Changes in CT Radiomic Features Associated with Lymphocyte Distribution Predict Overall Survival and Response to Immunotherapy in Non–Small Cell Lung Cancer

Mohammadhadi Khorrami1, Prateek Prasanna1, Amit Gupta2, Pradnya Patil3, Priya D. Velu4, Rajat Thawani5, German Corredor1, Mehdi Aililou1, Kaustav Bera1, Pingfu Fu6, Michael Feldman7, Vamsidhar Velcheti8, and Anant Madabhushi1,9

ABSTRACT

No predictive biomarkers can robustly identify patients with non–small cell lung cancer (NSCLC) who will benefit from immune checkpoint inhibitor (ICI) therapies. Here, in a machine learning setting, we compared changes ("delta") in the radiomic texture (DelRADx) of CT patterns both within and outside tumor nodules before and after two to three cycles of ICI therapy. We found that DelRADx patterns could predict response to ICI therapy and overall survival (OS) for patients with NSCLC. We retrospectively analyzed data acquired from 139 patients with NSCLC at two institutions, who were divided into a discovery set ($D_1 = 50$) and two independent validation sets ($D_2 = 62, D_3 = 27$). Intranodular and perinodular texture descriptors were extracted, and the relative differences were computed. A linear discriminant analysis (LDA) classifier was trained with 8 DelRADx features to predict RECIST-derived response. Association of delta-radiomic risk score (DRS) with OS was determined. The association of DelRADx features with tumor-infiltrating lymphocyte (TIL) density on the diagnostic biopsies ($n = 36$) was also evaluated. The LDA classifier yielded an AUC of 0.88 ± 0.08 in distinguishing responders from nonresponders in $D_1$, and 0.85 and 0.81 in $D_2$ and $D_3$. DRS was associated with OS [HR: 1.64; 95% confidence interval (CI), 1.22–2.11; $P = 0.0011$; C-index = 0.72]. Peritumoral Gabor features were associated with the density of TILs on diagnostic biopsy samples. Our results show that DelRADx could be used to identify early functional responses in patients with NSCLC.

Introduction

Immune checkpoint inhibitors (ICI) targeting the PD-1/PD-L1 axis are the standard of care for treatment of patients with advanced non–small cell lung cancer (NSCLC) without targetable genetic alterations (1). Susceptibilities in immune checkpoint pathways allow tumor cells to escape immunosurveillance, leading to tumor propagation. mAbs targeting these pathways reinvigorate the host immunity against tumor cells and have revolutionized the treatment of a range of malignancies due to their favorable toxicity profiles and their ability to produce durable clinical responses (2). However, the response rates to these drugs remain modest [27% in PD-L1− NSCLC in the first-line setting (3) and 45% in PD-L1high subgroup (4), and 19% in the second-line setting]. In addition, optimal patient selection for therapy remains a challenge because most available biomarkers such as PD-L1 expression are insufficient to accurately classify patients. Historically, clinical efficacy of antitumor therapy has been evaluated by using RECIST or World Health Organization (WHO) criteria, which incorporate two-dimensional measures of target lesions that are tracked over time (5). These criteria have been used to define responses to therapy in a majority of the clinical trials that has led to subsequent drug approval for years (6–8). As per RECIST, at least a 20% increase in the sum of the longest diameter of target lesions or development of new ones is considered disease progression. RECIST guidelines focus on changes in tumor size and do not take into account changes in tumor heterogeneity, which may be more indicative of tumor biology and evolution during therapy.

Radiomics is the science of quantifying patterns of tumor phenotypes on radiographic scans in a high-throughput manner and analyzing them with bioinformatics tools to build clinically relevant models that assess tumor and microenvironment heterogeneity. Indeed, radiomic features in regions around the tumor might reflect immune response in breast and lung cancers (9–11). In addition to intranodular heterogeneity in PD-L1 status (12–15), microenvironmental factors in the peritumoral region may be contributing to treatment failure. For instance, angiogenesis, a pathologic response to hypoxia, is routinely observed adjacent to the tumor. Impaired vasculature around the tumor can lead to regions of hypoxia in tumor microenvironment,
which diminish the efficacy of drugs (16, 17). With this as a basis, we hypothesized that delta-radiomic analysis (DelRADx) or changes in radiomic patterns from the intra- and perinodular regions of the CT scan could provide a more accurate characterization of tumor response to ICIs in advance of visually perceptible imaging changes.

