A prognostic risk model based on DNA methylation levels of genes and lncRNAs in lung squamous cell carcinoma

Weiqing Wang, Ming Xiang, Hui Liu, Xiao Chu, Zhaoyun Sun and Liang Feng

Department of Thoracic Surgery, The Fifth People's Hospital of Shanghai, Shanghai, China

ABSTRACT

Background: Recurrence is a risk factor for the prognosis of lung squamous carcinoma (LUSC). DNA methylation levels of RNAs are also associated with LUSC prognosis. This study aimed to construct a prognostic model with high performance in predicting LUSC prognosis using the methylation levels of lncRNAs and genes.

Methods: The differentially expressed RNAs (DERs) and differentially methylated RNAs (DMRs) between the recurrent and non-recurrent LUSC tissues in The Cancer Genome Atlas (TCGA; training dataset) were identified. Weighted correlation network analysis was performed to identify co-methylation networks. Differentially methylated genes and lncRNAs with opposite expression-methylation levels were used for the screening of prognosis-associated RNAs. The prognostic model was constructed and its performance was validated in the GSE39279 dataset.

Results: A total of 664 DERs and 981 DMRs (including 972 genes) in recurrent LUSC tissues were identified. Three co-methylation modules, including 226 differentially methylated genes, were significantly associated with LUSC. Among prognosis-associated RNAs, 18 DERs/DMRs with opposite methylation-expression levels were included in the methylation prognostic risk model. LUSC patients with high risk scores had a poor prognosis compared with patients who had low risk scores (TCGA: HR = 3.856, 95% CI [2.297–6.471]; GSE39279: HR = 3.040, 95% CI [1.435–6.437]). This model had a high accuracy in predicting the prognosis (AUC = 0.903 and 0.800, respectively), equivalent to the nomogram model inclusive of clinical variables.

Conclusions: Referring to the methylation levels of the 16-RNAs might help to predict the survival outcomes in LUSC.

INTRODUCTION

Lung squamous carcinoma or lung squamous cell carcinoma (LUSC) is a histologic type of lung cancer that ranks first in the rate of incidence and mortality (*Bray et al., 2018*). The number of newly diagnosed cases and cancer-related death of lung cancer in 2018 was approximately 2.1 million and 1.8 million, respectively (*Bray et al., 2018*). Non-small cell

Corresponding author
Liang Feng,
Fengliang@5thhospital.com

lung cancer (NSCLC) is the most predominant type of lung cancer (accounting for 80%) and LUSC composes ~30% of NSCLC (*Lemjabbar-Alaoui et al., 2015*).

Patients with recurrent LUSC have a very poor survival prognosis. Tumor recurrence is distinctly influenced by clinical characteristics including patients' age, histologic type, stage, treatment response, gene expression, and epigenetic regulation (*Dziedzic et al., 2016*; *Feng et al., 2015*; *Smith et al., 2010*). DNA methylation of genes is critical for tumorigenesis, progression, metastasis, recurrence, and resistance to therapy (*Castilho, Squarize & Almeida, 2017*; *Liu et al., 2017*; *Ma, Chen & Petersen, 2017*; *Sosa, Bernstein & Aguirre-Ghiso, 2017*. For instance, the promoter methylation levels of secreted frizzled-related protein genes have a high accuracy in diagnosing NSCLC (*Liu et al., 2017*). Also, DNA methylation levels of tumor suppressor genes, including death-associated protein kinase 1 (DAPK1) and retinoic acid receptor beta (RARB), are higher in lung cancer tissues than noncancerous lung tissues (*Daniunaite et al., 2020*). Also, the aberrant hypermethylation levels of RARB and DAPK1 promoters are strongly associated with the invasive progression in cervical cancer (*Tawe et al., 2020*). Furthermore, DNA methylation-mediated regulation of long non-coding RNAs (lncRNAs) also plays crucial roles in tumors by regulating the expression of genes and microRNAs (*Heilmann et al., 2017*; *Ma et al., 2017*)

The hypomethylated promoter of the angiotensin-converting enzyme 2 (ACE2) gene, a receptor of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), has been identified in lung adenocarcinoma tissues (*Chai et al., 2020*). Most of the coronavirus disease 2019 (COVID-19)-affected patients are patients who have malignancies (*Calabrò et al., 2020*; *Luo et al., 2020*). Also, lung adenocarcinoma patients are more susceptible to SARS-CoV-2 infection than LUSC patients (*Kong et al., 2020*). COVID-19 patients who have a history of radiation therapy for cancers appear to have a poor prognosis (*Robilotti et al., 2020*). COVID-19 pandemic has brought great challenges not only to lung cancer therapy (*Calabrò et al., 2020*; *Guckenberger et al., 2020*; *Passaro et al., 2020*) but also to the diagnosis and prognosis of lung cancer (*Maringe et al., 2020*; *Pruis et al., 2020*; *Robilotti et al., 2020*). These studies show a potential correlation between an increased risk of COVID-19 infection in patients with lung cancer. Therefore, the identification of more and potential biomarkers in lung cancer might provide additional information on making treatment strategies for lung cancer.

This study aimed to screen a prognostic risk score system based on DNA methylation levels of genes and lncRNAs between recurrent and non-recurrent LUSC using bioinformatics analysis. The performance of this model in predicting the prognosis of LUSC was validated using a microarray dataset. Also, a combined nomogram model to predict the prognosis in LUSC patients was constructed. This study might provide a reference for assessing the prognosis of patients with LUSC.

MATERIALS AND METHODS

Datasets

The Cancer Genome Atlas (TCGA; https://www.cancer.gov/about-nci/organization/ccg/research/structural-genomics/tcga) LUSC RNA-seq (Illumina HiSeq 2000 RNA

Sequencing; $n = 550$) and methylation profile data (Illumina Infinium Human Methylation 450 BeadChip; $n = 415$) were downloaded. After matching sample IDs, 293 samples with gene expression profiles, methylation profiles, and clinical recurrence information (with recurrence = 78; without recurrence = 215) were retained and used as the training dataset in this study.

The methylation microarray dataset GSE39279 was downloaded from the Gene Expression Omnibus (GEO) at the National Center for Biotechnology Information (NCBI; https://www.ncbi.nlm.nih.gov/geo/). It was selected using the following criteria: (1) methylation profiles from patients with LUSC; (2) inclusive of clinical recurrence information; (3) ≥150 samples. GSE39279 (GPL13534, Illumina HumanMethylation450 BeadChip) contains 444 samples, including 43 samples with recurrence information. This dataset was used as the validation dataset.

RNA annotation and identification of RNAs differentially expressed and methylated in recurrent LUSC

The annotation of lncRNAs and mRNAs in the expression and methylation files was performed using the HUGO Gene Nomenclature Committee (HGNC; http://www.genenames.org/). Ensembl IDs were converted to official gene symbols. The differentially expressed RNAs (DER) and differentially methylated RNAs (DMR), including lncRNAs and genes, between the samples with and without recurrence were identified in the TCGA training dataset. The limma package (version 3.34.7; https://bioconductor.org/packages/release/bioc/html/limma.html) in R was used for the identification of DERs and DMRs. To maximize the intersection of DERs and DMRs, we set $p < 0.05$, false discovery rate (FDR) < 0.05, and $|\log_2(\text{Fold change, FC})| > 0.263$ as the criteria for significant difference. The expression profiles of RNAs with differential expression and methylation levels were presented using the bidirectional hierarchical clustering heatmap by pheatmap (version 1.0.8; https://cran.r-project.org/web/packages/pheatmap/index.html). Besides, RNAs with both differential methylation and expression levels were identified using the Venn diagram.

Weighted correlation network analysis (WGCNA) of genes

The WGCNA networks identify gene modules associated with disease status based on the expression/methylation profiles of genes. The WGCNA package (version 1.63; https://cran.r-project.org/web/packages/WGCNA/index.html) in R was used to analyze the co-methylation networks for RNAs in the TCGA training dataset, irrespective of the expression and methylation level. The criteria for WGCNA module identification were: min size = 100, cutHeight = 0.995, $p < 0.05$, and enrichment fold > 1. Genes included in WGCNA modules were matched with the RNAs with differentially expression and methylation levels, and the overlapping items were used for further analysis.

Construction of methylation prognostic model

Genes and lncRNAs included in WGCNA co-methylation modules that had opposite levels of methylation and expression between the TCGA LUSC samples with and without

recurrence were used as candidates for screening prognosis-associated RNAs. The univariate Cox regression analysis in the R Survival package (version 2.41-1; http://bioconductor.org/packages/survivalr/) was used for the screening of RNAs associated with LUSC prognosis. The criterion was log-rank p value < 0.05. The optimal methylation combination was subsequently identified using the Cox-Proportional Hazards (Cox-PH) model (L1-penalized least absolute shrinkage and selection operator, LASSO) in the R penalized package (version 0.9-50, http://bioconductor.org/packages/penalized/). The methylation prognostic model was then established and the prognostic risk score of each individual was calculated using the following algorithm: Prognostic risk score = $\sum coef_{DMRs} \times Methylation_{DMRs}$, where coef represents the coefficient (LASSO coef) of each gene identified by the Cox-PH model and Methylation notes the methylation level of the gene in each sample, respectively. For the Kaplan–Meier survival curve analysis (version 2.41-1), samples in the training and validation datasets were grouped into the high and low-risk groups according to the median prognostic risk score. Also, the receiver operating characteristic (ROC) curve analysis was performed using the R pROC (version 1.14.0; https://cran.r-project.org/web/packages/pROC/index.html) to evaluate the accuracy of using this methylation prognostic model in predicting LUSC prognosis.

