



Identifying ^{18}F -FDG PET-metabolic radiomic signature for lung adenocarcinoma prognosis via the leveraging of prognostic transcriptomic module

Jin Li^{1#}, Yixin Liu^{1,2#^}, Wenlei Dong³, Yang Zhou^{1,3}, Jingquan Wu¹, Kuan Luan¹, Lishuang Qi^{4^}

¹College of Intelligent Systems Science and Engineering, Harbin Engineering University, Harbin, China; ²Basic Medicine College, Harbin Medical University, Harbin, China; ³Harbin Medical University Cancer Hospital, Harbin Medical University, Harbin, China; ⁴College of Bioinformatics Science and Technology, Harbin Medical University, Harbin, China

Contributions: (I) Conception and design: J Li, L Qi; (II) Administrative support: J Li, K Luan; (III) Provision of study materials or patients: Y Liu, W Dong; (IV) Collection and assembly of data: W Dong, Y Zhou; (V) Data analysis and interpretation: Y Liu, J Wu; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

[#]These authors contributed equally to this work.

Correspondence to: Lishuang Qi. College of Bioinformatics Science and Technology, Harbin Medical University, Harbin 150086, China. Email: qilishuang7@ems.hrbmu.edu.cn; Kuan Luan. College of Intelligent Systems Science and Engineering, Harbin Engineering University, Harbin 150001, China. Email: luankuan@hrbeu.edu.cn.

Background: Imaging with ^{18}F -fluorodeoxyglucose positron emission tomography (^{18}F -FDG PET), which identifies molecular and metabolic abnormalities within tumor cells, could support prognostic assessment of lung adenocarcinoma (LUAD). We aimed to develop a radiomic signature with the aid of a transcriptomic module for individualized clinical prognostic assessment of LUAD patients.

Methods: Using a gene expression profile consisting of 334 stage I–IIIA LUAD patients, prognostic-related gene coexpression modules were constructed via a weighted correlation network analysis algorithm. The robustness and prognostic performance of the coexpression modules were then tested across 2 gene expression datasets totaling 331 patients. Finally, using a discovery dataset with matched transcriptomic and ^{18}F -FDG PET radiomic data of 15 patients and multiple linear regression analysis, we developed a PET-metabolic radiomic signature that had optimal correlation with the expression of a robust prognostic module.

Results: We selected a superior coexpression module for LUAD prognosis in which the genes were significantly enriched in important biological processes associated with tumors (e.g., cell cycle, DNA replication and p53 signaling pathway). The prognostic performance of the module for overall survival (OS) and recurrence-free survival (RFS) was validated in 2 independent gene expression datasets (log-rank $P < 0.05$). Through the leveraging of this prognostic coexpression module, a radiomic signature consisting of 3 PET features associated with metabolic processes was developed in the discovery dataset. The radiomic signature was significantly associated with patients' OS and RFS in an independent PET dataset consisting of 72 LUAD patients (OS: log-rank $P = 0.0006$; RFS: log-rank $P = 0.0013$). Multivariate Cox analysis demonstrated that the radiomic signature was an independent prognostic factor for OS and RFS. Furthermore, the novel proposed radiomic nomograms for OS and RFS had significantly better performance (concordance indices) than did the clinicopathological nomograms.

Conclusions: The radiomic signature, which reflects biological processes in tumors (e.g., cell cycle and p53 signaling pathway), could noninvasively identify LUAD patients with poor prognosis who should receive

[^] ORCID: Yixin Liu, 0000-0002-5988-722X; Lishuang Qi, 0000-0002-2991-6544.

postoperative adjuvant treatment. The signature is suitable for clinical application and could be robustly applied at an individual level across multicenter cohorts.

Keywords: Lung adenocarcinoma (LUAD); prognosis; coexpression module; radiomic signature; positron emission tomography

Submitted Jul 12, 2021. Accepted for publication Nov 17, 2021.

doi: 10.21037/qims-21-706

View this article at: <https://dx.doi.org/10.21037/qims-21-706>

Introduction

Lung adenocarcinoma (LUAD) has a high mortality rate and is the most common type of lung cancer (1). Presently, surgical resection is the standard treatment for patients with stage I–IIIA LUAD. However, approximately 35–50% of patients who receive this treatment show recurrence or die within 5 years after surgery (2). This highlights the importance of identifying high-risk patients who require postoperative adjuvant treatment, while sparing low-risk patients its toxicity.

Imaging with ¹⁸F-fluorodeoxyglucose positron emission tomography (¹⁸F-FDG PET) can reflect metabolic differences between healthy tissue and tumor tissue and plays an important role in high accuracy staging and prognosis of LUAD (3–7). Conventional metabolic features of PET, such as maximum standardized uptake value (SUV_{max}), have been reported to be valuable prognostic factors for LUAD patients (8,9). However, the prognostic performance of these metabolic features still needs to be improved (10). More sophisticated tools based on existing imaging metrics need to be developed for the purpose of precision medicine.

Radiomics, a high-throughput method that converts medical images into quantitative features, is a newly prominent field that could be the vanguard of precision medicine (11–14). Given that medical images can provide information on the underlying pathophysiology that is associated with a patient's prognosis (15,16), considerable efforts have been made to identify prognostic signatures based on radiomic features in patients with cancer (17–23). At present, most radiomic signatures are constructed based on the tumor phenotypic radiomic features derived from computed tomography (CT) or PET/CT; however, recent studies have demonstrated that the values of these phenotypic features are sensitive to imaging protocols (24) and the platforms used to calculate them (25), which can result in different values among multicenter datasets.

Several studies have adopted data normalization to reduce the impact of multicenter effects (26); however, these signatures tend not to be universal in clinical settings because their application requires precollection of samples for normalization. Moreover, sample risk classification is influenced by the risk composition of other samples adopted for normalization (27). Radiomic signatures based on quantitative phenotypic features are therefore not suitable for clinical translation.

Another approach involves reconstructing radiomic features (e.g., SUV_{max}) derived from PET/CT using CT for attenuation correction and normalizing based on the patient's body weight. These kinds of features, reflecting metabolic abnormalities, may be more robust against multicenter effects (28). For example, the recommended cutoff value for SUV_{max} is typically 2.5 for diagnosis worldwide (29), suggesting the repeatability and reproducibility of metabolic features. By using the metabolic features of PET/CT to develop prognostic radiomic signatures for cancer, it may therefore be possible to circumvent the impact of multicenter effects. However, limited by small-scale samples of public PET image datasets, a single-center retrospective study could introduce selection bias and overfitting risk. Motivated by the phenomenon that metabolic radiomic features of PET can indicate molecular abnormalities (such as messenger RNA [mRNA] or protein dysregulation) within the tumor cells (5), we hypothesized the existence of an association between particular radiomic features and specific clinical outcomes.

In this study, we used the gene expression profile of LUAD patients to identify a robust prognostic coexpression module. We then leveraged this module to extract a radiomic signature from metabolic features of ¹⁸F-FDG PET images to predict the death and recurrent risk of patients with LUAD and validated the signature in an independent PET dataset. Finally, we assessed the incremental value of the radiomic signature relative to clinical factors for estimating patient death and recurrent risk.

Table 1 Baseline clinical characteristics of patients in the analyzed cohorts

Variables	GSE68465	GSE31210	GSE50081	TCGA-LUAD	NR
Sample	334	204	127	15	72
Age					
≤60 years	107	99	19	5	8
>60 years	227	105	108	10	64
Sex					
Female	167	109	62	10	19
Male	167	95	65	5	53
TNM stage					
I	230	162	92	6	57
II	65	42	35	1	10
III	39	0	0	6	5
Grade					
Low	50	–	–	–	24
Medium	157	–	–	–	36
High	123	–	–	–	12
Smoking status					
Non-smoker	33	105	23	–	14
Smoker	217	99	92	–	58
Lymphovascular invasion					
Absent	249	–	–	–	66
Present	83	–	–	–	6
Pleural invasion					
No	–	–	–	–	55
Yes	–	–	–	–	17
Average OS (months)	47.54	60.47	52.44	35.57	42.79
Average RFS (months)	–	55.85	43.30	15.90	38.20

NR, non-small cell lung cancer (NSCLC)–radiogenomics dataset; OS, overall survival; RFS, recurrence-free survival; TCGA-LUAD, The Cancer Genome Atlas–lung adenocarcinoma dataset.

Methods

Study population

In this study, 4 gene expression datasets and 2 ¹⁸F-FDG PET image datasets were sourced from public databases. The details of multicenter data as well as the clinical information of the patients are presented in *Table 1*. The inclusion criteria for the samples were as follows: (I) histologically confirmed LUAD, (II) TNM stage confirmed

as stage I–IIIA, (III) survival information and complete clinical information available, (IV) treated with curative resection alone, and (V) each gene expression profile dataset having more than 100 samples. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). These details and applications of the analyzed datasets are displayed in *Figure S1*.

The 4 public gene expression datasets of LUAD tissues used in this study were downloaded from Gene Expression

Omnibus (GEO; <http://www.ncbi.nlm.nih.gov/geo/>) and the Cancer Genome Atlas (TCGA; <https://portal.gdc.cancer.gov/>). The largest GSE68465 (30) dataset consisting of 334 patients with overall survival (OS) information was used to identify gene coexpression modules. Two independent datasets [GSE31210 (31), GSE50081 (32)] with OS and recurrence-free survival (RFS) information were used to test the robustness and prognostic performance of the coexpression modules. For the microarray datasets generated by Affymetrix platforms, the robust multi-array average algorithm (33) was used for preprocessing of the raw data. For TCGA data derived from Illumina HiSeq 2000 RNA Sequencing Version 2 (Illumina, San Diego, CA, USA), the normalized count values determined by the fragments per kilobase of exon per million fragments mapped (FPKM) method were obtained and log₂-transformed to yield the gene expression values. Probe IDs were mapped to gene IDs according to the corresponding platform files. Ensembl (ENSG) gene IDs or gene symbols were mapped to the Entrez gene IDs. The 11,804 genes commonly measured by the 3 platforms, Affymetrix U133 Plus 2.0, Affymetrix U133A and IlluminaHiSeq_RNASeqV2, were used for analysis.

¹⁸F-FDG PET digital imaging and communications in medicine (DICOM) image data of TCGA-LUAD (n=15) and non-small cell lung cancer (NSCLC)-radiogenomics (NR; n=72) datasets (34) were downloaded from the Cancer Imaging Archive (TCIA; <https://www.cancerimagingarchive.net/>, 2020). Both TCGA-LUAD and NR datasets contained OS and RFS information, and the TCGA-LUAD dataset was used as a discovery dataset to develop a prognostic radiomic signature. The NR dataset was then used as an independent validation dataset.

Acquisition, processing, segmentation, and quantitative feature extraction of ¹⁸F-FDG PET images

¹⁸F-FDG PET image acquisition included routine coverage of base of skull to midhigh with additional spot views where necessary. PET images were reconstructed using CT for attenuation correction with iterative ordered subset expectation maximization reconstruction. The standardized uptake value (SUV) of each patient was normalized based on the patient's body weight. Table S1 summarizes the PET/CT parameters (dose and uptake time) used to obtain PET images.

PET image processing, segmentation, and quantitative feature extraction were performed within 3D Slicer (Version 4.10.2; <https://download.slicer.org/>). First, a DICOM

image was loaded, and the SUV was normalized with the PET DICOM extension based on the patient's body weight from the DICOM image header. Next, the interest regions of the image were delineated and modified by 2 experienced radiologists using the semiautomated segmentation approach described by Beichel *et al.* (35). The final feature set included 5 conventional and 17 new metabolic radiomic features, which were calculated using the PET-IndiC extension (<https://www.slicer.org/wiki/Documentation/Nightly/Extensions/PET-IndiC>). The 17 new features were designed to characterize the fluorodeoxyglucose uptake pattern within the segmented lesion using descriptive statistics. The detailed descriptions and analytical definitions of these features are represented in Table S2.

