A Nomogram Based On A-To-I RNA Editing Predicting The Overall Survival of Patients With Lung Squamous Carcinoma

Li Liu

Guangzhou Medical University

Jun Liu

The Institute for Chemical Carcinogenesis, Guangzhou Medical University

Li Tu

Hospital of Changan

Zhuxiang Zhao

The Institute for Chemical Carcinogenesis, Guangzhou Medical University

Chenli Xie

The Institute for Chemical Carcinogenesis, Guangzhou Medical University

Lei Yang (✉ leiyang@gzhmu.edu.cn)

Guangzhou Medical University

Research Article

Keywords: LUSC, A-to-I RNA editing, overall survival, nomogram

Posted Date: May 26th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-500478/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background: A-to-I RNA editing has been recognized as a novel hallmark of cancer, while little is known about its contribution to cancer prognosis. We aimed to develop a prognostic nomogram with A-to-I editing events in lung squamous cell carcinoma (LUSC).

Methods: The TCGA A-to-I editing profile, corresponding clinical and gene expression data of LUSC were analyzed. Patients were randomly divided into a training (n = 134) and validation group (n = 94). An A-to-I risk biosignature was generated by univariate Cox regression and followed Lasso regression.

Results: We identified a seven A-to-I sites-based risk biosignature that includes TMEM120B_(A2588I), HMOX2_(A224I), CALCOCO2_(A2603I), MIR548AE2_(A113641I), ZNF440_(A3942I), CLCC1_(A2315I), and CHMP3_(A23735I). High risk was significantly associated with worse overall survival (OS) in both the training (HR = 7.30; 95%CI = 3.48-15.3) and validation groups (HR = 2.13; 95%CI = 1.09-4.15). Patients with advanced T ($P = 0.021$) or clinical stages ($P = 0.021$) had higher risk scores than the counterparts. We then developed a nomogram incorporating the A-to-I biosignature and clinicopathological features, which exerted well performance on predicting probability of LUSC OS with C-indexes as 0.808 and 0.685 in both sets. Moreover, the editing levels of ZNF440_(A3942I), CLCC1_(A2315I), HMOX2_(A224I) are correlated with expressions of their host genes, while the levels of TMEM120B_(A2588I), CLCC1_(A2315I), and CALCOCO2_(A2603I) differed between tumor tissues and normal tissues.

Conclusions: Our results for the first time provided insight into the development of A-to-I editing-based nomogram for predicting OS of LUSC patients.

Introduction

As the leading-cause of cancer-related death, lung cancer has resulted in 1,796,144 deaths in 2020, representing 18.0% of all global cancer-related deaths[1]. Over the past two decades, individualized targeted therapy have been practiced in lung cancer with appreciable benefits in some patients[2]. However, there are also a certain portion of patients who undergo non-response, acquired drug resistance, unexpected toxic effects after targeted therapy, underscoring the critical need for precisely prognostic models and more targeted agents. Hence, there is an urgent necessity to identify the underlying molecular alterations of lung cancer and characterize biosignature for cancer prognosis.

The leapfrog development of high-throughput sequencing technology and analytic tools of bioinformatics have deeply revealed the widespread DNA mutations and RNA editing in human genome and transcriptome during tumorigenesis, which are of great potential to be used as biomarkers for early detection, treatment selection and prognostic estimation, and targeted drug discovery of cancer. Unlike specifically expressed genes[3–7], somatic mutations[8], genetic variants[9], and DNA hypo- or hyper-methylation[10–13], all of which have been studied extensively for cancer prognostic prediction, little is known about the contribution of RNA editing to cancer prognosis. RNA editing is a molecular process

through which cells can make specific alterations in the chemical structure of RNA molecules occurring after DNA transcription and synthesis by the RNA polymerase enzyme. RNA editing events can result in missense codon changes in mRNA, modulation of alternative splicing in mRNA, or re-direction of mRNA-microRNA or Mrna-RNA binding protein interaction. Once thought rare in humans, RNA editing is now recognized as widespread in the human genome that is in the forefront of cancer research[14].

