



Original Article



Associating Preoperative MRI Features and Gene Expression Signatures of Early-stage Hepatocellular Carcinoma Patients using Machine Learning

Xiaoming Li^{1#}, Lin Cheng^{1#}, Chuanming Li¹, Xianling Hu¹, Xiaofei Hu¹, Liang Tan^{2,3}, Qing Li⁴, Chen Liu^{1*} and Jian Wang^{1*}

¹Department of Radiology, Southwest Hospital, Third Military Medical University (Army Military Medical University), Chongqing, China; ²Department of Neurosurgery, Southwest Hospital, Third Military Medical University (Army Military Medical University), Chongqing, China; ³Department of Electrical and Computer Engineering, Faculty of Science and Technology, University of Macau, Macau, China; ⁴MR Collaborations, Siemens Healthcare Ltd., Shanghai, China

Received: 14 January 2021 | Revised: 7 May 2021 | Accepted: 18 May 2021 | Published: 21 June 2021

Abstract

Background and Aims: The relationship between quantitative magnetic resonance imaging (MRI) imaging features and gene-expression signatures associated with the recurrence of hepatocellular carcinoma (HCC) is not well studied.

Methods: In this study, we generated multivariable regression models to explore the correlation between the preoperative MRI features and Golgi membrane protein 1 (GOLM1), SET domain containing 7 (SETD7), and Rho family GTPase 1 (RND1) gene expression levels in a cohort study including 92 early-stage HCC patients. A total of 307 imaging features of tumor texture and shape were computed from T2-weighted MRI. The key MRI features were identified by performing a multi-step feature selection procedure including the correlation analysis and the application of RELIEFF algorithm. Afterward, regression models were generated using kernel-based support vector machines with 5-fold cross-validation.

Results: The features computed from higher specificity MRI better described GOLM1 and RND1 gene-expression levels, while imaging features computed from lower specificity MRI data were more descriptive for the SETD7 gene. The GOLM1 regression model generated with three features demonstrated a moderate positive correlation ($p < 0.001$), and the RND1 model developed with five variables was positively associated ($p < 0.001$) with gene expression levels. Moreover, RND1 regression model integrating four features was moderately correlated with expressed RND1 levels ($p < 0.001$).

Keywords: Hepatocellular carcinoma; Gene-expression; MRI; Radiomics feature.

Abbreviations: FOS, first-order statistics; GLCM, gray-level co-occurrence matrix; GLRLM, gray-level run-length matrix; GLSZM, gray-level size-zone matrix; GOLM1, Golgi membrane protein 1; HCC, hepatocellular carcinoma; LBP, local binary patterns; MRI, magnetic resonance imaging; NGTDM, neighborhood gray-tone difference matrix; OS, overall survival; RND1, Rho family GTPase 1; ROI, region of interest; SETD7, SET domain containing 7.

*Contributed equally to this work.

Correspondence to: Chen Liu and Jian Wang, Department of Radiology, Southwest Hospital, Third Military Medical University, Shapingba District, Chongqing 400038, China. ORCID: <https://orcid.org/0000-0001-5149-2496> (CL), <https://orcid.org/0000-0003-1210-0837> (JW). Tel: +86-13108968808 (CL), +86-13883785811 (JW), Fax: +86-23-6546-3026, E-mail: liuchen@aifmri.com (CL), wangjian_811@yahoo.com (JW)

Conclusions: The results demonstrated that MRI radiomics features could help quantify GOLM1, SETD7, and RND1 expression levels noninvasively and predict the recurrence risk for early-stage HCC patients.

Citation of this article: Li X, Cheng L, Li C, Hu X, Hu X, Tan L, et al. Associating Preoperative MRI Features and Gene Expression Signatures of Early-stage Hepatocellular Carcinoma Patients using Machine Learning. J Clin Transl Hepatol 2022;10(1):63–71. doi: 10.14218/JCTH.2021.00023.

Introduction

HCC is among the leading causes of cancer-related death, with an estimated 42,030 new cases and 31,780 mortalities each year in the USA.¹ Although the overall survival (OS) has improved up to 18% in the last four decades, the mortality rate has doubled, and the incidence rate has tripled during the same period.² Hepatic resection and locoregional ablation are curative treatments that improve the OS of HCC patients. However, HCC recurrence, a key contributor to reducing the long-term survival rate, is observed up to 60–70% in the patients after surgical resection.^{3,4}

Genomic profiling of tumors has revealed a complex genetic landscape and characteristics of HCC; moreover, it has identified key genes that play roles in tumorigenesis and tumor growth.^{5–7} The identified genes can be used to discriminate heterogeneous responses to treatments and to predict HCC prognosis.^{8–10} Several genes are already identified as associated with the prognosis of HCC patients, as reported from previous studies in the literature.^{11–13} Komatsu et al.¹¹ linked poor prognosis of HCC patients with reduced Rho family GTPase 1 (RND1) gene expression, which was associated with increased HCC cell proliferation, invasion, and chemoresistance to cisplatin. Golgi membrane protein 1 (GOLM1) was associated with tumor growth and metastasis, while being implicated as a cause of early recurrence and poor survival.¹² Moreover, a previous study revealed the affiliation of decreased level of the SET domain containing 7 (SETD7) protein, which regulates cell cycle, with HCC recurrence and metastasis, and there-

Table 1. Demographic information of the patients regarding GOLM1 gene expression levels

	GOLM1 gene expression levels					
	1	2	3	4	5	6
Age in years	68	49±13.5	50.46±11.52	53±10.4	40±12	75
Gender						
Male	2	14	18	34	2	1
Female	0	1	6	6	1	0
Liver cirrhosis						
Positive	0	5	13	23	2	1
Negative	2	10	11	17	1	0
HBsAg						
Positive	1	12	20	34	2	1
Negative	1	3	4	6	0	0
Blood test						
AFP	2.65±1.32	197.46±318.81	160.63±262.56	121.8±228.2	266.7±461.9	800±0
ALP	90±1.24	97.2±26.60	92.08±41.43	88.8±30.9	89.7±31.8	101±0
GGT	38.5±19.09	40.87±25.04	50.83±45.08	67.7±56.2	42.3±32	32±0
AST	36.6±4.81	144.09±159.05	123.65±133.34	144.9±189.6	141.3±68.6	20±0
ALT	40.6±2.26	150.23±161.73	115.80±141.68	157.2±226.7	148.5±127.9	13±0
AFU	25.85±4.74	31.88±10.10	30.88±11.37	31±11.1	25.1±108	34.4±0
5'-NT	5.35±1.91	7.14±4.20	6.69±4.05	7.7±4.1	6.1±2.1	7.7±0
ADA	9.1±1.70	13.87±6.43	13.69±6.52	12.9±6.2	17.7±6	15.4±0