In this study, we evaluate the performance of DelRADx from within and outside thoracic lesions, in predicting response to multiple ICIs in patients with advanced NSCLC. Previous studies have begun to look at the role of radiomics in the context of predicting response to immunotherapy (18). Here, we look at delta-radiomic features (change in radiomic measures between pre- and posttherapy scans) to assess RECIST-derived response to ICIs and characterize survival differences. We evaluated the efficacy of DelRADx on a total of \( n = 139 \) cases collected across two different institutions, which included different ICI agents (nivolumab, atezolizumab, and pembrolizumab). To evaluate the morphologic basis of the DelRADx features, we also investigated associations between the predictive DelRADx features and computerized estimations of density of tumor-infiltrating lymphocytes (TIL) on corresponding digitized pretreatment diagnostic hematoxylin and eosin (H&E) biopsy scans (\( n = 36 \)). We also compared the efficacy of DelRADx against that of PD-L1 expression in predicting OS and response to ICIs. Our study demonstrated that dynamic textural changes, both inside and outside the nodule between baseline and posttreatment CT scans of patients with NSCLC following ICI therapy, correlate with clinical benefit. These subvisual morphologic changes could potentially be used to identify early clinical response before objective reduction in tumor burden. Because these delta-radiomic features are also associated with overall survival (OS), they could be used to identify patients who are likely to derive clinical benefit from therapy beyond functional responses. This could be valuable in the clinical decision-making process for patients with long-term disease stabilization or those with delayed radiographic responses due to immunotherapy.

**Materials and Methods**

**Datasets and patient selection**

The study was conducted in full accordance with the U.S. Common Rule and Health Insurance Portability and Accountability Act (HIPAA) regulations after approval from the Institutional Review Board (IRB 02-13-42C) at Case Western Reserve University (Cleveland, OH). The IRB waived the requirements for informed consent because the study was performed on retrospective, archived specimens. A total of \( n = 218 \) patients with advanced NSCLC treated with a PD-1/ PD-L1 ICI therapy (nivolumab/pembrolizumab/atezolizumab) from January 2012 to August 2017 in Cleveland Clinic Foundation (CCF) and University of Pennsylvania Health System (UPHS) were identified. Samples in which the board-certified radiologist (8 years of experience) could not isolate index lesions on CT scans or the CT scans with poor image quality, not suitable for radiomic feature extraction, led to exclusion of 76 of the 188 patients from CCF and 3 of the 30 patients from UPHS. The study included the remaining 112 patients from CCF and 27 patients from UPHS, for a total of 139 patients. Digitized histology scans of baseline biopsies were available for 36 cases from CCF. All patients underwent a baseline contrast CT imaging before starting treatment with ICI. All patients in the included study cohorts received ICI monotherapy. After two (for nivolumab) to three cycles (for atezolizumab/pembrolizumab) of ICI, a follow-up CT scan was acquired. The timeline of ICI therapy administration and corresponding CT acquisition is shown in [Fig. 1](#).

The entire cohort of 139 patients was divided into a discovery set, \( D_1 \), comprising 50 patients (25 responders and 25 nonresponders) from CCF. The remaining 62 cases (42 responders and 20 nonresponders) from CCF, not used in the discovery set, were used for independent validation (\( D_2 \)). Twenty-seven patients from UPHS (10 responders and 17 nonresponders) were used as the second independent validation set, \( D_3 \). The patients were selected in a manner that ensured an equal number of responders and nonresponders in the discovery set.

---

**Figure 1.**
ICI treatment timeline and flowchart of proposed delta-radiomics machine learning framework. Note that for some agents, there were three cycles instead of two.
Clinical endpoints
The primary endpoint of our study was response status (responder/nonresponder), as defined by RECIST v1.1. Patients who had progressive disease after treatment (PD, defined as at least 20% increase in the sum of largest diameters of target lesions with an absolute increase of ≥5 mm, counting any new lesions) were classified as “nonresponders,” and patients who had radiographic response, including complete response, partial response or stable disease, and clinical improvement, were classified as “responders” (5). The secondary endpoint was OS that was measured from the date from the posttreatment CT scan to the date of death, and censored at the date of last follow-up for survivors.

CT acquisition and segmentation
CT scans were acquired from all participants at baseline and immediately after two to three cycles (6–8 weeks) of ICI treatment. Scans were acquired using a multi-slice (Philips Healthcare, General Electric Healthcare, Siemens Healthcare) CT system with a tube voltage of 100 to 120 kVp, slice thickness (spacing) of 1 to 5 mm (mean = 2.82 mm, SD = 0.71 mm), and in-plane resolution of 0.75 × 0.75 mm. CT images were acquired with patients at inspiration breath-hold after contrast injection. All scans were acquired using the facilities’ CT chest protocol and standard image reconstruction.

Target lung nodules on pre- and posttreatment CT scans were first annotated with 3D SLICER software by a cardiothoracic radiologist (AG, 8 years’ experience). The segmented nodules were used to compute the intranodular texture and shape features. The imaging features were computed in 2D, whereas shape features were calculated as described in ref. 10. Tumor measurements were then reviewed to assess responses of each target lesion. The same thresholds for response and progression for the total tumor burden were applied for the lesion-based assessment. The nodule mask was then dilated out to a 30-mm perinodular radius, which was divided into 15 annular rings of 2 mm each. The choice of perinodular compartment size was determined on the basis of National Comprehensive Cancer Network guidelines (19) that suggests a resection margin of 2 cm or greater. During the peritumoral texture analysis of the lung parenchyma, care was taken to remove air (<900 HU) and mediastinal muscle pixels (≥100 HU) in peripheral tumor. To avoid edge artifacts that might arise during feature extraction, the “dead” pixels of the CT scan were substituted with the average pixel intensity from a 9 × 9 window surrounding the pixel of interest.

