


## ORIGINAL ARTICLE

# Parallel comparison and combining effect of radiomic and emerging genomic data for prognostic stratification of non-small cell lung carcinoma patients

Ki Hwan Kim<sup>1,2†</sup>, Jinho Kim<sup>3†</sup>, Hyunjin Park<sup>4,5†</sup>, Hankyul Kim<sup>1</sup>, Seung-hak Lee<sup>4</sup>, Insuk Sohn<sup>6</sup>, Ho Yun Lee<sup>1,7</sup>  & Woong-Yang Park<sup>3,7,8</sup>

1 Department of Radiology and Center for Imaging Science, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, South Korea

2 Department of Radiology, Myongji Hospital, Goyang, South Korea

3 Samsung Genome Institute, Biomedical Research Institute, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, South Korea

4 School of Electronic and Electrical Engineering, Sungkyunkwan University, Suwon, South Korea

5 Center for Neuroscience Imaging Research, Institute for Basic Science (IBS), Suwon, South Korea

6 Statistics and Data Center, Research Institute for Future Medicine, Samsung Medical Center, Seoul, South Korea

7 Department of Health Sciences and Technology, Samsung Advanced Institute for Health Science and Technology (SAIHST), Sungkyunkwan University, Seoul, South Korea

8 Department of Molecular Cell Biology, Sungkyunkwan University, Seoul, South Korea

## Keywords

CT; lung adenocarcinoma; quantitative imaging; survival stratification; texture analysis.

## Correspondence

Ho Yun Lee, Department of Radiology and Center for Imaging Science Samsung Medical Center, Sungkyunkwan University School of Medicine 50, Ilwon-Dong, Gangnam-Gu, Seoul 135-710, South Korea.

Tel: 822 3410 2502

Fax: 822 3410 0049

Email: hoyunlee96@gmail.com

<sup>†</sup>These authors contributed equally.

Received: 14 May 2020;

Accepted: 18 June 2020.

doi: 10.1111/1759-7714.13568

Thoracic Cancer **11** (2020) 2542–2551

## Abstract

**Background:** A single institution retrospective analysis of 124 non-small cell lung carcinoma (NSCLC) patients was performed to identify whether disease-free survival (DFS) achieves incremental values when radiomic and genomic data are combined with clinical information.

**Methods:** Using the least absolute shrinkage and selection operator (LASSO) Cox regression method, radiomic and genetic features were reduced in number for selection of the most useful prognostic feature. We created four models using only baseline clinical data, clinical data with selected genetic features, clinical data with selected radiomic features, and clinical data with selected genetic and radiomic features together. Multivariate Cox proportional hazards analysis was performed to determine predictors of DFS. Receiver operating characteristic (ROC) calculation was made to compare the discriminative performance for DFS prediction by four constructed models at the five-year time point.

**Results:** On precontrast scan, improved discrimination performance was obtained in a merging of selected radiomics and genetics (AUC = 0.8638), compared with clinical data only (AUC = 0.7990), selected genetic features (AUC = 0.8497), and selected radiomic features (AUC = 0.8355). On post-contrast scan, discrimination performance was improved (AUC = 0.8672) compared with the clinical variables (AUC = 0.7913), and selected genetic features (AUC = 0.8376) and selected radiomic features (AUC = 0.8399) were considered.

**Conclusions:** The combination of selected radiomic and genomic features improved stratification of NSCLC patients upon survival. Thus, integrating clinicopathologic model with radiomic and genomic features may lead to improved prognostic accuracy compared to conventional clinicopathological data alone.

**Key points/Significant findings of the study:** Receiver operating characteristic (ROC) calculation was made to compare the discriminative performance for disease-free survival (DFS). The discriminative performance for DFS was better when combining radiomic and genetic features compared to clinical data only, selected genetic features, and selected radiomic features.

**What this study adds:** The combination of selected radiomic and genomic features improved stratification of NSCLC patients upon survival. Thus, integrating a clinicopathological model with radiomic and genomic features may lead to improved prognostic accuracy compared to conventional clinicopathological data alone.

## Introduction

Lung cancer is the leading cause of cancer-related mortality worldwide, with non-small cell lung carcinoma (NSCLC) accounting for 85% of cases.<sup>1</sup> However, the widespread survival time that exists even after complete resection of NSCLC at the same stage indicates the vital importance of personalized medicine.<sup>2, 3</sup> Improvements in survival estimation have largely been made as a result of advances in biological and genomic technologies that enable the integration of survival-related biological or genetic signatures.<sup>4, 5</sup> However, the difficulty of obtaining comprehensive information on heterogeneous tumors is a limitation of such invasive methods.<sup>6, 7</sup> Notwithstanding, enabled by the possibilities of large-scale high-throughput next-generation sequencing (NGS) technologies along with the computational resources and tools to store, process and analyze the data, more powerful methods for the characterization of individual patients and tumor types became available.<sup>8</sup>

While substantial progress like NGS has been made in genetic information, new areas of attention have been created in the medical imaging field. Radiomics is a field of medical study that aims to extract a large amount of quantitative features from medical images using data-characterization algorithms.<sup>9–12</sup> Radiomics enables noninvasive profiling of tumor heterogeneity by extracting high-throughput quantitative descriptors from routinely gained computed tomography (CT) studies.<sup>9, 10, 13</sup> Recent advances in radiomics have provided insights into personalized medicine in oncologic practice in the areas of tumor detection, subtype classification, and response assessment to treatment.<sup>13–15</sup>

Despite recent developments in Genomics, there is still a need for surgical procurement of tissue. Radiographic imaging is routine in clinical practice, but is currently based on histopathology. If specific imaging traits and gene expression patterns that can predict underlying cellular pathophysiology can be correlated with each other through radiomics, radiologic data can be used as a molecular surrogate marker to monitor diagnosis, prognosis, and possibly gene-expression-associated treatment response of various human cancers.<sup>16</sup>

The aim of this study was to analyze disease-free survival (DFS) in patients with NSCLC through radiomic signature and to compare the results of radiomics with those of traditional staging systems or genetic analysis to determine if an incremental value could be obtained when they are combined.