5'-NT, 5'-Nucleotidase; ADA, Adenosine deaminase; AFP, alpha fetoprotein; ALP, alkaline phosphatase; ALT, alanine aminotransferase; AFU, a-L-fucosidase; AST, aspartate aminotransferase; GGT, gamma-glutamyltransferase; HBsAg, hepatitis B surface antigen.

fore poor prognosis.¹³

Radiomics has great potential to describe the structural characteristics of the tissue by quantitative analysis of non-invasive medical image data using different mathematical perspectives. In the previous studies, radiomics features were utilized to predict disease stages,¹⁴ clinical outcomes,¹⁵ treatment responses,¹⁶ tumor heterogeneities,¹⁷ and early detection of cancers.¹⁸ Recently, several studies utilized radiomics features to investigate the relationship between gene expression signatures and medical imaging data for several cancers, including glioblastoma multiforme,¹⁹ breast cancer,²⁰ non-small cell lung cancer,²¹ and HCC.²²⁻²⁴ Although significant correlations have been observed between imaging data and gene expression signatures, the correlation between quantitative magnetic resonance imaging (MRI) features and genes associated with disease recurrence in early-stage HCC patients still remains the primary focus in cancer research.

In this study, we developed multivariable regression models by interpreting underlying structural characteristics from T2-weighted MRI of tumor tissues and investigated whether radiomics analysis of preoperative MRI data can assist noninvasive quantification of gene expression signatures associated with disease recurrence in early-stage HCC patients.

Methods

Patients

Written consent was obtained from all study participants

using protocols approved by the Institutional Research Sub-panel on Human Studies at Southwest Hospital. The patients were selected according to pre-determined inclusion criteria: I) having Barcelona Clinic liver cancer stage 0 or A; II) having pathologically-verified HCC; III) having undergone preoperative MRI; IV) having undergone tumor surgical resection; and V) having gene expression signatures for resected tumor tissue. A total of 92 patients that visited our institution between October 2014 and September 2016 were included in this retrospective study.

Genomic profiling

The microarrays were generated from formalin-fixed, paraffin-embedded tissue specimens collected from surgically-resected tumors, and immunohistochemistry analysis was performed according to the commercial kits' manufacturers' protocols (GOLM1: A12584, Abclonal Technology, Woburn, MA, USA; SETD7: ab14820, Abcam, Cambridge, UK; RND1: ab206669, Abcam). The gene expression scores were evaluated by an expert pathologist and demographic information was collected from unidentified patient data. Gene expression signatures and other information of the patients are presented in Tables 1-3.

MRI image acquisitions

Patients were imaged in our institution using a 3T MRI scanner (MAGNETOM Trio; Siemens Healthcare, Erlangen, Germany) with a 18-ch body coil. Preoperative T2-weighted

Table 2. Demographic information of the patients regarding SETD7 gene expression levels

	SETD7 gene expression levels		
	0	2	4
Age in years	51.8±12	58.4±10.9	48±0
Gender			
Male	66	6	0
Female	11	1	2
Liver cirrhosis			
Positive	42	2	1
Negative	35	5	1
HBsAg			
Positive	64	5	1
Negative	12	2	1
Blood test			
AFP	171.1±277.1	6.7±5.8	402.3±562.5
ALP	90±32.2	102.7±40	100±19.8
GGT	56.8±49.2	50.1±37.3	29.5±14.9
AST	121.4±142.6	232±313.6	40.2±11.1
ALT	132.5±167.1	228.8±350.3	22.9±10
AFU	30.8±11	29±2.2	27.6±18
5'-NT	7.2±4.1	7±3	5±0.8
ADA	13.4±6.2	13.3±7.1	14.1±7.7

5'-NT, 5'-Nucleotidase; ADA, Adenosine deaminase; AFP, alpha fetoprotein; ALP, alkaline phosphatase; ALT, alanine aminotransferase; AFU, a-L-fucosidase; AST, aspartate aminotransferase; GGT, gamma-glutamyltransferase; HBsAg, hepatitis B surface antigen.

MRI data were acquired with a two-dimensional half Fourier acquisition single shot turbo spin echo sequence ("HASTE") under breath-holding with imaging parameters: TR/TE: 1,000/88 ms, slice thickness: 6 mm; section gap: 7.8 mm; matrix size: 244×320. Two radiologists (Observer 1: C Liu, with 10 years of experience in abdominal imaging; Observer 2: X Li, with 6 years of experience in abdominal imaging) were blinded to other clinical information, and they independently selected the slice with the biggest tumor diameter on T2-weighted images. Interobserver variation was measured with statistics. The tumor diameter was measured 3 times and recorded as mean value by Observer 1.

Feature extraction and selection

Afterward, tumor tissues were outlined on MRI slice to generate a region of interest (ROI) using ITK-SNAP (v.3.6.0), by the two observers. Observer 1 repeated tumor segmentation twice in a week and Observer 2 independently performed the segmentation to evaluate test-retest and interobserver reproducibility. The reproducibility was subject to the intraclass correlation coefficient. An example of T2-weighted MRI is presented in Figure 1 (panel A) with an enlarged HCC tumor shown in panel B. After tumors were outlined, the MRI image was resampled to 1×1 mm using a B-spline interpolation approach to match in-plane image resolution for consistency during tumor characterization with quantitative texture analysis.²⁵ Afterward, MRI signal intensity was standardized by performing a z-score normalization method. The standardized image intensities were discretized using a fixed-bin size approach as pre-

ferred by Image Biomarker Standardization Initiative.²⁶ Four different bin sizes (8, 16, 32, and 64), as the level of image specificity, were evaluated to find the discretization parameters that can perform the best interpretation of the association between tumor heterogeneity and gene expression levels. A representative MRI slice quantized with different levels (image specificity) is presented in Figure 1C-F.