Identification of clinical variables associated with LUSC prognosis

This study also identified clinical variables associated with the LUSC prognosis in the TCGA training dataset. The associations of prognostic risk score and clinical variables (including patient age, gender, pathologic TNM classification, tumor stage, and radiation/targeted therapy) with LUSC prognosis were assessed using the univariate and multivariate Cox regression analysis in the R Survival package (version 2.41-1). Independent variables that were significantly associated with LUSC prognosis were selected using the threshold of log-rank p value < 0.05.

Prognostic nomogram and index for survival

Prognostic nomogram and index are widely applied for estimated survival probability among patients with cancers and other conditions (*Barnholtz-Sloan et al., 2012*; *Gold et al., 2009*). Nomogram was established using the R "RMS" package (Version 5.1-2; https://cran.r-project.org/web/packages/rms/index.html). The score of each variable was ascribed according to its weight in the nomogram, and the individualized 3- and 5-year survival probabilities were then predicted according to total points.

Functional annotation for lncRNAs in the methylation prognostic model

Importantly, the functional annotations of differentially expressed lncRNAs including in the methylation prognostic model were identified to investigate the biological themes associated with lncRNAs. The correlation between expression profiles of differentially expressed genes and lncRNAs in recurrent LUSC samples in the TCGA training dataset were calculated using the Pearson correlation coefficient (r). LncRNA-gene pairs with confident coexpression levels ($r > 0.4$ or $r < -0.4$) were retained and the lncRNA-mRNA

regulatory network was constructed using the Cytoscape software (version 3.8.0; http://www.cytoscape.org/). Using the expression profiles of the lncRNA-associated genes in the TCGA training samples, Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways associated with lncRNAs were identified using the Gene Set Enrichment Analysis (GSEA) software (version 4.0.2; http://software.broadinstitute.org/gsea/index. jsp). The threshold for significant enrichment was a normal p value < 0.05.

RESULTS

Identification of DERs and DMRs

After annotation in the HGNC database, a total of 54 lncRNAs and 12,932 genes with available expression profiles were obtained in the TCGA dataset, including 664 DERs and 981 DMRs (including 972 genes and nine lncRNAs) between the LUSC samples with and without recurrence (Figs. 1A and 1B). The expression and methylation profiles of the DERs and DMRs in the LUSC samples are shown in Figs. 1C and 1D, respectively. Venn diagram showed that 155 DERs were differentially methylated in recurrent LUSC samples compared with samples without tumor recurrence (Fig. 2; Table S1).

WGCNA for DERs with differentially methylation levels in LUSC samples

Prior to the WGCNA module analysis, the scale-free topology prerequisite soft-thresholding power was identified. It was 7 when the scale-free topology model fit R^2 = 0.9 for the first time (Fig. 3A). The mean gene connectivity = 1 when soft-thresholding power = 7 (Fig. 3B), conforming to the scale-free network property. Based on the methylation profiles of all RNAs, 12 WGCNA modules (including 105-1327 RNAs) were subsequently identified in the TCGA training dataset according to the criteria we set before (soft threshold power = 7, min size = 100, and cutHeight = 0.995; Fig. 3C; Table 1). The module-trait relationship heatmap showing the correlations of modules with clinical variables is shown in Fig. 4. Also, the numbers of differentially methylated genes and fold enrichment value of each module are shown in Table 1. Three WGCNA co-methylation modules (green, pink, and turquoise), including 226 differentially methylated genes, had enrichment folds of >1 and an enrichment p value of <0.05.

Identification of LUSC prognosis-associated RNAs

After matching the above 226 WGCNA module genes with the 155 RNAs in Fig. 2, 41 common genes were identified (Table S3). Among them, 28 genes had opposite methylation-expression levels in the TCGA LUSC samples (Table S3). Besides, six of the nine lncRNAs listed in Table S1 had opposite methylation-expression levels. Among the 34 candidate DERs (including 28 genes and six lncRNAs), 26 DERs (including four lncRNAs and 22 genes) were identified as prognosis-associated RNAs by univariate Cox regression analysis and methylation profiles (Table S4).
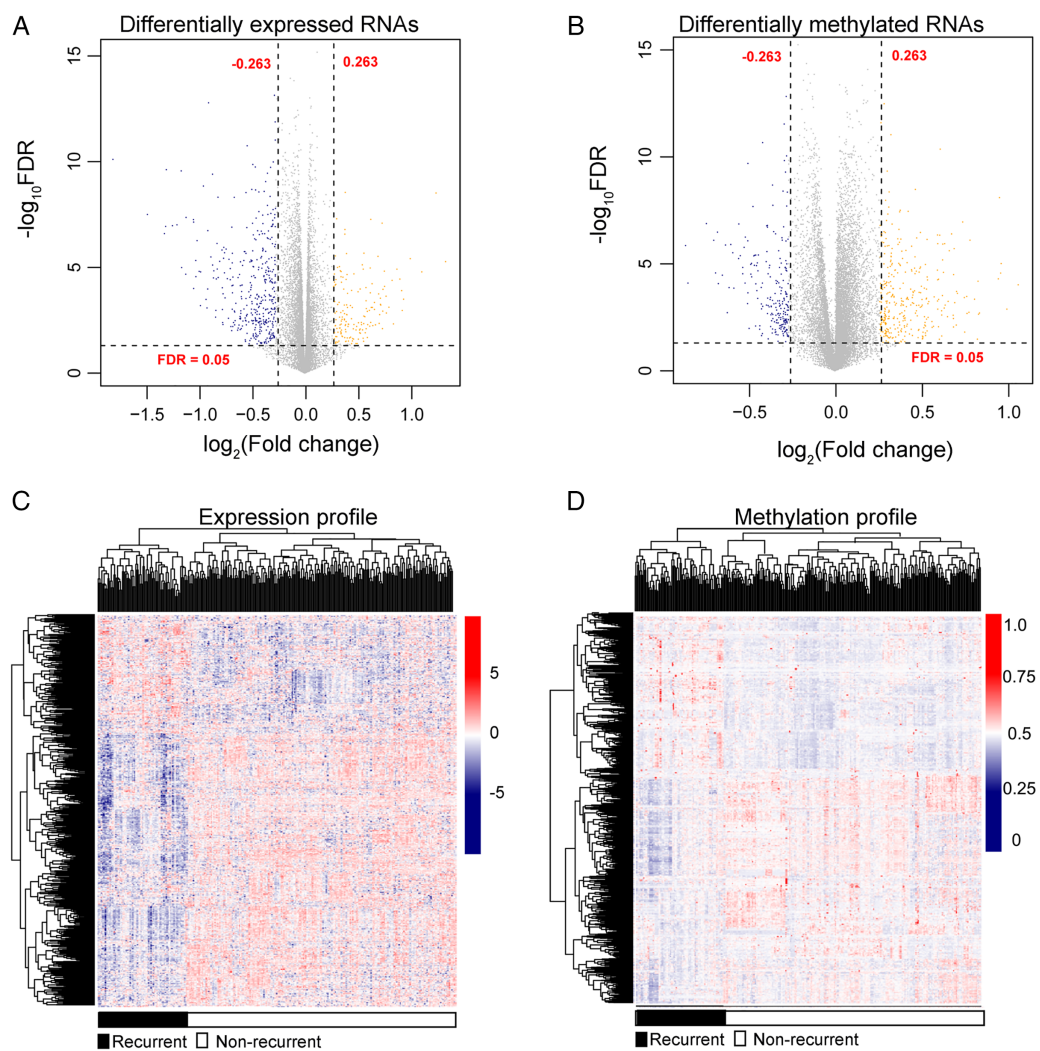
Figure 1 The expression and methylation levels of lncRNAs and genes in lung squamous carcinoma samples. (A & B) The Volcano plots indicating the differentially expressed and methylated RNAs between recurrent and non-recurrent tumors. (C & D) The bidirectional hierarchical clustering heatmap indicating the expression ($\log_2[\text{FPKM} + 1]$) and methylation profiles (beta value) of RNAs. FDR, false discovery rate.

Construction of the methylation prognostic model

An optimized methylation prognostic combination was identified using the Cox-PH model based on the 26 prognosis-associated DMRs. Eighteen DMRs, including two lncRNAs and 16 genes, were included in this methylation prognostic combination (Table 2). The associations of seven RNAs with LUSC prognosis by the multivariate Cox PH model (Table 2), including *DIRC3*, *RMST*, *ABCA12*, *ADH7*, *DGKA*, *NPHP3*, and *WFDC10B*, were inconsistent with those analyzed by the univariate Cox regression analysis in R package (Table S4). Accordingly, the methylation prognostic model was constructed and the prognostic risk score of each sample was calculated as: $(-0.0103) \times \text{Methylation}_{DIRC3} + (-0.0242) \times \text{Methylation}_{RMST} + 0.0866 \times \text{Methylation}_{LTF} + 0.0438 \times \text{Methylation}_{SGCG} + (-0.0476) \times \text{Methylation}_{ABCA12} + (-0.0171) \times \text{Methylation}_{ADH7} +$

Differentially expressed RNAs   Differentially methylated RNAs
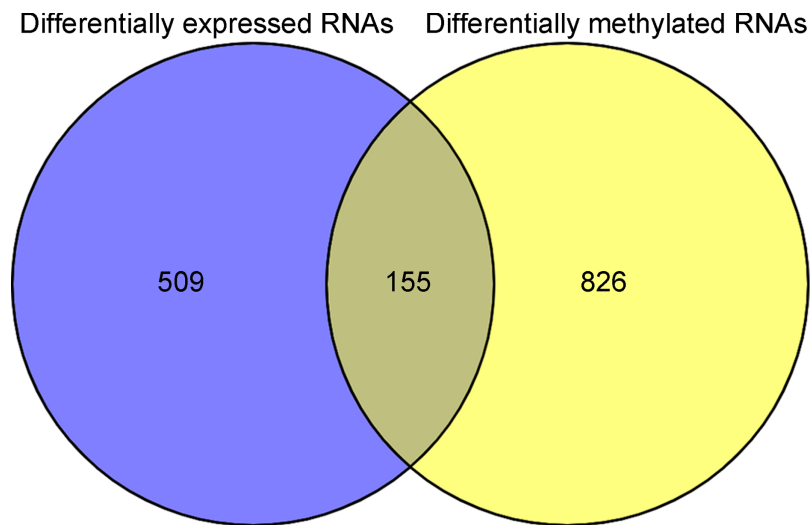
Figure 2 The Venn diagram indicating the RNAs with differentially expression and methylation levels in recurrent lung squamous carcinoma samples.

