lists the means and standard deviations of obtained AUCs. The highest average AUC was 0.641 with five features selected for prediction in each cross-validation trial. The Data Supplement provides details of the prediction and feature selection scheme and additional details on results, such as the most frequently selected features for prediction and their frequencies.

DISCUSSION

Using TCGA and TCIA data, we conducted an exploratory yet comprehensive imaging-genomic study. To our knowledge, this is the first study to integrate radiomic features of CT images with whole-genome measurements depicting multiple layers of the tumor molecular system in HNSCC. We report statistically significant associations between radiomic features characterizing multiple aspects of the tumor imaging phenotype and various genomic features (including transcriptional activity, CNV, DNA methylation, miRNA expression, and somatic mutation). The identified associations support existing knowledge related to HNSCC pathogenetic mechanisms and provide evidence for novel hypotheses on the relationship between tumor genomic mechanisms and subsequent tumor phenotypes, which can be validated in future studies. Also, we attempted to use radiomic features to predict important molecular biomarkers in HNSCC, such as HPV status and disruptive *TP53* mutation, with decent AUC values. These

results provide a basis for future investigations to establish the potential of using a noninvasive imaging approach to probe the genomic and molecular status of HNSCC.

Compared with pathway transcriptional activity, fewer statistically significant associations were identified for pathway CNVs and DNA methylation changes (Table 2; Fig 2). There could be two reasons for this. First, transcriptional activity is closer to phenotype formation than CNVs and DNA methylation in the process of the molecular system regulating the development of phenotype. Basically, transcriptional activity can more directly influence the generation of various phenotypes, whereas CNVs and DNA methylation changes may have to function through transcription. Secondly, DNA mutation events, such as CNVs and somatic mutations, are rarely shared across many patients, resulting in a small number of samples with the same mutation event, limiting the statistical power to identify potential associations.

Our study is based on CT images of 126 HNSCC samples and their multilayer whole-genome genomic data, forming a unique imaging-genomic data set that was not available before TCGA/TCIA era. This unique data set enabled our novel investigation of associations between tumor phenotypes and multiple molecular layers in HNSCC. To our knowledge, there is no other data set including matched imaging, multiomic genomic, and clinical data for HNSCC, as our data set does, on which we can repeat our analysis for validation. Our findings have been uploaded to <http://www.compgenome.org/Radiogenomics/> as a public resource to facilitate future research on HNSCC imaging-genomic associations. Future studies can either use our results as evidence to support their hypotheses or validate our findings through new analyses and experiments. Although unique and novel, our imaging-genomic data set is not large. Its sample size might limit the statistical power necessary to identify imaging-genomic associations and the accuracy to predict tumor molecular status on the basis of radiomic features. Nonetheless, we believe our study will pave the way for future HNSCC imaging-genomic investigations using more samples and advanced imaging technologies.

Bogowicz et al⁴² also used radiomic features to predict HPV status in HNSCC and achieved an AUC of 0.78, which is in a range similar to but higher than our HPV prediction accuracy (average AUC, 0.71). Multiple factors, such as different patient cohorts, could have contributed to the difference of prediction performance in the two studies. Considering the cohorts used in both studies were not large (sample size < 150), the obtained prediction performances indicate the potential of using imaging to probe tumor molecular status in the future, with the accumulation of imaging-genomic data and the development of imaging techniques.

More imaging-genomic analyses have been planned in HNSCC. One particularly interesting approach is to

TABLE 3. AUCs Obtained Through Two-Tier Five-Fold Cross-Validation Scheme

Prediction Target	AUC (No. of features)					
	All	100	50	20	10	5
HPV status						
Mean	0.701	0.71	0.697	0.7	0.71	0.706
SD	0.13	0.127	0.133	0.137	0.133	0.146
Disruptive <i>TP53</i> mutation						
Mean	0.587	0.594	0.624	0.627	0.62	0.641
SD	0.071	0.095	0.087	0.111	0.102	0.112

NOTE. Cross-validation scheme included 30 cross-validation trials when different numbers of radiomic features were selected for prediction in each cross-validation trial.

