Interobserver variability in tumor contouring affects the use of radiomics to predict mutational status

Qiao Huang
Lin Lu
Laurent Dercle
Philip Lichtenstein
Yajun Li
Qian Yin
Min Zong
Lawrence Schwartz
Binsheng Zhao

Interobserver variability in tumor contouring affects the use of radiomics to predict mutational status

Qiao Huang, Lin Lu, Laurent Dercle, Philip Lichtenstein, Yajun Li, Qian Yin, Min Zong, Lawrence Schwartz, and Binsheng Zhao*

Columbia University Medical Center, Department of Radiology, New York, New York, United States

Abstract. Radiomic features characterize tumor imaging phenotype. Nonsmall cell lung cancer (NSCLC) tumors are known for their complexity in shape and wide range in density. We explored the effects of variable tumor contouring on the prediction of epidermal growth factor receptor (EGFR) mutation status by radiomics in NSCLC patients treated with a targeted therapy (Gefitinib). Forty-six early stage NSCLC patients (EGFR mutant:wildtype = 20:26) were included. Three experienced radiologists independently delineated the tumors using a semiautomated segmentation software on a noncontrast-enhanced baseline and three-week post-therapy CT scan images that were reconstructed using 1.25-mm slice thickness and lung kernel. Eighty-nine radiomic features were computed on both scans and their changes (radiomic delta-features) were calculated. The highest area under the curves (AUCs) were 0.87, 0.85, and 0.80 for the three radiologists and the number of significant features (AUC > 0.8) was 3, 5, and 0, respectively. The AUCs of a single feature significantly varied among radiologists (e.g., 0.88, 0.75, and 0.73 for run-length primitive length uniformity). We conclude that a three-week change in tumor imaging phenotype allows identifying the EGFR mutational status of NSCLC. However, interobserver variability in tumor contouring translates into a significant variability in radiomic metrics accuracy.

© 2017 Society of Photo-Optical Instrumentation Engineers (SPIE) [DOI: [10.1117/1.JMI.5.1.011005](https://doi.org/10.1117/1.JMI.5.1.011005)]

Keywords: radiomics; epidermal growth factor receptor; nonsmall cell lung cancer; contouring; variability.

Paper 17193SSPR received Jun. 30, 2017; accepted for publication Sep. 21, 2017; published online Oct. 20, 2017.

1 Introduction

Lung cancer is the leading cause of cancer death in the USA and worldwide for men and women.¹ Nonsmall cell lung carcinoma (NSCLC) accounts for ~85% of cases. One mechanism of the tumorigenesis of NSCLC is the appearance of somatic mutations of the epidermal growth factor receptor (EGFR) gene that is overexpressed in NSCLC tissue as compared to the adjacent normal lung tissue.² When ligands bind to the EGFR tyrosine kinase receptor, the molecule is phosphorylated and activates a downstream signaling pathway that promotes NSCLC growth. Patients with NSCLC harboring a mutation in the EGFR tyrosine kinase receptor benefit from a personalized treatment strategy by EGFR tyrosine kinase inhibitors such as Gefitinib.

A mutation in the EGFR gene is a strong predictor of a prolonged progression-free survival with Gefitinib and is associated with a high response rate according to RECIST^{3–6} and an even higher rate of symptomatic improvement.^{7,8} However, EGFR mutational status only partially predicts sensitivity to Gefitinib^{9,10} and the identification of early predictors of response remains an area of active investigation. Independent prognostic factors associated with improved survival have been highlighted, such as female gender, never-smokers, good performance status, and adenocarcinoma histology.^{7,8,11}

The literature suggested that precision medicine guided by radiomics, which extracts imaging features for characterizing the imaging phenotype of NSCLC, could improve the

management of NSCLC patients. Test–retest and correlation analyses have indeed identified nonredundant computed tomography (CT) image features, which are minimally changed after a short-interval repeat scan, have high inpatient reproducibility and interpatient biological range.¹² Promising results have suggested that imaging features are associated with lung cancer histology,¹³ prognosis,^{12,14} EGFR-mutation status,¹⁵ distant metastasis risk,¹⁶ pathologic gross residual disease,¹⁷ disease-free survival,¹⁸ and cancer recurrence risk.¹⁹