## Methods

### Patient population

The Institutional Review Board approved this retrospective study which was exempt from the need to acquire informed consent. From May 2002 to January 2015, 171 patients with NSCLC had a surgical resection and agreed to genetic analysis of the specimen in our institution. After the review of medical records including imaging and pathologic studies, a total of 124 patients were included in this study (Fig 1). The exclusion of 47 patients was because genetically analyzed specimens had not been obtained from the lung ( $n = 20$ ), patient was lost to follow-up ( $n = 17$ ), insufficient image sequence for analysis ( $n = 5$ ), tumor too small to texture analysis ( $n = 3$ ) and metastatic pulmonary nodules ( $n = 2$ ). In our study, radiomics were all obtained by three-dimensional (3D) texture analysis, so it required at least three axial scan slices. Therefore, three cases whose axial scans were only one and two slices were included in the exclusion criteria.

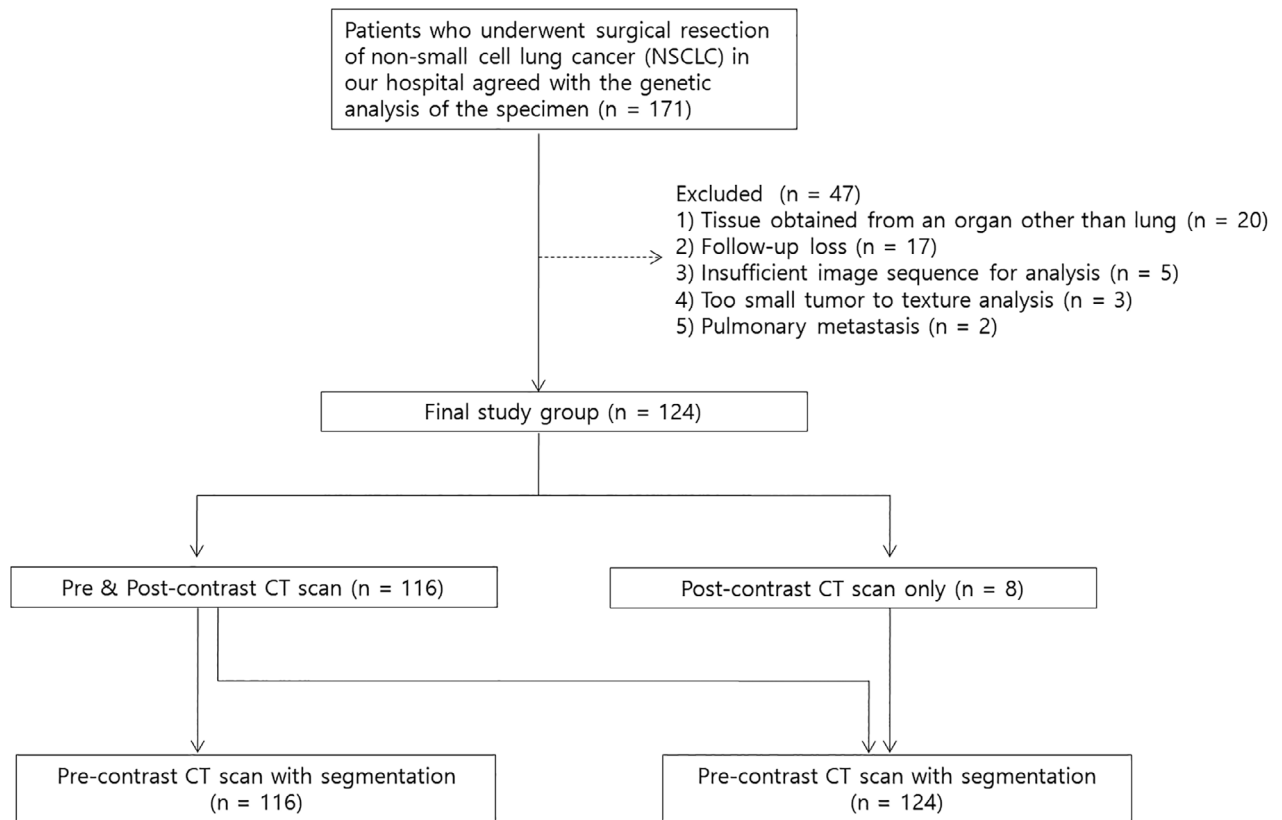
Research data collected through electronic medical records and clinical characteristics at the time of the diagnostic work-up were evaluated. Age, sex, smoking status, and American Joint Committee on Cancer (AJCC) stage were recorded. Pathology include histologic type and grade were recorded.

### Follow-up

The end point of this study was DFS, which was defined as the time from the surgical resection date until either the date of relapse (event), which refers to tumor recurrence within, or immediately adjacent to, the treated field, mediastinal relapse, distant relapse, or death, or until the date that the patient was last known to be free of relapse (censored). The minimum follow-up period to ascertain DFS was 18 months after surgical resection; while the maximum follow-up time was 175 months (median, 36 months). Following our institution's follow-up protocol, the patient was followed-up postoperatively by chest CT every 6–12 months for the first two years and annually thereafter.