Abbreviations: AUC, area under the receiver operating characteristic curve; HPV, human papillomavirus; SD, standard deviation.

integrate genomic, epigenomic, and proteomic data simultaneously with imaging data to provide a more comprehensive depiction of how the multilayer molecular system regulates and produces various tumor imaging

phenotypes. Graphical models can be powerful tools for studying such a complex relationship because of their ability to model conditional dependence and competing regulatory factors.⁴³

AFFILIATIONS

¹NorthShore University HealthSystem, Evanston, IL

²The University of Texas MD Anderson Cancer Center, Houston, TX

³Alexandria University, Alexandria, Egypt

⁴The University of Chicago, Chicago, IL

Preprint version available on <https://www.biorxiv.org/content/early/2017/11/05/214312>.

CORRESPONDING AUTHOR

Yuan Ji, PhD, NorthShore University HealthSystem, 1001 University Place, Evanston, IL 60201; e-mail: koaeraser@gmail.com.

EQUAL CONTRIBUTION

Y.Z. and A.S.R.M. contributed equally to this work.

SUPPORT

Supported by National Institutes of Health (NIH) Grant No. 2R01CA132897 (Y.J.); in part by philanthropic donations from the family of Paul W. Beach to Dr G. Brandon Gunn (M.K., H.E.); by Grants No. 1R01DE025248-01 and R56DE025248-01 from the National Institute for Dental and Craniofacial Research, NIH (A.S.R.M., S.Y.L., C.D.F.); by Grant No. NSF 1557679 from the Division of Mathematical Sciences, National Science Foundation (NSF), Joint NIH/NSF Initiative on Quantitative Approaches to Biomedical Big Data and Award No. 1R01CA214825-01 from the NIH Big Data to Knowledge Program of the National Cancer Institute (NCI) Early Stage Development of Technologies in Biomedical Computing, Informatics, and Big Data (A.S.R.M., C.D.F.); by NIH/NCI Head and Neck Specialized Programs of Research Excellence Developmental Research Program Award No. P50CA097007-10, Paul Calabresi Clinical Oncology Program Award No. K12CA088084-06, a Center for Radiation Oncology Research at Anderson Cancer Center Seed Grant, the Anderson Institutional Research Grant Program, and Elekta (C.D.F.); and by 2016-2017 Radiological Society of North America Education and Research Foundation Research Medical Student Grant Award No. RSNA RMS1618 (A.K.). C.D.F. is a Sabin Family Foundation Fellow.

Written on behalf of M.D. Anderson Head and Neck Cancer Quantitative Imaging Working Group, in concert with The Cancer Imaging Archive.

None of the listed funders or in-kind support providers were privy to the content of the manuscript or the data or analysis provided herein. They had no prior oversight/preauthorization capacity regarding the content of the report/repository or authors' decision to submit.

REFERENCES

1. Torre LA, Bray F, Siegel RL, et al: Global cancer statistics, 2012. *CA Cancer J Clin* 65:87-108, 2015
2. American Cancer Society: Cancer Facts & Figures 2016. <https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/annual-cancer-facts-and-figures/2016/cancer-facts-and-figures-2016.pdf>
3. Li R, Agrawal N, Fakhry C: Anatomical sites and subsites of head and neck cancer, in Fakhry C, D'Souza G (eds): HPV and Head and Neck Cancers. Berlin, Germany, Springer, 2015, pp 1-11
4. Fakhry C, Zhang Q, Nguyen-Tan PF, et al: Human papillomavirus and overall survival after progression of oropharyngeal squamous cell carcinoma. *J Clin Oncol* 32:3365-3373, 2014
5. Jiang W, Chamberlain PD, Garden AS, et al: Prognostic value of p16 expression in Epstein-Barr virus-positive nasopharyngeal carcinomas. *Head Neck* 38:E1459-E1466, 2016 (suppl 1)

AUTHOR CONTRIBUTIONS

Conception and design: Yitan Zhu, Abdallah S.R. Mohamed, Yuan Ji, Clifton D. Fuller

Financial support: Yuan Ji, Clifton D. Fuller

Administrative support: Yuan Ji, Clifton D. Fuller

Collection and assembly of data: Yitan Zhu, Abdallah S.R. Mohamed, Shengjie Yang, Lin Wei, Mona Kamal, Subhjit Sengupta, Hesham Elhalawani, Heath Skinner, Dennis S. Mackin, Jay Shiao, Jay Messer, Andrew Wong, Yao Ding, Lifei Zhang, Yuan Ji, Clifton D. Fuller

Data analysis and interpretation: Yitan Zhu, Abdallah S.R. Mohamed, Stephen Y. Lai, Aasheesh Kanwar, Lin Wei, Laurence Court, Yuan Ji, Clifton D. Fuller

Manuscript writing: All authors

Final approval of manuscript: All authors

Accountable for all aspects of the work: All authors

AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST

The following represents disclosure information provided by authors of this manuscript. All relationships are considered compensated. Relationships are self-held unless noted. I = Immediate Family Member, Inst = My Institution. Relationships may not relate to the subject matter of this manuscript. For more information about ASCO's conflict of interest policy, please refer to www.asco.org/rwc or ascopubs.org/jco/site/ifc.