However, a major drawback in the implementation of radiomics is that the extraction of imaging features is dependent on delineated tumor volume: the difficulties posed by volumetric measurements are significant. The first issue is that the manual contouring of a lesion throughout its entire volume is a time-consuming procedure for the radiologists. While automated systems are developed for computer-driven contouring of lesions, consistently accurate contours still require trained human input. The second problem is the possible amplification of interradiologist variability that comes from delineating the entire circumference of a tumor over each slice of a CT scan. Finally, the greatest problem is that observers vary in their judgment of the optimal part of the tumor to include (notably in ground glass opacities and atelectasis), which could make for arbitrary unidimensional measurements that have a profound effect on extraction of advanced imaging features derived from the tumor volumes.

Since EGFR-sensitizing mutations have been validated as highly predictive of improved progression-free survival on

*Address all correspondence to: Binsheng Zhao, E-mail: bz2166@cumc.columbia.edu

Gefitinib, we used a model studying the differential response of EGFR mutant and wild-type tumors to Gefitinib therapy. We hypothesized that advanced quantitative imaging features could be utilized as biomarkers to distinguish EGFR-resistant tumors from those harboring EGFR-sensitizing mutations. Therefore, imaging features could play a role in selecting patients who will potentially benefit from Gefitinib while identifying others who should be spared the toxicity. The aim of this study was to better understand the effect of interobserver variability of tumor-contouring on the quantification of changes in tumor imaging phenotype by radiomic analysis. To this end, we evaluated the diagnostic accuracy of imaging features extracted from NSCLC tumors by three different radiologists on predicting EGFR mutational status according to the change in imaging features between baseline and three weeks after Gefitinib treatment initiation.

2 Materials and Methods

2.1 Patient Clinical and Imaging Data

Patients' imaging and mutation status information were retrieved from a completed phase II trial correlating early radiographic response induced by Gefitinib with mutations in the protein-tyrosine kinase domain of the EGFR gene in patients with NSCLC.¹⁴ It was an institutional review board approved study conducted between July 2004 and March 2008. Fifty patients had stage I or II NSCLC and were deemed operable and resectable. Patients received Gefitinib daily for three weeks before surgery and discontinued it two days before their operations. At the time of resection, tumor tissue was snap frozen in liquid nitrogen and stored in a -80°C freezer. Representative areas of these specimens were pathologically reviewed to confirm the diagnosis and presence of tumor. Genomic DNA was analyzed for the most common EGFR-sensitizing mutations using previously described PCR-based methods.^{16–18}

Noncontrast-enhanced diagnostic chest CT imaging scans were performed within two weeks before starting Gefitinib (baseline scan) and after Gefitinib but before surgery (three-week follow-up scan), using the same imaging acquisition technique. High-resolution images with 1.25-mm slice thickness and a lung reconstruction kernel were collected per our volumetric imaging protocol. Four patients were excluded from this radiomic analysis either because 1.25-mm slice thickness or lung reconstruction were not done, leaving 46 patients (EGFR mutant:wildtype = 20:26) qualified for our analysis.

2.2 Tumor Segmentation

Tumor segmentation was performed using an imaging platform that we built based on an open source software, Weasis.^{20,21} Our semiautomated lung lesion segmentation algorithm and contour editing tool were integrated into this system.²² Three radiologists (R1, R2, and R3), with CT reading experiences of 23, 15, and 8 years respectively, used the image platform to independently delineate the tumors at baseline and three-week follow-up scans. Computer-generated tumor contours were then overlapped on the original baseline and follow-up images, side-by-side, for radiologist's review. If any contours were deemed suboptimal, radiologists were allowed to correct the contours using the editing tool.