$0.0846 \times \text{Methylation}_{BNIPL} + (-0.0273) \times \text{Methylation}_{DGKA} + 0.1114 \times \text{Methylation}_{FAM181B} + 0.0376 \times \text{Methylation}_{GNRH2} + 0.0575 \times \text{Methylation}_{HORMAD2} + (-0.7505) \times \text{Methylation}_{LIMCH1} + (-0.0164) \times \text{Methylation}_{NPHP3} + 0.0823 \times \text{Methylation}_{RTP1} + 0.1422 \times \text{Methylation}_{ST6GALNAC1} + 0.0667 \times \text{Methylation}_{THNSL2} + 0.0546 \times \text{Methylation}_{TRIM7} + (-0.0604) \times \text{Methylation}_{WFDC10B}$. Almost all these genes, except *LIMCH1* (expression logFC = 0.27, methylation logFC = −0.39), were downregulated and hypermethylated in recurrent LUSC samples compared with non-recurrent LUSC samples (Tables S1 and S2).

Accuracy of the methylation prognostic model in predicting prognosis in LUSC

To validate the predictive power of the methylation prognostic model in LUSC, samples in the TCGA training dataset (n = 293) and the GSE39279 validation dataset (n = 43) were separately grouped into the high and low-risk groups according to the median risk score of each sample. Kaplan–Meier survival curve analysis showed that there was a significant difference in the recurrence-free survival ratio between patients with high and low-risk scores in the TCGA training dataset (HR = 3.856, 95% CI [2.297–6.471], p = 2.648e − 08; Fig. 5A) and the GSE39279 validation dataset (HR = 3.040, 95% CI [1.435–6.437], p = 2.403e − 03; Fig. 5B). The ROC curve analysis showed that this methylation prognostic model had a high accuracy in predicting LUSC prognosis in the training dataset (area under the ROC curve, AUC = 0.903, 95% CI [0.902–0.946]; Fig. 5C) and in the validation dataset (AUC = 0.800, 95% CI [0.700–0.846]; Fig. 5D).

Clinical variables associated with LUSC prognosis

We also identified the clinical variables associated with the prognosis of LUSC in the TCGA cohort. Univariate and multivariate Cox regression analysis confirmed that pathologic stage (HR = 1.443, 95% CI [1.297–2.854], p = 2.92e − 02), radiation therapy (HR = 1.963, 95% CI [1.022–3.771], p = 4.27e − 02), and prognostic risk score (HR = 3.874,
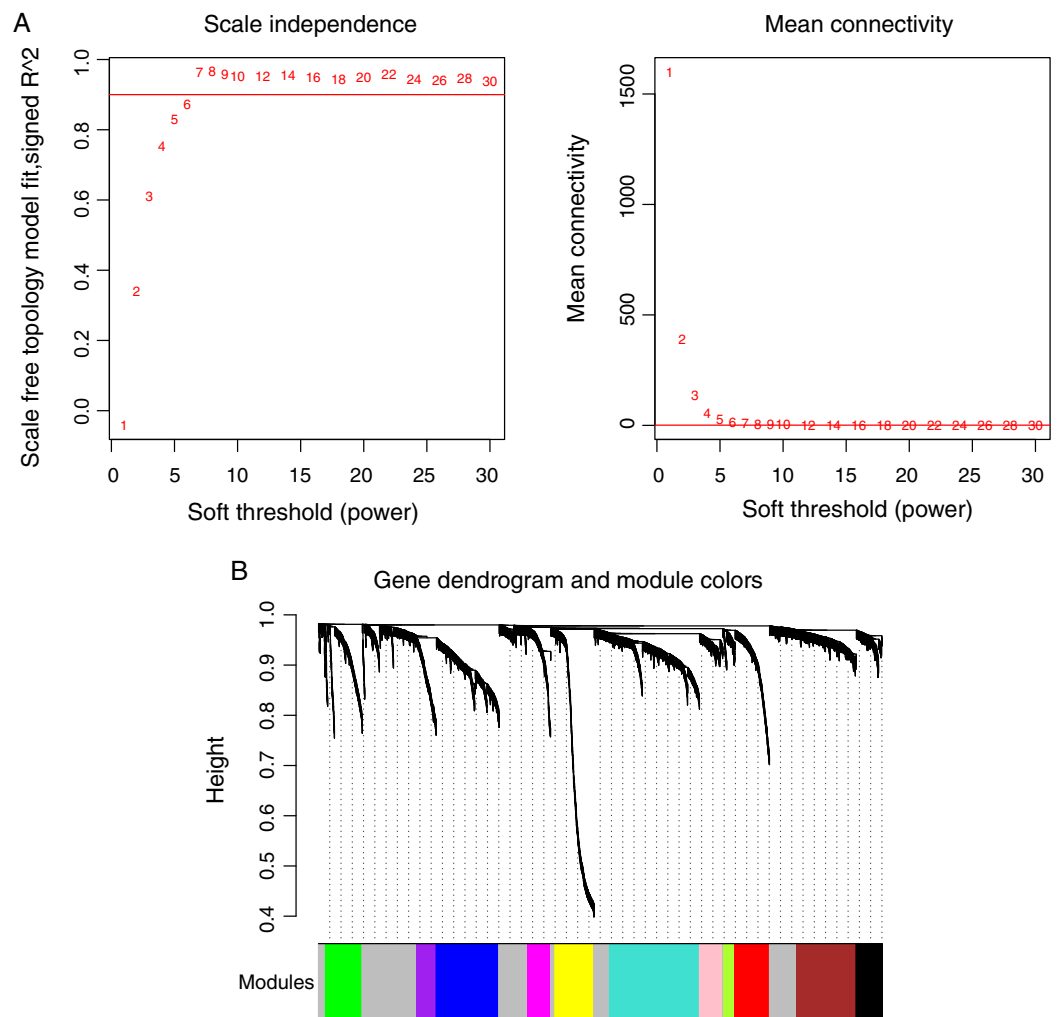
Figure 3 Weighted correlation network analysis (WGCNA) for RNAs with differentially methylation and expression levels. (A) Plots showing the selection of the soft-thresholding power (n = 7). (B) The gene dendrogram indicating the identification of WGCNA modules.

95% CI [2.256–6.653], p = 9.12e − 07) were associated with LUSC prognosis (Table 3). Kaplan–Meier curve analysis showed that patients with advanced stages and with radiotherapy had a lower recurrence-free survival ratio compared with controls (pathologic stage: HR = 2.604, 95% CI [1.570–4.318], p = 1.177e − 04; Fig. 6A, left; radiotherapy: HR = 2.054, 95% CI [1.145–3.686], p = 1.36e − 02; Fig. 6B, left). Subgroup survival analysis showed that patients with high prognostic risk scores (high-risk group) had a lower survival ratio compared with patients who had low scores (low-risk group), irrespective of pathologic stage and radiotherapy (Figs. 6A and 6B, middle and right).

Prognostic nomogram

The prognostic nomogram was constructed using the methylation prognostic model, pathologic stage (1–4), and radiation therapy (yes/no; Fig. 7A). The predicted 3- and 5-year survival probabilities based on the nomogram had high compliances with the actual

Table 1 Weighted correlation network analysis (WGCNA) module information.