### Imaging and texture analysis

Chest CT scan was used to assess the imaging characteristics of each lesion. Dedicated chest CT images were



**Figure 1** Flow diagram of the patient cohort.

obtained with several multidetector CTs including eight-, 16-, or 64-detector row CT scanners. Among 124 patients, 116 CT examinations consisted of pre- and post-contrast scan, while eight CT examinations consisted of post-contrast scan only. CT images were obtained with the following parameters: detector collimation was 1.25 or 0.625 mm, the tube peak potential energies ranged from 80 to 140 kVp, tube current ranged from 150 to 200 mA, and reconstruction interval ranged from 1 to 2.5 mm. All patients underwent chest CT at full inspiration through breath hold to minimize the effect of the tumor motion due to breathing. Chest CT scanning was obtained 90 seconds after the administration of contrast material. A total of 1.5 mL/kg (bodyweight) iomeron 300 (iomeprol, 300 mg iodine/mL; Bracco; Milan, Italy) was injected at an infusion rate of 3 mL/second using a power injector (MCT Plus; Medrad; Pittsburgh, PA, USA). All CT images were displayed at standard mediastinal (window width, 400 HU; window level, 20 HU) and lung (window width, 1500 HU; window level, -700 HU) window settings. In-plane resolution varied from 0.49 to 0.88 mm with a mean and standard deviation (SD) of 0.7 and 0.07, respectively. The mean slice thickness of images was 2.33 (range: 1–5 mm) and the SD was 0.98. Chest CT data were interfaced

directly with a picture archiving and communication system (PACS) (Path-Speed or Centricity 2.0; GE Healthcare, Mt. Prospect, IL, USA), which displayed all image data on two monitors (1536 × 2048 matrix, eight-bit viewable grayscale, 60-foot-lambert luminescence).

For quantitative CT analyses, our in-house software was used for lesion segmentation. Tumors were segmented by drawing a region of interest (ROI) with a semiautomatic approach using a MRicro (version 1.40, Chris Rorden, University of Nottingham, UK) that traced the edge of the tumor to include the largest area along all axial slice CT images until the entire tumor was covered, containing the lesion displayed in the lung window setting.

Segmentations were performed on precontrast CT scans (n = 116) and post-contrast CT scans (n = 124) each. Two radiologists carried out additional manual corrections to eliminate bronchovascular bundles and ground glass opacity boundary. After the nodule had been segmented, radiomic features were automatically calculated and extracted.

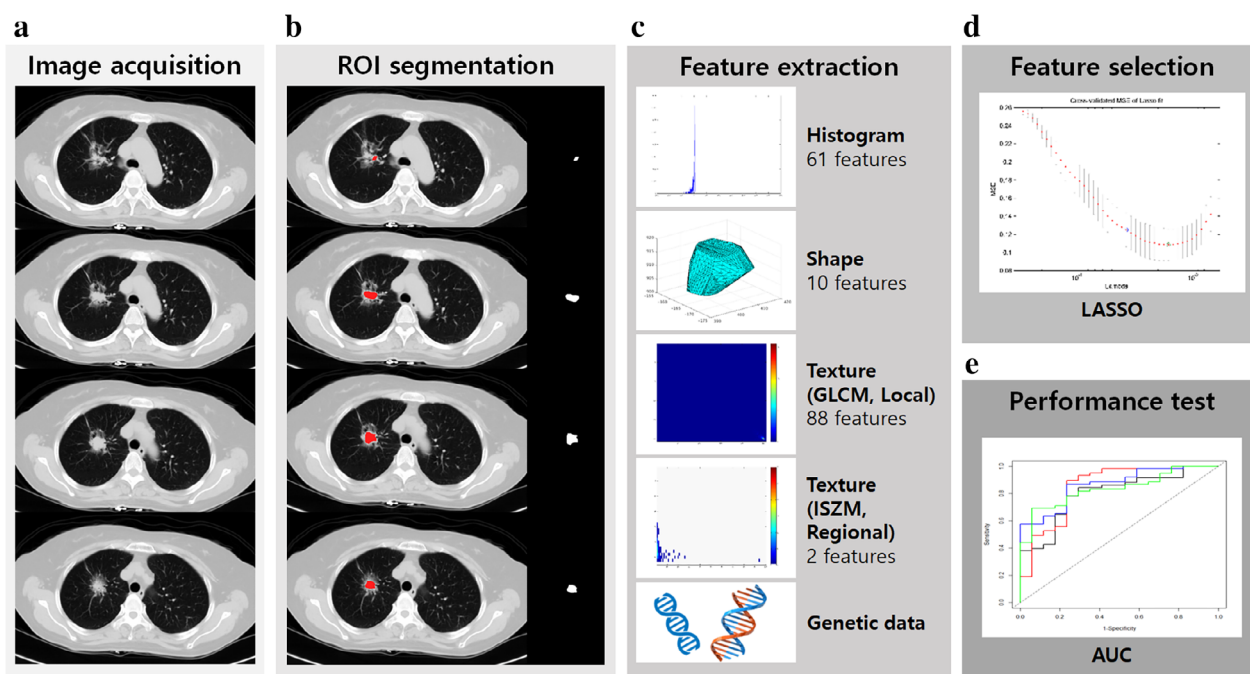
We calculated a total of 161 features of raw imaging over the given ROI using a combination of open-source (Pyradiomics) and in-house MATLAB code (MATLAB 2017b, Mathworks Inc., MA, USA). Our MATLAB code was used for features that were not calculated in

pyradiomics.<sup>13</sup> The features could be classified into four categories of histogram-based, shape-based, gray level co-occurrence matrix (GLCM)-based, and intensity size zone matrix (ISZM)-based. All of the features were verified for stability by image biomarker standardization initiative (IBSI).<sup>13, 17</sup> A detailed definition of the features adopted is given in Table S1. The workflow of radiomic feature extraction is illustrated in Figure 2.