2.3 Feature Extraction

For each segmented tumor on both baseline and follow-up scans, 89 well-defined radiomic features were extracted.²³ These features were developed to characterize tumor size (e.g., volume), shape (e.g., compactness factor), margin (e.g., Sigmoid function), low-order density statistics (e.g., histogram-derived standard deviation), and density texture distributions (e.g., gray level co-occurrence matrix, GLCM). Some features were computed in three dimensions (3-D) and some in two dimensions (2-D). The 2-D features were calculated on the automatically determined axial image where the lesion had the maximal diameter. More details on the definition and computation of 89 features can be found in supplemental materials in Zhao's publication.²³ Delta-features, the differences between baseline and follow-up features, were then calculated.

2.4 Reproducibility Test

To qualify as an imaging biomarker, a radiomic feature must be reproducible. The reproducibility of the 89 radiomic features was explored using a publicly available data set: the reference image database to evaluate therapy response (RIDER).²⁴ Briefly, the same-day repeat CT data set consisted of 32 lung cancer patients who underwent two repeat CT scans within 15 min, using the same imaging protocol (noncontrast-enhanced, 1.25-mm slice thickness, and lung reconstruction kernel). Each feature was calculated on both of the scan images and reproducibility was assessed by the concordance correlation coefficient (CCC) method as reported in Ref. 23. If either its baseline or follow-up feature was of CCCs < 0.85 on the RIDER data set, it would be regarded as a nonreproducible feature and excluded from the delta-feature set.

2.5 Redundancy Test

We used a clustering approach²⁵ to identify nonredundant features. First, we defined redundant feature subgroups within the delta-feature set. Features with correlation of Spearman's rho correlation coefficient greater than 0.5 were regarded as highly correlated features and gathered into one redundant feature subgroup. Second, we selected the delta-feature with the highest area under the curve (AUC) for predicting EGFR mutation as the representative feature for each redundant feature subgroup. Third, in each redundant feature subgroup, we only used the representative feature for analysis, excluding the other features within the same group.

2.6 Statistical Analysis

The AUC of the receiver operator characteristic was computed to assess the power of delta-features for predicting the EGFR mutational status of patients. 95% confidence interval of AUC was computed as well to indicate the stability of the predictions.

3 Results

3.1 Reproducibility and Nonredundancy

Using the CCC reproducibility cut-off value of 0.85 and the redundancy cut-off value of 0.5, 24, 19, and 20 out of the 89 features were deemed to be reproducible and nonredundant

Table 1 Diagnostic accuracy of reproducible and nonredundant radiomic delta-features for the prediction of the EGFR mutational status.

R1		R2		R3	
Feature name	AUC	Feature name	AUC	Feature name	AUC
Delta-volume	0.87	Delta-intensity_mean	0.85	Delta-compact_factor	0.80
Delta-run_short_run_emphasis	0.77	Delta-intensity_skewness	0.84	Delta-run_gray_level_uniformity	0.79
Delta-compact_factor	0.75	Delta-volume	0.81	Delta-intensity_skewness	0.78
Delta-GLCM_diff_variance	0.74	Delta-intensity_kurtosis	0.79	Delta-GLCM_sum_variance	0.76
Delta-GTDM_coarseness	0.74	Delta-compact_factor	0.78	Delta-GLCM_ASM	0.74
Delta-eccentricity	0.73	Delta-eccentricity	0.77	Delta-run_percentage	0.73
Delta-intensity_kurtosis	0.73	Delta-GLCM_sum_entropy	0.73	Delta-GLCM_sum_entropy	0.72
Delta-GLCM_entropy	0.70	Delta-GLCM_ASM	0.73	Delta-intensity_kurtosis	0.71
Delta-intensity_skewness	0.70	Delta-shape_SI7	0.72	Delta-solidity	0.70
Delta-GLCM_diff_entropy	0.70	Delta-run_short_run_emphasis	0.68	Delta-GTDM_coarseness	0.67
Delta-laws_1	0.68	Delta-DWT1_LHH	0.67	Delta-GTDM_strength	0.67
Delta-GLCM_sum_variance	0.67	Delta-GLCM_IMC2	0.64	Delta-shape_SI6	0.65
Delta-round_factor	0.66	Delta-solidity_2-D_maxdiameter	0.61	Delta-shape_SI7	0.63
Delta-sigmoid_offset	0.65	Delta-GLCM_IMC1	0.60	Delta-GLCM_IMC2	0.63
Delta-sigmoid_slope	0.65	Delta-DWT1_HLH	0.59	Delta-shape_SI2	0.62
Delta-shape_SI3	0.63	Delta-intensity_std	0.59	Delta-shape_SI5	0.62
Delta-shape_SI7	0.61	Delta-shape_SI5	0.58	Delta-fractal_mean	0.58
Delta-LoG_MGI	0.61	Delta-sigmoid_offset	0.57	Delta-eccentricity	0.58
Delta-shape_SI6	0.60	Delta-shape_SI2	0.57	Delta-sigmoid_slope	0.58
Delta-GLCM_IMC2	0.60			Delta-sigmoid_offset	0.57
Delta-DWT1_LHH	0.60				
Delta-shape_SI9	0.60				
Delta-DWT1_LLH	0.58				
Delta-shape_SI2	0.57				