Module	Size	DMGs	Enrichment information	
			Enrichment fold (95% CI)	P_{hyper}
Black	245	10	0.387 (0.182–0.729)	1.30E−03
Blue	601	58	0.914 (0.677–1.217)	5.75E−01
Brown	570	47	0.781 (0.561–1.068)	1.27E−01
Green	**352**	**99**	**2.665 (2.076–3.398)**	**8.41E−14**
Greenyellow	105	5	0.451 (0.143–1.093)	9.72E−02
Grey	1,327	135	0.964 (0.786–1.176)	7.65E−01
Magenta	220	21	0.905 (0.544–1.431)	7.38E−01
Pink	**224**	**36**	**1.523 (1.029–2.198)**	**3.16E−02**
Purple	186	4	0.204 (0.0547–0.533)	1.08E−04
Red	329	33	0.951 (0.637–1.378)	8.54E−01
Turquoise	**861**	**91**	**1.201 (1.046–2.267)**	**4.96E−02**
Yellow	372	28	0.713 (0.463–1.060)	9.28E−02

Notes:
DMG, differentially methylated genes; CI, confident interval; P_{hyper}, p value by hypergeometric algorithm.
Bold text notes significant modules and parameters.
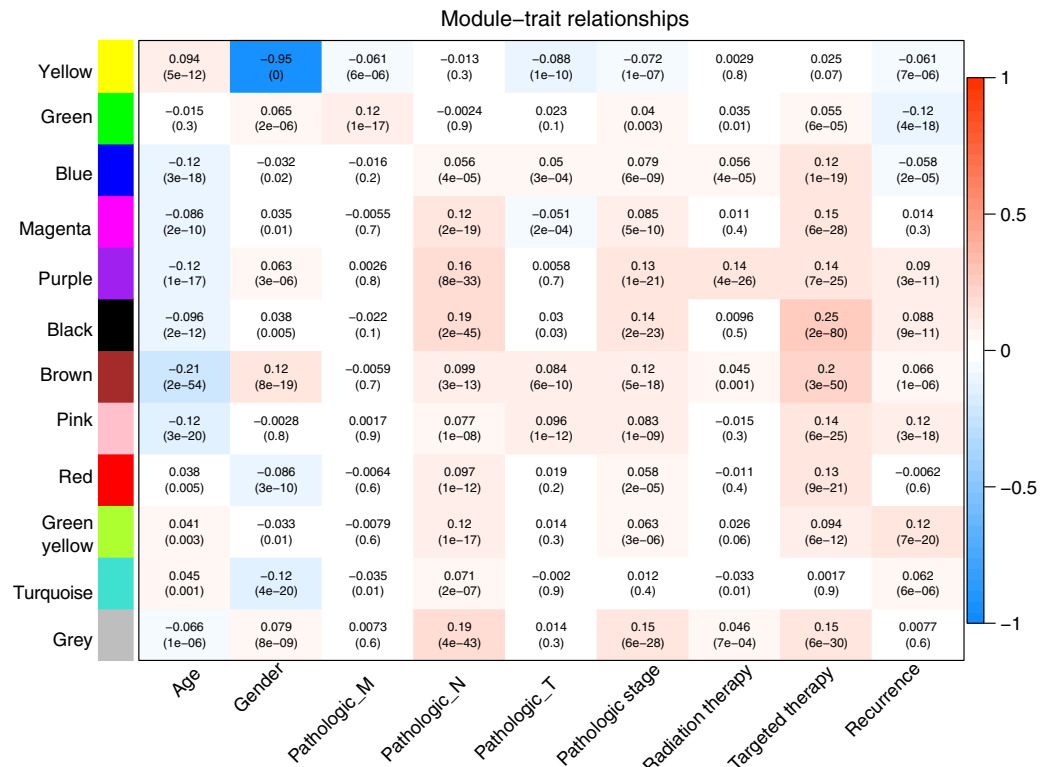
Figure 4 Heatmap showing the correlation of WGCNA modules with clinical variables. Red and blue notes positive (maximum value = 1) and negative (minimum value = −1) correlation, respectively. WGCNA, weighted correlation network analysis.

Table 2 RNAs used for the construction of the prognostic methylation model in this study.

Symbol	Type	Multivariate Cox regression analysis			LASSO coef
		HR	95% CI	p value	
DIRC3	lncRNA	0.932	0.779–0.996	3.880E−02	−0.0103
RMST		0.961	0.830–0.985	1.140E−02	−0.0242
LTF	mRNA	1.276	1.101–3.946	2.080E−02	0.0866
SGCG		1.198	1.054–5.930	4.170E−02	0.0438
ABCA12		0.925	0.769–0.986	4.120E−02	−0.0476
ADH7		0.965	0.797–0.976	4.140E-02	−0.0171
BNIPL		1.349	1.112–5.9131	2.630E−02	0.0846
DGKA		0.864	0.684–0.937	3.220E−02	−0.0273
FAM181B		1.585	1.199–3.864	2.550E−02	0.1114
GNRH2		1.194	1.029–3.888	2.690E−02	0.0376
HORMAD2		1.380	1.114–4.899	4.130E−02	0.0575
LIMCH1		0.401	0.257–0.624	5.280E−06	−0.7505
NPHP3		0.953	0.800–0.986	4.420E−02	−0.0164
RTP1		1.363	1.118–5.918	3.070E−02	0.0823
ST6GALNAC1		1.491	1.190–3.950	1.100E−02	0.1422
THNSL2		1.224	1.080–4.954	1.980E−02	0.0667
TRIM7		1.289	1.097–4.934	2.940E−02	0.0546
WFDC10B		0.895	0.728–0.912	3.640E−02	−0.0604

Note:
LASSO, L1-penalized least absolute shrinkage and selection operator; HR, hazard ratio; CI, confident interval; coef, coefficient.

situations (Fig. 7B; c-index = 0.747 and 0.759 for 3-year and 5-year prediction, respectively). For an example case with stage III tumor (points = 33.3), radiation therapy (points = 0), and a risk score of 0.2 (points = 37.7; total points = 71), the predicted 3- and 5-year survival probability of the case based on the nomogram model was about 38.7% and 23.9%, respectively. ROC curve analysis showed that the combination of the clinical variables with the methylation prognostic model had the highest AUC value (AUC = 0.948, 95% CI [0.894–0.958]) and c-index (0.755; $p = 2.22e-16$; Fig. 7C; Table 4), followed by the methylation prognostic model alone (AUC = 0.927, 95% CI [0.884–0.931], c-index = 0.739, $p = 6.66e-15$).

Pathways associated with the two lncRNAs in the methylation prognostic model

A total of 320 lncRNA-mRNA pairs related to the two lncRNAs (DIRC3 and RMST, downregulation and hypermethylation) in the methylation prognostic model were extracted ($r > 0.4$ or $r < -0.4$; Table S5). Accordingly, the lncRNA-mRNA network included 320 interactions and 281 nodes (two lncRNAs and 279 differentially expressed genes; Fig. 8).

Downregulated genes with hypermethylation levels, including LTF, ADH7, ST6GALNAC1, THNSL2, and WFDC10B, had positive correlations with RMST, genes
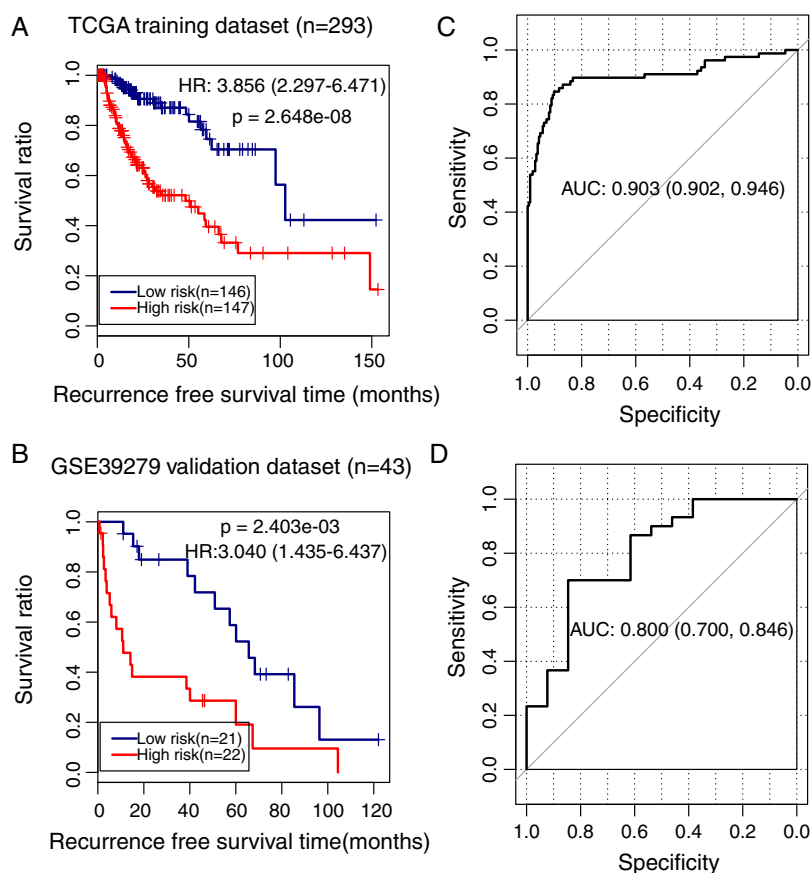
Figure 5 Analysis of the methylation prognostic model in lung squamous cell carcinoma. (A & B) The Kaplan–Meier curves showing the differences in recurrence-free survival ratios between patients with high- and low-risk scores in the training (TCGA) and validation (GSE39279) dataset, respectively. Prognostic risk scores were calculated based on the methylation prognostic model. (C & D) The receiver operating characteristic (ROC) curve analysis for this model in the training and validation cohort, respectively. AUC, area under the ROC curve.

including *ABCA12*, *DGKA*, *FAM181B*, and *TRIM7* were positively correlated with *DIRC3*, *LIMCH1* had negative correlation with both of *RMST* and *DIRC3*, and the other genes including *SGCG*, *BNIPL*, *GNRH2*, *HORMAD2*, *NPHP3*, and *RTP1* had positive correlations with both of *RMST* and *DIRC3* (Table S5). Based on the expression profiles of the 279 genes in the TCGA LUSC cohort, we identified that *DIRC3* and *RMST* were associated with five and four KEGG pathways, respectively (Table 5). *DIRC3* and *RMST* both were associated with the "Pathways in cancer" and "Endocytosis".