## Genetic analysis

We obtained genomic variant information by using CancerSCAN, a targeted sequencing platform designed by Samsung Medical Center. This customized platform provides the researchers and clinicians the flexibility to include selected genes in the literature required. Single nucleotide variation (SNV), small indels, copy number variation, and gene fusion were detected using both existing<sup>18</sup> and custom algorithms.<sup>19</sup> Samples were typically sequenced without paired normal tissue. Genomic deoxyribonucleic acid (DNA) was extracted from formalin-fixed, paraffin-embedded (FFPE) tissue or fresh tissue using a QIAamp DNA mini kit (Qiagen, Valencia, CA, USA) and Promega Maxwell 16 CSC DNA FFPE kit or QIAamp DNA after

the pathologist's diagnosis and tumor content were examined. DNA concentration and purity were determined using a Nanodrop 8000 UV-Vis spectrometer (Thermo Scientific, Waltham, MA, USA) and a Qubit 2.0 fluorescence meter (Life Technologies, Grand Island, NY, USA). The degree of DNA degradation was measured using a 200 TapeStation Instrument (Agilent Technologies, Santa Clara, Calif., USA) and real-time PCR (Agilent Technologies, Santa Clara, CA, USA). Genomic DNA was sheared using Covaris S220 (Covaris, Woburn, Massachusetts, USA). We used SureSelect-XT Reagent Kit, HSQ (Agilent Technologies, Santa Clara, CA, USA) to capture the target and create a bidirectional sequencing library using barcodes. After checking the quality of the library, sequencing was performed on a HiSeq 2500 with a 100 bp reading (Illumina, San Diego, California, USA). When sequencing is complete, the analytics workflow has started automatically. Pair-ended reads were aligned to the human reference genome (hg19) using BWA-MEM (v.0.7.5). SAMTOOLS (v0.1.18), GATK (v3.1-1), and Picard (v1.93) were used for reading alignment, local realignment and duplicate read removal, respectively. The baseline quality score was recalibrated using the GATK BaseRecalibrator based on known single nucleotide polymorphisms (SNPs) and indels of dbSNP138. SNV was identified using a



GLCM: Gray-level co-occurrence matrix; ISZM: Intensity-size zone matrix; LASSO: Least absolute shrinkage and selection operator; AUC: Area under the curve

**Figure 2** Workflow of extracting radiomic features. (a) A lung tumor is scanned in multiple slices. (b) Experienced radiologists contour the tumor areas on all CT slices. (c) Features are extracted from within the defined tumor contours on the CT images. (d) Feature selection by the LASSO Cox regression method. (e) Receiver operating characteristic (ROC) analysis with area under the curve (AUC) calculation was conducted to compare the discrimination performance for the prediction of DFS. (■) Clinical, (■) Clinical + SGI, (■) Clinical + radiomics, (■) Clinical + radiomics + SGI.

combination of the two algorithms MuTect<sup>20</sup> (v1.1.4) and LoFreq<sup>21</sup> (v0.6.1). We applied a custom-designed filtering step based on a regression model trained on the low variant allele frequency (VAF) variance of the normal sample to improve its accuracy.<sup>22</sup> Pindel was used to search insertions and deletions.<sup>23</sup> To calculate copy number variations, the read depth per exon was divided by the estimated normal reads per exon using in-house references. Gene fusion of the target site was identified using an in-house algorithm. A detailed analysis workflow has been described in a previous report.<sup>22</sup> An open SNP database such as the Exome Aggregation Consortium (EXAC),<sup>19</sup> Korean Reference Genome (KRG), and Korean Variant Archive (KOVA)<sup>24</sup> was used to differentiate somatic mutations from germline mutations. VAF of at least 0.1% in any of the SNP databases were considered as germ-line variants. For the remaining variants, a manually constructed internal database containing variants frequently found in the entire cohort was used for additional variant filtering. The variant was annotated by ANNOVAR.<sup>25</sup>

Mutation data was condensed into multiple binary values for each gene to represent functional effects. (i) Loss of function (LoF) mutation: Damaging mutations including frame-shift insertions and deletions and stop gain mutations; (ii) S1 mutation: Mutations that have been implicated in cancer development or therapy; (iii) missense (MS) mutation: Mutations in the coding region causing amino acid change other than S1 mutations; (iv) amplification (Amp): Gene amplifications with estimated copies  $\geq 4$ ; (v) deletion (Del): One or two copy gene deletions; and (vi) gene fusion (Fusion): Gene fusions inferred from the sequenced DNA reads that were specially designed to detect gene fusions in the CancerSCAN platform. Only driver gene fusions were included.

## Statistical analysis

Continuous variables were presented as means and standard deviations and categorical variables as numbers and percentages.

We created four models using only clinical data, clinical data with selected genetic features, clinical data with selected radiomic features, and clinical data with selected genetic and radiomic features together. We performed least absolute shrinkage and selection operator (LASSO) to remove redundancy within the radiomic and genetic information by selecting the most prognostic characteristics in radiomic and genetic analysis.<sup>18, 26</sup> LASSO is a method of selecting a few suitable features using L1-norm regularization and the method is frequently used for the feature selection in radiomics.<sup>27, 28</sup> We performed multivariate Cox proportional hazards analysis to determine predictors of DFS. The variables investigated included demographics,

pathological characteristics, and selected radiomic and genetic features.

Next, leave-one-out cross validation tests were performed to test the validity of our prediction model. We divided the data set into equal subgroups where the number of folds was equal to the number of instances in the data set. Subsequently, to compare the discrimination performance for DFS prediction by four constructed models at the five-year time point, the receiver operating characteristic (ROC) analysis with area under the curve (AUC) calculation was performed. Statistical analyzes were performed using SAS v9.4 (SAS Institute Inc., Cary, NC) and R 3.4.0 (Vienna, Austria; <http://www.R-project.org/>). *P*-values less than 0.05 indicated statistical significance.