Over 70% of RNA editing in human is adenosine-to-inosine (A-to-I) editing[15]. A-to-I editing is a post-transcriptional modification of pre-mRNA that converts adenosines to inosines, which results in recognition of inosine as guanosine by posttranscriptional regulatory machinery causing potential recoding events in amino acid sequences and binding site rearrangement of microRNAs or RNA binding proteins in untranslated region (UTR). A-to-I editing has recently emerged as an important mechanism in cancer biology via affecting protein variation and diversity[16, 17]. In cancer studies, a lot of A-to-I editing events such as A-to-I-edited GABRA3[18], AZIN1[19], and miRNA-379-5p[20] have significant effects on cancer survival. Other studies have also shown that A-to-I editing regulates cancer progression[21, 22], metastasis[23], tumorigenesis[24–26], drug resistance[27]. However, the extent to which A-to-I editing predicts cancer survival is not known. In the current study, we aimed to develop a nomogram based on A-to-I editing sites for predicting overall survival (OS) of individuals affected by lung squamous cell carcinoma (LUSC), which is one of the primary types of lung cancer. The whole A-to-I RNA editing profiles of LUSC and clinical data were obtained from the TCGA database, and a nomogram incorporating A-to-I editing risk biosignature and baseline clinicopathological features that has considerably high accuracy to predict probability of LUSC OS was constructed.

Material And Methods

Sample selection and data processing

We downloaded the TCGA A-to-I RNA editing dataset of LUSC from the synapse website (<https://www.synapse.org/#!/Synapse:syn4382524>), which was uploaded by Han L et al [17], and the corresponding clinical information from the TCGA database (<https://portal.gdc.cancer.gov/>). Only 228 samples owned A-to-I editing data were retained and randomly divided into a training group (i.e., 134 cases) and a validation group (i.e., 94 cases). A flowchart describing the data processing is provided in **Figure 1a**. We first performed the univariate Cox analysis to determine the OS-related A-to-I sites in the training group after removing the A-to-I sites with undetermined editing level in $\geq 50\%$ cases. We then incorporated the promising A-to-I sites with P value < 0.001 in a least absolute shrinkage and selection operator (Lasso) regression model to select the most prognostic A-to-I sites with penalty parameter tuning conducted by 10-fold cross-validation [28]. Before the Lasso analysis, A-to-I sites with editing level less than 5% were removed because very low level of editing is difficult to detect accurately. Seven A-to-I sites were selected to consistent risk biosignature and the coefficient for each was obtained through the Lasso analysis. A risk score formula was established as follows:

$$\sum_i \text{Coefficient}(A - to - I \text{ site}_i) \times \text{editing level}(A - to - I)_i$$

Development and validation of the Nomogram

We developed an OS prognostic nomogram incorporating the above risk biosignature and baseline clinicopathological features including T, N stages, age at diagnosis and gender using the Cox regression model and the “rms” package in R. The nomogram was based on proportionally converting each Cox regression coefficients in multivariate logistic regression to a 0- to 100-point scale. The predictive performance of the nomogram was measured by Harrell's C-index and calibration with 100 bootstrap samples [29]. For validation of the nomogram, the total points of each patient in the validation set were calculated according to the established nomogram, then Cox regression in this cohort was performed using the total points as a factor, and finally, the Harrell's C-index and calibration were analyzed. Finally, a decision curve analysis determined the clinical usefulness of the nomogram, A-to-I biosignature and clinicopathological features by quantifying the net benefits at different threshold probabilities [30].

Correlation between A-to-I editing level and host gene expression

Since 4 of the 7 prognostic A-to-I sites are located at the 3'-untranslated region (3'-UTR) and 1 is located at the 5'-UTR of host genes, we queried the expression data of each host gene from the TCGA database and analyzed the correlations between the editing levels of the A-to-I sites and expressions of their host genes.

