Physics Contribution

Robust Intratumor Partitioning to Identify High-Risk Subregions in Lung Cancer: A Pilot Study

Jia Wu, PhD,* Michael F. Gensheimer, MD,* Xinzhe Dong, MD,* Daniel L. Rubin, MD, MS,†,‡ Sandy Napel, PhD,†
Maximilian Diehn, MD, PhD,*,§ Billy W. Loo, Jr, MD, PhD,*,‖ and Ruijiang Li, PhD*,*

Departments of *Radiation Oncology, †Radiology, ‡Medicine (Biomedical Informatics Research), §Stanford Cancer Institute, and ‖Institute for Stem Cell Biology and Regenerative Medicine, Stanford University School of Medicine, Stanford, California

Received Jan 12, 2016, and in revised form Feb 16, 2016. Accepted for publication Mar 12, 2016.

Summary

We developed a robust, intratumor partitioning framework to identify clinically relevant, high-risk subregions by integrating 18F-fluorodeoxyglucose positron emission tomography (FDG-PET) and computed tomography (CT) imaging. In a population of 44 non-small cell lung cancer (NSCLC) patients, 3 different types of tumor subregions emerged, each with distinct imaging phenotypes. Among these, the volume of the most metabolically active and metabolically heterogeneous solid component of the tumor predicted overall survival (OS) and out-of-field progression (OFP) on the entire cohort, with a concordance index or CI of 0.66-0.67. When restricting the analysis to patients with stage III disease (n ≥ 32), the same subregion achieved an even higher CI of 0.75 (hazard ratio 3.93, log-rank P = .002) for predicting OS, and a CI of 0.76 (hazard ratio 4.84, log-rank P = .002) for predicting OFP. In comparison, conventional imaging markers, including tumor volume, maximum standardized uptake value, and metabolic tumor volume using threshold

Purpose: To develop an intratumor partitioning framework for identifying high-risk subregions from 18F-fluorodeoxyglucose positron emission tomography (FDG-PET) and computed tomography (CT) imaging and to test whether tumor burden associated with the high-risk subregions is prognostic of outcomes in lung cancer.

Methods and Materials: In this institutional review board–approved retrospective study, we analyzed the pretreatment FDG-PET and CT scans of 44 lung cancer patients treated with radiation therapy. A novel, intratumor partitioning method was developed, based on a 2-stage clustering process: first at the patient level, each tumor was over-segmented into many superpixels by k-means clustering of integrated PET and CT images; next, tumor subregions were identified by merging previously defined superpixels via population-level hierarchical clustering. The volume associated with each of the subregions was evaluated using Kaplan-Meier analysis regarding its prognostic capability in predicting overall survival (OS) and out-of-field progression (OFP).

Results: Three spatially distinct subregions were identified within each tumor that were highly robust to uncertainty in PET/CT co-registration. Among these, the volume of the most metabolically active and metabolically heterogeneous solid component of the tumor predicted overall survival and out-of-field failure in...
NSCLC patients treated with radiation therapy or concurrent chemoradiotherapy.

Introduction

Imaging with 18F-fluorodeoxyglucose positron emission tomography (FDG-PET) and computed tomography (CT) plays an important role in the clinical management of non-small cell lung cancer (NSCLC), including diagnosis, staging, radiation treatment planning, and response evaluation (1-3). Recent studies suggested that simple imaging characteristics (4-10), such as tumor size, maximum standardized uptake value (SUV_max), and metabolic tumor volume, may have prognostic value. However, the prediction accuracy of current prognostic imaging biomarkers in NSCLC is quite limited (4). For example, a recent study (11) reported a concordance index or c-index of 0.62 for predicting overall survival (OS) in locally advanced NSCLC on the basis of quantitative imaging features of FDG-PET that include texture features (c-index ranges from 0 to 1; random guess would give 0.5). Because accurate prognostication is crucial for tailoring therapy on the basis of an individual patient’s risk, there is an unmet need to identify reliable biomarkers according to noninvasive imaging characteristics.