Weighted gene coexpression network analysis

Statistical analyses in this study were performed using R software 3.5.3 (<http://www.r-project.org/>). Weighted gene coexpression network analysis (WGCNA) (36) was performed to identify distinct coexpression modules of prognostic-related genes in the GSE68465 dataset using the "WGCNA" R package (details in Supplementary Methods). The first principal component of each module was picked as the module eigengene, which represents the expression level of the module. For each module, the median expression value of all samples in the GSE68465 dataset was set as the cutoff point to distinguish the high and low expression groups. The homogeneity of each module was calculated by averaging all pairwise Pearson correlation coefficients of genes within the module (37).

Construction of a PET-metabolic radiomic signature

Using the matched radiomic and transcriptomic data of 15 patients in the discovery dataset (TCGA-LUAD), we first calculated the expression level of the prognostic coexpression module. PET-metabolic radiomic features with values that were significantly associated with the expression of the prognostic coexpression module were identified as candidate metabolic features. Based on the candidate metabolic features, we applied multiple linear regression analysis to obtain an optimal set that was most correlated with the expression of the prognostic coexpression module. The set was selected as the PET-metabolic radiomic signature for LUAD prognosis. The median value of regression scores of the signature in the discovery dataset was used as the cutoff value for dividing patients into high-risk (\geq median) and

low-risk (< median) groups.

Statistical analyses

OS was defined as the time from the date of initial surgical resection to the date of death or last contact (censored). RFS was defined as the time from the date of initial surgical resection to the first occurrence of disease progression or death from any cause. To avoid deviations in the patient follow-up time among the different datasets, patients' OS and RFS were truncated at 60 months. Survival curves were estimated using the Kaplan-Meier method and were statistically compared using the log-rank test (38). Multivariate Cox regression analysis was performed to test the independent prognostic value of the signature after the prognosis-associated clinical factors tested were adjusted by univariate Cox regression analysis. Hazard ratio (HR) and 95% CIs were generated using Cox proportional hazards models, and the concordance index (C-index) (39) was also used to estimate the predictive performance of a patient's survival-influencing factors.

To assess the complementary effect of the PET-metabolic radiomic signature on the clinical model in predicting patient prognosis, a radiomic nomogram was constructed using multivariable linear regression analysis ("rms" R package). Additionally, the performance of the radiomic nomogram was evaluated using the C-index, calibration curve, and decision curve analysis. The net reclassification improvement (NRI) index was then determined to quantify the radiomic signature's incremental prognosis improvement using the "nricens" R package.

The "clusterProfiler" R package was used to conduct the functional enrichment analysis of genes correlated with the radiomic features based on the current Kyoto Encyclopedia of Genes and Genomes (KEGG) database, in which a hypergeometric test was employed. Spearman's rank correlation analysis was used to estimate correlations between 2 factors. Wilcoxon rank sum test was used to examine the difference between different groups.

Herein, P values were adjusted using the Benjamini-Hochberg procedure (40) for multiple testing to control the false discovery rate (FDR). A 2-sided P value <0.05 or FDR <0.05 was considered indicative of statistical significance.

Results

Identification of a prognostic coexpression module for LUAD

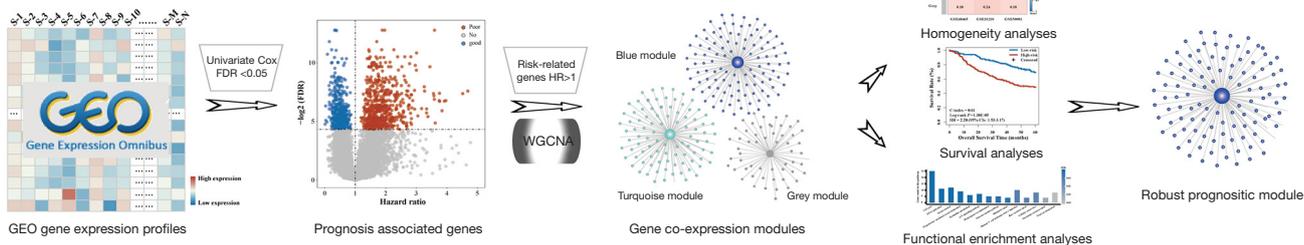
Figure 1 presents a flowchart of this study. First, we

extracted 647 risk-related genes whose overexpression were significantly associated with shorter OS of patients (univariate Cox regression, FDR <0.05 and HR >1; Figure 2A) and then identified 3 coexpression modules of the risk-related genes using WGCNA (blue, turquoise, and gray in Figure 2B,2C). Univariate Cox analysis showed that the expressions of the 3 coexpression modules were significantly associated with a shorter OS in patients (P<0.05 and HR >1; Table 2) in the GSE68465 dataset. Furthermore, for each of the 3 modules, the patients in the high expression group (\geq the median value) had significantly shorter OS than did the patients in the low expression group (< the median value) (Figure 2D-2F).

Next, we tested the homogeneity of the 3 coexpression modules in the 2 validation datasets (GSE31210 and GSE50081). The results (Figure 3A) showed that the blue module, which included 186 genes, had the highest homogeneity scores in the 2 validation sets (GSE31210 score =0.63; GSE50081 score =0.50) when compared with the homogeneity scores in the turquoise module (GSE31210 score =0.21; GSE50081 score =0.40) and the gray module (GSE31210 score =0.24; GSE50081 score =0.18). Furthermore, univariate Cox analysis (Table 2) showed that only the expression of the blue module was significantly associated with OS of patients in the 2 validation datasets (GSE31210: P=0.0006; GSE50081: P=0.0071). Similarly, the expression of the blue module was significantly associated with the RFS of patients in the GSE31210 dataset (P=5.79E-07; HR =1758, 95% CI: 93–32,917) and a similar tendency was observed in the GSE50081 dataset but without statistical significance (P=0.1380; HR =12.76, 95% CI: 0.44–368). Multivariate Cox analysis revealed that the OS was independently predicted by the expression of the blue module after the prognosis-associated clinical factors were adjusted for in the 2 validation cohorts (GSE31210: P=0.0288; GSE50081: P=0.0102; Table S3). The blue module remained significantly associated with the RFS of patients in the multivariate Cox analysis in the GSE31210 dataset (P=3.62E-05; Table S3), and a similar tendency was observed in the GSE50081 dataset but without statistical significance (P=0.2028; Table S3).

Thereafter, according to the cutoff value (-0.0044) of the blue module defined in the GSE68465 dataset, the patients were divided into high- and low-expression groups with significantly different OS in both the GSE31210 dataset (log-rank P=5.17E-05; HR =5.33, 95% CI: 2.15–13.21; C-index =0.69; Figure 3B) and the GSE50081 dataset (log-rank P=0.0010; HR =2.66, 95% CI: 1.45–4.89; C-index

Step I: Identification of a prognostic co-expression module for LUAD prognosis



Step II: Construction and validation of ¹⁸F-FDG PET-metabolize radiomic signature by leveraging prognostic gene co-expression module

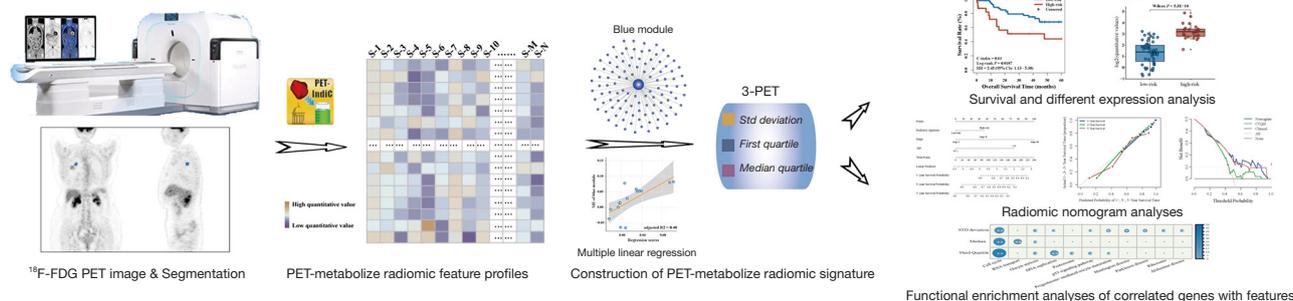


Figure 1 Flowchart showing the development and validation of ¹⁸F-fluorodeoxyglucose positron emission tomography (¹⁸F-FDG PET)-metabolic radiomic signature for stage I-IIIa lung adenocarcinoma (LUAD) prognosis.

=0.62; *Figure 3C*). Similar results were observed for the RFS analysis (*Table S3*); that is, the expression of the blue module was significantly and marginally significantly associated with RFS of patients in the GSE31210 dataset (log-rank $P=2.43E-06$; HR = 3.66, 95% CI: 2.06–6.53; C-index = 0.66; *Figure 3D*) and the GSE50081 dataset (log-rank $P=0.0826$; HR = 1.77, 95% CI: 0.92–3.41; C-index = 0.57; *Figure 3E*), respectively.

These results indicated that the blue module is a robust prognostic factor for LUAD patients. In addition, the functional enrichment analysis showed that the genes of the blue module were significantly enriched in 14 KEGG functional terms (hypergeometric test, FDR < 0.05; *Figure 3F*), including several functions related to LUAD prognosis such as cell cycle, DNA replication, p53 signaling pathway, and “mismatch repair. The results provided biological evidence in support of the ability of the blue module to predict LUAD prognosis.

Construction and validation of ¹⁸F-FDG PET-metabolic radiomic signature via the leveraging of the prognostic gene coexpression module

Using the gene expression profile collected in parallel with PET image data in the discovery dataset (TCGA-LUAD)

with 15 LUAD patients, we first validated the homogeneity score of the blue module and found that it maintained a high level of homogeneity (score = 0.63; *Figure 4A*) and was significantly higher than the 1000 random scores of the bootstrap approach ($P=0.0010$). The result indicated the robustness of transcriptomic characteristics within tumor cells in small-scale samples. Additionally, we estimated the homogeneity score of the turquoise (score = 0.33; *Figure 4A*) and gray (score = 0.19; *Figure 4A*) modules, which were not significantly higher than random ($P>0.05$). We then calculated the expression of the blue module for each sample and extracted 7 candidate metabolic radiomic features whose values were significantly associated with the expression of the blue module (Spearman’s rank correlation, FDR < 0.05). Furthermore, we performed multiple linear regression analysis for the candidate features to generate a PET-metabolic radiomic signature (denoted as 3-PET: third quartile, standard deviation, and median; *Figure S2*) consisting of the 3 features which had the highest correlation with the expression of the blue module (adjusted $R^2 = 0.48$; *Figure 4B*, *Supplementary Methods*). Using the median (0.0082) as the cutoff value, we divided the patients in the TCGA-LUAD discovery dataset into high- and low-risk patient groups, which yielded significant differences in OS (log-rank $P=0.0428$; HR = 4.64, 95% CI: 0.92–23.51;