Statistical analysis

In addition to the aforementioned data analysis, the effect of risk score generated by the prognostic A-to-I sites on OS was analyzed using the log-rank test, univariate or multivariate Cox regression model, with estimation of hazard ratio (HR) and 95% confidence intervals (CI). A stratification analysis with regard to baseline clinicopathological features was also performed using the Cox model. A multiple interaction analysis between the risk score and clinicopathological features was conducted with the Cox model, too. The difference in editing levels of the prognostic A-to-I sites between the cancer tissues and normal tissues was tested by the student's *t* test, and among different stages were analyzed by the ANOVA test. The correlation between editing level of each A-to-I site and expression of host gene was assessed using Pearson correlation test, in which we removed the non-editing samples. All tests were two-sided and evaluated by the Stata software version 16.0 or R software version 4.0.1. *P* < 0.05 was considered to be statistically significant.

Results

Robustness of the novel biosignature based on seven A-to-I editing sites

Demographic and clinicopathological features of TCGA LUSC cases who owned A-to-I editing data among the training group and validation group are presented in **Table 1**. There was no significant difference in age, gender, smoking status, TNM stages and survival status between the training group and

validation group. The multivariable Cox analysis showed that increase of age, male, advanced T stages were significantly associated with OS in the 228 LUSC cases (**Figure S1**).

Table 1. Frequency distribution of demographic and clinicopathological features of LUSC cases.

Variables	Training set (n = 134)	Validation set (n = 94)	<i>P</i> value
Age (Mean $\pm$ SD, y)	68.3 $\pm$ 7.6	67.5 $\pm$ 9.0	0.445 ^a
Gender			
Male	101 (73.4%)	67 (71.3%)	0.489 ^b
Female	33 (24.6%)	27 (28.7%)	
Smoking status			
Yes	124 (92.5%)	89 (94.5%)	0.520 ^b
No	10 (7.5%)	5 (5.3%)	
T stages			
1	33 (24.6%)	18 (19.1%)	0.324 ^b
2	84 (62.7%)	58 (61.8%)	
3+4	17 (12.7%)	18 (19.1%)	
N stages			
0	91 (67.9%)	57 (60.6%)	0.515 ^b
1	30 (22.4%)	25 (26.6%)	
2+3	13 (9.7%)	12 (12.8%)	
M stages			
0	133 (99.2%)	91 (96.8%)	0.166 ^b
1	1 (0.8%)	3 (3.2%)	
Clinical stages			
I	72 (53.7%)	48 (51.1%)	0.248 ^b
II	38 (28.4%)	21 (22.3%)	
III+IV	24 (17.9%)	25 (26.6%)	
Survival status			
Dead	53 (39.6%)	39 (41.5%)	0.678 ^c
Alive	81 (60.4%)	55 (58.5%)	

^a *P* value calculated by the student's *t* test.

^b *P* value calculated by the two-side χ^2 test.

^c *P* value calculated by the log-rank test.