DISCUSSION

Our study constructed a methylation prognostic model consisting of 18 DERs, including two lncRNAs (*DIRC3* and *RMST*) and 16 genes (*LTF*, *ADH7*, *ST6GALNAC1*, *LIMCH1*, *THNSL2*, *WFDC10B*, *SGCG*, *BNIPL*, *GNRH2*, *HORMAD2*, *ABCA12*, *DGKA*, *FAM181B*, *TRIM7*, *NPHP3*, and *RTP1*), based on the recurrent status in LUSC patients. Among the 18 RNAs, 17 DERs were downregulated with hypermethylation levels in recurrent LUSC tissues compared with non-recurrent tumors. The 18-RNA methylation prognostic model

Table 3 Cox regression analysis for clinical variables in The Cancer Genome Atlas (TCGA) lung squamous cell carcinoma cohort.

Clinical characteristics	TCGA (n = 293)	Univariate		Multivariate	
		HR (95% CI)	p value	HR (95% CI)	p value
Age (years, mean ± SD)	67.72 ± 8.63	0.988 (0.962–1.014)	3.51E−01		
Gender (Male/Female)	211/82	1.329 (0.783–2.257)	2.90E−01		
Pathologic M (0/1/−)	227/2/64	0.303 (0.151–1.536)	6.07E−01		
Pathologic N (0/1/2/−)	193/76/20/4	1.562 (1.142–2.138)	4.72E−03	1.011 (0.519–1.854)	9.53E−01
Pathologic T (1/2/3/4)	79/159/48/7	1.639 (1.209–2.220)	1.43E−03	1.223 (0.749–1.998)	4.21E−01
Pathologic stage (I/II/III/IV/−)	139/110/39/2/3	1.768 (1.326–2.356)	7.57E−05	1.443 (1.297–2.854)	2.92E−02
Radiation therapy (Yes/No/−)	31/251/11	2.054 (1.145–3.686)	1.37E−02	1.963 (1.022–3.771)	4.27E−02
Targeted therapy (Yes/No/−)	94/188/11	1.238 (0.774–1.981)	3.72E−01		
PS status (High/Low)	146/147	3.856 (2.297–6.471)	2.65E−08	3.874 (2.256–6.653)	9.12E−07
Recurrence (Yes/No)	78/215				
RFS time (months, mean ± SD)	28.28 ± 27.45				

Note:

CI, confident interval; HR, hazard ratio; SD, standard deviation; RFS, recurrence-free survival; PS, prognostic risk score.
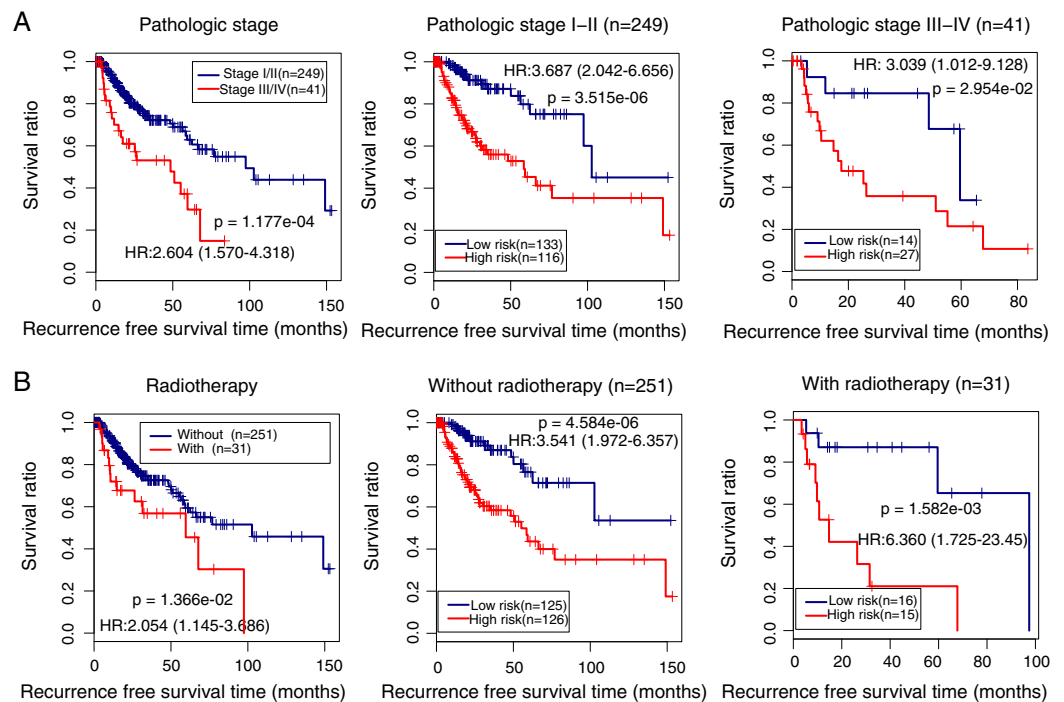
Figure 6 Subgroup analysis showing the performance of the methylation prognostic model in predicting prognosis in lung squamous carcinoma. (A & B) The differences in the survival ratio between patients with different clinical variables (left) and between patients with high- and low-risk scores (middle and right). HR, hazards ratio.

and the nomogram model consisting of the 18-RNA methylation signature and clinical variables both had high performance in predicting LUSC prognosis.

The inconsistency that seven RNAs, including *DIRC3*, *RMST*, *ABCA12*, *ADH7*, *DGKA*, *NPHP3*, and *WFDC10B*, by the multivariate Cox PH model was inconsistent with the
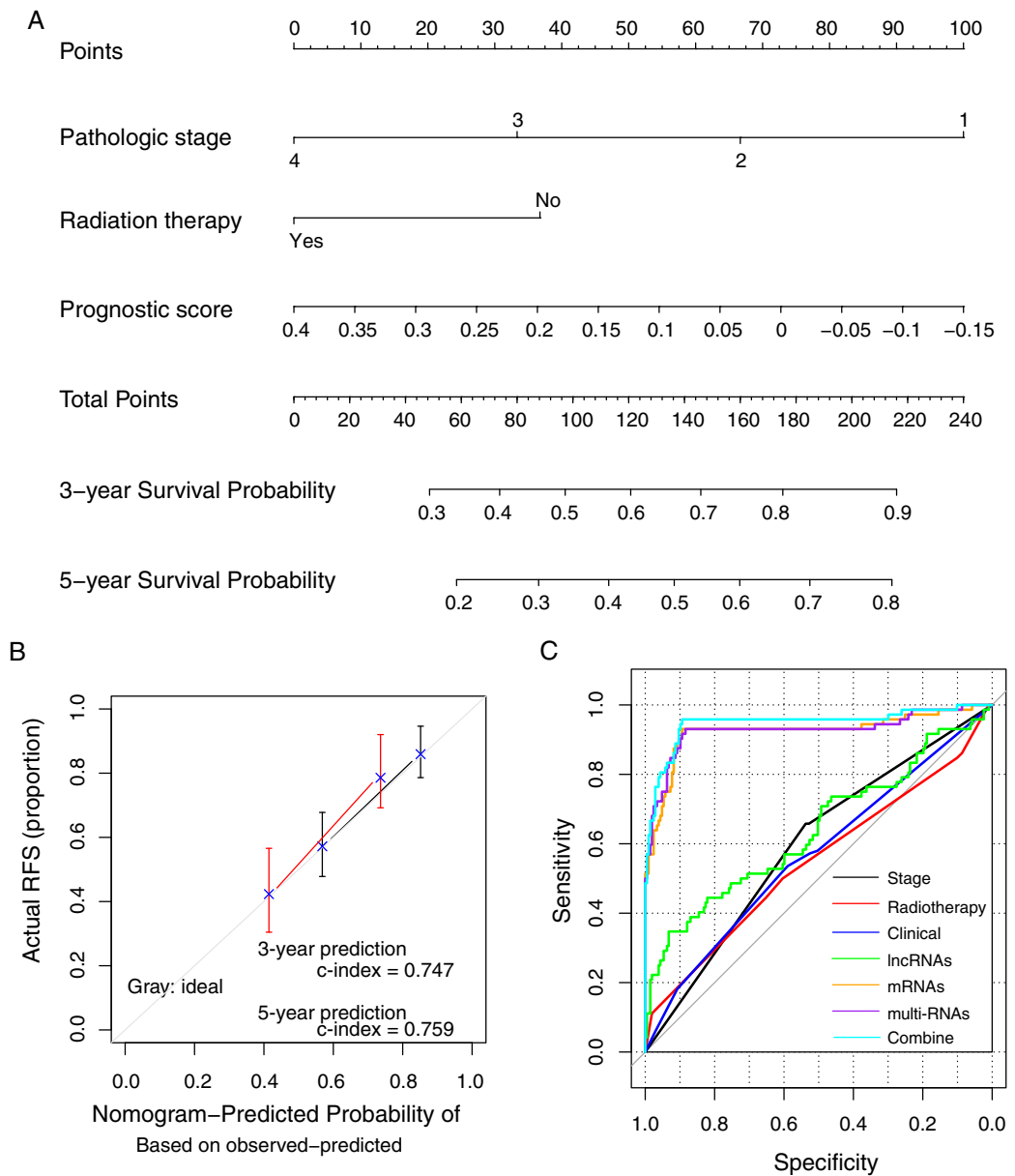
Figure 7 The prognostic nomogram and index for recurrence-free survival. (A) The weight of clinical factors and prognostic risk score in predicting the prognosis of lung squamous carcinoma. (B) The c-index analysis for the predicted survival probability based on the nomogram model. (C) The receiver operating characteristic (ROC) curve analysis for all models. RFS, recurrent-free survival.

univariate Cox regression analysis in R package was identified in our study.