A total of 307 radiomic features regarding intensity, texture, pattern, and shape characteristics of the images was computed from ROIs of the selected slice in T2-weighted MRI using 10 different feature extraction approaches and two transformations (wavelet and gradients) using MATLAB (v.9.1; MathWorks, Natick, MA, USA). The feature extraction methods were: first-order statistics (FOS), with six features; gray-level co-occurrence matrix (GLCM), with nine features; gray-level run-length matrix (GLRLM), with thirteen features; gray-level size-zone matrix (GLSZM), with thirteen features; neighborhood gray-tone difference matrix (NGTDM), with five features; local binary patterns (LBP), with ten features; fractal analysis, with one feature; (8) shape metrics, with nine features; and moments, with forty-three features.^{16,27-29} In addition, six FOS features and one hundred ninety-two features (FOS, GLCM, GLRLM, GLSZM, variance, and power) were computed from oriented gradients histogram and four sub-band wavelet images, respectively. GLCM and GLRLM features were computed through four main directions (0°, 45°, 90°, and 135°) and combined by averaging. Afterward, all the features were normalized to the range of 0 and 1 by utilizing min-max normalization. The complete list of features is presented in Supplementary Table 1.

The trivial features were identified by performing a multi-

Table 3. Demographic information of the patients regarding RND1 gene expression levels

	RND1 gene expression levels					
	0	1	2	3	4	5
Age in years	50.9±10.6	53.9±11.7	54.7±11.4	47±1.4	37.5±13.4	63
Gender						
Male	19	18	19	2	3	1
Female	8	1	3	0	1	0
Liver cirrhosis						
Positive	11	11	15	1	1	1
Negative	16	8	7	1	3	0
HBsAg						
Positive	22	15	16	2	3	1
Negative	5	4	6	0	0	0
Blood test						
AFP	237.9±328.5	94.4±195.1	127.3±238.1	51.9±68	3.74±3.5	9.9
ALP	90.2±32.1	94.6±34	88.3±30.1	62.5±19.1	75.9±44.9	110
GGT	41.5±25.8	85.7±80.9	51.7±35.6	50.5±48.8	54±22.6	64
AST	87.9±129	164.8±218.1	104.9±106.7	407.7±185.8	106.7±73	37
ALT	90.2±141.5	176.8±231.9	105.4±109.5	323.4±78.6	101.9±86.1	41
AFU	28.3±7.9	35±13.8	28.3±7.5	28.6±3.7	37.8±17.3	33
5'-NT	5.7±2.3	8.7±5.3	6.7±4.6	4.7±3	7.4±1.4	8.1
ADA	12.9±5.9	15.1±5.8	13.3±4.9	8.5±8.9	13.9±8.3	13.7

5'-NT, 5'-Nucleotidase; ADA, Adenosine deaminase; AFP, alpha fetoprotein; ALP, alkaline phosphatase; ALT, alanine aminotransferase; AFU, a-L-fucosidase; AST, aspartate aminotransferase; GGT, gamma-glutamyltransferase; HBsAg, hepatitis B surface antigen.

step feature selection process. Initially, features were selected by their correlation with tumor size ($r > 0.75$). Afterward, features were clustered based on their cross-similarity

($r > 0.75$) measured with Pearson correlation coefficient to minimize artificially increased numbers of candidate features. In each cluster, correlations of each feature pair were

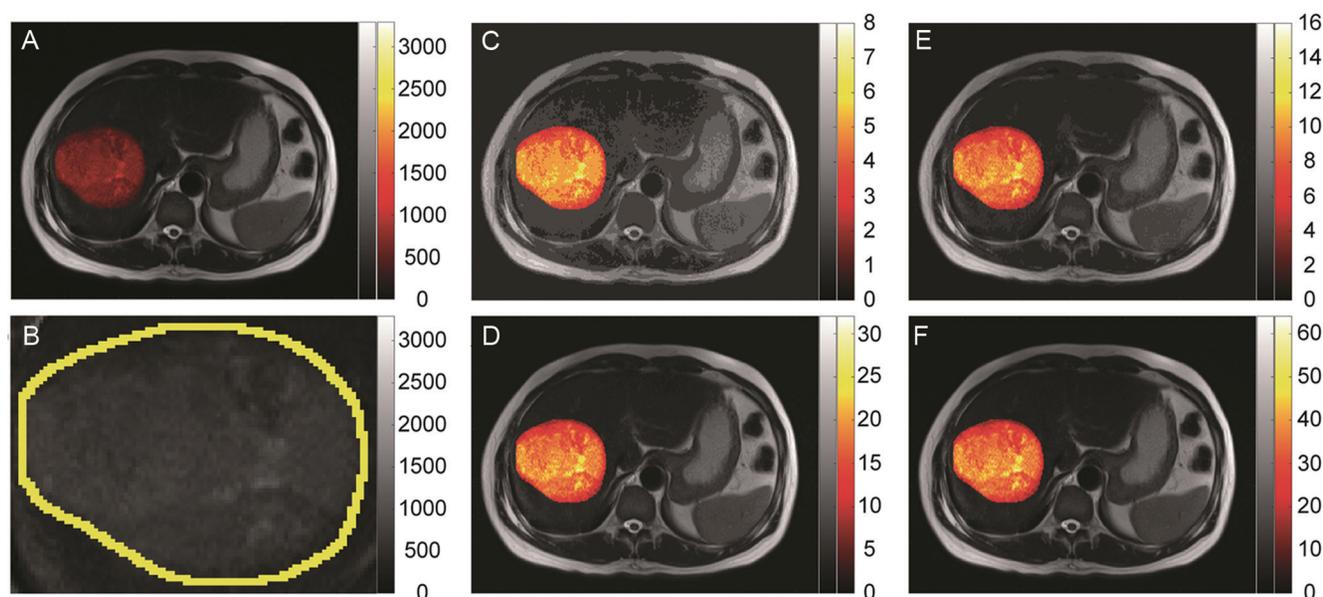


Fig. 1. A representative MRI slice for an early-stage HCC patient. (A) Original MRI slice (arrow). (B) Enlarged tumor. (C–D) Effect of image quantization performed before feature extraction. The same MRI slice is quantized with 8 (C), 16 (D), 32 (E), and 64 (F) levels (arrow). HCC, hepatocellular carcinoma; MRI, magnetic resonance imaging.