As shown in **Fig. 1b**, thirty-two A-to-I editing sites were identified with significant associations with OS of LUSC ($P < 0.001$) and were included in the Lasso regression model in the training set. After removing such A-to-I editing sites with very low editing level ($< 5\%$) and the Lasso analysis, there were finally 7 A-to-I sites to be included in the prognostic risk biosignature (**Fig. 1c**). They are TMEM120B_(A2588I), HMOX2_(A224I), CALCOCO2_(A2603I), MIR548AE2_(A113641I), ZNF440_(A3942I), CLCC1_(A2315I), CHMP3_(A23735I). The number in parentheses refers to the position, where the A-to-I editing occurs at the cDNA sequences of host gene, except for CHMP3_(A23735I) and MIR548AE2_(A113641I). The numbers of CHMP3_(A23735I) and MIR548AE2_(A113641I) refer to the position, where the A-to-I editing occurs at the DNA sequences, because the sites are located at introns. All HRs of these A-to-I sites are > 1 , indicating that high editing levels of them were associated with unfavorable OS. A risk score incorporating the risk effects of these A-to-I sites is calculated by: Risk score = $(11.58 \times \text{Editing level of HMOX2}_{(A224I)}) + (4.19 \times \text{Editing level of CALCOCO2}_{(A2603I)}) + (8.21 \times \text{Editing level of MIR548AE2}_{(A113641I)}) + (12.46 \times \text{Editing level of ZNF440}_{(A3942I)}) + (2.63 \times \text{Editing level of CLCC1}_{(A2315I)}) + (4.47 \times \text{Editing level of CHMP3}_{(A23735I)}) + (7.69 \times \text{Editing level of TMEM120B}_{(A2588I)})$. Taking the median risk score as the cutoff point, LUSC cases were divided into a high-risk group and a low-risk group. Distribution of the risk score based on the seven A-to-I sites, survival status, and editing levels of the seven A-to-I sites are shown in **Fig. 2A-D**. The low-risk group had almost quadrupled in a median survival time (MST, 98.0 months) when compared to the high-risk group (24.3 months, log-rank test: $P < 0.001$), and the latter exerted obviously increased death rate (multivariate Cox regression model: HR = 7.30; 95%CI = 3.48–15.3) in the training set. A consistent finding was observed in the validation set that the high-risk group exerted shorter OS time and higher death rate than the low-risk group (HR = 2.13; 95%CI = 1.09–4.15). In stratification analysis (**Fig. 2E**), we merged the training set and validation set to increase the study power. The hazard effect of high-risk was prominent in most sub-groups, except in patients with T3/4 stages, cases with lymphatic metastasis, and those in stage II, which may be due to the limited sample size. Stratification analyses with regard to smoking and distant metastases were omitted because of the extremely low sample size. Moreover, no significant interaction was observed between the risk biosignature and demographic and clinicopathological features.

Correlation between the signature and clinicopathological features

An analysis to test correlations between the risk score and clinicopathological features, including T, N stages, clinical stages, and age at diagnosis was performed. The risk score was significantly increased along with the T stages ($P = 0.021$, **Figure 2F**) but not the N stages ($P = 0.297$, **Figure 2G**). Being consistent, cases with advanced clinical stages exerted higher risk scores ($P = 0.002$, **Figure 2H**). However, no significant association was observed for the risk biosignature and diagnostic age (**Figure 2I**).

Predictive performance of the established nomogram

Based on the risk biosignature and clinicopathological features including T, N stages, gender and age at diagnosis, we developed a nomogram to predict OS of LUSC in the training set (**Figure 3A**) and validated the nomogram in the validation set. The calibration plots presented a superior agreement in both the training set and the validation set between the observed OS rate and nomogram-predicted OS rate of 1-, 3-, and 5-year (**Figure 3B, C**). Besides, the Harrell's C-indexes were 0.808 (95%CI = 0.770-0.845) in the training set and 0.685 (95%CI = 0.638-0.733) in the validation set, showing a noticeable value on predicting probability of LUSC OS. Moreover, the decision curve analysis showed that the nomogram-produced point exerted a higher net benefit than the clinicopathological features and the pure A-to-I biosignature in both the training set and the validation set at 1 and 3 years (**Figure 3D, E; Figure S2**).