Stephen Y. Lai

Consulting or Advisory Role: Navidea, Cardinal Health

Travel, Accommodations, Expenses: Navidea

Dennis S. Mackin

Patents, Royalties, Other Intellectual Property: Patent pending related to range verification in proton therapy (unrelated to research in this report), "Techniques for Producing an Image of Radioactive Emissions Using a Compton Camera," Application No. 15/738597

Yuan Ji

Stock and Other Ownership Interests: Laiya Consulting

Clifton D. Fuller

Honoraria: Elekta

Research Funding: Elekta (Inst), RaySearch Laboratories (Inst)

Patents, Royalties, Other Intellectual Property: Patent application pending on unrelated medical device (Inst)

Travel, Accommodations, Expenses: Elekta

No other potential conflicts of interest were reported.

6. Maasland DH, van den Brandt PA, Kremer B, et al: Alcohol consumption, cigarette smoking and the risk of subtypes of head-neck cancer: Results from the Netherlands Cohort Study. *BMC Cancer* 14:187, 2014
7. Sturgis EM, Cinciripini PM: Trends in head and neck cancer incidence in relation to smoking prevalence: An emerging epidemic of human papillomavirus-associated cancers? *Cancer* 110:1429-1435, 2007
8. Nutting CM, Morden JP, Harrington KJ, et al: Parotid-sparing intensity modulated versus conventional radiotherapy in head and neck cancer (PARSPORT): A phase 3 multicentre randomised controlled trial. *Lancet Oncol* 12:127-136, 2011
9. Dahiya K, Dhankhar R: Updated overview of current biomarkers in head and neck carcinoma. *World J Methodol* 6:77-86, 2016
10. Skinner HD, Sandulache VC, Ow TJ, et al: TP53 disruptive mutations lead to head and neck cancer treatment failure through inhibition of radiation-induced senescence. *Clin Cancer Res* 18:290-300, 2012
11. Zhu Y, Li H, Guo W, et al: Deciphering genomic underpinnings of quantitative MRI-based radiomic phenotypes of invasive breast carcinoma. *Sci Rep* 5:17787, 2015
12. Guo W, Li H, Zhu Y, et al: Prediction of clinical phenotypes in invasive breast carcinomas from the integration of radiomics and genomics data. *J Med Imaging (Bellingham)* 2:041007, 2015
13. Li H, Zhu Y, Burnside ES, et al: Quantitative MRI radiomics in the prediction of molecular classifications of breast cancer subtypes in the TCGA/TCIA data set. *NPJ Breast Cancer* 2:16012, 2016
14. Li H, Zhu Y, Burnside ES, et al: MRI radiomics signatures for predicting the risk of breast cancer recurrence as given by research versions of gene assays of MammaPrint, Oncotype DX, and PAM50. *Radiology* 281:382-391, 2016
15. Burnside ES, Drukker K, Li H, et al: Using computer-extracted image phenotypes from tumors on breast magnetic resonance imaging to predict breast cancer pathologic stage. *Cancer* 122:748-757, 2016
16. Gevaert O, Xu J, Hoang CD, et al: Non-small cell lung cancer: Identifying prognostic imaging biomarkers by leveraging public gene expression microarray data: Methods and preliminary results. *Radiology* 264:387-396, 2012
17. Aerts HJ, Velazquez ER, Leijenaar RT, et al: Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun* 5:4006, 2014 [Erratum: *Nat Commun* 5:4644, 2014]
18. Jamshidi N, Diehn M, Bredel M, et al: Illuminating radiogenomic characteristics of glioblastoma multiforme through integration of MR imaging, messenger RNA expression, and DNA copy number variation. *Radiology* 270:1-2, 2014
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* 270:464-471, 2014
20. Yang YS, Guccione S, Bednarski MD: Comparing genomic and histologic correlations to radiographic changes in tumors: A murine SCC VII model study. *Acad Radiol* 10:1165-1175, 2003