Note: All *P*-values were significant and the 95% confidence interval of the AUCs ranged from 0.015 to 0.025.

for radiologists 1, 2, and 3's measurements and used for further analysis (see Table 1).

3.2 Variability in the Prediction Performances of the EGFR Status

Figure 1 shows an example of two tumors contoured by the three radiologists on baseline and follow-up scan images. The corresponding performances for the prediction of EGFR status are presented in Table 1. We can see that the mutation prediction performances varied due to different tumor contours drawn by the three radiologists. The highest AUCs of the three radiologists were 0.87, 0.85, and 0.80, respectively. We call a feature

Fig. 1 Two tumors segmented by three radiologists, on (a) baseline and (b) follow-up CT images. Different colors represent different radiologists' contouring results.

the significant feature if its predictive power $AUC > 0.8$ (bold features in Table 1).

3.3 Top Three Imaging Features

For R1, the “volume” was the best feature and was the only significant feature. For R2, in addition to the “volume,” “intensity_mean” and “intensity_skewness” were also significant features. In contrast to R1, the best feature for R2 was “intensity_mean,” a surrogate marker of tumor intensity (density). For R3, there was no significant feature and the best feature was “compact_factor,” a feature characterizing tumor shape. Overall, the top three imaging features for the prediction of EGFR mutational status in radiologists R1, R2, and R3 were delta-volume ($AUC = 0.87, 0.81, \text{ and } 0.73$), delta-intensity mean ($AUC = 0.67, 0.82, \text{ and } 0.74$), and delta-compact factor ($AUC = 0.75, 0.78, \text{ and } 0.79$). The group of NSCLC tumors with a mutation in EGFR tended to have a significant decrease in their volume, their compact factor, and intensity mean ($P < 10^{-3}$).

3.4 Bias Among Observers

The Bland–Altman plots and box plots in Figs. 2–4 show that there is a systematic measurement bias between radiologists 1, 2, and 3 at baseline and follow-up for the estimation of the top three imaging features. However, there is no significant bias using delta-feature since the bias is the same at baseline and follow-up. As a result, 85 out of the 89 delta features have no significant bias ($p < 0.05$).

4 Discussion

In this study, we evaluated the early three-week changes in tumor imaging phenotype on CT scan in NSCLC patients treated with Gefitinib. Patients were divided into two groups according to a tissue biomarker extracted from surgical specimen: EGFR-mutant ($n = 20$) versus EGFR-wildtype ($n = 26$). We tested the accuracy of imaging features to distinguish between tumors with and without sensitizing mutations. Strikingly, we observed that early changes in imaging features allowed differentiating NSCLC according to their mutational status. Indeed, EGFR-mutant had a significant early decrease in tumor volume, tumor mean intensity (density of tumor tissue), and tumor compactness (a shape feature) after treatment initiation. Targeting a single molecular pathway is thus responsible for changes in tumor cell biology that translates into early three-week changes in imaging phenotype that could be used as a tool for early dichotomizing sensitive and resistant tumors. Interestingly, we also demonstrated the significant effect of interobserver variability of tumor contouring with a significant bias among observers: radiologist’s interpretation of tumor boundary did influence the performance of EGFR mutation prediction, the AUC values.