Possible functions of the seven A-to-I sites

As a novel biosignature for cancer prognostic prediction, nothing is known about the functions of the seven A-to-I editing sites. We evaluated the differences in the editing levels of these A-to-I sites between cancer tissues and normal tissues in LUSC, and assessed correlations between editing levels of these sites and expressions of their host genes. We overlooked such effects for CHMP3_(A23735I) and MIR548AE2_(A113641I), since the editing sites are located at introns. As shown in **Figure 4A-E**, the editing levels of CLCC1_(A2315I) ($P = 0.004$) and CALCOCO2_(A2603I) ($P < 0.001$) but not ZNF440_(A3942I) ($P = 0.942$) and HMOX2_(A224I) ($P = 0.442$) were significantly different between cancer tissues and normal tissues. Besides, since the sample size of normal tissues were too little, the editing level of TMEM120B_(A2588I) displayed a barely detectable statistically significant difference ($P = 0.051$). Furthermore, as shown in **Figure 4F-J**, the correlation analysis revealed negative correlations between ZNF440_(A3942I) and ZNF440 ($r = -0.312$, $P < 0.001$), CLCC1_(A2315I) and CLCC1 ($r = -0.282$, $P < 0.001$), and HMOX2_(A224I) and HMOX2 ($r = -0.163$, $P = 0.073$). However, no obvious correlation was observed for TMEM120B_(A2588I) and TMEM120B ($r = -0.146$, $P = 0.232$), and CALCOCO2_(A2603I) and CALCOCO2 ($r = 0.050$, $P = 0.492$).

Discussion

To explore a prognostic nomogram of tumor is not only critical to know of the probability of the survival outcome but also the selection appropriate treatment options. Despite the progress made in lung cancer OS predictive models, most of which are constructed by gene expression, DNA methylation, radiographic, genome mutation or variant, new predictive rule based on novel biomarkers still need to be done to discover a more precise forecast model independently or by combining with previous nomograms. A-to-I editing has been widely implicated in cancer as reported recently, which produces a great potential for cancer prognostic prediction. Our study successfully developed an A-to-I editing-based nomogram for LUSC. The C-index, calibration plot and decision curve demonstrated a good predictive performance of this nomogram in both the training group and validation group. To the best of knowledge, this is the first study reporting a nomogram for cancer with regard to A-to-I editing.

Our study first identified OS-related A-to-I sites for LUSC according to transcriptome-wide A-to-I editing analysis based on the TCGA data, following by Lasso regression to determine an A-to-I biosignature including TMEM120B_(A2588I), HMOX2_(A224I), CALCOCO2_(A2603I), MIR548AE2_(A113641I), ZNF440_(A3942I), CLCC1_(A2315I), and CHMP3_(A23735I), which was significantly associated with OS of LUSC patients. Besides, this biosignature was positively correlated with T stages and clinical stages. With regard to the corresponding genes the seven A-to-I sites betiding, HMOX2 is a biomarker for tumor initiating cells of lung cancer and a potentially therapeutic target, suppression of which will significantly increase cancer survival[31]. CALCOCO2 is an autophagy receptor that contributes to autophagy addiction in K-Ras driven lung cancer[32, 33]. CLCC1 acts as a cell-surface biomarker for tumor initiating cells[34]. CHMP3 is a possible tumor suppressor with downregulated expression across a wide range of human cancers and its high level predicts a better survival outcome of breast cancer patients[35]. These evidences supported the functional underpinnings of the A-to-I risk biosignature. Nevertheless, the biological effects of TMEM120B, MIR548AE2 and ZNF440 on LUSC have not been studied.

A-to-I editing may lead to non-synonymous amino acid mutations, mis-regulation of alternative splicing, disturbance codon preference, and microRNA-mRNA redirection or RNA-binding protein-mRNA redirection, thereby influencing gene expression and function[36]. Among the seven A-to-I editing sites, TMEM120B_(A2588I), CALCOCO2_(A2603I), ZNF440_(A3942I) and CLCC1_(A2315I) are located in the 3'-UTR of their host genes, and HMOX2_(A224I) is located in the 5'-UTR of HMOX2, thus they have a great potential to affect expression of host genes. Indeed, the editing level of ZNF440_(A3942I) and CLCC1_(A2315I) were negatively correlated with the expressions of ZNF440 and CLCC1, respectively. HMOX2_(A224I) editing level might be also related to HMOX2 expression with a clear trend to be statistically significant. These results suggested that the three editing sites are functional. However, no association was observed between the expression and the editing level of other two genes (TMEM120B and CALCOCO2), suggesting an unconventional regulatory mechanism across them. In future, we will focus on verifying the biological functions of these A-to-I sites. Wonderingly, MIR548AE2_(A113641I) and CHMP3_(A23735I) are located in the intron of MIR548AE2 and CHMP3, respectively. We guessed they may belong to pre-mRNA sequences or non-coding RNA molecules that have not been identified by now.