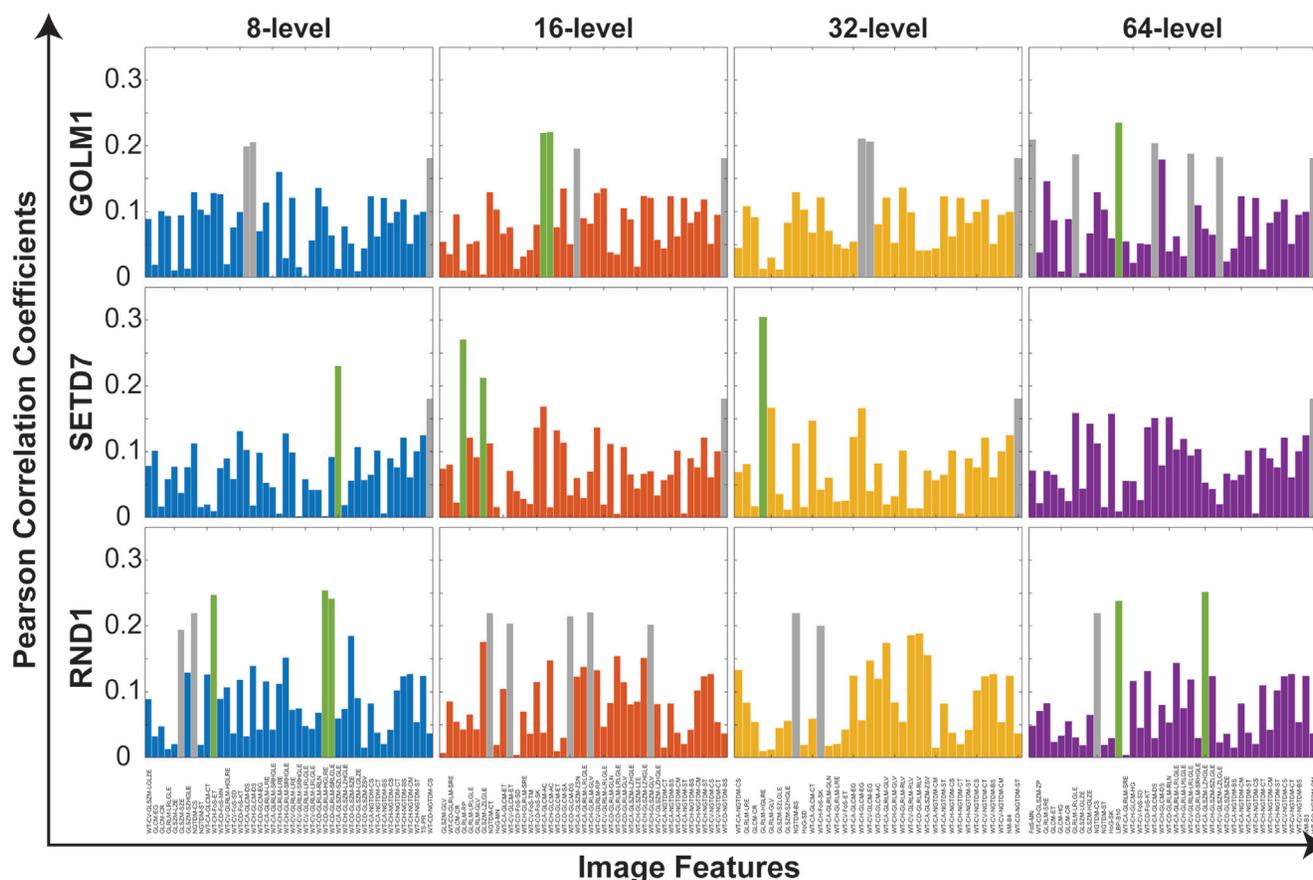


Fig. 2. Correlation of gene expression signatures of GOLF1, SETD7, and RND1 and textural features extracted from MRI data with different quantization levels. The statistically significant features are highlighted in green and features with a *p*-value between 0.05 and 0.10 are shown in gray. Multivariate analysis is required to improve the association between gene expression signatures and MRI image features. GOLF1, Golgi membrane protein 1; MRI, magnetic resonance imaging; RND1, Rho family GTPase 1; SETD7, SET domain containing 7.

computed and the one with the largest average similarity was selected as the representative of the cluster. Later, the interactions of representative features were assessed using the RELIEFF algorithm with 10 nearest neighbors to determine feature importance/rank.³⁰ The top-ranked 20 features were selected as the final candidate-feature set to associate with gene expression signatures. The feature selection procedure was repeated separately for each level of intensity (8, 16, 32, and 64) and key variables were determined separately for each gene.

Statistical analysis

The identified features with the RELIEFF algorithm were further examined in the exhaustive search experiment, while the features were applied to generate regression models using a support vector machines approach with a radial basis function kernel and 5-fold cross-validation. Spearman’s rank correlation coefficient was used to evaluate the performance of the generated regression models. Wilcoxon rank-sum test was utilized to analyze continuous variables. Meanwhile the chi-squared test was performed to analyze dichotomous variables. Moreover, correlation analysis was performed to determine the association between patient clinical characteristics and gene expression signatures. The continuous ($\mu \pm \sigma$) and $p < 0.05$ was ap-

proved as statistically significant.