Using 39 randomly selected patients, the reliability and reproducibility of tumor segmentation and feature extraction was attained by comparing the intraclass correlation coefficient (ICC)<sup>29</sup> values between both radiologists.

## Results

### Patient group characteristics

Table 1 shows baseline clinicopathologic data such as age, sex, smoking status, AJCC stage, histologic type and histologic grade. A model using only clinical factors was designed with these variables. A total of 62 women (50%) and 62 (50%) men (median age, 56.6 years) were enrolled. We analyzed a total of 124 measurable lesions from 124 patients. There were 77 (62.1%) patients who were never-smokers, and 47 (37.9%) were ever-smokers. Among 124 lesions, 114 (91.9%) were adenocarcinomas, five were squamous cell carcinomas, two were pleomorphic carcinoma, two were large cell neuroendocrine carcinoma, and one was adenosquamous cell carcinoma. As of the last follow-up, 59 patients (34.4%) had experienced a confirmed disease relapse. The mean DFS was 27.66 months, and the median DFS was 23 months.

### Feature selection and diagnostic performance of constructed models

Using the LASSO method, radiomic and genetic features were reduced in number. On precontrast CT scan, two radiomic features (IMC1\_val\_out and IMC1\_val\_delta\_sub2) and six genetic features (ALK\_Fusion, ATRX\_Amp, MET\_MS, MLH1\_MS, MTOR\_MS, and STK11\_MS) were selected. On post-contrast CT scan, 12 radiomic features (Kurtosis\_val\_in, mean\_delta, Skewness\_delta, Sphericity\_val, Contrast\_glcml\_val\_delta\_sub2, Entropy\_glcml\_val\_in, Energy\_glcml\_val\_delta\_sub2, IMC1\_val\_delta, Max\_probability\_val\_in\_sub2, Max\_probability\_val\_out\_sub2, Variance\_glcml\_val\_delta, and Variance\_glcml\_val\_delta\_sub2) and seven genetic features

**Table 1** Patient group characteristics

Characteristics	No. of patients ( <i>n</i> = 124)
Age at diagnosis (year)	56.6 ± 11.9
Sex	
Female	62 (50)
Male	62 (50)
Smoking history	
Never	77 (62.1)
Ever	47 (37.9)
Pathology	
Adenocarcinoma	114 (91.9)
Squamous cell carcinoma	5 (4.0)
Pleomorphic carcinoma	2 (1.6)
Large cell neuroendocrine carcinoma	2 (1.6)
Adenosquamous cell carcinoma	1 (0.7)
Histologic grade	
Well-differentiated	3 (2.4)
Moderately-differentiated	89 (71.8)
Poorly-differentiated	32 (25.8)
AJCC stage at diagnosis	
I	53 (42.7)
II	25 (20.2)
III	35 (28.2)
IV	11 (8.9)
Follow-up time (months)	
Median (quartile)	36 ± (28.6–58.9)
Maximum	175

(ALK\_Fusion, ATRX\_Amp, MET\_MS, MLH1\_MS, MTOR\_MS, NF1\_LoF, and STK11\_MS) were selected. Radiomic and genetic features obtained from pre- and post-contrast scans were arranged and are shown in Tables S1 and S2.

Tables 2 and 3 summarize the survival prognosis performance of the three models. Using radiometric data in a precontrast CT scan, when selected radiomic and genetic features were integrated with clinical variables, the discrimination performance improved (AUC = 0.8638) compared with consideration of only clinical variables (AUC = 0.7990), selected genomic features incorporated with clinicopathological data (AUC = 0.8497), and selected radiomic features incorporated with clinicopathological data (AUC = 0.8355). When selected radiographic and genetic features were integrated with clinical variables using radiomic data on post-contrast CT scan, the discrimination performance improved (AUC = 0.8672) compared with consideration of only clinical variables

**Table 2** Diagnostic performances of models using clinical/genetic/radiomic data on precontrast CT

Model	AUC	<i>P</i> -value
Clinical data only	0.7990	Reference
Clinical + genetic data	0.8497	0.1923
Clinical + radiomic data	0.8355	0.2793
Clinical + genetic + radiomic data	0.8638	0.0523

(AUC = 0.7913), selected genomic features incorporated with clinicopathological data (AUC = 0.8376), and selected radiomic features incorporated with clinicopathological data (AUC = 0.8599).

### Reliability and reproducibility of tumor segmentation and feature extraction

The ICC values ranged from 0.484 to 1.000 with a mean

**Table 3** Diagnostic performances of models using clinical/genetic/radiomic data on post-contrast CT

Model	AUC	<i>P</i> -value
Clinical data only	0.7913	Reference
Clinical + genetic data	0.8376	0.1929
Clinical + radiomic data	0.8599	0.1994
Clinical + genetic + radiomic data	0.8672	0.2092

value of 0.929, representing a higher reliability level of agreement.

### Discussion

Recently, several researchers have succeeded in extracting radiomic features in the oncology field by computing enormous amounts of quantitative variables from conventional medical images.<sup>30–32</sup> Although previous studies have reported the prognostic value of some radiomic or genomic features, the question remains whether these new radiomic and genomic features have any incremental value above the current clinicopathologic model. Identifying the predictive power and clinical relevance of the numerous radiomic and genomic features for lung cancer survival remains an important issue.<sup>33, 34</sup>

In this study, we compared four models to evaluate selected radiomic and genomic feature increments along with clinicopathologic data using: (i) only clinicopathologic data; (ii) selected genomic features incorporated with clinicopathologic data; (iii) selected radiomic features incorporated with clinicopathologic data; and (iv) selected radiomic and genomic features incorporated with clinicopathologic data. The merging of selected radiomic and genomic features improved stratification in lung cancer patients upon survival. Thus, our results show that integration of the current clinicopathologic model with radiomic and genomic features may lead to improved prognostic accuracy compared to conventional clinicopathologic data alone.