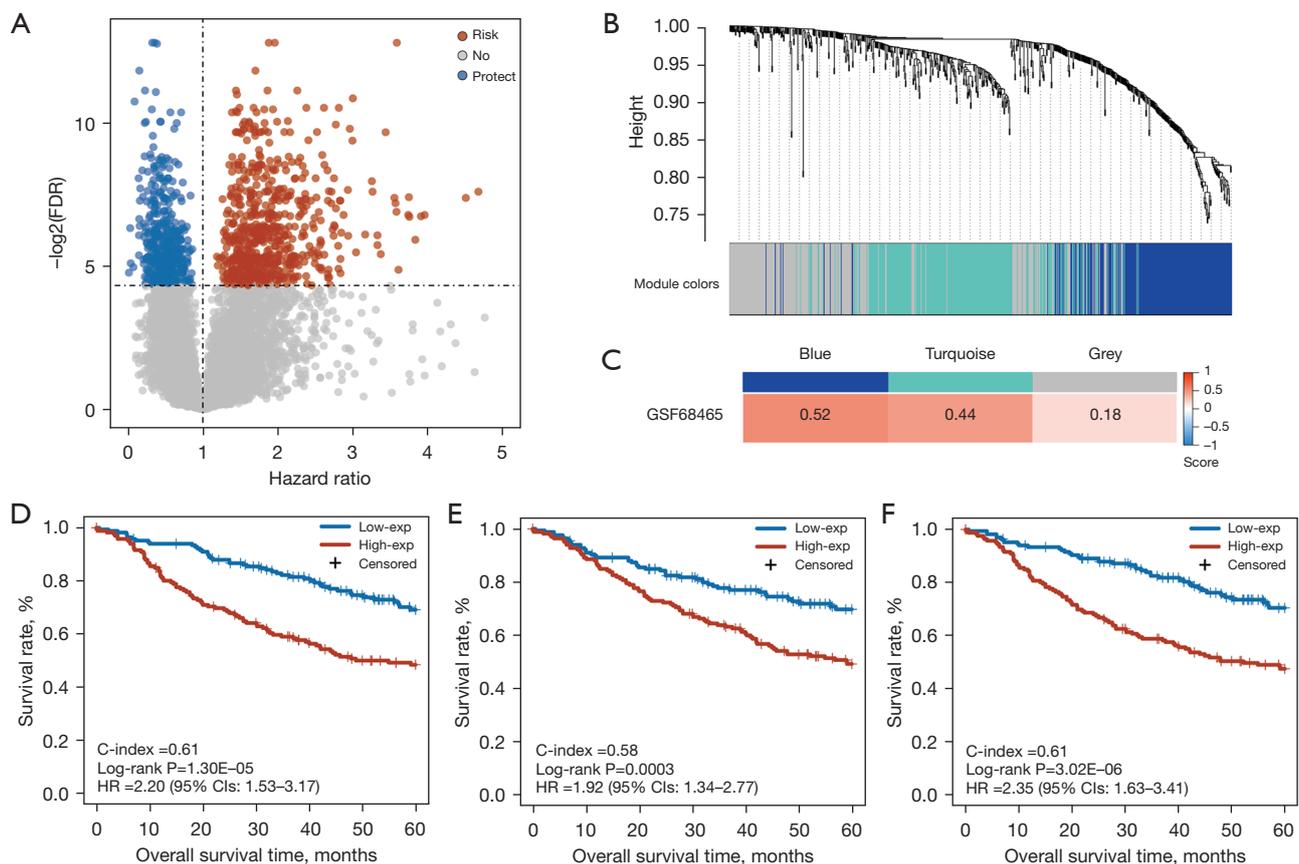


Figure 2 Identification of gene coexpression modules in the GSE68465 dataset. (A) Volcano plot showing the prognosis-associated genes in the GSE68465 dataset. The risk-related genes are highlighted in red [hazard ratio (HR) >1]; the protect-related genes are highlighted in blue (HR <1). (B) Gene coexpression modules identified by weighted gene coexpression network analysis (WGCNA). Three color-coded modules were obtained and are shown as the branches of the clustering dendrogram. Genes that could not be assigned to a module are labeled in gray. (C) The homogeneity scores of 3 coexpression modules in the GSE68465 dataset. (D-F) Kaplan-Meier curves for the 3 modules in the GSE68465 dataset (left to right: blue, turquoise, and gray modules).

C-index =0.73; *Figure 4C*) and RFS (log-rank $P=0.0004$; HR =18.98, 95% CI: 2.26–159.30; C-index =0.80; *Figure 4D*).

Subsequently, the prognostic performance of 3-PET was validated in an independent PET imaging dataset (the NR dataset) consisting of 72 stage I–IIIA LUAD patients treated with curative resection alone. According to the trained cutoff (0.0082) of 3-PET, 19 patients were classified into the high-risk group and had significantly shorter OS (log-rank $P=0.0006$; HR =3.64, 95% CI: 1.66–7.98; C-index =0.64; *Figure 5A*) and RFS (log-rank $P=0.0013$; HR =4.96, 95% CI: 1.69–14.57; C-index =0.66; *Figure 5B*) than did the 53 patients classified into the low-risk group. In the multivariate Cox model, 3-PET remained significantly associated with patients' OS ($P=0.0365$; HR =2.62, 95% CI: 1.06–6.45; *Figure 5C*) and RFS ($P=0.0175$; HR =4.02, 95% CI: 1.28–12.66;

Figure 5D) after the prognosis-associated clinical factors (TNM stage or lymphovascular invasion) were adjusted for estimated in the univariate Cox model (*Table S4*).

Incremental value and biological interpretation of 3-PET

The multivariate Cox analysis revealed that 3-PET and tumor TNM stage could be used to independently predict OS and RFS in the NR dataset. Consequently, 2 radiomic nomograms for OS and RFS that incorporated a clinical risk factor (TNM stage) and 3-PET were generated using the NR dataset (*Figure 6A,6B*). The discrimination performance of the radiomic nomograms exhibited significantly higher C-indices relative to those of the clinical nomograms (*Figure S3A,S3B*) and 3-PET (OS: C-index =0.67; RFS: C-index

Table 2 Univariate Cox analysis of the expression of the coexpression module in the analyzed gene expression datasets

Variables	Overall survival		Recurrence-free survival	
	HR (95% CI)	P value	HR (95% CI)	P value
GSE68465				
Blue	1774.82 (90.00–34,999)	8.75E-07	–	–
Turquoise	6,754.30 (200.9–227,041)	8.79E-07	–	–
Gray	196,687.70 (6,470–5,979,021)	2.61E-12	–	–
GSE31210				
Blue	1,685 (24.82–114,404)	0.0006	1,758 (93–32,917)	5.79E-07
Turquoise	1,721.51 (19.20–154,374)	0.0012	2,304 (99.89–53,133)	1.33e-06
Gray	17,142.30 (178.3–1,648,098)	2.85E-05	10,986 (439.9–274,371)	1.45e-08
GSE50081				
Blue	60.96 (3.06–1213)	0.0071	12.76 (0.44–368)	0.1380
Turquoise	0.61 (0.02–18.38)	0.7760	0.06 (0.01–2.01)	0.1150
Gray	24.51 (0.87–690.9)	0.0604	5.78 (0.14–235.1)	0.3540

HR, hazard ratio.

=0.72; *Table 3*) based on the NRI indices for OS and RFS (*Figure S3C,S3D*), respectively. The calibration curves corresponding to the radiomic nomogram at 1-, 3-, and/or 5-year OS and RFS showed good agreement between the estimations and the clinical outcomes in the NR dataset (*Figure 6C,6D*). Furthermore, the decision-curve analysis showed that the radiomic nomogram exhibited superior performance compared with the clinical nomogram across the majority of the range of reasonable threshold probabilities in the NR dataset (*Figure 6E,6F*).

To clarify the underlying biological processes associated with 3-PET, we further analyzed the gene expression profiles collected along with PET image data corresponding to the TCGA-LUAD dataset. First, we identified 1,833, 1,822, and 1,782 genes significantly correlated with the values of the standard deviation, median, and third quartile features (Spearman's rank correlation, $P < 0.05$), respectively. We found these genes were significantly enriched in several functional pathways related to LUAD progress (hypergeometric test, $FDR < 0.05$; *Figure 7A*), including cell cycle, p53 signaling pathway, and DNA replication. Then, we analyzed the correlation of 3-PET with the conventional PET prognostic index (SUV_{max}) and found that the high-risk patients stratified by 3-PET had significantly higher values of SUV_{max} than did the low-risk patients (Wilcoxon rank sum, $P = 5.30E-10$; *Figure 7B*), which indicated the

good prognostic ability of 3-PET.

Discussion

LUAD is a clinically heterogeneous disease with large variations in clinical outcomes even among patients with the same TNM stage (41). It is therefore necessary to develop a novel signature to improve the prediction of death or recurrence risk in patients with LUAD. Radiomics is an emerging technique that converts conventional medical images into high dimensional features, providing valuable data to develop noninvasive signatures for cancer prognosis (12,42,43). Compared with tumor phenotypic radiomic features derived from CT or PET/CT, which are sensitive to the multicenter effect (28), metabolic features are more repeatable and reproducible across multicenter cohorts and are more suitable for clinical application. In this study we have therefore developed a noninvasive radiomic signature based on PET-metabolic features with the aid of a transcriptomic module, which could be individually applied in a clinical setting to identify patients with LUAD who have a high risk of death and recurrence.

Thus far, the small-scale of ¹⁸F-FDG PET datasets (TCGA, $n = 15$; NR, $n = 78$) has limited the development of a robust PET-metabolic radiomic signature for LUAD prognosis. Additionally, the survival information (OS and

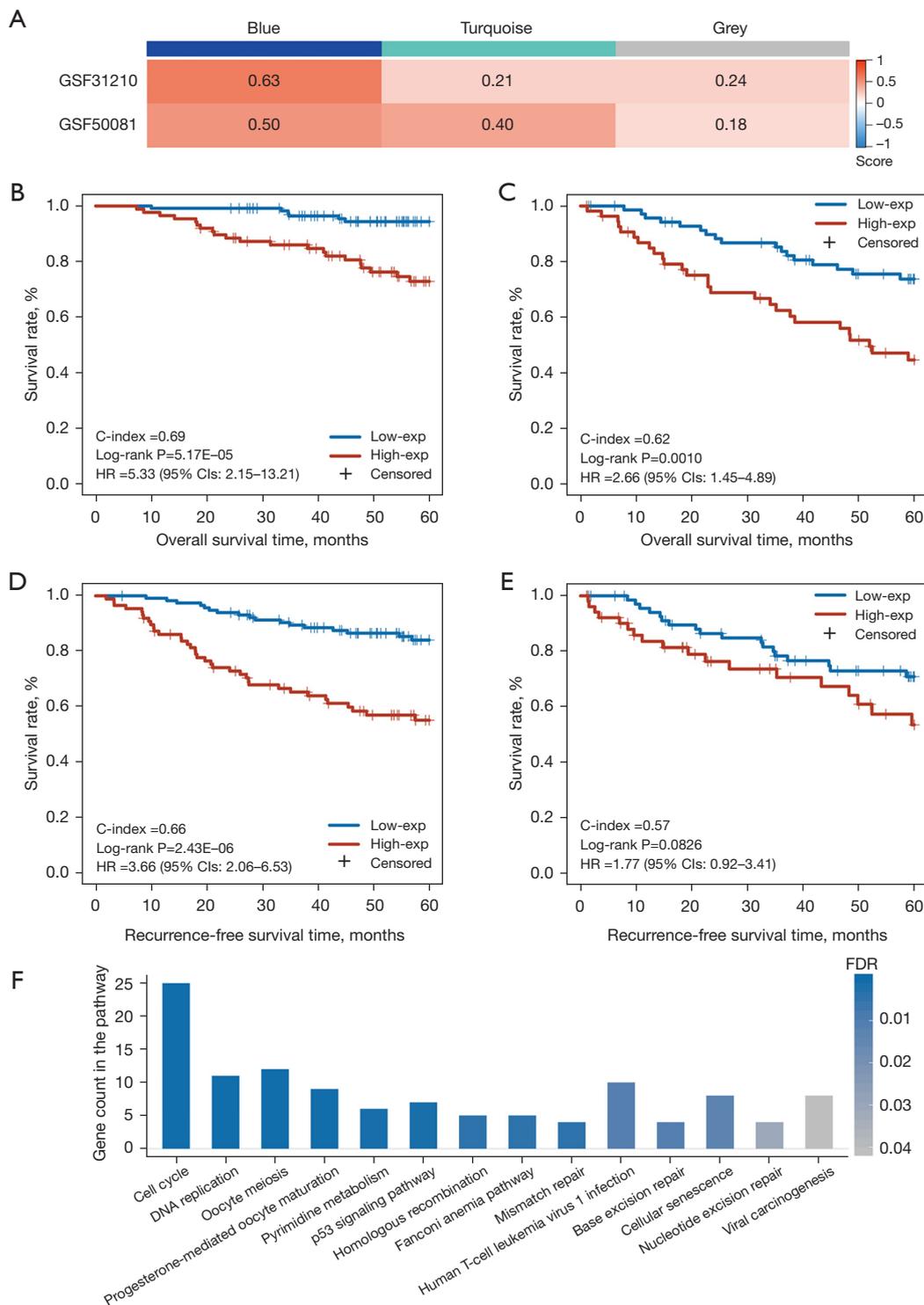


Figure 3 Validation of coexpression module for lung adenocarcinoma (LUAD) prognosis in the GSE31210 and GSE50081 datasets. (A) The homogeneity scores of 3 coexpression modules (top to bottom: GSE31210 and GSE50081). (B,C) Kaplan-Meier curves of overall survival for blue module (left to right: GSE31210 and GSE50081). (D,E) Kaplan-Meier curves of recurrence-free survival for the blue module (left to right: GSE31210 and GSE50081). (F) Kyoto Encyclopedia of Genes and Genomes (KEGG) functional enrichment analyses of the 186 genes in the blue module.