A number of studies have focused on extracting advanced, quantitative image features from clinical images to predict lung cancer prognosis (12-17). Given that FDG-PET and CT provide complementary information (metabolic vs anatomic), the image features based on either modality likely represent a partial view of the tumor. Another, perhaps more fundamental limitation of previous approaches is the lack of considering spatial information, whereby the imaging features were calculated as aggregate measures on the entire tumor volume. Such approaches implicitly assume that the tumor is homogeneous, or heterogeneous but uniformly mixed throughout the whole volume. In reality, however, regional variations within the tumor are often apparent on imaging (18-20). Some parts of the tumor are biologically more aggressive than others and may play a dominant role in disease progression. Failure to explicitly account for these regional variations may obscure the subtle differences among tumors and limit our ability to discover useful imaging biomarkers (18).

There is emerging evidence suggesting that explicitly incorporating spatial information on CT (21) (eg, necrosis in the core and proliferation along the periphery) may lead to improved prediction of survival in NSCLC. However, the simple division into tumor core and boundary regions does not take into account other kinds of spatial variations in tumors that may occur, and the assumption that these regions have different underlying biology may not hold true for all tumors. More sophisticated, potentially more effective image analysis techniques are needed to incorporate tumor-specific regional variations by leveraging the differential information of multimodality PET/CT imaging.

In this work we propose a novel approach to developing prognostic imaging signatures for NSCLC. On the basis of the hypothesis that regional differences within a tumor reflect differing fundamental biological processes, we performed intratumor partition through an integrated analysis of PET and CT images routinely acquired for radiation treatment planning. We show that multiple, spatially distinct subregions exist within tumors and that they can be reliably extracted in a robust fashion across patients. Further, we demonstrate that the volume of one particular subregion predicts OS and metastasis, demonstrating superior performance over existing prognostic imaging metrics in lung cancer.

Methods and Materials

Patient population

Under approval from the institutional review board, we retrospectively reviewed records and images of 44 lung cancer patients who were treated with definitive radiation therapy or concurrent chemoradiotherapy at our institution from October 2005 through December 2009. A majority of the patients (n = 32) had stage III disease, and others had stage I, II, or IV disease. In our institution, we follow all NSCLC patients at regular intervals (usually every 3 months) after treatment. During the follow-up 32 patients (72.7%) died, and 23 patients (52.3%) had an out-of-field failure. Patient characteristics are summarized in Table 1.

PET/CT imaging

All patients underwent PET/CT imaging for radiation treatment planning before radiation therapy. Regarding the imaging protocol, after an 8-hour fast, patients were injected with 10 to 18 mCi of FDG, with imaging 60 minutes later on a GE Discovery PET/CT scanner (GE Medical Systems, Milwaukee, WI). The CT images were collected in the helical acquisition mode. In the same scan...
locations and generally spanning around 7-bed positions, PET data were acquired in 2-dimensional mode, with 3 to 5 minutes of acquisition time per bed position. The PET images were then reconstructed with an ordered set expectation maximization algorithm, with the CT data for attenuation correction. A separate non-contrast-enhanced CT scan was performed at the end of the natural exhale. The complete PET/CT examination requires approximately 1.5 hours, including patient setup, tracer uptake, and image acquisition. The spatial resolution of the original PET image was $2.34 \times 2.34 \times 3.27$ mm$^3$, and the spatial resolution of the original CT image was $0.98 \times 0.98 \times 1.25$ mm$^3$.