We employed a cancer panel sequencing platform to explore relationships between cancer genomic features and radiomic characteristics. There have been many debates on

using whole-genome sequencing (WGS) versus whole-exome sequencing (WES) versus gene panels.<sup>35, 36</sup> For clinical sequencing, we previously reported that many hotspot variants have a low VAF variant,<sup>22</sup> which make it difficult to detect variants from a clinical sample at a standard coverage of WGS or WES platform (100–200x). The CancerSCAN panel sequencing platform aims at 1000x coverage to detect low VAF variants such as 2% VAF. This capability enables the detection of variants that only a small subclone contains, even in a low purity tumor sample. Thus our data represents near complete variant profiles in the panel genes, which other studies who employed WGS or WES would have missed.

Most lung adenocarcinoma genomic expression appears as a mixed-subtype such as polyclonal composition of two or more different pathologic subtypes.<sup>37, 38</sup> Recent NGS and bioinformatic studies have studied different types of tumor heterogeneity including intratumor heterogeneity in which heterogeneity between tumor cells exists due to the presence of multiple subclones within the tumor in addition to interpatient or intertumor heterogeneity.<sup>39, 40</sup> Based on this concept, recently published large databases characterizing the molecular features of human tumors are attempting to change the determination of each cancer type from the conventional histopathological classification to a new classification based on genetic identity.<sup>41–44</sup> Such changes could better explain the symptoms of partial tumor response to treatment, the emergence of drug-resistant malignant cells, or the dissemination of metastatic cells with the emergence of genomically distinct minor subclones of malignant cells more practically.<sup>45–47</sup> Several observations have demonstrated that the presence of minor subclones affects tumor progression.<sup>48–50</sup>

However, performing multiple tumor biopsies to precisely determine the clonal composition of a tumor is not a simple or practical solution in the context of clinical research. In addition, the results obtained with a single tumor-biopsy sample may also underestimate the overall genomic landscape. The error rate of cancer histopathology can be as high as 23%, and it is estimated that even the most clinically relevant solid tumors become very heterogeneous at phenotypic, physiologic, and genomic levels over time.<sup>11</sup> Phylogenetic reconstruction has been reported to show a branching evolutionary tumor growth in which almost 70% of all somatic mutations were not detected across all tumor regions.<sup>51</sup> In comparison, the medical image has the advantage of a quantified comprehensive macroscopic picture in estimating the entire tumor plus surrounding tissue relatively safely and less inconveniently for patients.

Radiogenomics focuses on defining relationships between radiomic phenotypes and genomic information to interconnect macroscopic imaging and subcellular characteristics to

as oncologic diagnosis changes from traditional tissue-based approaches to molecular stratification.<sup>52</sup> Although it is still poorly understood, many studies have attempted to find radiographic tumor phenotypes using radiomics and to find correlations with specific genetic expressions, and some have been successful.<sup>53–55</sup> Radiomics may function as complementary “virtual biopsy” information that can be monitored more frequently than the invasive method during an entire course of cancer treatment, but it will still be difficult to completely replace conventional biopsy, which allows more detailed genomic analysis.

Of the 14 radiomics selected in our study, 10 were GLCM-based features, one was shape, and three were histogram-based features, and we conclude that the results reflect intratumoral heterogeneity well. GLCM texture features contains information about the positions of pixels having similar gray level values. They describe the high-order statistical spatial distributions of the voxel intensities characterizing heterogeneity with spatial information within a tumor or ROI. Therefore, GLCM features extracted from the raw image can act as an effective texture descriptor to reflect lesion heterogeneity. Representative features extracted from GLCM included correlation, cluster, contrast, energy, and entropy.

In addition, in our study, the features that were finally selected through LASSO were generally partial ROI associated features (in, out, and subsampled) than features related to whole ROI. In a study conducted on breast cancer, Wu *et al.*<sup>56</sup> reported that the subregional radiomic analysis method had an advantage in quantifying the tumor subregion which was more correlated with the tumor growth or aggressiveness. Similarly, in our research, most of the radiomic features finally selected were partial ROI related features correlated with subregion microenvironment, and we can assume that these features were accompanied with risk group stratification.

We focused on CT-derived radiomic features in this study. CT is the most popular and routinely performed modality in the lung cancer field, and CT-derived radiomic features are practical for clinical application. We included both precontrast-enhanced CT scan and enhanced CT scan in this study. Theoretically, precontrast-enhanced CT-derived features are indicative of cellularity of the tumor and contrast-enhanced CT-derived features are indicative of vascularity of the tumor, reflecting both characteristics may represent tumor heterogeneity and complexity more accurately. As we utilized both precontrast scan CT-derived features and post-contrast CT scan-derived features, it broadens the applicability of our study, given the different imaging protocols of each hospital.

Despite the advantages of the radiomic approach to lung cancer prognostication, our study had some limitations. First, there was a small number of subjects included in our

study, and they were collected from a single institute. The second limitation is that our study was retrospectively performed and based on previously acquired gene-expression data. Therefore, our findings will need to be further characterized and validated in future studies by patient groups receiving the same treatment to generate gene-expression data prospectively.

Our results indicate that integrating the current clinicopathologic model with selected radiomic and genomic features may lead to improved prognostic accuracy compared to conventional clinicopathologic data alone.

## Acknowledgments

The authors would like to offer our special thanks to Insuk Sohn, PhD (Statistics and Data Center, Research Institute for Future Medicine, Samsung Medical Center, Seoul, Korea) for his assistance with the statistics used in this report. This research was supported by the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), which was funded by the Ministry of Health & Welfare (HI17C0086) and the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIP; Ministry of Science, ICT & Future Planning) (No. NRF-2016R1A2B4013046 and NRF-2017M2A2A7A02018568). The funders had no role in the design of the study, data collection, data analysis, interpretation, or the writing of this report.