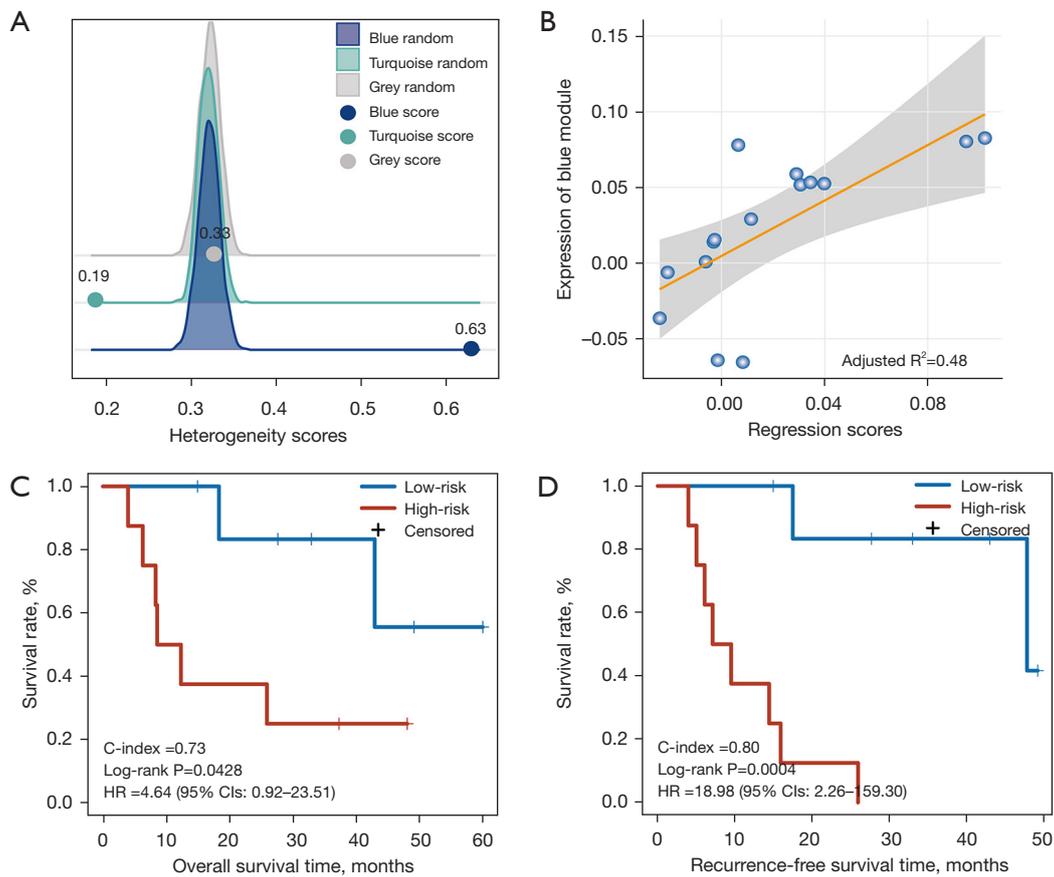


Figure 4 Construction of 3-PET in the The Cancer Genome Atlas–lung adenocarcinoma dataset (TCGA-LUAD) discovery dataset. (A) The homogeneity scores of 3 coexpression modules and their distribution of 1,000 random scores in the TCGA-LUAD gene expression dataset. The dots of blue, turquoise, and gray represent the actual homogeneity scores of the 3 modules. (B) Scatter plots of correlations between the expression of the blue module and regression score of 3-PET. (C) Kaplan-Meier curves of overall survival for the 15 patients in the TCGA-LUAD discovery dataset. (D) Kaplan-Meier curves of recurrence-free survival for the 15 patients in the TCGA-LUAD discovery dataset.

RFS) of patients in TCGA dataset cannot not fully reflect the patients' malignancy degree because they receive different therapeutic strategies after surgery (44,45). Therefore, we hypothesized that using robust molecular information (such as gene expression patterns) related to the clinical outcomes of interest would be an effective method in identifying suitable radiomic features. In this study, we first identified 3 coexpression modules (blue, turquoise, and gray) in a gene expression profile consisting of 334 patients with stage I–IIIA LUAD (GSE68465) and found that the blue module had the highest homogeneity scores in 2 independent expression datasets (GSE31210 and GSE50081). Moreover, the blue module demonstrated better performance in predicting OS and RFS in the 2 independent datasets. The

functional enrichment analyses showed that the genes of the blue module were significantly enriched in pathways related to the prognosis of LUAD, such as cell cycle, DNA replication, and p53 signaling pathway. The results showed that the blue module is a robust prognostic factor for LUAD patients and is a worthwhile intermediary in the development of a noninvasive radiomic signature for clinical application.

To confirm our hypothesis that robust molecular information (gene expression patterns) can be used to identify radiomic features related to specific clinical outcomes, we first validated the homogeneity score of the blue module (i.e., the prognostic coexpression module). We found that the homogeneity score was significantly higher

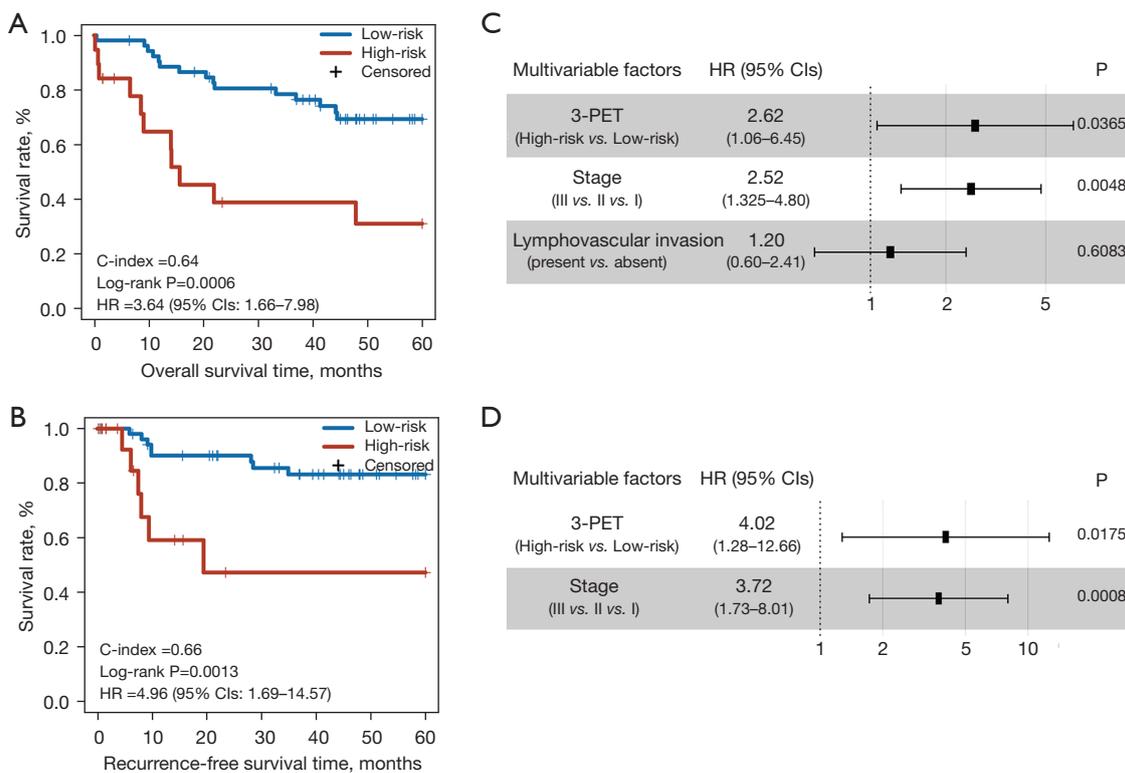


Figure 5 The validation of 3-PET in the non-small cell lung cancer (NSCLC)–radiogenomics (NR) dataset. (A) Kaplan-Meier curves of overall survival for the 72 patients. (B) Kaplan-Meier curves of recurrence-free survival for the 72 patients. (C) Multivariate Cox analyses of overall survival for 3-PET after adjustment for the prognosis-associated clinical factors. (D) Multivariate Cox analyses of recurrence-free survival for 3-PET after adjustment for the prognosis-associated clinical factors.

than random in the TCGA-LUAD dataset with only 15 samples, supporting the robustness of transcriptomic characteristics within tumor cells in small-scale samples. In the TCGA-LUAD dataset, we further identified 7 PET-metabolic radiomic features whose levels were significantly associated with the expression of the blue module (Spearman’s rank correlation, FDR <0.05). On the other hand, there were no features whose levels were significantly associated with patients’ OS or RFS in the TCGA-LUAD dataset (univariate Cox analysis, P>0.05). Thereafter, by leveraging the expression of the blue module, we developed a PET-metabolic radiomic signature (3-PET) for LUAD prognosis in the TCGA-LUAD discovery dataset. 3-PET could be directly applied to an individual patient to determine the patient’s death or recurrence risk based on the trained cutoff value without any cohort normalization. The prognostic performance of 3-PET was effectively validated in an independent dataset (NR dataset) in which both the 5-year survival and recurrence rate of the high-risk patients predicted by 3-PET was significantly lower than

that predicted for the low-risk patients (5-year survival: high-risk vs. low-risk = 0.31 vs. 0.69; 5-year recurrence: high-risk vs. low-risk = 0.47 vs. 0.83). These results indicate good the prognostic performance and robustness of 3-PET in identifying LUAD patients with poor prognosis. Additionally, we compared 3-PET with a conventional metabolic prognostic index (SUV_{max}) and a published PET phenotypic-based signature (Ahn’s model) (42) in the NR dataset. Although 3-PET showed significant association with SUV_{max}, the survival analyses in the NR dataset demonstrated the superior prognostic performance of 3-PET (OS: Cox P=0.0012, C-index =0.64; RFS: Cox P=0.0036, C-index =0.66) as compared to SUV_{max} with grouping by risk cutoff (2.5), with the latter being only potentially associated with OS and RFS of patients (OS: Cox P=0.1500, C-index =0.58; RFS: Cox P=0.0740, C-index =0.63). Similar results were observed for the other risk cutoff value (5.9) for SUV_{max} (OS: Cox P=0.0023, C-index = 0.63; RFS: Cox P=0.1030, C-index =0.60), which has also been used in several studies (46). Next, for Ahn’s model,