Radiomic feature extraction
After CT acquisition and segmentation, radiomic features were extracted using MATLAB 2018b (Mathworks) with an in-house developed toolbox (9, 19). Radiomic features were extracted from within and around the voxels of the segmented nodules from the CT scans. Ninety-nine texture features, from 4 feature groups (25 Laws Energy, 48 multi-scale multi-orientation Gabor, 13 Haralick, and 13 CoLIAGe features) were extracted from each region. Statistics (mean, median, SD, skewness, and kurtosis) for each feature were computed within the tumor and each perinodular annulus, resulting in 495 statistical features per region. A total of 24 shape features were also automatically extracted from the annotated nodules and investigated in the study. All of these features have been used in previous radiomic studies for discriminating cancer phenotypes (20–23). The extracted radiomic feature intensities were then normalized to adjust values to a notionally common scale between −1 and 1. To find which features were most distinct between responders and nonresponders, the percentage change of feature statistics between baseline and after 6 to 8 week scans was calculated to yield the DelRADx feature set.

Feature selection
To construct generalizable classifiers, the features identified as predictive must be consistent across sites and scanners. To identify which of the predictive features identified during the feature selection step were also stable and reproducible, we evaluated the DelRADx features in the context of the test-retest RIDER lung CT dataset (24). The RIDER dataset includes 31 patients, each patient having been scanned twice on a CT scanner with an interval of approximately 15 minutes. For each radiomic feature identified via feature selection, the intraclass correlation coefficient (ICC) was calculated to quantify reproducibility between the test–retest scans. Because ICC describes similarity of units in the same group, features with high ICC values are thus more reproducible and potentially more robust to variations in CT scanners and acquisition parameters. We used an ICC cutoff >0.8 to identify stable and reproducible features. Within all stable features, Wilcoxon rank sum (WLX) feature selection method was used to select predictive features.

Statistical analysis
Classification
A linear discriminant analysis (LDA) classifier was employed in conjunction with the stable and discriminating DelRADx features identified during the feature selection step. Within the discovery set D1, the classifier was trained with top selected features on 70% of data and validated on the remaining 30%. The procedure was iterated over 100 runs. The performance of the DelRADx feature classifier was assessed from the ROC curve as AUC. In addition, we employed unsupervised clustering to measure the efficacy of the identified DelRADx features. A heatmap clustergram was used to display unsupervised hierarchical clustering using the intranodal and perinodular DelRADx feature sets. A consensus clustering approach (Kmax = 10, Pearson distance, hierarchical clustering) was also used to determine the number and affiliation of possible clusters within the patients in the discovery set. To do this, the similarity between different nodules based on distance in the space of the top ranked features was first determined. Nodules belonging to different clusters will have lower correlation, whereas nodules within a cluster are likely to have high intraclass correlation.

There may possibly be a difference in tumor texture phenotype in patients who get two cycles as compared with three cycles of therapy. To ascertain whether there was a significant difference between two groups, an LDA classifier trained on the discovery set was evaluated in two validation sets, corresponding to patients who got two cycles (N1) and three cycles of therapy (N2), respectively. The number of correct and incorrect classifications decisions was calculated, and Fisher exact test (25) was performed to assess for any significant differences.

Because CT studies have a range of slice thicknesses (from 1 to 5 mm), the impact of slice thickness on the performance of the classifier was also evaluated.

Survival analysis
The Kaplan–Meier survival analysis and log-rank statistical tests (26) were performed to assess the univariable discriminative ability of the features. The prognostic value of DelRADx features on OS and independent “added-value” of the imaging features over the traditional clinical factors was estimated by using the DelRADx risk score (DRS) signature. To build the multivariable DRS, a Cox
regression model was trained on D1 for the selected imaging variables (top eight) and the prediction by this model was validated on D2 and D3. Backward elimination feature selection was performed on selected prognostic features in multivariable analysis. Intermediate models were tested by repeated random subsampling-based cross-validation with 100 iterations on the discovery set. Once the mean concordance index (C-index) of the shrinking model dropped, a linear validation with 100 iterations on the discovery set. Once the mean concordance index (C-index) of the shrinking model dropped, a linear combination of the selected features with their corresponding coefficients was calculated as the final DRS model for predicting OS. The patients were stratified on the basis of median DRS into high- and low-risk groups. A multivariable Cox proportional hazards model was employed to evaluate the ability of the DRS in predicting OS (by using survfit and coxph functions, respectively, in R, version 3.1.3). In addition, relative HRs with 95% confidence intervals (CI) were calculated using the Wald test and the G-rho rank test, respectively.