## Disclosure

The authors declare that they have no conflicts of interests.

## References

- Brawley OW. Avoidable cancer deaths globally. *CA Cancer J Clin* 2011; **61** (2): 67–8.
- Chansky K, Sculier J-P, Crowley JJ *et al.* The International Association for the Study of Lung Cancer staging project: Prognostic factors and pathologic TNM stage in surgically managed non-small cell lung cancer. *J Thorac Oncol* 2009; **4** (7): 792–801.
- Scott WJ, Howington J, Feigenberg S *et al.* Treatment of non-small cell lung cancer stage I and stage II. *Chest* 2007; **132** (3): 234S–42S.
- Halabi S, Lin C-Y, Kelly WK *et al.* Updated prognostic model for predicting overall survival in first-line chemotherapy for patients with metastatic castration-resistant prostate cancer. *J Clin Oncol* 2014; **32** (7): 671–7.
- Zhang J-X, Song W, Chen Z-H *et al.* Prognostic and predictive value of a microRNA signature in stage II colon cancer: A microRNA expression analysis. *Lancet Oncol* 2013; **14** (13): 1295–306.
- Tran B, Dancey JE, Kamel-Reid S *et al.* Cancer genomics: Technology, discovery, and translation. *J Clin Oncol* 2012; **30** (6): 647–60.
- Hofman V, Ilie M, Long E *et al.* Immunohistochemistry and personalised medicine in lung oncology: Advantages and limitations. *Bull Cancer* 2014; **101** (10): 958–65.
- Horak P, Fröhling S, Glimm H. Integrating next-generation sequencing into clinical oncology: Strategies, promises and pitfalls. *ESMO Open* 2016; **1** (5): e000094.
- Lambin P, Rios-Velazquez E, Leijenaar R *et al.* Radiomics: Extracting more information from medical images using advanced feature analysis. *Eur J Cancer* 2012; **48** (4): 441–6.
- Kumar V, Gu Y, Basu S *et al.* Radiomics: The process and the challenges. *Magn Reson Imaging* 2012; **30** (9): 1234–48.
- Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images are more than pictures, they are data. *Radiology* 2016; **278** (2): 563–77.
- Parekh V, Jacobs MA. Radiomics: A new application from established techniques. *Expert Rev Precis Med Drug Dev* 2016; **1** (2): 207–26.
- Aerts HJWL, Velazquez ER, Leijenaar RTH *et al.* Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun* 2014; **5** (1): 4006.
- Altazi B, Fernandez D, Zhang G, Biagioli M, Moros E, Moffitt HL. SU-E-J-258: Prediction of cervical cancer treatment response using radiomics features based on F18-FDG uptake in PET images. *Med Phys* 2015; **42** (6Part11): 3326–6.
- Li H, Lan L, Drukker K, Perou C, Giger M. TU-AB-BRA-08: Radiomics in the analysis of breast cancer heterogeneity on DCE-MRI. *Med Phys* 2015; **42** (6Part31): 3588–8.
- Rutman AM, Kuo MD. Radiogenomics: Creating a link between molecular diagnostics and diagnostic imaging. *Eur J Radiol* 2009; **70** (2): 232–41.
- Zwanenburg A, Vallières M, Abdalah MA *et al.* The image biomarker standardization initiative: Standardized quantitative radiomics for high-throughput image-based phenotyping. *Radiology* 2020; **295** (2): 328–38.
- Lee H, Lee K-W, Lee T *et al.* Performance evaluation method for read mapping tool in clinical panel sequencing. *Genes Genomics* 2018; **40** (2): 189–97.
- Lek M, Karczewski KJ, Minikel EV *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature* 2016; **536** (7616): 285–91.
- Cibulskis K, Lawrence MS, Carter SL *et al.* Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat Biotechnol* 2013; **31** (3): 213–9.
- Wilm A, Aw PPK, Bertrand D *et al.* LoFreq: A sequence-quality aware, ultra-sensitive variant caller for uncovering cell-population heterogeneity from high-throughput sequencing datasets. *Nucleic Acids Res* 2012; **40** (22): 11189–201.
- Shin H-T, Choi Y-L, Yun JW *et al.* Prevalence and detection of low-allele-fraction variants in clinical cancer samples. *Nat Commun* 2017; **8** (1): 1377.