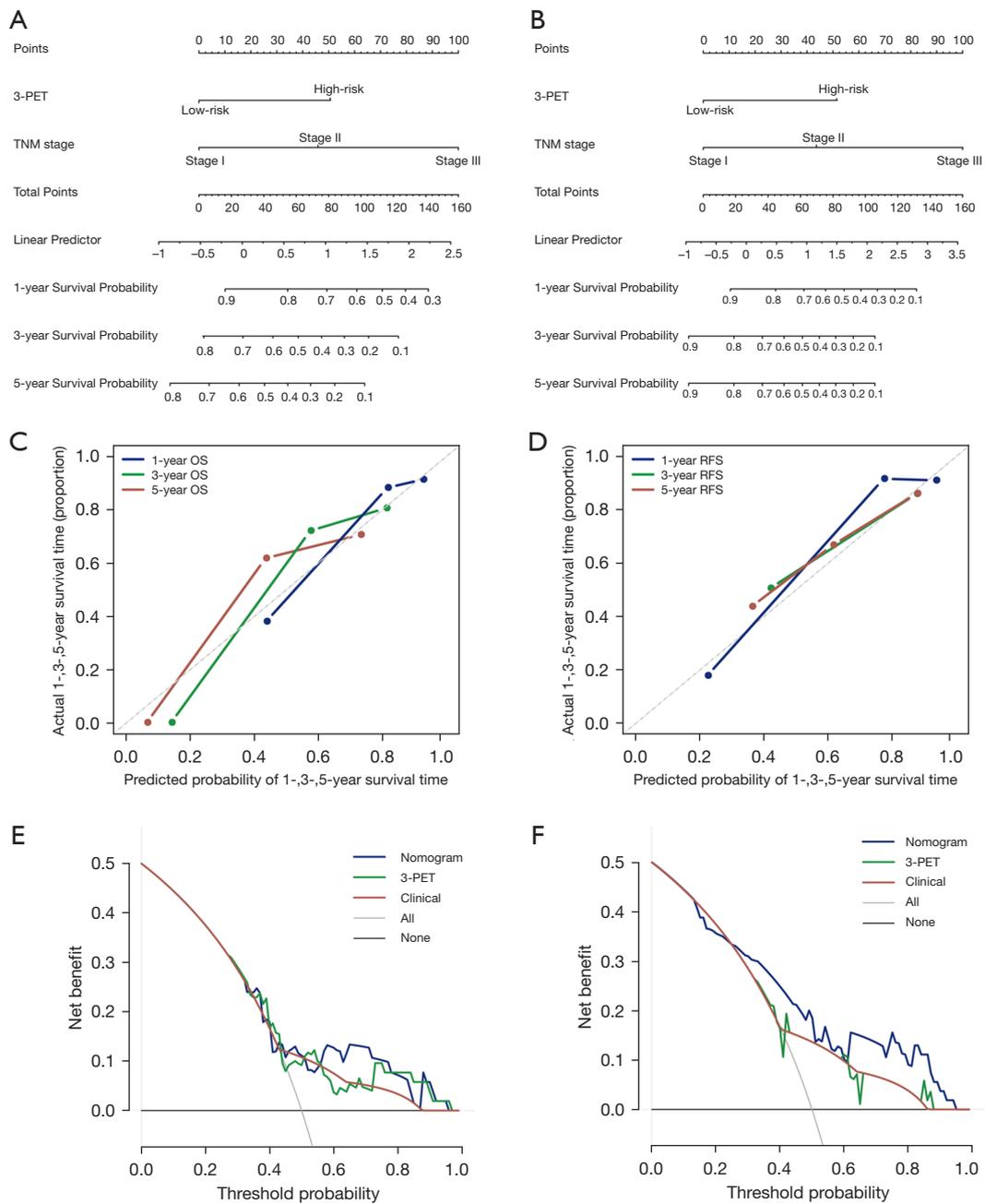


Figure 6 A lung adenocarcinoma (LUAD) survival radiomic nomogram and its performance in the non-small cell lung cancer (NSCLC)-radiogenomics (NR) dataset. (A,B) The LUAD survival radiomic nomogram with 3-PET and a clinicopathological factor (TNM stage) trained in the NR dataset (left to right: radiomic nomogram for overall survival and recurrence-free survival). The points of the radiomic signature and clinical factor were obtained based on the top points bar with a scale of 0–100. Then, the total points were calculated by summing the 2 points, and a line was drawn downward to the survival axis to determine the likelihood of 1-, 3-, or 5-year survival. (C,D) The calibration curves for the radiomic nomogram (left to right: calibration curves for overall survival and recurrence-free survival). The diagonal gray line represents an ideal evaluation. (E,F) The decision curves for the nomogram for the radiomic nomogram (left to right: decision curves for overall survival and recurrence-free survival).

Table 3 Prognostic performances of models for OS and RFS

Parameters	C-index (95% CI)	
	OS	RFS
Radomic nomogram	0.67 (0.57–0.77)	0.72 (0.58–0.87)
3-PET	0.64 (0.55–0.73)	0.66 (0.54–0.79)
Clinical nomogram	0.62 (0.54–0.72)	0.67 (0.54–0.80)

OS, overall survival; RFS, recurrence-free survival.

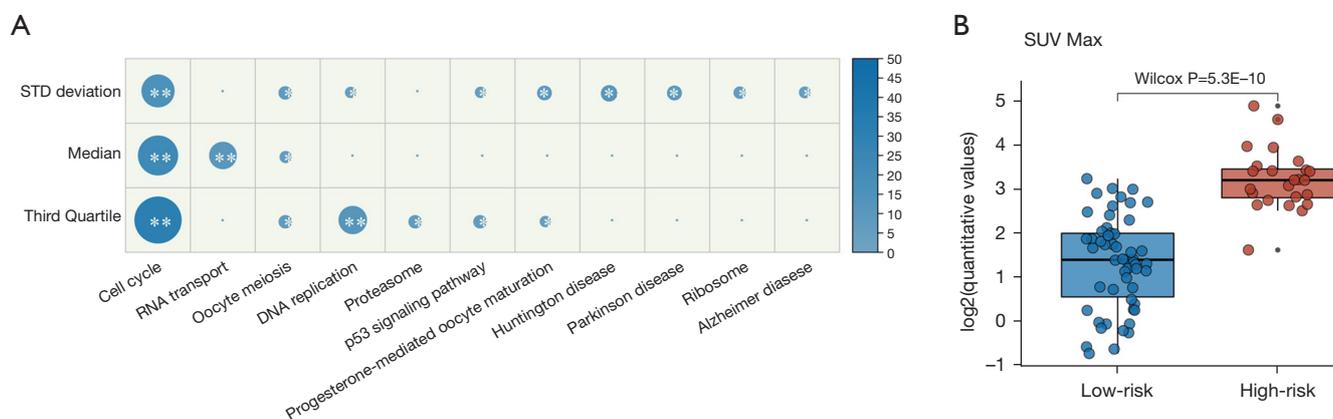


Figure 7 Molecular characteristics associated with 3-PET in lung adenocarcinoma and conventional metabolic features differential analysis. (A) Gene-enrichment analysis of correlated genes with the 3 PET-metabolic radiomic features in 3-PET based on the Kyoto Encyclopedia of Genes and Genomes (KEGG) database in the non-small cell lung cancer (NSCLC)–radiogenomics (NR) dataset. The corresponding significance is shown by asterisks in the circles (*FDR <0.05; **FDR <0.01). (B) The box plots of SUVmax between the 2 risk groups in the NR cohort.

the risk scores of LUAD patients were calculated based on a single PET phenotypic-based feature (neighborhood-gray-level-difference-matrix contrast) in the model. The risk scores were not significantly associated with patients' OS and RFS in the NR dataset (OS: Cox $P=0.1036$, C-index =0.59; RFS: Cox $P=0.3939$, C-index =0.58). More importantly, according to the risk cutoff (0.001) all patients were classified as high risk; this suggests that the risk cutoff of a PET/CT phenotypic-based radiomic signature derived from its discovery cohort cannot be directly applied to independent cohorts due to multicenter effects.

Furthermore, combining the clinical TNM stage with 3-PET in a radiomic nomogram significantly improved the prognostic performance of the clinical staging system in the NR dataset. Our results indicated that 3-PET could provide additional prognostic information for patients within the same clinical stage. Therefore, it is worthwhile to develop 3-PET as a noninvasive individual prognostic

tool for clinical application. Here, we tentatively performed subgroup analyses and showed that 3-PET also could predict the prognosis of patients with T1 stage LUAD (Figure S4). A similar trend was observed in patients with N0 stage LUAD; that is, the high-risk patients had shorter OS and RFS than did the low-risk patients, but not significantly so. This needs further large-scale validation with multicenter clinical trials.

A radiomic signature that is accurately linked to the underlying biological progression of a disease is favorable for clinical application. Therefore, we also identified the genes corresponding to each feature within 3-PET and found that these correlated genes were also significantly enriched in cell cycle, p53 signaling pathway, and DNA replication functional pathways (Table S5). For example, all 3 features showed strongly positive correlations with genes involved in the cell cycle pathway, including cyclin B1 (*CCNB1*), cyclin B2 (*CCNB2*), and cyclin E1 (*CCNE1*), which have

been reported to promote the occurrence and development of LUAD (47,48). These results indicate that 3-PET could reflect the underlying biological characteristics within tumor cells that are strongly associated with clinical outcomes.

Conclusions

In conclusion, the PET-metabolic radiomic signature was able to identify LUAD patients with high risk of death and recurrence who required further adjuvant therapy after surgery, thereby improving prognosis for high risk patients and sparing low risk patients complications associated with adjuvant therapy. This noninvasive approach can provide a more precise diagnosis from imaging studies and may serve as a tool to guide the personalized treatment of patients.

Acknowledgments

The authors thank Professor Zheng Guo (College of Bioinformatics Science and Technology, Harbin Medical University, Harbin, China) for providing significant insights.

Funding: This work was supported by the National Natural Science Foundation of China (61773134, 61803117, and 81872396) and the National Undergraduate Innovation and Entrepreneurship Training Program (202110226018).

Footnote

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at <https://dx.doi.org/10.21037/qims-21-706>). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). No ethical approval or written informed consent was required for this retrospective study.

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the

formal publication through the relevant DOI and the license). See: <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

References

1. Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer Statistics, 2021. *CA Cancer J Clin* 2021;71:7-33.
2. Esposito L, Conti D, Ailavajhala R, Khalil N, Giordano A. Lung Cancer: Are we up to the Challenge? *Curr Genomics* 2010;11:513-8.
3. Chetan MR, Gleeson FV. Radiomics in predicting treatment response in non-small-cell lung cancer: current status, challenges and future perspectives. *Eur Radiol* 2021;31:1049-58.
4. Kim HR, Kim DJ, Lee WW, Jheon S, Sung SW. The significance of maximum standardized uptake values in patients with stage I pulmonary adenocarcinoma. *Eur J Cardiothorac Surg* 2009;35:712-6; discussion 716-7.
5. Na KJ, Choi H. Tumor Metabolic Features Identified by ¹⁸F-FDG PET Correlate with Gene Networks of Immune Cell Microenvironment in Head and Neck Cancer. *J Nucl Med* 2018;59:31-7.
6. Dissaux G, Visvikis D, Da-Ano R, Pradier O, Chajon E, Barillot I, Duvergé L, Masson I, Abgral R, Santiago Ribeiro MJ, Devillers A, Pallardy A, Fleury V, Mahé MA, De Crevoisier R, Hatt M, Schick U. Pretreatment ¹⁸F-FDG PET/CT Radiomics Predict Local Recurrence in Patients Treated with Stereotactic Body Radiotherapy for Early-Stage Non-Small Cell Lung Cancer: A Multicentric Study. *J Nucl Med* 2020;61:814-20.
7. Silva Y, Riedinger JM, Chrétien ML, Caillot D, Corre J, Guillen K, Cochet A, Tabouret-Viaud C, Loffroy R. Comparison between tumour metabolism derived from ¹⁸F-FDG PET/CT and accurate cytogenetic stratification in newly diagnosed multiple myeloma patients. *Quant Imaging Med Surg* 2021;11:4299-309.
8. Inal A, Kucukoner M, Kaplan MA, Urakci Z, Karakus A, Komek H, Dostbil Z, Isikdogan A. Prognostic value of fluorine-18 fluorodeoxyglucose positron emission tomography in patients with advanced non-small cell lung cancer: single center experience. *J BUON* 2012;17:724-8.
9. Mattonen SA, Davidzon GA, Bakr S, Echegaray S, Leung ANC, Vasanawala M, Horng G, Napel S, Nair VS. ¹⁸F FDG Positron Emission Tomography (PET) Tumor and Penumbra Imaging Features Predict Recurrence in Non-Small Cell Lung Cancer. *Tomography* 2019;5:145-53.
10. Jiang Y, Yuan Q, Lv W, Xi S, Huang W, Sun Z, Chen H, Zhao L, Liu W, Hu Y, Lu L, Ma J, Li T, Yu J, Wang