### Table 1: Demographic data for the study cohort

<table>
<thead>
<tr>
<th>Parameter</th>
<th>N = 44</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (y), median (range)</td>
<td>69.5 (47.2-89.7)</td>
</tr>
<tr>
<td>Gender (male/female)</td>
<td>24 (54.5)/20 (45.5)</td>
</tr>
<tr>
<td>Stage (1a/1b/2a/2b/3a/3b/4)</td>
<td>4 (9.0)/3 (6.8)/1 (2.3)/1 (2.3)/12 (27.3)/20 (45.5)/3 (6.8)</td>
</tr>
<tr>
<td>Concurrent chemotherapy (yes/no)</td>
<td>34 (77.3)/10 (22.7)</td>
</tr>
<tr>
<td>Radiation therapy dose (Gy), median (range)</td>
<td>68 (40-74)</td>
</tr>
<tr>
<td>Karnofsky performance status, median (range)</td>
<td>80 (50-90)</td>
</tr>
<tr>
<td>Follow-up time (mo), median (range)</td>
<td>20.2 (1.8-76.7)</td>
</tr>
<tr>
<td>Death (yes/no)</td>
<td>32 (72.7)/12 (27.3)</td>
</tr>
<tr>
<td>Out-of-field failure (yes/no)</td>
<td>23 (52.3)/21 (47.3)</td>
</tr>
</tbody>
</table>

Values are number (percentage) unless otherwise noted.

---

**Intratumor partitioning by integrated analysis of PET/CT imaging**

Here we propose a multiparametric intratumor partitioning method to consistently divide the entire tumor into multiple spatially distinct subregions. Figure 1 summarizes the complete pipeline that contains 3 key steps. The first step is to accurately align the PET and CT images for each patient. The second step aims to independently segment the entire tumor into a large number of superpixels, where neighboring voxels with similar imaging characteristics are grouped together. Finally, similar superpixels within the same tumor are merged to form larger, dominant subregions via population-level clustering. This process also establishes the correspondence of the subregions so that they can be compared across patients.

In more detail, PET and CT images of the same patient were first rigidly registered using Elastix software (Elastix, Utrecht, Netherlands, http://elastix.isi.uu.nl/) (22). The primary tumor was contoured on the CT image by an experienced radiation oncologist as part of the radiation treatment planning. In addition to the original voxel intensity on CT and PET, the local entropy (9 neighborhood) of CT and PET, which measures intratumor heterogeneity locally, was also computed at each voxel and incorporated into the analysis, leading to a 4-dimensional feature vector. For patient-level tumor oversegmentation, the k-means clustering algorithm with spatial constraint (23) was used to group voxels with similar features together for each tumor, with squared Euclidean distance as the distance measure.

In the population-level clustering step, each superpixel was characterized by the mean feature vector (the 4
Sensitivity of intratumor partition with respect to PET/CT registration

Although modern PET/CT scanners acquire both PET and CT that are co-registered, the CT component is used for attenuation correction purposes and its image quality is suboptimal, with large slice thickness and lower mAs. Here, we are using the high-quality CT used for radiation treatment planning, which is not automatically registered with the PET. Moreover, in lung cancer, tumor motion between scans usually brings about additional uncertainty that would require a careful co-registration between PET and CT. We investigated the sensitivity of the intratumor partition results given the uncertainty of registration between PET and CT. We simulated the registration uncertainty by intentionally shifting the original CT image away from the PET image for 3 mm (26) along a randomly picked direction among all 6 directions in 3 dimensions. Then we quantitatively assessed the consistency between the clustering results with and without the random shift. The in-group proportion (IGP) (27) was selected to test the reproducibility of the clustering results. In brief, the IGP indicates whether one cluster in one data set is similar to another cluster in another data set. When IGP is 100%, the clusters from 2 data sets are identical. More importantly, the robustness of the main prognostic factors tested in this study (ie, the volumes associated with distinct tumor subregions) were evaluated against registration uncertainty.

Prognostic evaluation and statistical analysis

The tumor volume associated with each dominant subregion was identified as a putative biomarker of treatment outcomes for further validation. The Cox proportional hazard model was used to evaluate each imaging feature in terms of their prognostic capability for both OS and freedom from out-of-field progression (OFP). We used the median value of the predicted hazard/risk to divide the cohort into high- versus low-risk groups. Concordance index (CI) was used to assess the prognostic capability of the imaging biomarkers. Kaplan-Meier analysis was used to estimate the probability of a certain event. The statistical significance of CIs and whether one CI was higher than another were assessed via bootstrapping. In detail, for testing the statistical significance of the CI, the original survival data were firstly permuted. Then both the original and permuted survival data were bootstrapped to a sample size 50 with 100 replications. Last, the CI distributions from both bootstrapped samples were compared via the Wilcoxon signed rank test. A P value <.05 was considered statistically significant. The Cox survival modeling and statistical analysis were performed in R (www.r-project.org).