The methylations of seven RNAs were associated with poor prognosis in LUSC patients by the univariate Cox regression analysis, but the Multivariate Cox PH model showed that they were associated with good prognosis in LUSC. However, further analyses showed that there were confident correlations among the two lncRNAs *RMST* and *DIRC3* and the other 16 genes, which could be the cause of the above inconsistency. Also, the lncRNA-mRNA network showed that the genes might be regulated by the two lncRNAs. These results

Table 4 The receiver operating characteristic (ROC) curve analysis for the clinical variables and prognostic methylation model.

Type	AUC	C-index	p value
Stage	0.596 (0.537, 0.658)	0.653	4.64E−06
Radiotherapy	0.548 (0.604, 0.500)	0.550	6.41E−02
Clinical	0.598 (0.568, 0.536)	0.656	8.38E−06
lncRNAs	0.640 (0.732, 0.547)	0.542	2.77E−01
mRNAs	0.924 (0.894, 0.931)	0.732	7.04E−14
multi-RNAs	0.927 (0.884, 0.931)	0.739	6.66E−15
Combine	0.948 (0.894, 0.958)	0.755	2.22E−16

Note:
AUC, area under the ROC curve.
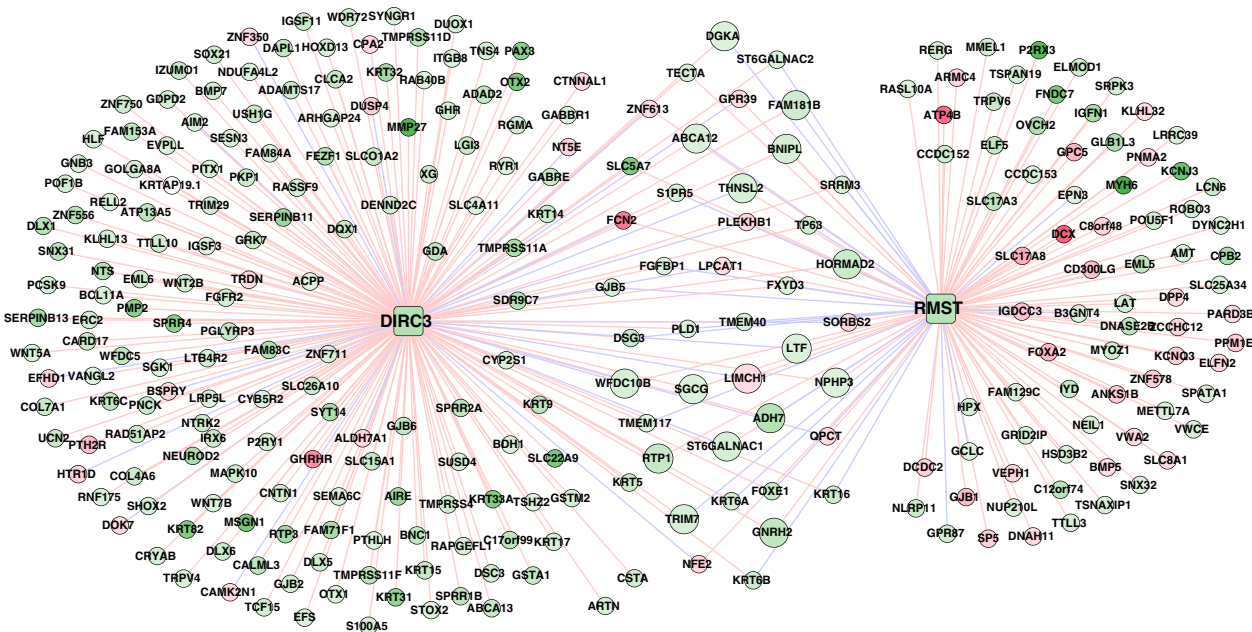
Figure 8 The lncRNA–mRNA network of *DIRC3* and *RMST*. This network was constructed using the differentially expressed genes that had confident correlations ($r > 0.4$ or $r < -0.4$) with the two lncRNAs in lung squamous carcinoma samples. Green and red color indicate downregulation and upregulation, respectively. Red and blue lines note positive ($r > 0.4$) and negative ($r < -0.4$) correlations, respectively.

showed that the two lncRNAs *RMST* and *DIRC3* might be confounding factors for the other genes.

In the methylation prognostic model, LIM and calponin-homology domains 1 (*LIMCH1*) is the only gene downregulated and positively correlated with the LUSC prognosis by univariate Cox regression analysis (HR = 7.96E−03). *LIMCH1* regulates the activity of an important motor protein monmuscle myosin II and cell migration and growth in lung cancer cells (*Lin et al., 2017*; *Zhang, Zhang & Xu, 2019*). It partially restored the invasive phenotype of cancer cells (*Bersini et al., 2020*). However, *Halle et al. (2021)* showed that high *LIMCH1* expression was significantly associated with poor survival

Table 5 Pathways associated with lncRNA DIRC3 and RMST.

Name	Size	ES	NES	NOM p-value	Gene
lncRNA DIRC3					
Metabolism of xenobiotics by cytochrome P450	4	0.82	1.65	8.14E−03	*CYP2S1, GSTM2, GSTA1, ADH7*
Drug metabolism cytochrome P450	3	0.82	1.51	2.75E−02	*GSTM2, GSTA1, ADH7*
Endocytosis	3	0.74	1.50	3.37E−02	*GRK7, PLD1, FGFR2*
MAPK signaling pathway	2	0.64	1.31	4.14E−02	*NTRK2, FGFR2*
Pathways in cancer	4	0.59	1.44	4.82E−02	*WNT2B, WNT5A, COL4A6, FGFR2*
lncRNA RMST					
Pathways in cancer	5	−0.62	−1.64	5.03E−03	*WNT7B, WNT5A, COL4A6, PLD1, FGFR2*
Neurotrophin signaling pathway	2	−0.76	−1.37	1.20E−02	*NTRK2, CALML3*
Endocytosis	4	−0.67	−1.37	1.55E−02	*EPN3, GRK7, PLD1, FGFR2*
Wnt signaling pathway	2	−0.52	−1.18	2.88E−02	*WNT7B, WNT5A*

Note:
ES, enrichment score; NES, normalized enrichment score; Nom p-value, normal p value.

of cervical cancer ($p = 0.004$, HR = 3.17). Also, our study showed that high *LIMCH1* methylation level contributed to a low risk score and was significantly associated with a good prognosis of LUSC. These results showed that the contribution of expression and methylation levels of *LIMCH1* and other RNAs in the methylation prognostic model to LUSC prognosis should be validated in more cohorts.

The HORMA Domain Containing 2 (*HORMAD2*) gene is a host gene of HIV proviruses and it has one provirus integration, which is essential for synapsis surveillance during host cell meiosis (*Fennessey et al., 2017*; *Symons, Cameron & Lewin, 2018*). Also, this increases the difficulty of virus removal and provides an implication for virus management (*Symons, Cameron & Lewin, 2018*). The tripartite motif-containing 7 (*TRIM7*) gene is a tumor suppressor gene that encodes glycogenin-interacting proteins and an E3 ubiquitin ligase, which suppresses the progression of hepatocellular carcinoma (*Jin et al., 2020*; *Zhu et al., 2020*). E3 ubiquitin ligases are involved in various cellular functions, including innate immunity and inflammation (*Giraldo et al., 2020a*). Loss of N6-Methyladenosine modification of *TRIM7* promoter was observed in osteosarcoma tissues (*Zhou et al., 2020*). Evidence shows that E3 ubiquitin ligase positively regulates toll-like receptor 4 (TLR4)-mediated immune response in macrophages (*Lu et al., 2019*). Of special interest are recent reports showing their roles in virus replication (*Giraldo et al., 2020a*; *Stukalov et al., 2020*).

TRIM7 can either promote virus pathogenesis or protect against infection, depending on the context of virus infection (*Giraldo et al., 2020a*; *Orchard et al., 2019*). It promotes herpes virus and Zika virions infection (*Giraldo et al., 2020b*) but may play an antiviral role against norovirus (*Orchard et al., 2019*). TRIM7-mediated ubiquitination of the envelope protein of Zika virions promotes the release of virus from infected cells and binding of envelope protein to host cellular receptors, and enhances viral entry and replication (*Orchard et al., 2019*). A recent study by *Stukalov et al. (2020)* highlighted that the TRIM7 protein binding to the SARS-CoV-2 M phosphorylation site to drive

ubiquitination. However, TRIM7's effect on SARS-CoV-2 replication is not clear now. Our present study found that the DNA methylation levels of *HORMAD2* and *TRIM7* promoter were upregulated in recurrent LUSC tissues compared with non-recurrent tissues, and *HORMAD2* and *TRIM7* expression were downregulated in recurrent LUSC tissues. The deregulation and methylation levels of *HORMAD2* and *TRIM7* in recurrent LUSC tissues might provide references for exploring the potential association of LUSC prognosis with virus infection.

LncRNAs rhabdomyosarcoma 2 associated transcript (*RMST*) and disrupted in renal carcinoma 3 (*DIRC3*) both, especially the former, are known as tumor suppressors (*Coe et al., 2019*; *Kong, Liu & Kong, 2018*; *Liu et al., 2019*; *Wang et al., 2018*). The depletion and loss of function mutation of the two lncRNAs eliminate the threat of malignant transformation, promote tumor cell proliferation, migration, and invasion (*Coe et al., 2019*; *Liu et al., 2019*; *Wang et al., 2018*). *RMST* could enhance tumor cell apoptosis, block the G0/G1 phase and cell proliferation, and restrain cell invasion and migration in triple-negative breast cancer (*Wang et al., 2018*). *RMST* (*Peng et al., 2020*) positively regulates the DNA methyltransferase 3B (*DNMT3B*) and negatively regulates *DNMT3* expression (*Peng et al., 2020*). *RMST* knockout suppresses *DNMT3* by promoting the binding of HuR protein to DNMT3B and enhancing DNMT3B stability (*Peng et al., 2020*). What's more, the deletion of *Dnmt3a* in mouse promotes lung tumor progression (*Gao et al., 2011*), suggesting the critical roles of *RMST* in regulating tumor progression and development.