Results

Patients

A total of 92 patients were incorporated in our retrospective study, which includes 14 female patients (51.36 ± 13.45 years) and 78 male patients (52.17 ± 11.67 years) with early-stage HCC. The average size of the lesions was 3.9 ± 1.5 cm (range: 1.1–6.3cm). The correlation analysis demonstrated that patient characteristics including blood markers were not strongly correlated with any of the gene expression signatures ($|r| < 0.254$). For the GOLF1 gene, cirrhosis had the highest correlation value ($r = 0.234, p = 0.027$) and gender had the strongest association with the SETD7 gene signature ($r = -0.254, p = 0.018$). Moreover, AFP was negatively correlated with the expression level of the RND1 gene ($r = -0.232, p = 0.052$). The agreement between the two observers was excellent ($\kappa = 0.78$) for the chosen slice with the biggest diameter of tumor tissue.

Feature selection

The key features were identified in a multistep feature selec-

tion procedure. First, features with high stability (intraclass correlation coefficient >0.75) from both observers were kept for further analysis. An average of 33 variables (35 to 31 features), strongly associated with tumor size ($r \geq 0.75$), was removed from the feature set. With decreasing number of features correlated with tumor size, the image specificity was improved, except for a bin size of 64 in which 32 features were included. Later, remaining features were grouped into an average of 41 clusters (44 to 35 groups) based on their pairwise similarity measured by Pearson correlation coefficients. However, we did not observe the same behavior with features computed from MRI data quantized with 64 levels (40 groups). The correlations between representative features of each cluster and gene expression signatures were visualized in Figure 2, which shows the need for multivariate models. The correlation between features and gene expression were seen in Supplementary Table 5.

The inter-relationship of representative features is demonstrated in Figure 3A. The representative features were ranked by the RELIEFF algorithm via analysis of gene expression signatures. The top-ranked 20 features were selected for multivariable regression models. The candidate variables computed from MRI with the lowest image specificity for GOLM1, SETD7, and RND1 genes are presented in Figure 3 (panels B–D). Other candidate features are presented in Supplementary Tables 2–4.

Model evaluation

The image specificity to predict gene expression levels were evaluated by measuring the association between the selected features and gene expression signatures using multivariable regression models. In Figure 4, the relationship between the overall behavior of the models and the number of features was presented separately for three genes. For the training data, the correlation between multivariate regression models' response and gene expression scores was improved with an increasing number of features. However, the correlation did not improve similar level for the validation data. r values for training and validation differed with increasing model complexity. Thus, the final regression models were generated with features determined by the overall performance of the circular training and validation experiments. Besides, there was no statistically significant difference between performances of the regression models with different image specificities ($p > 0.3$). The final multivariate regression models were generated with key features and optimal level of image quantization levels (32-level for GOLM1 and RND1, and 8-level for SETD7). Therefore, results for the identified image quantization levels are reported in this study.

Spearman correlation coefficients of GOLM1, SETD7, and RND1 genes improved from 0.38 to 0.79, 0.3 to 0.53, and 0.33 to 0.68 in the training cohort and from 0.37 to 0.56, 0.3 to 0.53, and 0.32 to 0.64 in the validation cohort with an increasing number of variables in the regression model. The results in the r -value for training and validation of the final regression model were seen in Table 4. The r -value of the validation were all greater than 0.5, and the RND model was the best ($r = 0.67$).

Discussion

We investigated the potential value of radiomics analysis to determine the correlation between imaging features from structural MRI data and expression signatures of GOLM1, SETD7, and RND1 genes associated with HCC recurrence by building multivariable regression models using a machine

learning method. The quantitative texture features from T2-weighted MRI showed a positive correlation with gene expression levels in early-stage HCC patients. These results demonstrated that preoperative MRI data can reveal gene expression signatures noninvasively, which could be utilized to predict recurrence risk for HCC patients instead of using the invasive and expensive genomic tests.

Detection of gene expression profiles and characteristics of the malignant tissues has great potential to reveal complex mechanisms of cancer biology and improve the prediction of disease recurrence or other types of outcomes for each patient. The previous genetic studies have demonstrated the potential values of GOLM1, SETD7, and RND1 genes in the evaluation of tumor growth, metastasis, and prognosis for HCC patients.^{11–13} Preoperative biopsy is a gold-standard clinical method to measure gene expression profiles. However, the costly and invasive biopsy cannot represent the whole tumor structure.^{22,31,32} Therefore, the association between noninvasive imaging characteristics and gene expression profiles becomes an active research topic for cancer studies during recent years.^{33,34}

Several studies investigated the relationship between quantitative features of noninvasive imaging data and gene expression levels associated with the prognosis of HCC disease. Taouli *et al.*³⁵ examined the correlation between 13 previously reported HCC gene signatures and characteristic features (11 qualitative and 4 quantitative) of noninvasive imaging (computed tomography or MRI) data. The results from 38 HCC patients showed a correlation between phenotypic imaging parameters and gene signatures of aggressive HCC disease. Hectors *et al.*³⁶ investigated the heterogeneity of HCC tumors utilizing multiparametric MRI data and also evaluated the correlations between quantitative MRI parameters and gene expression signatures. Despite the observed significant correlation between expression levels of GLUL, FGFR4, EPCAM, and KRT19 and advanced MRI method parameters. Reproducibility is one of the main challenges that limits the generalization of the results. Xia *et al.*²³ identified radiomics features extracted from contrast-enhanced CT data for the prognosis of HCC patients. Eight radiomics features selected from imaging data were significantly correlated with genes associated with HCC prognosis in 38 patients. Besides, they also found two features with significant correlation with OS. However, the authors stated the low number of patient cohorts limited the study in reproducibility.

In this study, we explored the value of quantitative MRI features for noninvasive extraction of gene information via correlating the MR features with gene expression levels of GOLM1, SET7, and RND1 associated with HCC recurrence using multivariable regression models. The quantitative T2-weighted MRI radiomics features were extracted and evaluated by performing a multi-step procedure. The developed regression models with the identified important variables showed a significant correlation for SETD7 and RND1 gene expression levels. Meanwhile, the GOLM1 model had a moderate correlation with a p -value closer to the significance level. The obtained results demonstrated the potential of quantitative analysis of conventional MRI data as a noninvasive tool for evaluation of the recurrence of the disease in HCC patients.