Results

Three dominant intratumor subregions emerged

The intratumor partitioning method identified 3 major clusters or tumor subregions that show distinct imaging phenotypes (Fig. 2; Figs. E1 and E2; available online at www.redjournal.org). Here, 3 clusters showed the optimal gap metric among 2 to 10 clusters. The first cluster, labeled as cluster A, was associated with high PET SUV, high CT number, and high PET entropy. This particular region contains the most metabolically active and heterogeneous solid tumor component. The second cluster, labeled as cluster B, was associated with low PET SUV, low PET entropy, and high CT number, which contains the...
metabolically nonactive solid component of the tumor. The third cluster, labeled as cluster C, contains low CT number regions (ie, nonsolid component). Two outliers with a negligibly small volume (labeled as clusters E1 and E2) were excluded for further analysis.

**Intratumor partitioning was robust against PET/CT misalignment**

For patient-level over-segmentation, the median of IGP was 95.5%, with a range from 80.0% to 100%, which indicates highly similar clusters between the k-means clustering results with and without random shift of 3 mm. For population-level labeling, there were still 3 dominant subregions (refer to Fig. E3; available online at www.redjournal.org, for the corresponding dendrogram and heat map), and the IGPs for these 3 main clusters were 98.3%, 98.2%, and 98.6%. More importantly, Figure E4 (available online at www.redjournal.org) shows a very strong linear correlation ($R^2 > 0.95$, $P < 0.001$) of the derived tumor volume of all 3 tumor clusters with and without shift. Together these data show that the proposed framework was able to identify robust intratumor regions despite uncertainty in PET/CT registration.

**Volume of high-risk subregion was predictive of outcomes**

In the entire cohort ($n=44$) the tumor volume associated with cluster A was prognostic of OS, with a CI of 0.67 ($P=2.67e-32$) and hazard ratio (HR) 2.79 (log-rank $P = 0.04$) for separating low- versus high-risk patients, as shown in Figure 3a. In terms of predicting OFP, the same subregion achieved a good CI of 0.66 ($P=3.78e-26$), with HR 2.07 (log-rank $P = 0.08$), as shown in Figure 3b. These results compared favorably ($P<0.001$) with conventional imaging metrics, whereby total tumor volume, SUV$_{max}$, and metabolic tumor volume (MTV) using threshold of 50% standardized uptake value (SUV) maximum (MTV$_{50}$) were not prognostic of OS or OFP (Figs. E5-E7; available online at www.redjournal.org).

When restricting our analysis to stage III patients ($n = 32$), the prognostic capability of the high-risk tumor volume was further improved. For predicting OS it

---

**Fig. 3.** Kaplan-Meier curves of using tumor volume of the high-risk subregion (ie, cluster A) to predict: (a) overall survival, (b) out-of-field progression, for the entire cohort ($n=44$), and (c) overall survival, (d) out-of-field progression, for patients with stage III disease ($n=32$). The cutoff between high- versus low-risk patients was determined according to the median of predicted risk. Abbreviation: HR = hazard ratio.
achieved a CI of 0.75 ($P = 2.33e-34$) and HR of 3.93 (log-rank $P = .002$), as shown in Figure 3c, and for predicting OFP it achieved a CI of 0.76 ($P = 2.88e-34$) and HR of 4.84 (log-rank $P = .002$), as shown in Figure 3d. Again, these results significantly ($P < .001$) improved the prognostic capability compared with conventional imaging metrics (Figs. E5-E7; available online at www.redjournal.org).