Morphologic association with predictive radiomic features
A set of context-based features that attempt to model the tumor environment and the relationships between lymphocytes and their surrounding cells (lymphocytes and non-lymphocytes) were explicitly defined. A watershed-based algorithm (27) was first applied to segment nuclei on the image. Considering that lymphocyte nuclei are generally distinguished from other cell nuclei by their smaller size, more rounded shape, and darker homogeneous staining, we classified the segmented nuclei into either lymphocytes or non-lymphocytes (mainly, tumor cells) using nuclei texture, shape, and color features (28). Subsequently, density-related features were extracted that characterized the lymphocyte distribution in local neighborhoods. The density features included ratio between the number of TILs and the tissue area, ratio between the area covered by TILs to the total area of the corresponding region of interest, ratio between the number of TILs and the number of non-TILs within a region of interest, and a grouping value indicating how close the TILs are to each other (computed as the sum of the inverse distances between TILs). Consequently, 76 features quantifying density or compactness of TILs were extracted from the surgical specimens. To investigate the DelRADx-TIL associations, a pairwise Spearman correlation was performed between each of the top DelRADx and TIL compactness measures.

Results
Clinicopathologic features of the patient cohorts
The Eastern Cooperative Oncology Group performance status and tumor–node–metastasis (TNM) stage per the American Joint Committee on Cancer staging system (29, 30) were used in this study. Clinical tumor histology (adenocarcinoma or squamous cell carcinoma) was also performed according to Union for International Cancer Control 7th classification. Detailed demographics and clinical characteristics for the 112 patients from CCF and 27 cases from UPHS are summarized in Tables 1 and 2, respectively.

DelRADx discrimination of responders from nonresponders after ICI therapy
The top eight selected intranodular and annular perinodular DelRADx texture features are listed in Table 3. Figure 2A and B illustrates the discriminability of the intranodular Haralick entropy DelRADx feature for representative nonresponder and responder patients before and after two cycles of ICI therapy. We observed an elevated expression of Haralick entropy posttherapy in the nonresponder as compared with the responder. Figure 2C and D illustrates the discriminability of a perinodular low-frequency Gabor filter DelRADx feature (24–26 mm) for one nonresponder and one responder patient, before and after two cycles of ICI therapy. In the postradiation scan, the responder showed greater feature expression than the nonresponder. This trend is also reflected in the box and whisker plots of the skewness of Haralick entropy, illustrated in Fig. 3A.

Table 1. Demographics and clinical characteristics for the 112 cases from CCF, categorized by responders and nonresponders.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>All patients (n = 112)</th>
<th>Responders (n = 67)</th>
<th>Nonresponders (n = 45)</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>58 (52%)</td>
<td>31</td>
<td>27</td>
<td>0.57</td>
</tr>
<tr>
<td>Female</td>
<td>54 (48%)</td>
<td>36</td>
<td>18</td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median (range)</td>
<td>65 (42–83)</td>
<td>65 (42–81)</td>
<td>65 (46–83)</td>
<td>0.27</td>
</tr>
<tr>
<td>Race</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>92 (82%)</td>
<td>55</td>
<td>37</td>
<td>0.99</td>
</tr>
<tr>
<td>Black</td>
<td>20 (18%)</td>
<td>12</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>Smoking</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never</td>
<td>16 (14%)</td>
<td>7</td>
<td>9</td>
<td>0.18</td>
</tr>
<tr>
<td>Former or current</td>
<td>96 (86%)</td>
<td>60</td>
<td>36</td>
<td></td>
</tr>
<tr>
<td>Histology</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adenocarcinoma</td>
<td>80 (71%)</td>
<td>51</td>
<td>29</td>
<td>0.39</td>
</tr>
<tr>
<td>Squamous cell carcinoma</td>
<td>24 (22%)</td>
<td>12</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>Others</td>
<td>8 (7%)</td>
<td>4</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>EGFR mutation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>7</td>
<td>5</td>
<td>2</td>
<td>0.7</td>
</tr>
<tr>
<td>No</td>
<td>105</td>
<td>62</td>
<td>43</td>
<td></td>
</tr>
<tr>
<td>ALK mutation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wild type</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TNM staging</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Distant metastasis (M1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
We also used clustering approaches to illustrate the discriminability of the identified features. Figure 3B shows the feature expression-based clustergram using the most discriminating intranodular and perinodular DelRADx features. As illustrated in Fig. 3B, a number of the texture features show differential expression between responders and nonresponders. Figure 3C illustrates the consensus clustering plot of features identified to be both stable and discriminating of response to ICI on the discovery set. Clustering performed using only intranodular texture features, had a preponderance of responders (75%) and nonresponders (84%), respectively, and had noticeably stronger consensus within clusters, as compared with the clusters using just intranodular texture features. The combination of top identified DelRADx features yielded an AUC of 0.88 ± 0.09 (n = 50) in the discovery set, D1. On the validation sets D2 and D3, the DelRADx classifier yielded an AUC of 0.85 (n = 62) and 0.81 (n = 27), respectively. The corresponding response prediction accuracy for the DelRADx classifier was 88%, 80%, and 69%, for patients treated with nivolumab, atezolizumab, and pembrolizumab in D2, and 75%, 70%, and 75% for patients in D3, respectively.

Supplementary Table S1 shows the performance of the model in terms of AUC mapped out as a function of slice thickness. As may be observed, the AUC of classifier for the radiomic features decreased slightly with increasing slice thickness.