- Q, Li G. Radiomic signature of 18F fluorodeoxyglucose PET/CT for prediction of gastric cancer survival and chemotherapeutic benefits. *Theranostics* 2018;8:5915-28.
11. Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, van Stiphout RG, Granton P, Zegers CM, Gillies R, Boellard R, Dekker A, Aerts HJ. Radiomics: extracting more information from medical images using advanced feature analysis. *Eur J Cancer* 2012;48:441-6.
 12. Aerts HJ, Velazquez ER, Leijenaar RT, Parmar C, Grossmann P, Carvalho S, Bussink J, Monshouwer R, Haibe-Kains B, Rietveld D, Hoebers F, Rietbergen MM, Leemans CR, Dekker A, Quackenbush J, Gillies RJ, Lambin P. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun* 2014;5:4006.
 13. Chen B, Yang L, Zhang R, Luo W, Li W. Radiomics: an overview in lung cancer management-a narrative review. *Ann Transl Med* 2020;8:1191.
 14. Li J, Zhou Y, Wang P, Zhao H, Wang X, Tang N, Luan K. Deep transfer learning based on magnetic resonance imaging can improve the diagnosis of lymph node metastasis in patients with rectal cancer. *Quant Imaging Med Surg* 2021;11:2477-85.
 15. Lambin P, Leijenaar RTH, Deist TM, Peerlings J, de Jong EEC, van Timmeren J, Sanduleanu S, Larue RTHM, Even AJG, Jochems A, van Wijk Y, Woodruff H, van Soest J, Lustberg T, Roelofs E, van Elmpt W, Dekker A, Mottaghy FM, Wildberger JE, Walsh S. Radiomics: the bridge between medical imaging and personalized medicine. *Nat Rev Clin Oncol* 2017;14:749-62.
 16. Liu Z, Wang S, Dong D, Wei J, Fang C, Zhou X, Sun K, Li L, Li B, Wang M, Tian J. The Applications of Radiomics in Precision Diagnosis and Treatment of Oncology: Opportunities and Challenges. *Theranostics* 2019;9:1303-22.
 17. Liu Z, Meng X, Zhang H, Li Z, Liu J, Sun K, Meng Y, Dai W, Xie P, Ding Y, Wang M, Cai G, Tian J. Predicting distant metastasis and chemotherapy benefit in locally advanced rectal cancer. *Nat Commun* 2020;11:4308.
 18. Lu H, Arshad M, Thornton A, Avesani G, Cunnea P, Curry E, Kanavati F, Liang J, Nixon K, Williams ST, Hassan MA, Bowtell DDL, Gabra H, Fotopoulou C, Rockall A, Aboagye EO. A mathematical-descriptor of tumor-mesoscopic-structure from computed-tomography images annotates prognostic- and molecular-phenotypes of epithelial ovarian cancer. *Nat Commun* 2019;10:764.
 19. Trebeschi S, Drago SG, Birkbak NJ, Kurilova I, Călin AM, Delli Pizzi A, Lalezari F, Lambregts DMJ, Rohaan MW, Parmar C, Rozeman EA, Hartemink KJ, Swanton C, Haanen JBAG, Blank CU, Smit EF, Beets-Tan RGH, Aerts HJWL. Predicting response to cancer immunotherapy using noninvasive radiomic biomarkers. *Ann Oncol* 2019;30:998-1004.
 20. Vaidya P, Bera K, Gupta A, Wang X, Corredor G, Fu P, Beig N, Prasanna P, Patil P, Velu P, Rajiah P, Gilkeson R, Feldman M, Choi H, Velcheti V, Madabhushi A. CT derived radiomic score for predicting the added benefit of adjuvant chemotherapy following surgery in Stage I, II resectable Non-Small Cell Lung Cancer: a retrospective multi-cohort study for outcome prediction. *Lancet Digit Health* 2020;2:e116-28.
 21. Xie D, Wang TT, Huang SJ, Deng JJ, Ren YJ, Yang Y, Wu JQ, Zhang L, Fei K, Sun XW, She YL, Chen C. Radiomics nomogram for prediction disease-free survival and adjuvant chemotherapy benefits in patients with resected stage I lung adenocarcinoma. *Transl Lung Cancer Res* 2020;9:1112-23.
 22. Yang B, Ji H, Zhong J, Ma L, Zhong J, Dong H, Zhou C, Duan S, Zhu C, Tian J, Zhang L, Wang F, Zhu H, Lu G. Value of 18F-FDG PET/CT-Based Radiomics Nomogram to Predict Survival Outcomes and Guide Personalized Targeted Therapy in Lung Adenocarcinoma With EGFR Mutations. *Front Oncol* 2020;10:567160.
 23. Du D, Gu J, Chen X, Lv W, Feng Q, Rahmim A, Wu H, Lu L. Integration of PET/CT Radiomics and Semantic Features for Differentiation between Active Pulmonary Tuberculosis and Lung Cancer. *Mol Imaging Biol* 2021;23:287-98.
 24. Midya A, Chakraborty J, Gönen M, Do RKG, Simpson AL. Influence of CT acquisition and reconstruction parameters on radiomic feature reproducibility. *J Med Imaging (Bellingham)* 2018;5:011020.
 25. Fornaçon-Wood I, Mistry H, Ackermann CJ, Blackhall F, McPartlin A, Faivre-Finn C, Price GJ, O'Connor JPB. Reliability and prognostic value of radiomic features are highly dependent on choice of feature extraction platform. *Eur Radiol* 2020;30:6241-50.
 26. Shi Z, Zhovannik I, Traverso A, Dankers FJWM, Deist TM, Kalendralis P, Monshouwer R, Bussink J, Fijten R, Aerts HJWL, Dekker A, Wee L. Distributed radiomics as a signature validation study using the Personal Health Train infrastructure. *Sci Data* 2019;6:218.
 27. Qi L, Chen L, Li Y, Qin Y, Pan R, Zhao W, Gu Y, Wang H, Wang R, Chen X, Guo Z. Critical limitations of prognostic signatures based on risk scores summarized from gene expression levels: a case study for resected stage I non-small-cell lung cancer. *Brief Bioinform* 2016;17:233-42.
 28. Vargas HA, Kramer GM, Scott AM, Weickhardt A, Meier AA, Parada N, et al. Reproducibility and Repeatability

- of Semiquantitative 18F-Fluorodihydrotestosterone Uptake Metrics in Castration-Resistant Prostate Cancer Metastases: A Prospective Multicenter Study. *J Nucl Med* 2018;59:1516-23.
29. Wu Y, Li P, Zhang H, Shi Y, Wu H, Zhang J, Qian Y, Li C, Yang J. Diagnostic value of fluorine 18 fluorodeoxyglucose positron emission tomography/computed tomography for the detection of metastases in non-small-cell lung cancer patients. *Int J Cancer* 2013;132:E37-47.
 30. Shedden K, Taylor JM, Enkemann SA, Tsao MS, Yeatman TJ, et al. Gene expression-based survival prediction in lung adenocarcinoma: a multi-site, blinded validation study. *Nat Med* 2008;14:822-7.
 31. Okayama H, Kohno T, Ishii Y, Shimada Y, Shiraishi K, Iwakawa R, et al. Identification of genes upregulated in ALK-positive and EGFR/KRAS/ALK-negative lung adenocarcinomas. *Cancer Res* 2012;72:100-11.
 32. Der SD, Sykes J, Pintilie M, Zhu CQ, Strumpf D, Liu N, Jurisica I, Shepherd FA, Tsao MS. Validation of a histology-independent prognostic gene signature for early-stage, non-small-cell lung cancer including stage IA patients. *J Thorac Oncol* 2014;9:59-64.
 33. Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, Speed TP. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* 2003;4:249-64.
 34. Bakr S, Gevaert O, Echegaray S, Ayers K, Zhou M, Shafiq M, Zheng H, Benson JA, Zhang W, Leung ANC, Kadoch M, Hoang CD, Shrager J, Quon A, Rubin DL, Plevritis SK, Napel S. A radiogenomic dataset of non-small cell lung cancer. *Sci Data* 2018;5:180202.
 35. Beichel RR, Van Tol M, Ulrich EJ, Bauer C, Chang T, Plichta KA, Smith BJ, Sunderland JJ, Graham MM, Sonka M, Buatti JM. Semiautomated segmentation of head and neck cancers in 18F-FDG PET scans: A just-enough-interaction approach. *Med Phys* 2016;43:2948-64.
 36. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 2008;9:559.
 37. Zhou M, Leung A, Echegaray S, Gentles A, Shrager JB, Jensen KC, Berry GJ, Plevritis SK, Rubin DL, Napel S, Gevaert O. Non-Small Cell Lung Cancer Radiogenomics Map Identifies Relationships between Molecular and Imaging Phenotypes with Prognostic Implications. *Radiology* 2018;286:307-15.
 38. Bland JM, Altman DG. The logrank test. *BMJ* 2004;328:1073.
 39. Harrell FE Jr, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med* 1996;15:361-87.
 40. Dong D, Fang MJ, Tang L, Shan XH, Gao JB, Giganti F, Wang RP, Chen X, Wang XX, Palumbo D, Fu J, Li WC, Li J, Zhong LZ, De Cobelli F, Ji JF, Liu ZY, Tian J. Deep learning radiomic nomogram can predict the number of lymph node metastasis in locally advanced gastric cancer: an international multicenter study. *Ann Oncol* 2020;31:912-20.
 41. Devarakonda S, Morgensztern D, Govindan R. Genomic alterations in lung adenocarcinoma. *Lancet Oncol* 2015;16:e342-51.
 42. Ahn HK, Lee H, Kim SG, Hyun SH. Pre-treatment 18F-FDG PET-based radiomics predict survival in resected non-small cell lung cancer. *Clin Radiol* 2019;74:467-73.
 43. Wang L, Li T, Hong J, Zhang M, Ouyang M, Zheng X, Tang K. 18F-FDG PET-based radiomics model for predicting occult lymph node metastasis in clinical N0 solid lung adenocarcinoma. *Quant Imaging Med Surg* 2021;11:215-25.
 44. Uramoto H, Tanaka F. Recurrence after surgery in patients with NSCLC. *Transl Lung Cancer Res* 2014;3:242-9.
 45. Zhang Z, Zhang S, Li X, Zhao Z, Chen C, Zhang J, Li M, Wei Z, Jiang W, Pan B, Li Y, Liu Y, Cao Y, Zhao W, Gu Y, Yu Y, Meng Q, Qi L. Reference genome and annotation updates lead to contradictory prognostic predictions in gene expression signatures: a case study of resected stage I lung adenocarcinoma. *Brief Bioinform* 2021;22:bbaa081.
 46. Liu J, Dong M, Sun X, Li W, Xing L, Yu J. Prognostic Value of 18F-FDG PET/CT in Surgical Non-Small Cell Lung Cancer: A Meta-Analysis. *PLoS One* 2016;11:e0146195.
 47. Wang X, Xiao H, Wu D, Zhang D, Zhang Z. miR-335-5p Regulates Cell Cycle and Metastasis in Lung Adenocarcinoma by Targeting CCNB2. *Onco Targets Ther* 2020;13:6255-63.
 48. Wang F, Chen X, Yu X, Lin Q. Degradation of CCNB1 mediated by APC11 through UBA52 ubiquitination promotes cell cycle progression and proliferation of non-small cell lung cancer cells. *Am J Transl Res* 2019;11:7166-85.

Cite this article as: Li J, Liu Y, Dong W, Zhou Y, Wu J, Luan K, Qi L. Identifying ¹⁸F-FDG PET-metabolic radiomic signature for lung adenocarcinoma prognosis via the leveraging of prognostic transcriptomic module. *Quant Imaging Med Surg* 2022;12(3):1893-1908. doi: 10.21037/qims-21-706

Weighted gene coexpression network analysis

The “pickSoftThreshold” function was used to determine the proper soft-thresholding power (β) that fit the criteria of the approximate scale-free topology of the network. An adjacency matrix was then built with a soft-thresholding power of 6 in this study. The correlation matrix was transformed into an adjacency matrix (matrix of connection strength) using the power function, and pairwise topologic overlap between genes was calculated. Hierarchical clustering was constructed to identify distinct modules using the “blockwiseModules” function with the parameters minModuleSize and mergeCutHeight set to 50 and 0.25, respectively, in the weighted gene coexpression network analysis. Genes that were not assigned to any module were labeled in gray color.