The first important finding of this study is that there is a trend of interobserver variability in tumor contouring: trained radiologists tend to focus on the solid component of a tumor as opposed to the ground glass opacity (GGO) component. However, we will validate this finding by asking more radiologists to perform this task again for our future study. Indeed, the estimation of imaging features was associated with the experience of the radiologist. An increased experience of the radiologist ($R1 > R2 > R3$) was associated with a smaller tumor volume ($R1 < R2 < R3$) and a higher tumor mean

intensity ($R1 > R2 > R3$). Consequently, these results suggest that the most experienced radiologist (R1) tends to restrict the contouring to the dense solid component of the tumor, whereas less experienced radiologists (R2 and R3) tend to include the less dense GGO component of the tumor. As a result, in the tumors with a mixed component and important part of GGO (the tumors with the lowest average mean intensity), the interobserver difference is higher than in the tumor with mostly solid component. These results support the hypothesis that the GGO component plays a key role in the interobserver variability of contouring and that the radiomic output is dependent on the radiologists. Consequently, our findings may provide valuable insight into the proper delineation of tumors for further radiomic studies in lung cancer.

There has been little research to address the issue of interobserver variability in tumor contouring and thus in the measurement of quantitative imaging biomarkers in a clinical setting. We demonstrated that this significant interobserver variability in tumor contouring translated into significant differences in the accuracy of radiomic prediction of EGFR mutational status in NSCLC. We believe that with the rapid growth of the field of radiogenomics, our findings are valuable because they increase awareness of interobserver variations in the performance of radiomic features.

Despite the interobserver variability, we identified three delta-features that seemed robustly associated with EGFR mutational status in NSCLC treated with Gefitinib: delta-volume, delta-compact factor, and delta-intensity mean. Remarkably, we demonstrated that delta-tumor volume is the best predictor of tumor EGFR status and was not outperformed by other imaging features. This is explained by the fact that EGFR targeted molecular therapy is more efficient in the EGFR-mutant subpopulation of patients, leading to a greater tumor shrinkage. In addition, we demonstrated that the estimation of tumor volume is biased: senior radiologists tend to significantly delineate a smaller volume than junior radiologists ($P < 0.05$) that included less GGO (as demonstrated by the fact that the mean intensity tended to be higher in the senior radiologists). The assumed explanation is that trained radiologists tend to prefer the contouring of the most dominated part of the tumor: the solid portion. This illustrates the difficulty of a robust delineation of tumor contours in NSCLC, since GGO is a nonspecific imaging feature that might not be with certainty attributed to a tumor process. This point is important to consider for the implementation of radiomics features or volumetric measures in clinical routine: in the case of manual delineation of tumor contours, the baseline and follow-up exams should be performed by the same observer. Additionally, it further demonstrates that volumetric tumor measurement is the best biomarker and the use of volume-based response assessment for the development of tissue biomarkers could reduce contamination between sensitive and resistant tumor populations.²⁶ More importantly, volume measurement captures the change of the entire volume and is reproducible: the variability on same-day repeat CT scans in NSCLC is low (95% CI: $-12.1\% \text{ to } +13.4\%$).²⁷

Delta-compact factor appeared to be a good predictor of EGFR status after the initiation of Gefitinib for all three radiologists ($AUC = 0.75, 0.78, \text{ and } 0.79, p = 0.02$). Compact factor quantifies the compactness of a tumor in 3-D and is a function of the tumor volume and its surface. A tumor is considered compact if it is a sphere, and its compactness has the highest value compared to other shapes. Notably, we showed that

Fig. 2 (a)–(c) Measurement of the volume (baseline, follow-up, and delta) by three radiologists. We see a significant bias among radiologists: R1 tended to draw a smaller volume than R2 and R3. However, the delta volume was not significantly different since the same bias was observed at baseline and follow-up.