There are several limitations in our study. First, single-modality MRI data were analyzed to characterize tumor structure to predict gene expression levels. The correlation of image features with gene expression levels might be improved utilizing multiparametric MRI data. Second, we only analyzed three genes for predicting the risk of HCC recurrence in early-stage HCC patients. Even though these three genes have been applied in clinical practice as a prediction for the risk of HCC recurrence. The radiomics analysis for early prediction of the recurrence risk could benefit from

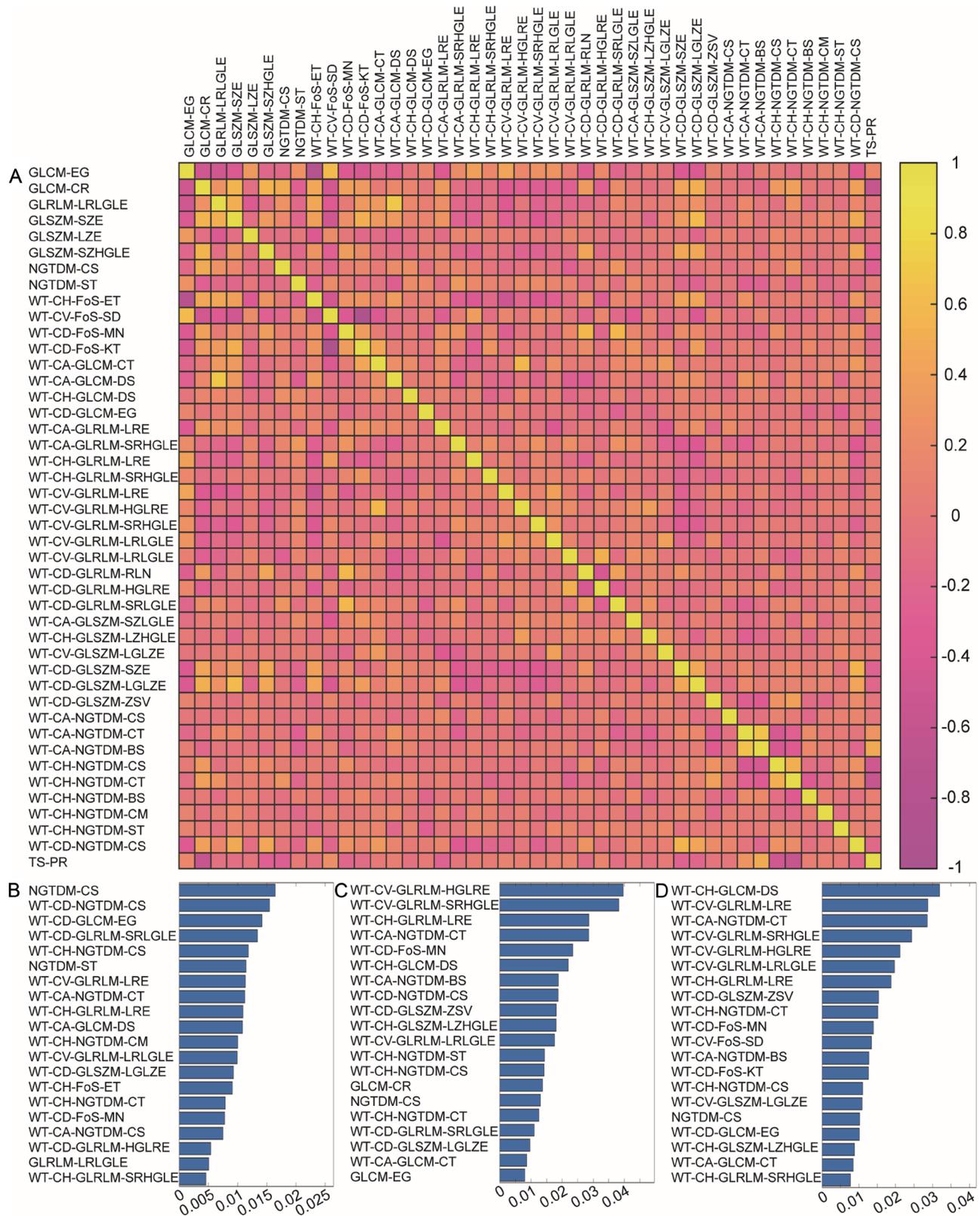


Fig. 3. Evaluation of the features to determine the candidate feature list for multivariate regression models. (A) Pairwise relationship of the representative features computed from 8-level MRI data. (B–D) The top-ranked 20 features of 8-level MRI data that was determined for GOLM1 (B), SETD7 (C), and RND1 (D) genes. GOLM1, Golgi membrane protein 1; MRI, magnetic resonance imaging; RND1, Rho family GTPase 1; SETD7, SET domain containing 7.

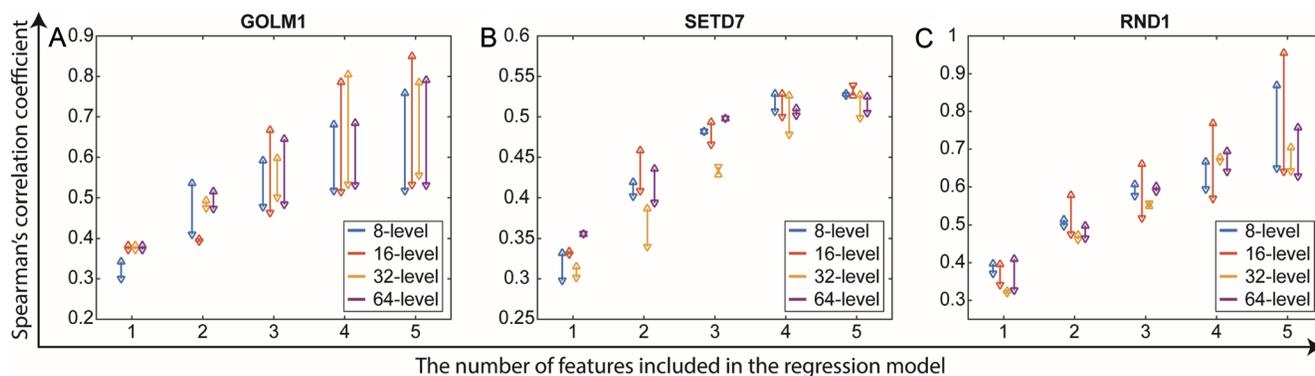


Fig. 4. Behavior of the generated multivariable regression model to predict gene expression signatures. The association between regression model outcome and gene expression signatures is improving for training and validation. Even though the correlation difference between training and validation is increasing, the model is getting more complex. Therefore, the final regression models were generated with key variables selected after analysis of the model performances.