Calculation formula for the radiomic signature

Risk score = $0.12310108 \times (\text{third quartile}) - 0.09250096 \times (\text{SD}) - 0.11141270 \times (\text{median}) - 0.03370519$.

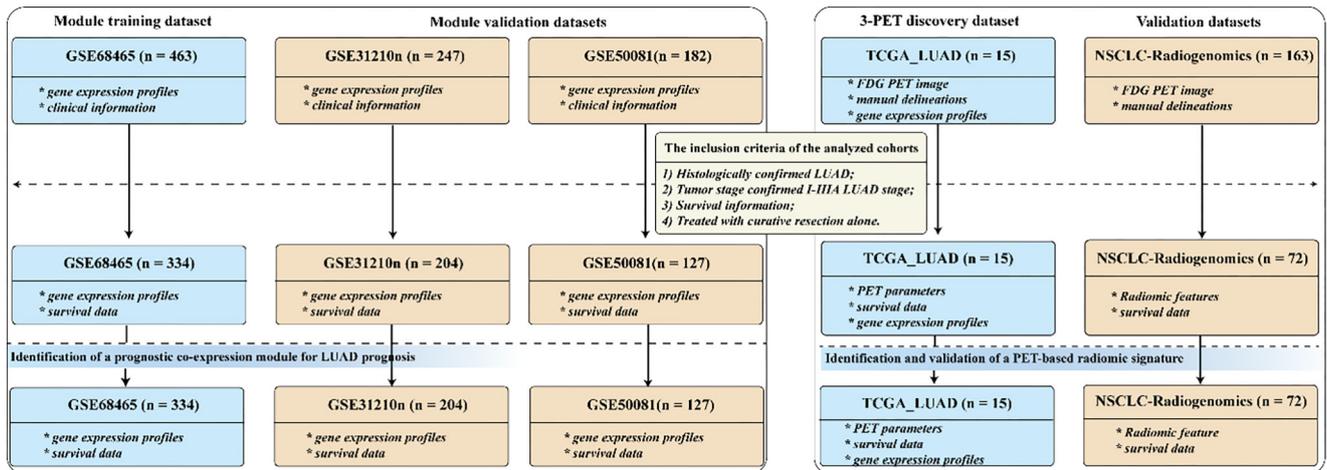


Figure S1 The inclusion criteria and flowchart of the analyzed datasets. FDG, fluorodeoxyglucose; LUAD, lung adenocarcinoma; NSCLC, non-small cell lung cancer; PET; positron emission tomography; TCGA, The Cancer Genome Atlas.

Table S1 Summary of key PET/CT parameters in the analyzed dataset

Parameter	Value
FDG dose (MBq)	138.90–572.25
FDG uptake time (min)	23.08–128.90

FDG, fluorodeoxyglucose; PET, positron emission tomography; CT, computed tomography.

Table S2 Overview of the image-derived features that were used

Feature name	Description	Feature type
SUV _{mean}	SUV mean of ROI	C
SUV _{min}	SUV minimum of ROI	N
SUV _{max}	SUV maximum of ROI	C
SUV _{peak}	Maximum average gray value that is calculated from a 1 cm ³ sphere placed within the ROI	C
MTV	Metabolic tumor volume	C
TLG	Total lesion glycolysis	C
SD	Standard deviation of ROI	N
First quartile	25th percentile value in ROI	N
Median	50th percentile value in ROI	N
Third quartile	75th percentile value in ROI	N
Upper adjacent	First value in ROI not greater than 1.5 times the interquartile range	N
RMS	Root mean square value in ROI	N
Glycolysis Q1	Lesion glycolysis calculated from the 1st quarter of the grayscale range within the ROI	N
Glycolysis Q2	Lesion glycolysis calculated from the 2nd quarter of the grayscale range within the ROI	N
Glycolysis Q3	Lesion glycolysis calculated from the 3rd quarter of the grayscale range within the ROI	N
Glycolysis Q4	Lesion glycolysis calculated from the 4th quarter of the grayscale range within the ROI	N
Q1 distribution	Percent of gray values that fall within the 1st quarter of the grayscale range within the ROI	N
Q2 distribution	Percent of gray values that fall within the 2nd quarter of the grayscale range within the ROI	N
Q3 distribution	Percent of gray values that fall within the 3rd quarter of the grayscale range within the ROI	N
Q4 distribution	Percent of gray values that fall within the 4th quarter of the grayscale range within the ROI	N
SAM	Standardized added metabolic activity	N
SAM background	Local background estimator near ROI	N

ROI, region of interest; C, conventional; N, new.

Table S3 Univariate and multivariate Cox regression analysis in the analyzed gene expression datasets

Variables	Univariate		Multivariate	
	HR (95% CI)	P value	HR (95% CI)	P value
GSE31210				
Overall survival analysis				
Expression of blue module	1,685 (24–114,404)	0.0006	233 (1.76–31,077)	0.0288
Stage (III vs. II vs. I)	4.07 (1.90–8.70)	0.0003	3.00 (1.37–6.55)	0.0058
Age (>60 vs. ≤60 years)	1.24 (0.58–2.65)	0.5770	–	–
Sex (male vs. female)	2.24 (1.02–4.88)	0.0438	1.28 (0.44, 3.72)	0.6462
Smoking status (smoker vs. non-smoker)	2.52 (1.13–5.62)	0.0238	1.48 (0.49, 4.50)	0.4866
GSE31210				
Recurrence-free survival analysis				
Expression of blue module	1,758 (93–32,917)	50.79E-07	845 (34.53–20,725)	30.62e-05
Stage (III vs. II vs. I)	3.55 (2.03–6.21)	80.48E-06	2.73 (1.55–4.82)	0.0005
Age (>60 vs. ≤60 years)	1.53 (0.88–2.65)	0.1320		
Sex (male vs. female)	1.46 (0.85–2.50)	0.1730		
Smoking status (smoker vs. non-smoker)	1.48 (0.86–2.55)	0.1530		
GSE50081				
Overall survival analysis				
Expression of blue module	60.96 (3.06–1213)	0.0071	56.42 (2.61–1,221)	0.0102
Stage (III vs. II vs. I)	2.19 (1.19–4.04)	0.0114	2.09 (1.14–3.85)	0.0174
Age (>60 vs. ≤60 years)	2.03 (0.73–5.68)	0.1780		
Sex (male vs. female)	1.53 (0.83–2.81)	0.1690		
Smoking status (smoker vs. non-smoker)	2.49 (0.89–7.03)	0.0832		
GSE50081				
Recurrence-free survival analysis				
Expression of blue module	12.76 (0.44–368)	0.1380	9.68 (0.29–318.54)	0.2028
Stage (III vs. II vs. I)	2.57 (1.32–5.01)	0.0054	2.48 (1.27–4.83)	0.0077
Age (>60 vs. ≤60 years)	1.01 (0.42–2.43)	0.9810		
Sex (male vs. female)	1.27 (0.66–2.46)	0.4730		
Smoking status (smoker vs. non-smoker)	0.97 (0.44–2.13)	0.9300		

HR, hazard ratio.

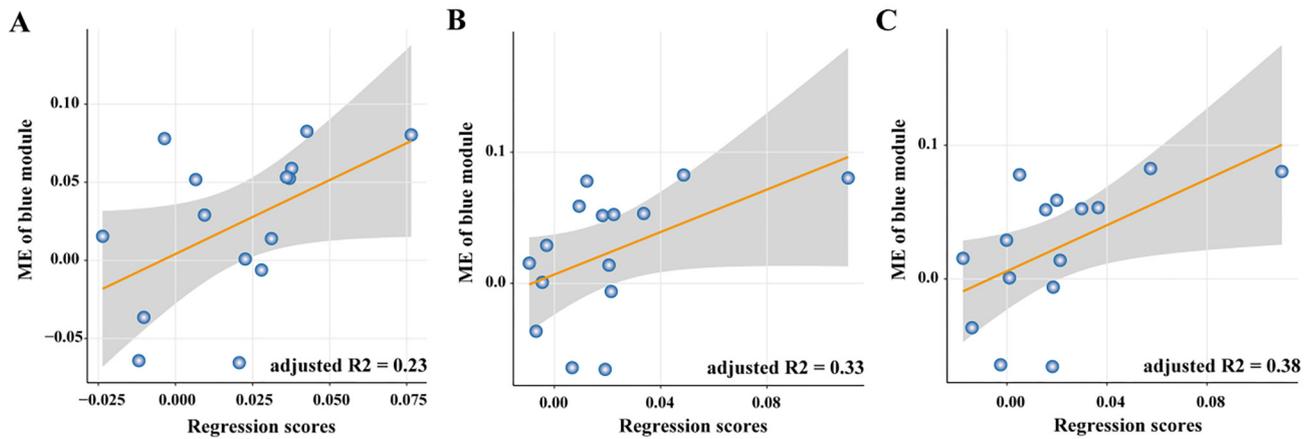


Figure S2 Scatter plots of correlations between the expression of blue module and PET parameters in the 3-PET. (A) Standard deviation. (B) Median. (C) Third quartile. ME, module eigengene; PET, positron emission tomography.

Table S4 Univariate Cox regression analysis in the NR validation datasets

Variables	Overall survival		Recurrence-free survival	
	HR (95% CI)	P value	HR (95% CI)	P value
3-PET (high risk vs. low risk)	3.64 (1.66–7.98)	0.0012	4.96 (1.69–14.57)	0.0036
TNM stage (III vs. II vs. I)	3.25 (1.81–5.86)	80.6e-05	4.19 (2.03–8.64)	0.0001
Age (>60 vs. ≤60 years)	1.56 (0.37–6.60)	0.5460	1.80 (0.23–13.75)	0.5720
Sex (male vs. female)	2.04 (0.70–5.92)	0.1900	0.94 (0.29–2.99)	0.9140
Grade (high vs. medium vs. low)	1.38 (0.77–2.47)	0.2840	1.80 (0.82–3.95)	0.1420
Smoking status (smoker vs. non-smoker)	1.32 (0.45–3.84)	0.6090	0.91 (0.25–3.27)	0.8870
Lymphovascular invasion (present vs. absent)	1.86 (1.01–3.43)	0.0470	1.86 (0.72–4.78)	0.1990
Pleural invasion (yes vs. no)	2.10 (0.90–4.88)	0.0852	1.91 (0.60–6.08)	0.2760
SUV _{max} (>2.5 vs. ≤2.5)	1.96 (0.78–4.88)	0.1500	3.92 (0.88–17.57)	0.0740

NR, non-small cell lung cancer (NSCLC)–radiogenomics; HR, hazard ratio.