21. Pickering CR, Shah K, Ahmed S, et al: CT imaging correlates of genomic expression for oral cavity squamous cell carcinoma. *AJNR Am J Neuroradiol* 34:1818-1822, 2013
22. Cancer Genome Atlas Network: Comprehensive genomic characterization of head and neck squamous cell carcinomas. *Nature* 517:576-582, 2015
23. Clark K, Vendt B, Smith K, et al: The Cancer Imaging Archive (TCIA): Maintaining and operating a public information repository. *J Digit Imaging* 26:1045-1057, 2013
24. Zhang L, Fried DV, Fave XJ, et al: IBEX: An open infrastructure software platform to facilitate collaborative work in radiomics. *Med Phys* 42:1341-1353, 2015
25. Zhu Y, Qiu P, Ji Y: TCGA-Assembler: Open-source software for retrieving and processing TCGA data. *Nat Methods* 11:599-600, 2014
26. Subramanian A, Tamayo P, Mootha VK, et al: Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA* 102:15545-15550, 2005
27. Kanehisa M, Goto S, Sato Y, et al: KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res* 40(D1):D109-D114, 2012
28. Ho TK: The random subspace method for constructing decision forests. *IEEE Trans Pattern Anal Mach Intell* 20:832-844, 1998
29. Mehta CR, Patel NR: Algorithm 643. FEXACT: A Fortran subroutine for Fisher's exact test on unordered $r \times c$ contingency tables. *ACM Trans Math Softw* 12:154-161, 1986
30. Clarkson DB, Fan Y, Joe H: A remark on algorithm 643: FEXACT—An algorithm for performing Fisher's exact test in $r \times c$ contingency tables. *ACM Trans Math Softw* 19:484-488, 1993
31. Wong RS: Apoptosis in cancer: From pathogenesis to treatment. *J Exp Clin Cancer Res* 30:87, 2011
32. Markwell SM, Weed SA: Tumor and stromal-based contributions to head and neck squamous cell carcinoma invasion. *Cancers (Basel)* 7:382-406, 2015
33. Bian Y, Hall B, Sun ZJ, et al: Loss of TGF- β signaling and PTEN promotes head and neck squamous cell carcinoma through cellular senescence evasion and cancer-related inflammation. *Oncogene* 31:3322-3332, 2012
34. Xie N, Wang C, Zhuang Z, et al: Decreased miR-320a promotes invasion and metastasis of tumor budding cells in tongue squamous cell carcinoma. *Oncotarget* 7:65744-65757, 2016
35. Tu HF, Lin SC, Chang KW: MicroRNA aberrances in head and neck cancer: Pathogenetic and clinical significance. *Curr Opin Otolaryngol Head Neck Surg* 21:104-111, 2013
36. Iorio MV, Croce CM: MicroRNA dysregulation in cancer: Diagnostics, monitoring and therapeutics—A comprehensive review. *EMBO Mol Med* 4:143-159, 2012
37. Sethi N, Wright A, Wood H, et al: MicroRNAs and head and neck cancer: Reviewing the first decade of research. *Eur J Cancer* 50:2619-2635, 2014
38. Ramdas L, Giri U, Ashorn CL, et al: miRNA expression profiles in head and neck squamous cell carcinoma and adjacent normal tissue. *Head Neck* 31:642-654, 2009
39. Yu BL, Peng XH, Zhao FP, et al: MicroRNA-378 functions as an onco-miR in nasopharyngeal carcinoma by repressing TOB2 expression. *Int J Oncol* 44:1215-1222, 2014
40. Martin D, Abba MC, Molinolo AA, et al: The head and neck cancer cell oncogenome: A platform for the development of precision molecular therapies. *Oncotarget* 5:8906-8923, 2014
41. Sok JC, Lee JA, Dasari S, et al: Collagen type XI $\alpha 1$ facilitates head and neck squamous cell cancer growth and invasion. *Br J Cancer* 109:3049-3056, 2013
42. Bogowicz M, Riesterer O, Ikenberg K, et al: Computed tomography radiomics predicts HPV status and local tumor control after definitive radiochemotherapy in head and neck squamous cell carcinoma. *Int J Radiat Oncol Biol Phys* 99:921-928, 2017
43. Zhu Y, Xu Y, Helseth DL Jr, et al: Zodiac: A comprehensive depiction of genetic interactions in cancer by integrating TCGA data. *J Natl Cancer Inst* 107:djv129, 2015