Among the downregulated and hypermethylated genes associated with both *RMST* and *DIRC3* (including *BNIPL*, *HORMAD2*, and *NPHP3*), *HORMAD2* is the only gene related to lung cancer (*Liu et al., 2012*). *Liu et al. (2012)* showed that *HORMAD2* mRNA expression was rarely expressed in the lung and HORMAD2 protein was detected more frequently in early-stage lung adenocarcinoma compared with advanced cancers. The nephronophthisis 3 (*NPHP3*) gene is associated with nephronophthisis and is necessary for the formation of primary cilia formation (*Abdullah et al., 2017*; *Lee, Kim & Moon, 2019*). BNIP-2-like (BNIPL) is an apoptosis-associated protein that interacts with cell proliferation-related proteins including BCL-2 and Cdc42GAP (*Qin et al., 2003*). However, our present study showed that these genes and lncRNAs were downregulated in recurrent LUSC specimens, along with increased methylation levels. Besides, the inclusion of them in the prognostic model had high performance in predicting the prognosis in LUSC patients, suggesting the novel and crucial roles of these genes in LUSC progression.

CONCLUSIONS

In summary, our study identified a prognostic model based on the methylation levels of lncRNAs and genes in LUSC patients with and without recurrence. This methylation had high performance in predicting the prognosis as well as the 3-year and 5-year survival probabilities in LUSC patients. Of special interest is that most of these RNAs had not been reported in lung cancers and two genes (*HORMAD2* and *TRIM7*) might associate with virus infection. Referring to the methylation levels of these RNAs might help to predict the survival outcomes in LUSC.

Funding

This work was supported by the Minhang District natural science research project (2020MHZ038). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Grant Disclosures

The following grant information was disclosed by the authors:
Minhang District Natural Science Research Project: 2020MHZ038.

Competing Interests

The authors declare that they have no competing interests.

Author Contributions

- Weiqing Wang conceived and designed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, and approved the final draft.
- Ming Xiang analyzed the data, authored or reviewed drafts of the paper, and approved the final draft.
- Hui Liu conceived and designed the experiments, authored or reviewed drafts of the paper, and approved the final draft.
- Xiao Chu analyzed the data, authored or reviewed drafts of the paper, and approved the final draft.
- Zhaoyun Sun analyzed the data, authored or reviewed drafts of the paper, and approved the final draft.
- Liang Feng conceived and designed the experiments, authored or reviewed drafts of the paper, and approved the final draft.

Data Availability

The original LUSC RNA-seq (Illumina HiSeq 2000 RNA Sequencing) and methylation profile data (Illumina Infinium Human Methylation 450 BeadChip) are available at The Cancer Genome Atlas (TCGA LUSC; https://www.cancer.gov/about-nci/organization/ccg/research/structural-genomics/tcga) and the original microarray data is available at the Gene Expression Omnibus: GSE39279.

Supplemental Information

Supplemental information for this article can be found online at http://dx.doi.org/10.7717/peerj.13057#supplemental-information.

REFERENCES

Abdullah U, Farooq M, Fatima A, Tauseef W, Sarwar Y, Nuri M, Tommerup N, Baig SM. 2017. Homozygous mutation in the NPHP3 gene causing foetal nephronophthisis. *Nephrology* **22(10)**:818–820 DOI 10.1111/nep.13097.

Barnholtz-Sloan JS, Yu C, Sloan AE, Vengoechea J, Wang M, Dignam JJ, Vogelbaum MA, Sperduto PW, Mehta MP, Machtay M. 2012. A nomogram for individualized estimation of survival among patients with brain metastasis. *Neuro-Oncology* **14(7)**:910–918 DOI 10.1093/neuonc/nos087.

Bersini S, Lytle NK, Schulte R, Huang L, Wahl GM, Hetzer MW. 2020. Nup93 regulates breast tumor growth by modulating cell proliferation and actin cytoskeleton remodeling. *Life Science Alliance* **3(1)**:e201900623 DOI 10.26508/lsa.201900623.

Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. 2018. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians* **68(6)**:394–424 DOI 10.3322/caac.21492.

Calabrò L, Peters S, Soria J-C, Di Giacomo AM, Barlesi F, Covre A, Altomonte M, Vegni V, Gridelli C, Reck M, Rizvi N, Maio M. 2020. Challenges in lung cancer therapy during the COVID-19 pandemic. *Lancet Respiratory Medicine* **8(6)**:542–544 DOI 10.1016/S2213-2600(20)30170-3.

Castilho RM, Squarize CH, Almeida LO. 2017. Epigenetic modifications and head and neck cancer: implications for tumor progression and resistance to therapy. *International Journal of Molecular Sciences* **18(7)**:1506 DOI 10.3390/ijms18071506.

Chai P, Yu J, Ge S, Jia R, Fan X. 2020. Genetic alteration, RNA expression, and DNA methylation profiling of coronavirus disease 2019 (COVID-19) receptor ACE2 in malignancies: a pan-cancer analysis. *Journal of Hematology & Oncology* **13(1)**:1–5 DOI 10.1186/s13045-020-00883-5.

Coe EA, Tan JY, Shapiro M, Louphrasitthiphol P, Bassett AR, Marques AC, Goding CR, Vance KW. 2019. The MITF-SOX10 regulated long non-coding RNA DIRC3 is a melanoma tumour suppressor. *PLOS Genetics* **15(12)**:e1008501 DOI 10.1371/journal.pgen.1008501.

Daniunaite K, Sestokaite A, Kubiliute R, Stuopelyte K, Kettunen E, Husgafvel-Pursiainen K, Jarmalaite S. 2020. Frequent DNA methylation changes in cancerous and noncancerous lung tissues from smokers with non-small cell lung cancer. *Mutagenesis* **35(5)**:373–379 DOI 10.1093/mutage/geaa022.

Dziedzic DA, Rudzinski P, Langfort R, Orlowski T. 2016. Risk factors for local and distant recurrence after surgical treatment in patients with non-small-cell lung cancer. *Clinical Lung Cancer* **17**:e157–e167 DOI 10.1016/j.cllc.2015.12.013.

Feng H, Yu Z, Tian Y, Lee Y-Y, Li MS, Go MY, Cheung Y-S, Lai PB, Chan AM, To K-F. 2015. A CCRK-EZH2 epigenetic circuitry drives hepatocarcinogenesis and associates with tumor recurrence and poor survival of patients. *Journal of Hepatology* **62(5)**:1100–1111 DOI 10.1016/j.jhep.2014.11.040.

Fennessey C, Pinkevych M, Immonen T, Camus C, Del Prete G, Estes J, Lifson J, Davenport M, Keele B. 2017. Assessing individual viral reactivations of the latent reservoir using a novel barcoded virus. *Journal of the International AIDS Society* **20(Suppl 5)**:22253 DOI 10.7448/IAS.20.6.22253.

Gao Q, Steine EJ, Barrasa MI, Hockemeyer D, Pawlak M, Fu D, Reddy S, Bell GW, Jaenisch R. 2011. Deletion of the de novo DNA methyltransferase Dnmt3a promotes lung tumor progression. *Proceedings of the National Academy of Sciences* **108(44)**:18061–18066 DOI 10.1073/pnas.1114946108.

Giraldo MI, Hage A, van Tol S, Rajsbaum R. 2020a. TRIM proteins in host defense and viral pathogenesis. *Current Clinical Microbiology Reports* **7**(4):1–14 DOI 10.1007/s40588-020-00150-8.

Giraldo MI, Xia H, Aguilera-Aguirre L, Hage A, van Tol S, Shan C, Xie X, Sturdevant GL, Robertson SJ, McNally KL. 2020b. Envelope protein ubiquitination drives entry and pathogenesis of Zika virus. *Nature* **585**(7825):414–419 DOI 10.1038/s41586-020-2457-8.

Gold JS, Gönen M, Gutiérrez A, Broto JM, García-del-Muro X, Smyrk TC, Maki RG, Singer S, Brennan MF, Antonescu CR. 2009. Development and validation of a prognostic nomogram for recurrence-free survival after complete surgical resection of localised primary gastrointestinal stromal tumour: a retrospective analysis. *Lancet Oncology* **10**(11):1045–1052 DOI 10.1016/S1470-2045(09)70242-6.

Guckenberger M, Belka C, Bezjak A, Bradley J, Daly ME, DeRuysscher D, Dziadziuszko R, Faivre-Finn C, Flentje M, Gore E. 2020. Practice recommendations for lung cancer radiotherapy during the COVID-19 pandemic: an ESTRO-ASTRO consensus statement. *Radiotherapy and Oncology* **146**:223–229 DOI 10.1016/j.radonc.2020.04.001.

Halle MK, Sødal M, Forsse D, Engerud H, Woie K, Lura NG, Wagner-Larsen KS, Trovik J, Bertelsen BI, Haldorsen IS. 2021. A 10-gene prognostic signature points to LIMCH1 and HLA-DQB1 as important players in aggressive cervical cancer disease. *British Journal of Cancer* **124**(10):1690–1698 DOI 10.1038/s41416-021-01305-0.