**DeiRADx features were associated with OS**

The median OS post therapy for D1 was 16 months (range: 1–45 months). Supplementary Table S2 shows the univariable Cox regression analysis on D1 with the top eight DelRADx features and six clinicopathologic variables. The univariable Cox regression model based on the intranodular Haralick entropy texture feature gave the highest HR among all other features in the discovery set (HR for survival in patients with decrease in Haralick entropy after therapy: 1.54; 95% CI, 1.19–2.27; P = 0.0023; C-index = 0.57). There was a statistically significant difference in survival curves for patients (in the D2 and D3 validation sets) who had Haralick entropy values below the median, as compared with those above the median (D2: P = 0.0027, D3: P = 0.0018). Also, in D2, there was no significant difference in OS with age, gender, race, smoking status, histology, or EGFR mutation (age: P = 0.55, gender: P = 0.35, race: P = 0.81, smoking: P = 0.53, histology: P = 0.26, and EGFR mutation: P = 1).

The DRS was associated with OS in the discovery set, D1 (HR: 1.64; 95% CI, 1.22–2.21; P = 0.0011; C-index = 0.72) and yielded similar

<table>
<thead>
<tr>
<th>Table 3. The most stable and discriminative DeiRADx features.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Feature family</strong></td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>5</td>
</tr>
<tr>
<td>6</td>
</tr>
<tr>
<td>7</td>
</tr>
<tr>
<td>8</td>
</tr>
</tbody>
</table>
results in D2 (HR: 1.66; 95% CI, 1.25–2.0; \( P = 0.0074 \); C-index = 0.69) and D3 (HR: 1.71; 95% CI, 1.36–2.11; \( P = 0.010 \); C-index = 0.68), respectively. Supplementary Table S3 shows multivariable Cox regression analysis for OS in discovery set D1 and in test sets D2 and D3.

The corresponding Kaplan–Meier survival curves showed a significant difference in OS between patients with low and high DRS (D2: \( P = 0.00039 \), D3: \( P = 0.00056 \)). Kaplan–Meier survival curves for D2 and D3 are shown in Fig. 4A and B, respectively.

Stratified PD-L1 expression before therapy as a predictor of OS

PD-L1 expression score was only available for 25 patients in our study cohort from CCF. We performed a subset analysis to compare the performance of our radiomic model against that of PD-L1 expression in predicting OS and response to ICIs. In the 25 patients included in this analysis, median OS posttherapy was 12.5 months (range: 2–19 months). The patients were divided into high PD-L1 and low PD-L1 score categories based on the different cutoff criteria (PD-L1 < 1% and >1%; PD-L1 < 10% and >10%; PD-L1 < 50% and >50%) shown in Supplementary Table S4.

PD-L1 expression based on the third criteria (PD-L1 < 50% and >50%) could discriminate patients into two groups with different OS (\( P = 0.031 \)), but no statistically significant difference in HR was observed (HR for survival in patients with low PD-L1 score: 0.35, 95% CI, 0.12–1.1; \( P = 0.065 \); C-index = 0.57). PD-L1 expression was also not correlated with response to treatment using Spearman correlation analysis.

When the perinodular Gabor DeRADX feature \((f = 0, \theta = \pi/8)\) was combined with PD-L1 score in a Cox model, the resulting PD-L1_Rad score was significantly associated with OS (HR: 0.26; 95% CI, 0.1–0.8; \( P = 0.021 \); C-index = 0.80). The corresponding Kaplan–Meier survival curves showed a significant difference in OS between patients with low and high PD-L1_Rad score (log-rank test, \( P = 0.0046 \), \( n = 25 \)). The Kaplan–Meier survival curves for this subset analysis are shown in Supplementary Fig. S1.

Predictive value for OS of volumetric changes of tumor during therapy

The median tumor volume in D1 was 4.65 mL (range, 0.137–166.83 mL) before ICI administration, and 5.58 mL (range, 0.07–370.41 mL) after two/three cycles of ICI. An LDA classifier trained with tumor volume change in discovery set (D1) could discriminate responders from nonresponders with an AUC of 0.70 and 0.78 in D2 and D3, respectively. The volume change was associated with OS in D1 (HR for survival in patients with less shrinkage in tumor volume after therapy: 0.58; 95% CI, 0.41–0.81; \( P = 0.0018 \); C-index = 0.63), with similar statistics observed in D2 (HR: 0.55; 95% CI, 0.48–0.78; \( P = 0.011 \); C-index = 0.61) and D3 (HR: 0.57; 95% CI, 0.41–0.88; \( P = 0.025 \); C-index = 0.60), respectively.
A multivariable Cox regression model using changes in tumor volume and DelRADx signature indicated that DelRADx signature is the only independent biomarker associated with OS in D1 (DRS: HR: 1.68; 95% CI, 1.23–1.98; \( P = 0.0001 \); volumetric changes: HR: 0.6; 95% CI, 0.24–1.63; \( P = 0.35 \); C-index = 0.84), D2 (DRS: HR: 1.67; 95% CI, 1.12–1.76; \( P = 0.0031 \); volumetric changes: HR: 0.51; 95% CI, 0.31–1.52; \( P = 0.21 \); C-index = 0.75), and D3 (DRS: HR: 1.74; 95% CI, 1.54–2.16; \( P = 0.0075 \); volumetric changes: HR: 0.58; 95% CI, 0.34–1.68; \( P = 0.43 \); C-index = 0.73), respectively.