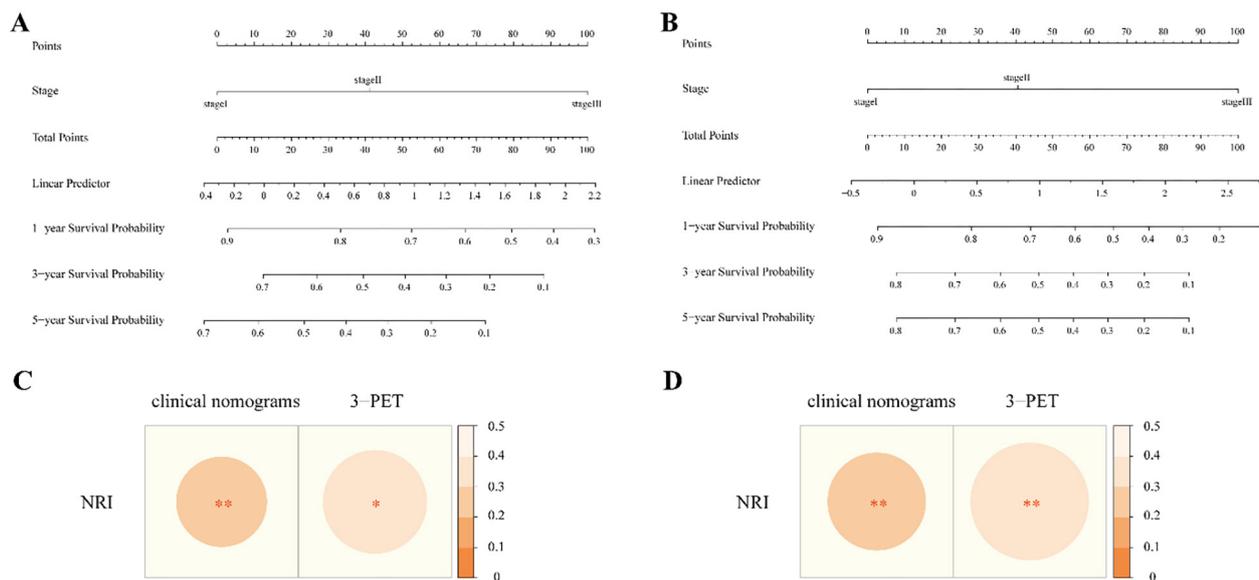


Figure S3 The clinical nomogram and performance improvement evaluation of the radiomic nomogram. (A) The clinical nomogram for overall survival; (B) the clinical nomogram for recurrence-free survival; (C) performance improvement of the radiomic nomogram compared with the clinical nomogram model and 3-PET estimated for overall survival by net reclassification improvement (NRI) methods; (D) performance improvement of the radiomic nomogram compared with clinical nomogram model and 3-PET estimated for recurrence-free survival by net reclassification improvement (NRI) methods. The corresponding significance is shown by asterisks (* $P < 0.05$ and ** $P < 0.01$).

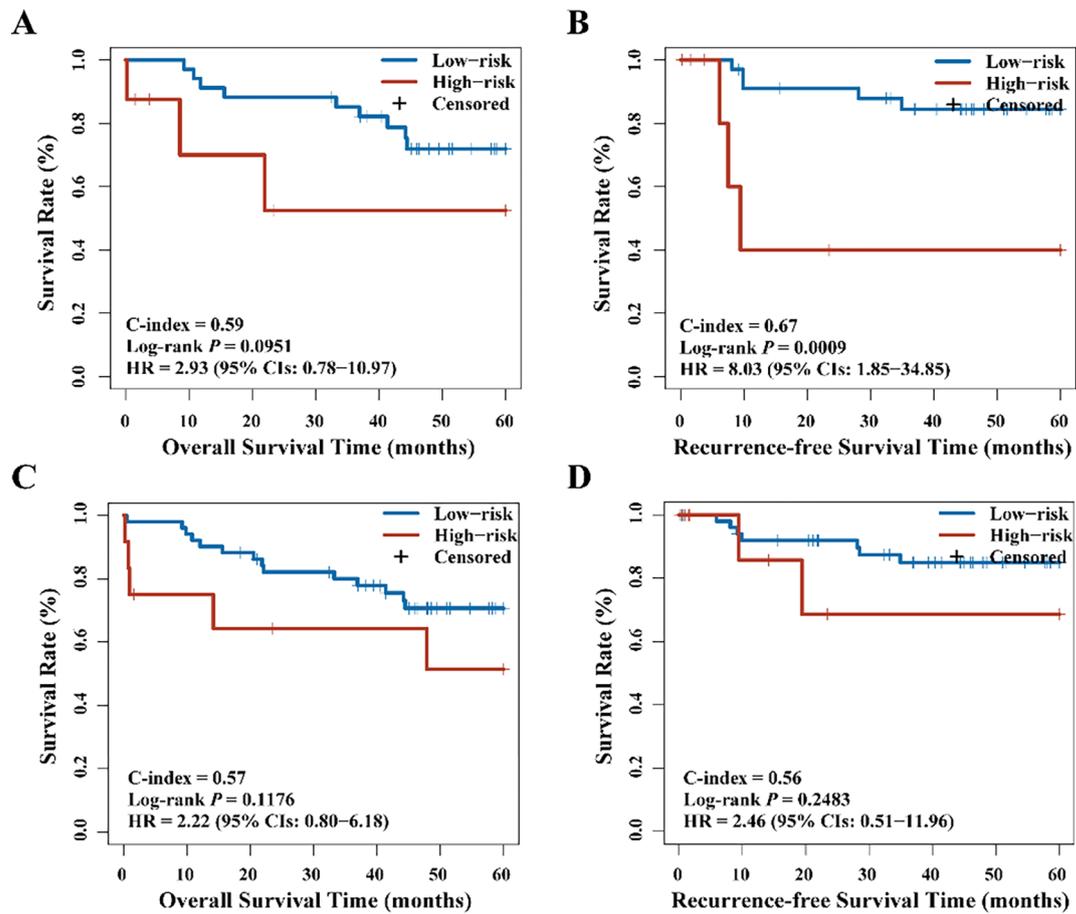


Figure S4 Kaplan-Meier curves for patients with early-stage disease in the non-small cell lung cancer (NSCLC)–radiogenomics (NR) dataset. (A) Kaplan-Meier curves for overall survival of 45 patients with stage T1 disease; (B) Kaplan-Meier curves for recurrence-free survival of 45 patients with stage T1 disease; (C) Kaplan-Meier curves for overall survival of 63 patients with stage N0 disease; (D) Kaplan-Meier curves for recurrence-free survival of 63 patients with stage N0 disease.

Table S5 Gene enrichment analysis of significantly correlated genes with radiomic features in 3-PET based on the KEGG database

Feature	Pathway	FDR	Gene symbol
Standard deviation	Cell cycle	2.66E-05	<i>TTK, SKP1, PRKDC, CDC6, PLK1, CDC20, MCM6, CCND1, CDC25C, MAD2L1, YWHAQ, CCNE1, CCNA2, CDC16, MAD2L2, TFDP1, HDAC1, CCNB2, CDK1, CCNB3, MCM4, ORC1, CCNB1, ANAPC11, BUB1B</i>
Standard deviation	Huntington's disease	0.0009	<i>PSMB5, POLR2I, TUBA1C, POLR2F, PSMA7, CYCS, ATP5PF, WIPI2, UQCRH, PSMC3, PIK3C3, CASP3, NDUFA6, UQCRHL, BBC3, GPX8, MAP3K5, PSMC4, PSMB2, KLC3, PSMB7, HDAC1, ATG2A, NDUFB6, DNAH1, TUBA1B, HAP1, UQCRFS1, CLTA, NDUFA4, VDAC2, PSMC1, SDHB, UQCRQ, TUBB2B, PLCB2, POLR2D, COX6B1</i>
Standard deviation	Parkinson's disease	0.0037	<i>PSMB5, TUBA1C, PSMA7, CYCS, ATP5PF, UQCRH, PSMC3, CASP3, NDUFA6, UQCRHL, PARK7, MAP3K5, PSMC4, PSMB2, KLC3, GNAI3, PSMB7, EIF2S1, MAOB, NDUFB6, SNCA, TUBA1B, UBE2L3, UQCRFS1, NDUFA4, VDAC2, PSMC1, SDHB, UQCRQ, TUBB2B, COX6B1</i>
Standard deviation	Progesterone-mediated oocyte maturation	0.004	<i>RPS6KA2, PLK1, SPDYE2, CDC25C, MAD2L1, CCNA2, CDC16, MAD2L2, GNAI3, AURKA, CCNB2, CDK1, MOS, CCNB3, CCNB1, ANAPC11, ADCY9</i>
Standard deviation	Oocyte meiosis	0.0105	<i>RPS6KA2, SKP1, PLK1, CDC20, SPDYE2, PPP2CA, CDC25C, MAD2L1, YWHAQ, CCNE1, CDC16, MAD2L2, AURKA, CCNB2, CDK1, MOS, CCNB1, ANAPC11, ADCY9</i>
Standard deviation	Ribosome	0.022	<i>RPL38, RPL9, MRPS16, RPL27, RPS8, RPL26L1, RPS17, RPS21, RPL17-C18orf32, RPL35, MRPL11, MRPL15, MRPL30, RPL11, RPL21, MRPS14, RPL4, MRPS7, MRPL12</i>
Standard deviation	Alzheimer's disease	0.0343	<i>ATP2A1, PSMB5, TUBA1C, PSMA7, CYCS, ATP5PF, WIPI2, UQCRH, PSMC3, PIK3C3, CASP3, NDUFA6, CACNA1F, UQCRHL, RTN3, AXIN1, NRAS, MAP3K5, PSMC4, PSMB2, KLC3, PSMB7, EIF2S1, ATG2A, NDUFB6, SNCA, TUBA1B, UQCRFS1, NDUFA4, CACNA1S, VDAC2, PSMC1, SDHB, UQCRQ, TUBB2B, PLCB2, COX6B1</i>
Standard deviation	DNA replication	0.0462	<i>RNASEH2B, RFC5, POLE2, RFC3, RNASEH2A, MCM6, RFC2, MCM4</i>
Standard deviation	P53 signaling pathway	0.0483	<i>CYCS, CASP3, GTSE1, RRM2, BBC3, CCND1, CCNE1, TNFRSF10A, TNFRSF10B, CCNB2, CDK1, CCNB1</i>
Median	Cell cycle	1.57E-07	<i>YWHAB, ANAPC1, CUL1, PRKDC, CDC6, CDK2, PLK1, STAG1, SMC3, CDC20, MCM6, CDC25C, ESPL1, CDKN1B, GADD45G, CCNA2, TFDP1, CCNB2, E2F4, CD-C25A, GADD45B, MCM2, MCM4, MCM3, CDK6, ORC1, CCNB1, BUB1B, CDC23</i>
Median	RNA transport	0.0007	<i>TGS1, UPF1, ALYREF, NUP188, EIF4EBP2, XPO5, TACC3, EIF3B, XPOT, POP1, PABPC3, NCBP2, THOC6, EIF2S1, EIF2B2, NUP214, POM121C, RPP38, EIF5B, EIF4G1, PRMT5, UPF3B, RANGAP1, NUP93, NUP50, NUP153, EIF4A2</i>
Median	Oocyte meiosis	0.0295	<i>YWHAB, ANAPC1, CUL1, CALM2, CDK2, PLK1, SMC3, CDC20, CDC25C, ESPL1, AURKA, CCNB2, PPP2R5D, MOS, PPP2R1B, PPP2R5E, PPP2CB, CCNB1, CDC23</i>
Third quartile	Cell cycle	3.37E-10	<i>YWHAB, CUL1, TTK, PRKDC, CDC6, CDC45, CDK2, PLK1, CDC20, MCM6, CDC25C, MAD2L1, ESPL1, CCNE1, GADD45G, CCNA2, CDC16, TFDP1, HDAC1, CCNB2, CDK1, BUB1, RBL1, CDC25A, GADD45B, MCM2, MCM4, MCM3, ORC1, CCNB1, BUB1B</i>
Third quartile	DNA replication	0.0004	<i>RFC5, FEN1, POLE2, RFC3, RNASEH2A, PRIM1, MCM6, MCM2, RFC2, MCM4, MCM3</i>
Third quartile	Proteasome	0.0128	<i>PSMB5, PSMA7, PSMC3, ADRM1, PSMC4, PSMB2, PSMD2, PSMB7, POMP, PSMC1</i>
Third quartile	P53 signaling pathway	0.0128	<i>ADGRB1, CDK2, GTSE1, SESN2, RRM2, BBC3, CCNE1, GADD45G, TNFRSF10B, CCNB2, CDK1, GADD45B, CCNB1</i>
Third quartile	Oocyte meiosis	0.0128	<i>YWHAB, CUL1, RPS6KA2, CDK2, PLK1, CDC20, CDC25C, MAD2L1, ESPL1, CCNE1, CDC16, AURKA, CCNB2, CDK1, BUB1, MOS, PPP2CB, CCNB1</i>
Third quartile	Progesterone-mediated oocyte maturation	0.0453	<i>RPS6KA2, CDK2, PLK1, CDC25C, MAD2L1, CCNA2, CDC16, AURKA, CCNB2, CDK1, BUB1, MOS, CDC25A, CCNB1</i>

FDR, false discovery rate; KEGG, Kyoto Encyclopedia of Genes and Genomes; PET, positron emission tomography.