EGFR-mutant NSCLC tumors have a significant decrease in tumor volume as well as in compactness: response in tumor was associated with a less compact shape. Our hypothesis is that Gefitinib is an inhibitor of the EGFR tyrosine kinase that induces retractile remodeling (hypothetically fibrotic process) in EGFR-mutant NSCLC. Indeed, a decrease in compact factor was observed in all patients and compact factor was not highly correlated and redundant with the tumor volume.

We believe that our findings could have a major impact on the application of radiomics in patients with NSCLC. Indeed, CT scans guide decision-making throughout the course of NSCLC disease, including the screening,^{28–30} characterization of lung nodules,³¹ TNM staging, biopsy guiding, radiation treatment planning, and response assessment.

Finally, the accuracy of these radiomic features is an important point to consider since they outperformed the current

Fig. 3 (a)–(c) Measurement of the mean intensity (baseline, follow-up, and delta) by three radiologists. We see a significant bias among radiologists: R1 tended to draw a smaller volume than R2 and R3 and restricted the contours to the most dense part of the tumor. Consequently, the differences among observers were more important in tumors with low-density and high-GGO component. However, the delta intensity mean was not significantly different since the same bias was observed at baseline and follow-up.

reference standard: the measurement of tumor unidimensional size per RECIST criteria, which was one of the 89 delta features studied in this paper and not picked up in Table 1. This basic imaging feature has been a key metric for the management of

patients with NSCLC. As a matter of fact, tumor unidimensional size is integral for T-staging, predicting patient outcome, and is the reference standard for response evaluation criteria (e.g., RECIST is defined by the change in unidimensional size of

Fig. 4 (a)–(c) Measurement of the compact factor (baseline, follow-up, and delta) by three radiologists. We see a significant bias among radiologists: R1 tended to draw a smaller volume than R2 and R3 leading to higher compact-factor. However, the delta compact-factor was not significantly different since the same bias was observed at baseline and follow-up.

target lesions). Nonetheless, the optimal measurement of this apparently simple metric—the primary tumor size—remains an area of active investigation since the measurement of a tumor with both solid and GGO components can sometimes be made arbitrarily in regards to what constitutes tumor versus normal reactive lung tissue. Follow-up measurements should not be biased by interobserver variability, and it is important that decisions be kept consistent along all scans and time points. Therefore, ensuring reproducible tumor measurements across the treatment sequence is essential.

Radiomic delta-features can be used as potential imaging biomarkers to make an early prediction of the gene mutational status of patients and its response to targeted molecular therapies. However, tumor delineation produced differences in predicting EGFR mutations and warrants further investigation.

Disclosures

No conflicts of interest, financial or otherwise, are declared by the authors.

Acknowledgments

This work was supported by Grant Nos. R01 CA149490 and U01 CA140207 from the National Cancer Institute (NCI). The content is solely the responsibility of the authors and does not necessarily represent the funding sources.