Figure 4 shows the process of using the proposed method to analyze 2 representative patients. Both patients had stage IIIb lung cancer, with similar tumor volume but very different outcomes: patient 1 was still alive and free from OFP during a 4-year follow-up, whereas patient 2 died after 3 months. The $SUV_{\text{max}}$ and $MTV_{50}$ of patient 1 were higher than those of patient 2, contrary to prior studies indicating that higher SUV and MTV correlate with poor survival. On the other hand, using our proposed method, patient 1 actually had a smaller volume of high-risk subregion (ie, cluster A) than patient 2, consistent with the clinical outcomes.

**Multivariate analysis**

To evaluate whether the proposed imaging biomarker provided independent prognostic information, we performed multivariate analyses to predict OS for the entire cohort as well as for the stage III cohort. Both the high-risk tumor volume and the gross tumor volume were included, which were further adjusted for clinical factors (ie, Karnofsky performance status, radiation therapy dose, stage, concurrent chemotherapy status, age, and gender). After adjustment, the volume of high-risk subregion remained statistically significant as an independent predictor of OS for both the entire cohort ($P = .03$) and the stage III patient cohort ($P = .001$), as shown in Table 2. In comparison, gross tumor volume was not an independent predictor of OS in either cohort according to multivariate analysis. Because the high-risk tumor volume was the only prognostic predictor for OFP, we did not perform a multivariate analysis for that.

**Discussion**

In this study we proposed a 2-stage multiparametric clustering approach to identify clinically relevant, high-risk subregions through an integrated analysis of PET/CT images in lung cancer. We found 3 intratumor subregions with distinct imaging features, and their volumes were robust against uncertainty in registration between PET and CT. In addition, they were robust to the initial number of over-
segmented tumor regions (both 20 and 10 led to almost identical results), largely owing to their merging during the population-level clustering. The tumor volume associated with the most metabolically active and metabolically heterogeneous solid component was prognostic of OS and metastasis, which consistently outperformed 3 conventional imaging metrics. Biologically, the high-risk subregion we identified may represent the most proliferative and hypoxic portions of the tumor (28, 29), explaining its prognostic value.

Intra-tumor heterogeneity is reflected on multiple spatial scales (at molecular, cellular, and tissue levels) and has been associated with aggressive disease and poor prognosis (30-34). Although regional differences are often observed on imaging and are an important manifestation of intratumor heterogeneity, their clinical significance has not been systematically investigated. It is plausible that some parts of the tumor are more aggressive than others, and the tumor volume associated with these high-risk regions could determine disease progression and ultimately patient survival, and thus may be more reliable biomarkers of clinical outcomes. In this work we have used local entropy as a simple measurement of phenotypic complexity of the tumor. In principle, more texture features that capture additional aspects of tumor heterogeneity could be incorporated into the proposed framework.

This study is among the first to explicitly extract subregions from heterogeneous tumors and explore their clinical significance in disease prognosis (21, 35). Besides the obvious differences in disease sites (NSCLC vs glioblastoma and head and neck cancer) and imaging modalities (FDG-PET/CT vs magnetic resonance imaging), the proposed method differs from prior studies in several key aspects: (1) our method is based on a 2-stage clustering process, whereby the first patient-level over-segmentation generates superpixels that are robust to image noise, and the second population-level clustering produces consistently defined subregions across the entire cohort, whereas prior methods were based on direct single-stage clustering; (2) our method does not assume a predefined number of intratumor clusters but rather learns this number from all available data; (3) instead of only considering voxel values, we incorporate both magnitude of image signal and its spatial heterogeneity map (measured with local entropy) in defining the intratumor clusters. Another recent study (36) proposed a 3-class fuzzy locally adaptive Bayesian (3-FLAB) for the delineation of inhomogeneous tumors. However, the method was based on a single image modality (FDG-PET) limited to a fixed number of (ie, 2) intratumor subregions. The clinical significance (eg, prognostic value) of either subregion defined as such was unclear from the study (36).