Heilmann K, Toth R, Bossmann C, Klimo K, Plass C, Gerhauser C. 2017. Genome-wide screen for differentially methylated long noncoding RNAs identifies Esrp2 and lncRNA Esrp2-as regulated by enhancer DNA methylation with prognostic relevance for human breast cancer. *Oncogene* **36**(46):6446–6461 DOI 10.1038/onc.2017.246.

Jin J, Lu Z, Wang X, Liu Y, Han T, Wang Y, Wang T, Gan M, Xie C, Wang J. 2020. E3 ubiquitin ligase TRIM7 negatively regulates NF-kappa B signaling pathway by degrading p65 in lung cancer. *Cellular Signalling* **69**(6):109543 DOI 10.1016/j.cellsig.2020.109543.

Kong X, Liu W, Kong Y. 2018. Roles and expression profiles of long non-coding RNAs in triple-negative breast cancers. *Journal of Cellular and Molecular Medicine* **22**(1):390–394 DOI 10.1111/jcmm.13327.

Kong Q, Xiang Z, Wu Y, Gu Y, Guo J, Geng F. 2020. Analysis of the susceptibility of lung cancer patients to SARS-CoV-2 infection. *Molecular Cancer* **19**:1–5 DOI 10.1186/s12943-020-01209-2.

Lee J-W, Kim HS, Moon E-Y. 2019. Thymosin β-4 is a novel regulator for primary cilium formation by nephronophthisis 3 in HeLa human cervical cancer cells. *Scientific Reports* **9**(1):1–11 DOI 10.1038/s41598-019-43235-1.

Lemjabbar-Alaoui H, Hassan OU, Yang Y-W, Buchanan P. 2015. Lung cancer: biology and treatment options. *Biochimica et Biophysica Acta (BBA)—Reviews on Cancer* **1856**(2):189–210 DOI 10.1016/j.bbcan.2015.08.002.

Lin Y-H, Zhen Y-Y, Chien K-Y, Lee I-C, Lin W-C, Chen M-Y, Pai L-M. 2017. LIMCH1 regulates nonmuscle myosin-II activity and suppresses cell migration. *Molecular Biology of the Cell* **28**(8):1054–1065 DOI 10.1091/mbc.e15-04-0218.

Liu S, Chen X, Chen R, Wang J, Zhu G, Jiang J, Wang H, Duan S, Huang J. 2017. Diagnostic role of Wnt pathway gene promoter methylation in non small cell lung cancer. *Oncotarget* **8**(22):36354–36367 DOI 10.18632/oncotarget.16754.

Liu M, Chen J, Hu L, Shi X, Zhou Z, Hu Z, Sha J. 2012. HORMAD2/CT46. 2, a novel cancer/testis gene, is ectopically expressed in lung cancer tissues. *Molecular Human Reproduction* **18**(12):599–604 DOI 10.1093/molehr/gas033.

Liu C, Peng Z, Li P, Fu H, Feng J, Zhang Y, Liu T, Liu Y, Liu Q, Liu Q. 2019. LncRNA RMST functions as enhancer of SUMOylation to suppress GBM cells mitophagy. *The Lancet Preprint* DOI 10.2139/ssrn.3401970.

Lu M, Zhu X, Yang Z, Zhang W, Sun Z, Ji Q, Chen X, Zhu J, Wang C, Nie S. 2019. E3 ubiquitin ligase tripartite motif 7 positively regulates the TLR4-mediated immune response via its E3 ligase domain in macrophages. *Molecular Immunology* **109**:126–133 DOI 10.1016/j.molimm.2019.01.015.

Luo J, Rizvi H, Preeshagul IR, Egger JV, Hoyos D, Bandlamudi C, McCarthy CG, Falcon CJ, Schoenfeld AJ, Arbour KC. 2020. COVID-19 in patients with lung cancer. *Annals of Oncology* **31(10)**:1386–1396 DOI 10.1016/j.annonc.2020.06.007.

Ma Y, Chen Y, Petersen I. 2017. Expression and epigenetic regulation of cystatin B in lung cancer and colorectal cancer. *Pathology—Research and Practice* **213(12)**:1568–1574 DOI 10.1016/j.prp.2017.06.007.

Ma X, Yu L, Wang P, Yang X. 2017. Discovering DNA methylation patterns for long non-coding RNAs associated with cancer subtypes. *Computational Biology and Chemistry* **69(8)**:164–170 DOI 10.1016/j.compbiolchem.2017.03.014.

Maringe C, Spicer J, Morris M, Purushotham A, Nolte E, Sullivan R, Rachet B, Aggarwal A. 2020. The impact of the COVID-19 pandemic on cancer deaths due to delays in diagnosis in England, UK: a national, population-based, modelling study. *Lancet Oncology* **21(8)**:1023–1034 DOI 10.1016/S1470-2045(20)30388-0.

Orchard RC, Sullender ME, Dunlap BF, Balce DR, Doench JG, Virgin HW. 2019. Identification of antinorovirus genes in human cells using genome-wide CRISPR activation screening. *Journal of Virology* **93(1)**:e01324-18 DOI 10.1128/JVI.01324-18.

Passaro A, Addeo A, Von Garnier C, Blackhall F, Planchard D, Felip E, Dziadziuszko R, de Marinis F, Reck M, Bouchaab H. 2020. ESMO management and treatment adapted recommendations in the COVID-19 era: lung cancer. *ESMO Open* **5**:e000820 DOI 10.1136/esmoopen-2020-000820.

Peng W-X, Koirala P, Zhang W, Ni C, Wang Z, Yang L, Mo Y-Y. 2020. lncRNA RMST enhances DNMT3 expression through interaction with HuR. *Molecular Therapy* **28(1)**:9–18 DOI 10.1016/j.ymthe.2019.09.024.

Pruis MA, Hussain B, Bakker M, Hoek RA, Miedema JR, Dingemans A-MC, Paats MS. 2020. A lung cancer patient with dyspnea: diagnostic difficulties during the COVID-19 pandemic. *Cancer Cell* **38(5)**:611–612 DOI 10.1016/j.ccell.2020.10.005.

Qin W, Hu J, Guo M, Xu J, Li J, Yao G, Zhou X, Jiang H, Zhang P, Shen L. 2003. BNIPL-2, a novel homologue of BNIP-2, interacts with Bcl-2 and Cdc42GAP in apoptosis. *Biochemical and Biophysical Research Communications* **308(2)**:379–385 DOI 10.1016/S0006-291X(03)01387-1.

Robilotti EV, Babady NE, Mead PA, Rolling T, Perez-Johnston R, Bernardes M, Bogler Y, Caldararo M, Figueroa CJ, Glickman MS. 2020. Determinants of COVID-19 disease severity in patients with cancer. *Nature Medicine* **26(8)**:1218–1223 DOI 10.1038/s41591-020-0979-0.

Smith JJ, Deane NG, Wu F, Merchant NB, Zhang B, Jiang A, Lu P, Johnson JC, Schmidt C, Bailey CE. 2010. Experimentally derived metastasis gene expression profile predicts recurrence and death in patients with colon cancer. *Gastroenterology* **138(3)**:958–968 DOI 10.1053/j.gastro.2009.11.005.

Sosa MS, Bernstein E, Aguirre-Ghiso JA. 2017. Epigenetic regulation of cancer dormancy as a plasticity mechanism for metastasis initiation. In: *Tumor Dormancy and Recurrence*. Cham: Springer, 1–16.

Stukalov A, Girault V, Grass V, Bergant V, Karayel O, Urban C, Haas DA, Huang Y, Oubraham L, Wang A. 2020. Multi-level proteomics reveals host-perturbation strategies of SARS-CoV-2 and SARS-CoV. *bioRxiv* DOI 10.1101/2020.1106.1117.156455.

Symons J, Cameron PU, Lewin SR. 2018. HIV integration sites and implications for maintenance of the reservoir. *Current Opinion in HIV and AIDS* 13(2):152–159 DOI 10.1097/COH.0000000000000438.

Tawe L, Grover S, Zetola N, Robertson E, Gaseitsiwe S, Moyo S, Kasvosve I, Paganotti GM, Narasimhamurthy M. 2020. Aberrant hypermethylation of gene promoters (RARB, CADM1, DAPK1 and PAX1) is strongly associated with invasive cervical cancer in Botswana. *The Lancet Preprint* DOI 10.2139/ssrn.3582726.

Wang L, Liu D, Wu X, Zeng Y, Li L, Hou Y, Li W, Liu Z. 2018. Long non-coding RNA (LncRNA) RMST in triple-negative breast cancer (TNBC): expression analysis and biological roles research. *Journal of Cellular Physiology* 233(10):6603–6612 DOI 10.1002/jcp.26311.

Zhang Y, Zhang Y, Xu H. 2019. LIMCH1 suppress the growth of lung cancer by interacting with HUWE1 to sustain p53 stability. *Gene* 712(18):143963 DOI 10.1016/j.gene.2019.143963.

Zhou C, Zhang Z, Zhu X, Qian G, Zhou Y, Sun Y, Yu W, Wang J, Lu H, Lin F. 2020. N6-Methyladenosine modification of the TRIM7 positively regulates tumorigenesis and chemoresistance in osteosarcoma through ubiquitination of BRMS1. *EBioMedicine* 59(2):102955 DOI 10.1016/j.ebiom.2020.102955.

Zhu L, Qin C, Li T, Ma X, Qiu Y, Lin Y, Ma D, Qin Z, Sun C, Shen X. 2020. The E3 ubiquitin ligase TRIM7 suppressed hepatocellular carcinoma progression by directly targeting Src protein. *Cell Death & Differentiation* 27(6):1819–1831 DOI 10.1038/s41418-019-0464-9.