References

1. R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics, 2016," *CA Cancer J. Clin.* **66**(1), 7–30 (2016).
2. V. Rusch et al., "Differential expression of the epidermal growth factor receptor and its ligands in primary non-small cell lung cancers and adjacent benign lung," *Cancer Res.* **53**(10 Suppl.), 2379–2385 (1993).
3. T. S. Mok et al., "Gefitinib or carboplatin-paclitaxel in pulmonary adenocarcinoma," *N. Engl. J. Med.* **361**(10), 947–957 (2009).
4. R. Rosell et al., "Screening for epidermal growth factor receptor mutations in lung cancer," *N. Engl. J. Med.* **361**(10), 958–967 (2009).
5. D. M. Jackman et al., "Impact of epidermal growth factor receptor and KRAS mutations on clinical outcomes in previously untreated non-small cell lung cancer patients: results of an online tumor registry of clinical trials," *Clin. Cancer Res.* **15**(16), 5267–5273 (2009).
6. L. V. Sequist et al., "First-line Gefitinib in patients with advanced non-small-cell lung cancer harboring somatic EGFR mutations," *J. Clin. Oncol.* **26**(15), 2442–2449 (2008).
7. M. G. Kris et al., "Efficacy of Gefitinib, an inhibitor of the epidermal growth factor receptor tyrosine kinase, in symptomatic patients with non-small cell lung cancer: a randomized trial," *J. Am. Med. Assoc.* **290**(16), 2149–2158 (2003).
8. V. A. Miller et al., "Bronchioloalveolar pathologic subtype and smoking history predict sensitivity to Gefitinib in advanced non-small-cell lung cancer," *J. Clin. Oncol.* **22**(6), 1103–1109 (2004).
9. J. G. Paez et al., "EGFR mutations in lung cancer: correlation with clinical response to Gefitinib therapy," *Science* **304**(5676), 1497–1500 (2004).
10. T. J. Lynch et al., "Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to Gefitinib," *N. Engl. J. Med.* **350**(21), 2129–2139 (2004).
11. P. A. Janne et al., "Outcomes of patients with advanced non-small cell lung cancer treated with Gefitinib (ZD1839, 'Iressa') on an expanded access study," *Lung Cancer* **44**(2), 221–230 (2004).
12. Y. Balagurunathan et al., "Test-retest reproducibility analysis of lung CT image features," *J. Digital Imaging* **27**(6), 805–823 (2014).
13. W. Wu et al., "Exploratory study to identify radiomics classifiers for lung cancer histology," *Front. Oncol.* **6**, 71 (2016).
14. D. V. Fried et al., "Prognostic value and reproducibility of pretreatment CT texture features in stage III non-small cell lung cancer," *Int. J. Radiat. Oncol. Biol. Phys.* **90**(4), 834–842 (2014).
15. H. J. Aerts et al., "Defining a radiomic response phenotype: a pilot study using targeted therapy in NSCLC," *Sci. Rep.* **6**, 33860 (2016).
16. T. P. Coroller et al., "CT-based radiomic signature predicts distant metastasis in lung adenocarcinoma," *Radiother. Oncol.* **114**(3), 345–350 (2015).
17. T. P. Coroller et al., "Radiomic phenotype features predict pathological response in non-small cell lung cancer," *Radiother. Oncol.* **119**(3), 480–486 (2016).
18. Y. Huang et al., "Radiomics signature: a potential biomarker for the prediction of disease-free survival in early-stage (I or II) non-small cell lung cancer," *Radiology* **281**(3), 947–957 (2016).
19. N. Emaminejad et al., "Fusion of quantitative image and genomic biomarkers to improve prognosis assessment of early stage lung cancer patients," *IEEE Trans. Biomed. Eng.* **63**(5), 1034–1043 (2016).
20. Weasis, <https://dcm4che.atlassian.net/wiki/display/WEA/Home>.
21. Weasis-pacs-connector, <https://sourceforge.net/projects/dcm4chee/files/Weasis/weasis-pacs-connector>.
22. Y. Tan, L. H. Schwartz, and B. Zhao, "Segmentation of lung lesions on CT scans using watershed, active contours, and Markov random field," *Med. Phys.* **40**(4), 043502 (2013).
23. B. Zhao et al., "Reproducibility of radiomics for deciphering tumor phenotype with imaging," *Sci. Rep.* **6**, 23428 (2016).
24. B. Zhao et al., "Evaluating variability in tumor measurements from same-day repeat CT scans of patients with non-small cell lung cancer 1," *Radiology* **252**(1), 263–272 (2009).
25. L. Lu et al., "Assessing agreement between radiomic features computed for multiple CT imaging settings," *PLoS One* **11**(12), e0166550 (2016).
26. B. Zhao et al., "A pilot study of volume measurement as a method of tumor response evaluation to aid biomarker development," *Clin. Cancer Res.* **16**(18), 4647–4653 (2010).
27. B. Zhao et al., "Evaluating variability in tumor measurements from same-day repeat CT scans of patients with non-small cell lung cancer," *Radiology* **252**(1), 263–272 (2009).
28. National Lung Screening Trial Research Team, "Reduced lung-cancer mortality with low-dose computed tomographic screening," *N. Engl. J. Med.* **365**(5), 395–409 (2011).
29. National Lung Screening Trial Research Team, "Results of initial low-dose computed tomographic screening for lung cancer," *N. Engl. J. Med.* **368**(21), 1980–1991 (2013).
30. J. Ma et al., "Annual number of lung cancer deaths potentially avertable by screening in the United States," *Cancer* **119**(7), 1381–1385 (2013).
31. H. MacMahon et al., "Guidelines for management of small pulmonary nodules detected on CT scans: a statement from the Fleischner Society," *Radiology* **237**(2), 395–400 (2005).