Table 4. The models resulting in an r -value for training and validation cohorts.

Model	Training		Validation	
	Spearman correlation coefficient	p value	Spearman correlation coefficient	p value
GOLM1	0.60	<0.001	0.50	0.06
STED7	0.53	<0.001	0.53	0.037
RND1	0.68	<0.001	0.67	0.046

an improved number of gene types. Third, we did not analyze the co-expression of the genes due to the small set of patient cohorts. Further studies including a larger number of the patient cohort with specialized patient selection procedure will improve interpretation of the imaging data for associating tumor characteristics and gene expression signatures.

In conclusion, the generated multivariable regression models obtained a good correlation with gene expression signatures associated with the recurrence of HCC in early-stage HCC patients. Our results indicate that quantitative analysis of MRI images had the potential to serve as a non-invasive approach for predicting gene expression levels by interpreting the tumor characteristics for early-stage HCC patients.

Funding

This study was supported by the National Key Research and Development Program of China (No. 2016YFC0107101 and No. 2016YFC0107109).

Conflict of interest

The authors have no conflict of interests related to this publication.

Author contributions

Study concept and design (XL, LC), acquisition of the data (CmL, XIH), analysis and interpretation of the data (CL, XfH), drafting of the manuscript (XL), critical revision of the manuscript for important intellectual content (JW), and study supervision (LT). All the authors read and approved the final manuscript.

Data sharing statement

All data generated or analyzed in this study are available from the corresponding author for the reasonable request.

References

- [1] Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. *CA Cancer J Clin* 2019;69(1):7–34. doi:10.3322/caac.21551.
- [2] American Cancer Society. Global Cancer Facts & Figures 2019. Available from: <https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/annual-cancer-facts-and-figures/2019/cancer-facts-and-figures-2019.pdf>.
- [3] Portolani N, Coniglio A, Ghidoni S, Giovanelli M, Benetti A, Tiberio GA, *et al*. Early and late recurrence after liver resection for hepatocellular carcinoma: prognostic and therapeutic implications. *Ann Surg* 2006;243(2):229–235. doi:10.1097/01.sla.0000197706.21803.a1.
- [4] Xu XF, Xing H, Han J, Li ZL, Lau WY, Zhou YH, *et al*. Risk factors, patterns, and outcomes of late recurrence after liver resection for hepatocellular carcinoma: a multicenter study from China. *JAMA Surg* 2019;154(3):209–217. doi:10.1001/jamasurg.2018.4334.
- [5] Woo HG, Park ES, Thorgerirsson SS, Kim YJ. Exploring genomic profiles of hepatocellular carcinoma. *Mol Carcinog* 2011;50(4):235–243. doi:10.1002/mc.20691.
- [6] Hoshida Y, Villanueva A, Kobayashi M, Peix J, Chiang DY, Camargo A, *et al*. Gene expression in fixed tissues and outcome in hepatocellular carcinoma. *N Engl J Med* 2008;359(19):1995–2004. doi:10.1056/NEJMoa0804525.
- [7] Skawran B, Steinemann D, Weigmann A, Flemming P, Becker T, Flik J, *et al*. Gene expression profiling in hepatocellular carcinoma: upregulation of genes in amplified chromosome regions. *Mod Pathol* 2008;21(5):505–516. doi:10.1038/modpathol.3800998.
- [8] Villanueva A, Hoshida Y, Battiston C, Tovar V, Sia D, Alsinet C, *et al*. Combining clinical, pathology, and gene expression data to predict recurrence of hepatocellular carcinoma. *Gastroenterology* 2011;140(5):1501–1512.e2. doi:10.1053/j.gastro.2011.02.006.
- [9] Li N, Li L, Chen Y. The identification of core gene expression signature in hepatocellular carcinoma. *Oxid Med Cell Longev* 2018;2018:3478305. doi:10.1155/2018/3478305.
- [10] Itzel T, Spang R, Maass T, Munker S, Roessler S, Ebert MP, *et al*. Random gene sets in predicting survival of patients with hepatocellular carcinoma. *J Mol Med (Berl)* 2019;97(6):879–888. doi:10.1007/s00109-019-01764-2.
- [11] Komatsu H, Iguchi T, Masuda T, Hirata H, Ueda M, Kidogami S, *et al*. Attenuated RND1 expression confers malignant phenotype and predicts poor prognosis in hepatocellular carcinoma. *Ann Surg Oncol* 2017;24(3):850–859. doi:10.1245/s10434-016-5573-9.