The Kaplan–Meier curve based on percentage change in tumor volume in D2 and D3 showed that patients with high percentage shrinkage in tumor volume have a higher survival probability compared with patients with lower percentage shrinkage in tumor volume. This difference was statistically significant (\( P = 0.0022 \) for patients in D2 and 0.015 in D3). Figure 4A and B shows the Kaplan–Meier curves for this model in D2 and D3, respectively. High and low percentage shrinkage in tumor volume was defined on the basis of the median percentage change in tumor volume.

**Morphologic association of DelRADx features with TIL density measures**

To investigate associations between the DelRADx features and density of tumor-infiltrating lymphocytes (TIL), we explored a subset of 36 cases from CCF. Toward this end, we employed automated nuclei detection (31) followed by density estimation of TILs in diagnostic digitized H&E images. To quantify density or compactness of TILs, 76 features were extracted from biopsied tissue specimens. A pairwise Spearman correlation was performed between each radiomic and TIL compactness measure. The 10 DelRADx features that could best separate ICI therapy responders from nonresponders were used for unsupervised hierarchical clustering. The TIL-density feature

---

**Figure 4.**

A, Box and whisker plots for the top eight features that best distinguish ICI therapy responders from nonresponders, as identified using WLCX feature selection scheme. B, Unsupervised hierarchical clustering of intranodular and perinodular delta-radiomic features. X-axis represents top selected features. Y-axis represents the training set of patients (\( n = 50 \)). The number on the x-axis corresponds to each feature in Table 3. A number of identified features showed differential expression between the responders and nonresponders for the patients in the discovery set. C, Left, consensus clustering using combined intranodular texture and shape features; right, consensus clustering using combined shape, intranodular texture, and perinodular texture features.
Figure 4. A, Kaplan–Meier survival curves on the independent validation set D2 for different clinical variables and risk score. B, Kaplan–Meier survival curves on the independent validation set D3 for different clinical variables and risk score.
demonstrating the highest correlation with the radiomic features was overlaid on the cluster-plot as a high/low expression based on its median value.

TIL density (TIL convex hull area/total convex hull area) was found to be significantly ($\rho = -0.5, P < 0.05$) correlated with a peritumoral Gabor filter DelRADx feature from the first annular ring outside the nodule. Supplementary Figure S2A shows examples of digitized histology samples from responder and nonresponder cases and illustrates the abundance of TILs (pink dots) in the responder, as compared with the nonresponder patient. The two hierarchical clusters that emerged using the CT radiomic features were associated with differential TIL density expressions. One cluster had a majority of responder cases (90%), whereas the other had a majority of nonresponders (69%; Supplementary Fig. S2B). The nonresponder cluster had lower compactness (high TIL convex hull area/total convex hull area), as compared with the second cluster.

Comparison of DelRADx with baseline radiomic features

An LDA classifier trained with a combination of 2 baseline Intra-nodular, 6 annular ring baseline peritumoral texture, and one shape feature in the discovery set yielded an AUC of 0.81 ± 0.05 and corresponding AUCs of 0.79 and 0.74 for the patients within $D_2$ and $D_3$ validation sets, respectively. An LDA classifier trained with baseline tumor volume in the discovery set could discriminate responders from nonresponders with an AUC of 0.64 and 0.63 in $D_2$ and $D_3$, respectively. However, the baseline tumor volume was not associated with OS in the discovery (HR: 1.17; 95% CI, 0.82–1.68; $P = 0.42$; C-index = 0.56) nor the validation sets. The risk score generated by backward elimination on predictive baseline texture features was associated with differential baseline tumor volume in the discovery set (HR: 1.25; 95% CI, 1.13–2.56; $P = 0.025$; C-index = 0.69) and yielded similar results in $D_2$ (HR: 1.18; 95% CI, 1.1–2.05; $P = 0.044$; C-index = 0.66). However, it was not significant in $D_3$ (HR: 1.25; 95% CI, 0.86–1.94; $P = 0.21$; C-index = 0.57). Supplementary Table S5 shows the differences in predicting response to treatment after two versus three cycles of ICI. The Fisher exact test yielded a $P$ value of 0.21 between correct and incorrect decisions by classifier, suggesting no significant differences between the two groups.

Discussion

ICIIs have changed the landscape of treatment for patients with advanced NSCLC without any targetable mutations (32). Despite advances in immuno-oncology, response rates to ICI agents are modest. A problem in immuno-oncology is the lack of objective response assessment tools to characterize responses to immunotherapy. In this study, we used routinely obtained CT scans to build DelRADx, a measure of radiomic features that quantifies image changes in the tumor and peritumoral regions between the baseline and the posttherapy scans. We tested DelRADx for its value in predicting response to ICI therapy.