- 23 Ye K, Schulz MH, Long Q, Apweiler R, Ning Z. Pindel: A pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. *Bioinformatics* 2009; **25** (21): 2865–71.
- 24 Lee S, Seo J, Park J *et al.* Korean variant archive (KOVA): A reference database of genetic variations in the Korean population. *Sci Rep* 2017; **7** (1): 4287.
- 25 Wang K, Li M, Hakonarson H. ANNOVAR: Functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* 2010; **38** (16): e164.
- 26 Tibshirani R. Regression shrinkage and selection via the lasso: A retrospective. *J R Stat Soc Ser B* 2011; **73** (3): 273–82.
- 27 Zheng BH, Liu LZ, Zhang ZZ *et al.* Radiomics score: A potential prognostic imaging feature for postoperative survival of solitary HCC patients. *BMC Cancer* 2018; **18** (1): 1148.
- 28 Huang Y, Liu Z, He L *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* 2016; **281** (3): 947–57.
- 29 Bartko JJ. The intraclass correlation coefficient as a measure of reliability. *Psychol Rep* 1966; **19** (1): 3–11.
- 30 Lee G, Lee HY, Ko ES, Jeong WK. Radiomics and imaging genomics in precision medicine. *Precis Futur Med* 2017; **1** (1): 10–31.
- 31 Zhu Y, Mohamed ASR, Lai SY *et al.* Imaging-genomic study of head and neck squamous cell carcinoma: Associations between radiomic phenotypes and genomic mechanisms via integration of the cancer genome atlas and the cancer imaging archive. *JCO Clin Cancer Inform* 2019; **3** (3): 1–9.
- 32 Zhou M, Leung A, Echegaray S *et al.* Non-small cell lung cancer radiogenomics map identifies relationships between molecular and imaging phenotypes with prognostic implications. *Radiology* 2018; **286** (1): 307–15.
- 33 Bodalal Z, Trebeschi S, Beets-Tan R. Radiomics: A critical step towards integrated healthcare. *Insights Imaging* 2018; **9** (6): 911–4.
- 34 Keek SA, Leijenaar RT, Jochems A, Woodruff HC. A review on radiomics and the future of theranostics for patient selection in precision medicine. *Br J Radiol* 2018; **91** (1091): 20170926.
- 35 Yu Y, Wu BL, Wu J, Shen Y. Exome and whole-genome sequencing as clinical tests: A transformative practice in molecular diagnostics. *Clin Chem* 2012; **58**: 1507–9.
- 36 Meienberg J, Bruggmann R, Oexle K, Matyas G. Clinical sequencing: Is WGS the better WES? *Hum Genet* 2016; **135** (3): 359–62.
- 37 Lee HY, Jeong JY, Lee KS *et al.* Solitary pulmonary nodular lung adenocarcinoma: Correlation of histopathologic scoring and patient survival with imaging biomarkers. *Radiology* 2012; **264** (3): 884–93.
- 38 Motoi N, Szoke J, Riely GJ *et al.* Lung adenocarcinoma: Modification of the 2004 WHO mixed subtype to include the major histologic subtype suggests correlations between papillary and micropapillary adenocarcinoma subtypes, EGFR mutations and gene expression analysis. *Am J Surg Pathol* 2008; **32** (6): 810–27.
- 39 Jamal-Hanjani M, Quezada SA, Larkin J, Swanton C. Translational implications of tumor heterogeneity. *Clin Cancer Res* 2015; **21** (6): 1258–66.
- 40 Burrell RA, McGranahan N, Bartek J, Swanton C. The causes and consequences of genetic heterogeneity in cancer evolution. *Nature* 2013; **501** (7467): 338–45.
- 41 Curtis C, Shah SP, Chin SF *et al.* The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature* 2012; **486** (7403): 346–52.
- 42 Heist RS, Engelman JA. SnapShot: Non-small cell lung cancer. *Cancer Cell* 2012; **21** (3): 448.e2.
- 43 Hammerman PS, Voet D, Lawrence MS *et al.* Comprehensive genomic characterization of squamous cell lung cancers. *Nature* 2012; **489** (7417): 519–25.
- 44 Collisson EA, Campbell JD, Brooks AN *et al.* Comprehensive molecular profiling of lung adenocarcinoma: The cancer genome atlas research network. *Nature* 2014; **511** (7511): 543–50.
- 45 Baldus SE, Schaefer KL, Engers R, Hartleb D, Stoecklein NH, Gabbert HE. Prevalence and heterogeneity of KRAS, BRAF, and PIK3CA mutations in primary colorectal adenocarcinomas and their corresponding metastases. *Clin Cancer Res* 2010; **16** (3): 790–9.
- 46 Kidd EA, Grigsby PW. Intratumoral metabolic heterogeneity of cervical cancer. *Clin Cancer Res* 2008; **14** (16): 5236–41.
- 47 Huyge V, Garcia C, Alexiou J *et al.* Heterogeneity of metabolic response to systemic therapy in metastatic breast cancer patients. *Clin Oncol* 2010; **22** (10): 818–27.
- 48 Engelman JA, Settleman J. Acquired resistance to tyrosine kinase inhibitors during cancer therapy. *Curr Opin Genet Dev* 2008; **18** (1): 73–9.
- 49 Kosaka T, Yatabe Y, Endoh H *et al.* Analysis of epidermal growth factor receptor gene mutation in patients with non-small cell lung cancer and acquired resistance to gefitinib. *Clin Cancer Res* 2006; **12** (19): 5764–9.
- 50 Turke AB, Zejnullahu K, Wu YL *et al.* Preexistence and clonal selection of MET amplification in EGFR mutant NSCLC. *Cancer Cell* 2010; **17** (1): 77–88.
- 51 Gerlinger M, Rowan AJ, Horswell S *et al.* Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N Engl J Med* 2012; **366** (10): 883–92.
- 52 Kuo MD, Jamshidi N. Behind the numbers: Decoding molecular phenotypes with radiogenomics- guiding principles and technical considerations. *Radiology* 2014; **270** (2): 320–5.
- 53 Nair VS, Gevaert O, Davidzon G *et al.* Prognostic PET 18F-FDG uptake imaging features are associated with major oncogenomic alterations in patients with resected non-small cell lung cancer. *Cancer Res* 2012; **72** (15): 3725–34.
- 54 Yoon HJ, Sohn I, Cho JH *et al.* Decoding tumor phenotypes for ALK, ROS1, and RET fusions in lung adenocarcinoma using a radiomics approach. *Medicine (Baltimore)* 2015; **94** (41): e1753.

- 55 Jeong CJ, Lee HY, Han J *et al.* Role of imaging biomarkers in predicting anaplastic lymphoma kinase-positive lung adenocarcinoma. *Clin Nucl Med* 2015; **40** (1): e34–9.
- 56 Wu J, Cao G, Sun X *et al.* Intratumoral spatial heterogeneity at perfusion MR imaging predicts recurrence-free survival in locally advanced breast cancer treated with neoadjuvant chemotherapy. *Radiology* 2018; **288** (1): 26–35.

## Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website:

**Table S1** Selected radiomic features **Table S2** Selected genomic features