Qiao Huang received his Dr Eng degree from Shanghai Jiao Tong University in 2014. Since 2015, he has been in the Computational Imaging Analysis Laboratory, Columbia University Medical Center, where he is currently a postdoctoral research scientist. His research interests include pattern recognition, image processing, and computer vision.

Lin Lu graduated from Shanghai Jiao Tong University and received his PhD in biomedical engineering in 2013. He is an associate research scientist in the Department of Radiology, Columbia University Medical Center. His research interests include image processing, data mining, and machine learning. He has published more than 25 research papers in peer-reviewed journals, including *Journal of IEEE Transactions on Biomedical Engineering*, *Medical Physics*, *Proteome Research*, *Computational Chemistry*, etc.

Laurent Dercle received his MD awarded a gold medal and MSc degree in signal processing. He is currently a PhD candidate in oncology. He is a postdoctoral research fellow at Columbia University Medical Center. His main research interest is the development of precision medicine approach guided by quantitative imaging biomarkers in oncology.

Philip Lichtenstein received his MD degree from Chicago University. He is a radiologist resident at Columbia University Medical Center. He is a research scientist in the Computational Image Analysis Laboratory. His main research interest is in the implementation of radiomic metrics in lung cancer.

Yajun Li received her medical bachelor degree in 1993 and her medical master degree in 2002 from Xiangya Medical University of Central South University (CSU). She has been a radiologist for twenty-four years in the Radiology Department of Second Xiangya Hospital at the CSU. She became an associate professor in 2006. She has accumulated rich experience in imaging diagnosis, CT, and MRI image postprocessing, specialize in imaging diagnosis of the nervous system and head and neck.

Qian Yin received her doctor of medicine degree from the Medical School of Xi'an Jiaotong University, Xi'an, China, in 2001, and her PhD from the Fourth Military Medical University, Xi'an, China, in 2009. She is a radiologist and an assistant professor at the Fourth Military Medical University, Xi'an, China. Her main research interests include pulmonary microvascular environment and fibrosis.

Min Zong received his doctor of MD degree in 2010 and PhD in 2016 from Nanjing Medical University. He is a radiologist at Jiangsu Province Hospital and an assistant professor at Nanjing Medical University. His main research interests include the clinical aspects

of musculoskeletal Imaging, tumor texture analysis, and small animal imaging. He has published more than 15 journal papers.

Lawrence Schwartz has been working on cancer therapeutic response assessment using multiple imaging modalities for the past two decades. His research interests include improving response assessment methodologies by using innovative quantitative imaging biomarkers derived by state-of-the-art technologies of computer-aided image analysis, radiomics, and potentially deep learning.

Binsheng Zhao has been working for the past two decades on quantitative image analysis methods for multiple clinical applications in radiology and oncology. Her research interests include the development of reliable and efficient software tools to assist obtaining imaging biomarkers and optimization (including standardizations of imaging acquisition and measurement techniques) and validation of image-based methods to improve tumor response prediction and assessment in the era of personalized medicine.