Biomarkers that can reliably predict treatment failure on an individualized basis will have tremendous value in guiding the optimal treatment of lung cancer. The need for predictive biomarkers is highlighted by a recent randomized phase 3 trial (Radiation Therapy Oncology Group protocol 0617) that failed to show a survival benefit of high-dose versus standard-dose radiation therapy in locally advanced NSCLC (37). Instead of delivering intensive

<table>
<thead>
<tr>
<th>Predictors for OS</th>
<th>Univariate</th>
<th>Multivariate</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HR</td>
<td>95% CI</td>
</tr>
<tr>
<td>All patients (n=44)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>High-risk tumor volume</td>
<td>2.79</td>
<td>1.36-5.75</td>
</tr>
<tr>
<td>Gross tumor volume</td>
<td>1.73</td>
<td>0.86-3.48</td>
</tr>
<tr>
<td>Gender</td>
<td>1.02</td>
<td>0.51-2.05</td>
</tr>
<tr>
<td>Age</td>
<td>1.37</td>
<td>0.68-2.75</td>
</tr>
<tr>
<td>Stage (I + II/III + IV)</td>
<td>1.52</td>
<td>0.70-3.34</td>
</tr>
<tr>
<td>Concurrent chemotherapy</td>
<td>1.37</td>
<td>0.63-2.99</td>
</tr>
<tr>
<td>Radiation therapy dose</td>
<td>1.17</td>
<td>0.59-2.35</td>
</tr>
<tr>
<td>KPS</td>
<td>2.95</td>
<td>1.45-5.99</td>
</tr>
<tr>
<td>Stage III patients (n=32)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>High-risk tumor volume</td>
<td>3.94</td>
<td>1.56-9.92</td>
</tr>
<tr>
<td>Gross tumor volume</td>
<td>2.75</td>
<td>1.13-6.72</td>
</tr>
<tr>
<td>Gender</td>
<td>1.09</td>
<td>0.46-2.59</td>
</tr>
<tr>
<td>Age</td>
<td>0.78</td>
<td>0.33-1.85</td>
</tr>
<tr>
<td>Stage (3a/3b)</td>
<td>2.10</td>
<td>0.81-5.43</td>
</tr>
<tr>
<td>Concurrent chemotherapy</td>
<td>1.64</td>
<td>0.48-5.60</td>
</tr>
<tr>
<td>Radiation therapy dose</td>
<td>1.21</td>
<td>0.51-2.86</td>
</tr>
<tr>
<td>KPS</td>
<td>3.77</td>
<td>1.50-9.48</td>
</tr>
</tbody>
</table>

Abbreviations: CI = confidence interval; HR = hazard ratio; KPS = Karnofsky performance status; OS = overall survival.

* P <.01.
† P <.05.

Nonsignificant clinical factors were not presented for multivariate analysis.
therapy to an unselected population, reliable biomarkers could be used to identify those patients at highest risk of distant progression, who may benefit from additional systemic therapy.

Our study has several limitations. This was a retrospective study with a relatively small cohort. To apply the integrated analysis, accurate co-registration between PET and CT is required, although we showed that the volume of high-risk subregion was robust against this kind of uncertainty of a reasonable amount (3 mm). Another technical limitation is 3-dimensional PET acquisition in the presence of respiratory motion, which could lead to distortion of the true metabolic activity, especially for tumors located close to the diaphragm. Respiratory-gated and 4-dimensional PET acquisition (38) may help improve the quantitative accuracy of PET imaging.

Finally, the prognostic value of the new imaging biomarker needs to be evaluated on large prospective cohorts in future studies. The volume of the high-risk subregion before treatment did not predict local failure (or in-field progression) in our study population. It would be interesting to include longitudinal scans (39-41) during follow-up and investigate whether the changes in the imaging feature could be useful in predicting local failure.

Conclusions

We have developed a robust, intratumor partitioning framework to identify clinically relevant subregions within a tumor by integrating FDG-PET and CT images. The volume of a high-risk subregion characterized by the most metabolically active and metabolically heterogeneous solid component of the tumor is predictive of survival and OFP in NSCLC. Its prognostic value needs to be confirmed in prospective large cohorts.

References