- [12] Ye QH, Zhu WW, Zhang JB, Qin Y, Lu M, Lin GL, *et al*. GOLM1 modulates EGFR/RTK cell-surface recycling to drive hepatocellular carcinoma metastasis. *Cancer Cell* 2016;30(3):444–458. doi:10.1016/j.ccell.2016.07.017.
- [13] Chen Y, Yang S, Hu J, Yu C, He M, Cai Z. Increased expression of SETD7 promotes cell proliferation by regulating cell cycle and indicates poor prognosis in hepatocellular carcinoma. *PLoS One* 2016;11(5):e0154939. doi:10.1371/journal.pone.0154939.
- [14] Li Y, Eresen A, Lu Y, Yang J, Shangguan J, Velichko Y, *et al*. Radiomics signature for the preoperative assessment of stage in advanced colon cancer. *Am J Cancer Res* 2019;9(7):1429–38.
- [15] Banerjee S, Wang DS, Kim HJ, Sirlin CB, Chan MG, Korn RL, *et al*. A computed tomography radiogenomic biomarker predicts microvascular invasion and clinical outcomes in hepatocellular carcinoma. *Hepatology* 2015;62(3):792–800. doi:10.1002/hep.27877.
- [16] Eresen A, Yang J, Shangguan J, Li Y, Hu S, Sun C, *et al*. MRI radiomics for early prediction of response to vaccine therapy in a transgenic mouse model of pancreatic ductal adenocarcinoma. *J Transl Med* 2020;18(1):61. doi:10.1186/s12967-020-02246-7.
- [17] Jeong WK, Jamshidi N, Felker ER, Raman SS, Lu DS. Radiomics and radiogenomics of primary liver cancers. *Clin Mol Hepatol* 2019;25(1):21–29. doi:10.3350/cmh.2018.1007.
- [18] Książek W, Abdar M, Acharya UR, Pławiak P. A novel machine learning approach for early detection of hepatocellular carcinoma patients. *Cognitive Systems Research* 2019;54:116–127. doi:10.1016/j.cogsys.2018.12.001.
- [19] Gevaert O, Mitchell LA, Achrol AS, Xu J, Echegaray S, Steinberg GK, *et al*. Glioblastoma multiforme: exploratory radiogenomic analysis by using quantitative image features. *Radiology* 2014;273(1):168–174. doi:10.1148/radiol.14131731.
- [20] Guo W, Li H, Zhu Y, Lan L, Yang S, Drukker K, *et al*. Prediction of clinical phenotypes in invasive breast carcinomas from the integration of radiomics and genomics data. *J Med Imaging (Bellingham)* 2015;2(4):041007. doi:10.1117/1.JMI.2.4.041007.
- [21] Li S, Han H, Sui D, Hao A, Qin H. A novel radiogenomics framework for genomic and image feature correlation using deep learning. 2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM) 2018: 899–906. doi:10.1109/BIBM.2018.8621432.
- [22] Lo Gullo R, Daimiel I, Morris EA, Pinker K. Combining molecular and imaging metrics in cancer: radiogenomics. *Insights Imaging* 2020;11(1):1. doi:10.1186/s13244-019-0795-6.
- [23] Xia W, Chen Y, Zhang R, Yan Z, Zhou X, Zhang B, *et al*. Radiogenomics of hepatocellular carcinoma: multiregion analysis-based identification of prognostic imaging biomarkers by integrating gene data—a preliminary study. *Phys Med Biol* 2018;63(3):035044. doi:10.1088/1361-6560/aaa609.
- [24] Hectors SJ, Wagner M, Bane O, Besa C, Lewis S, Remark R, *et al*. Quantification of hepatocellular carcinoma heterogeneity with multiparametric magnetic resonance imaging. *Sci Rep* 2017;7(1):2452. doi:10.1038/s41598-017-02706-z.
- [25] Lehmann TM, Gönner C, Spitzer K. Survey: interpolation methods in medical image processing. *IEEE Trans Med Imaging* 1999;18(11):1049–1075. doi:10.1109/42.816070.
- [26] Zwanenburg A, Leger S, Vallières M, Löck S. Image biomarker standardisation initiative. Available from: <https://ui.adsabs.harvard.edu/abs/2016arXiv161207003Z>.
- [27] Imamoglu N, Yagci A, Eresen A, Özbayoglu A. Performance analysis of moments in invariant object classification. In: *Intelligent Engineering Systems through Artificial Neural Networks* 2009. ASME Press, pp. 447–454. doi:10.1115/1.802953.paper56.
- [28] Amadasun M, King R. Textural features corresponding to textural properties. *IEEE Transactions on Systems, Man, and Cybernetics* 1989;19(5):1264–1274. doi:10.1109/21.44046.
- [29] Thibault G, Fertit B, Navarro CL, Pereira S, Cau P, Lévy N, *et al*. Texture indexes and gray level size zone matrix. Application to cell nuclei classification. 10th International Conference on Pattern Recognition and Information Processing, PRIP 2009; Minsk, Belarus.
- [30] Kononenko I, Simec E, Robnik-Šikonja M. Overcoming the Myopia of Inductive Learning Algorithms with RELIEFF. *Applied Intelligence* 1997;7(1):39–55. doi:10.1023/A:1008280620621.
- [31] Tam AL, Lim HJ, Wistuba II, Tamrazi A, Kuo MD, Ziv E, *et al*. Image-guided biopsy in the era of personalized cancer care: proceedings from the society of interventional radiology research consensus panel. *J Vasc Interv Radiol* 2016;27(1):8–19. doi:10.1016/j.jvir.2015.10.019.
- [32] Pritzker KPH, Nieminen HJ. Needle biopsy adequacy in the era of precision medicine and value-based health care. *Arch Pathol Lab Med* 2019;143(11):1399–1415. doi:10.5858/arpa.2018-0463-RA.
- [33] Hectors SJ, Cherny M, Yadav KK, Beksaç AT, Thulasidass H, Lewis S, *et al*. Radiomics features measured with multiparametric magnetic resonance imaging predict prostate cancer aggressiveness. *J Urol* 2019;202(3):498–505. doi:10.1097/JU.0000000000000272.
- [34] Wang T, Gong J, Duan HH, Wang LJ, Ye XD, Nie SD. Correlation between CT based radiomics features and gene expression data in non-small cell lung cancer. *J Xray Sci Technol* 2019;27(5):773–803. doi:10.3233/XST-190526.
- [35] Taouli B, Hoshida Y, Kakite S, Chen X, Tan PS, Sun X, *et al*. Imaging-based surrogate markers of transcriptome subclasses and signatures in hepatocellular carcinoma: preliminary results. *Eur Radiol* 2017;27(11):4472–4481. doi:10.1007/s00330-017-4844-6.
- [36] Hectors SJ, Wagner M, Bane O, Besa C, Lewis S, Remark R, *et al*. Quantification of hepatocellular carcinoma heterogeneity with multiparametric magnetic resonance imaging. *Sci Rep* 2017;7(1):2452. doi:10.1038/s41598-017-02706-z.