The most discriminating texture features for prediction of response to ICI therapy were (i) DelRADx Haralick features (33), which reflect the changes in cooccurrence intensity statistics; (ii) DelRADx Gabor features (34), which capture changes in multi-frequency gradient orientation patterns; (iii) DelRADx Laws features (35), which capture changes in wavy, ripple, and spot-like patterns; and (iv) DelRADx CoLIAge features (36), which capture local changes in structural texture. The Haralick entropy increment after therapy possibly arises from regions of hypoxia and acidosis (16, 17). Acidity and hypoxia can block T-cell activation and are a barrier to the efficacy of immunotherapy drugs (37–39). A previous study (40) has shown that Haralick features are highly associated with tumor hypoxia on FLAIR and Gd-T1w MR sequences in glioblastoma multiforme. We observed an increase in expression of peritumoral DelRADx Gabor features after therapy. This might be reflective of higher numbers of tumor-associated inflammatory cells in the peritumoral compartment, similar to our previous findings in ref. 41.

Increase in peritumoral DelRADx Laws (R5 × R5) features was associated with a decrease in OS. These features potentially capture peritumoral vascular invasion and neovascularization around the tumor, which manifest as changes in ripple patterns. This was consistent with our previous findings in patients with NSCLC treated with chemotherapy, where time to progression in patients with elevated peritumoral Laws texture features is lower than those with decreased peritumoral Laws features (10). Previous studies have also shown the role of blood vessels, lymphatic vessels, and angiogenesis in immunotheraphy response (42, 43).

We observed an association between PD-L1 expression and OS, when PD-L1 expression was stratified by 50% criteria ($P = 0.031$). Previous studies have shown that in patients with advanced NSCLC and PD-L1 expression on at least 50% of tumor cells, PD-1/PD-L1 targeting therapy is associated with longer progression-free survival and OS (13, 44–49). However, across a range of clinical trials, the performance of PD-L1 expression as a biomarker has been found to be suboptimal (50–52): in clinical practice, variations in thresholds of expression, scoring systems, and the intrinsically variable nature of PD-L1 expression have conspired to reduce the value of PD-L1 as a biomarker. In fact, although only about a quarter of patients with high PD-L1 scores respond to ICI (in the first line), about the same proportion of patients have shown measurable response even with low PD-L1 expression (44, 46, 53).

Our results identified perinodular Gabor DelRADx combined with PD-L1 score as a better biomarker for OS than PD-L1 score alone. Previous study by Mazzaschi and colleagues (54) showed that PD-L1 amounts were not prognostic in patients with resected NSCLC. However, patients with higher CD8+ lymphocytes lacking PD-1 inhibitory receptor had a longer OS. Our finding corroborates this result because perinodular Gabor DelRADx features associated with TIL density around the tumor were also prognostic of OS.

Finally, we evaluated the radio-pathomic association of TILs with predictive DelRADx features. Density of TILs is correlated with disease progression, survival outcomes, and response to therapy in a wide range of malignancies (55–57). The presence of an immune infiltration is more likely to manifest via unique textural patterns outside the tumor (58). In our study, TIL density was found to be correlated with a peritumoral Gabor feature from the first annular ring. Previous studies have reported an overexpression of TILs in patients who have responded to immunotherapy (59–61). Gabor features from outside the nodule had a higher expression in adenocarcinomas as compared with granulomas: H&E images showed densely packed TILs and tumor-associated macrophages (20). Most tumors trigger an immune response modulated by TILs. Indeed, Gabor features from the 0- to 3-mm peritumoral region in breast cancer DCE-MRI were associated with the density of TILs (41).

Previous studies have also used radiomics to predict response to ICIs (15, 24, 62, 63). Response to immunotherapy was negatively correlated with tumor convexity and positively correlated with edge-to-core size ratio on CT scans of patients with NSCLC (64). In a study by Tang and colleagues (63), favorable outcome group of immune response characterized by low CT intensity and high heterogeneity exhibited low PDL1 and high CD3 infiltration, suggestive of a favorable immune activated state. Patients with immune-activated tumors had...
the highest 5-year OS rate, and the patients with immune-inhibited tumors had the lowest. Radiomic signature of CD8+ cells has been shown to predict the immune phenotype of solid tumors and clinical outcomes for patients with cancer who had been treated with anti-PD-1 and PD-L1 therapies (65).

Our study differed from those previous studies in a few ways. First, in addition to radiomic features from within the confines of the nodules, we interrogated the radiomic features extracted from the annular perinodular regions. Our group has previously shown the utility of perinodular radiomic features for lung nodule diagnosis (20), predicting response to chemotherapy (10, 11, 66), and predicting OS on baseline CT in immunotherapy (67). Second, we validated our results in two independent test sets, accrued from different sites across different immunotherapy agents. The machine classifier trained with signatures obtained from cases treated with nivolumab yielded promising results when validated on cases treated with different ICI agents. Third, our study attempted to define the tissue morphology underpinning the radiomic features that were identified to be associated with response to ICI therapy. Specifically, we studied the association between the identified predictive radiomic descriptors and TIL density obtained from diagnostic biopsies. We provided a morphologic rationale by determining the association of the radiomic attributes with distribution of TILs and corresponding PD-L1 expression.

We acknowledge that our study did have its limitations. The sizes of our cohorts, both for discovery and validation, were relatively small although comparable with cohort sizes from other published studies (68, 69). Second, radiomic feature expressions might be sensitive to lesion annotation accuracy. In this study, nodule annotation was done by one single reader. However, because we used first-order statistics of the features within regions of interest, and not the pixel-wise feature values, we do not anticipate this impacting our findings, although further evaluation is need to ensure this. We did not consider the influence of convolution kernels, reconstruction algorithms, and slice thickness on the extracted radiomic features. Despite the limitations, our results across two sites and multiple ICI agents demonstrate the predictive value of the DelRADx features in assessing treatment response and the utility of perinodular regions in characterizing tumor microenvironment.

In conclusion, we presented an approach, DelRADx, that quantifies dynamic changes in certain textural radiomic attributes between baseline and posttreatment CT scans following ICI therapy in patients with advanced NSCLC. DelRADx features were (i) predictive of response to ICI therapy; (ii) prognostic of improved OS; and (iii) associated with TIL density on corresponding diagnostic biopsy samples. Incorporating DelRADx into personalized decision-making for patients with advanced NSCLC could help characterize distinct phenotypes of the disease and serve as a noninvasive approach to identify candidates who will benefit the most from immunotherapy while sparing others of harmful side effects. Additional validation of DelRADx features is warranted to accurately define their role as predictive and prognostic biomarkers in patients treated with immunotherapy.

Disclosure of Potential Conflicts of Interest

A. Madabhushi is a consultant/advisory board member for Inspirata and Elucida Biomaging, reports receiving a commercial research grant from Philips, and has ownership interest (including patents) in Inspirata. V. Velcheti is a consultant/advisory board member for Genentech, AstraZeneca, Bristol-Myers Squibb, Foundation Medicine, Novartis, and Celgene. No potential conflicts of interest were disclosed by the other authors.

Disclaimer

The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH, the U.S. Department of Veterans Affairs, the U.S. Department of Defense (DOD), or the U.S. Government.

Authors’ Contributions

Conception and design: M. Khorrami, P. Prasanna, P. Patil, R. Thawani, M. Alliò, K. Bera, M. Feldman, V. Velcheti, A. Madabhushi

Development of methodology: M. Khorrami, P. Prasanna, P. Patil, G. Correddo, M. Alliò, K. Bera, A. Madabhushi

Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): A. Gupta, P. Patil, P.D. Velu, R. Thawani, K. Bera, V. Velcheti

Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): M. Khorrami, P. Prasanna, A. Gupta, P. Patil, G. Correddo, K. Bera, P. Fu, M. Feldman, V. Velcheti, A. Madabhushi

Writing, review, and/or revision of the manuscript: M. Khorrami, P. Prasanna, A. Gupta, P. Patil, P.D. Velu, R. Thawani, G. Correddo, K. Bera, P. Fu, M. Feldman, V. Velcheti, A. Madabhushi

Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): M. Khorrami, P. Patil, P.D. Velu, R. Thawani, K. Bera, M. Feldman, V. Velcheti

Study supervision: P. Prasanna, A. Gupta, A. Madabhushi

Acknowledgments

Research reported in this publication was supported by the NCI of the NIH under award numbers 1U24CA199374-01, R01CA20272-01A1, R01CA208236-01A1, R01CA216579-01A1, and R01CA220581-01A1. This work was also supported by National Center for Research Resources under award number 1C06 RR12463-01, VA Merit Review Award I BX004121A from the U.S. Department of Veterans Affairs Biomedical Laboratory Research and Development Service, the U.S. DOD Prostate Cancer Idea Development Award (W81XWH-15-1-0558), the U.S. DOD Lung Cancer Investigator-Initiated Translational Research Award (W81XWH-16-1-0329), the Ohio Third Frontier Technology Validation Fund, the Wallace H. Coulter Foundation Program in the Department of Biomedical Engineering, and the Clinical and Translational Science Award Program (CTSA) at Case Western Reserve University.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received June 24, 2019; revised September 4, 2019; accepted November 5, 2019; published first November 12, 2019.

References


Khorrami et al.


Cancer Immunology Research

Changes in CT Radiomic Features Associated with Lymphocyte Distribution Predict Overall Survival and Response to Immunotherapy in Non–Small Cell Lung Cancer
Mohammadhadi Khorrami, Prateek Prasanna, Amit Gupta, et al.
Cancer Immunol Res  Published OnlineFirst November 12, 2019.

Updated version  Access the most recent version of this article at: doi:10.1158/2326-6066.CIR-19-0476
Supplementary Material  Access the most recent supplemental material at: http://cancerimmunolres.aacrjournals.org/content/suppl/2019/11/12/2326-6066.CIR-19-0476.DC1

E-mail alerts  Sign up to receive free email-alerts related to this article or journal.
Reprints and Subscriptions  To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.
Permissions  To request permission to re-use all or part of this article, use this link http://cancerimmunolres.aacrjournals.org/content/early/2019/12/04/2326-6066.CIR-19-0476. Click on “Request Permissions” which will take you to the Copyright Clearance Center’s (CCC) Rightslink site.