Moreover, a difference in the editing level of TMEM120B_(A2588I), CLCC1_(A2315I), and CALCOCO2_(A2603I) was observed between tumor tissues and normal tissues, suggesting a role of these sites on affecting the occurrence of LUSC. Therefore, these sites may be alternative biomarkers for LUSC diagnosis.

To construct a prognostic model for LUSC, we developed a nomogram incorporating the seven A-to-I sites, age at diagnosis, gender, and TN stages. The nomogram exerted a medium accuracy on predicting OS of LUSC, which is similar with those were reported in other studies. Besides, the nomogram had better overall net benefit than the T, N staging system at 1 and 3 years. Although the nomogram showed well predictive performance in our study, to apply the nomogram to the "real-world" clinic, more LUSC cohorts, especially the prospective cohorts, are needed to assess the robustness. In addition, the current study have several limitations. First, we lacked an external group to validate the A-to-I biosignature and the

nomogram. Second, therapeutic schedule is important to show the application of prognostic nomogram for selection of appropriate treatment options, but the data is unavailable in the TCGA database. Finally, there may be some bias in the process of subjects' recruitment and data analysis.

In conclusion, we found the first risk prognostic biosignature based on A-to-I RNA editing in LUSC, which was associated with OS and clinical stages of LUSC patients. We further constructed a nomogram incorporating the A-to-I biosignature and clinicopathological features. The nomogram exerted well predictive performance and reliability for LUSC OS. Large prospective cohorts are warranted to validate the robustness of this model to assess the application value in "real-world" clinic.

Declarations

Ethics approval and consent to participate: NOT APPLICABLE.

Consent for publication: NOT APPLICABLE.

Availability of data and materials: ALL data were downloaded from the synapse website (<https://www.synapse.org/#!Synapse:syn4382524>), and the corresponding clinical information from the TCGA database (<https://portal.gdc.cancer.gov/>). All these data are publicly available.

Competing interests: The authors declare no competing interests.

Funding: This study was supported by the National Natural Science Foundation of China grants 82073628, 81871876, 81672303 (L. Yang). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Authors' contributions: LY conceived and designed the study with support from CLX, and ZXZ. LY wrote the first draft of the manuscript. All authors interpreted the results, revised and approved the manuscript for submission. LL and JL performed the data analysis. LT downloaded the TCGA data.

Acknowledgments: We thank Shizhen Chen and Liming Lu for helping with data checking.

References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, Bray F: **Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries.** *CA: a cancer journal for clinicians* 2021.
2. Cheng B, Xiong S, Li C, Liang H, Zhao Y, Li J, Shi J, Ou L, Chen Z, Liang P *et al*: **An annual review of the remarkable advances in lung cancer clinical research in 2019.** *Journal of thoracic disease* 2020, **12**(3):1056-1069.
3. Jiang W, Xu J, Liao Z, Li G, Zhang C, Feng Y: **Prognostic Signature for Lung Adenocarcinoma Patients Based on Cell-Cycle-Related Genes.** *Frontiers in cell and developmental biology* 2021,

4. Mu L, Ding K, Tu R, Yang W: **Identification of 4 immune cells and a 5-lncRNA risk signature with prognosis for early-stage lung adenocarcinoma.** *Journal of translational medicine* 2021, **19**(1):127.
5. Shi R, Bao X, Unger K, Sun J, Lu S, Manapov F, Wang X, Belka C, Li M: **Identification and validation of hypoxia-derived gene signatures to predict clinical outcomes and therapeutic responses in stage I lung adenocarcinoma patients.** *Theranostics* 2021, **11**(10):5061-5076.
6. Wang X, Xiao Z, Gong J, Liu Z, Zhang M, Zhang Z: **A prognostic nomogram for lung adenocarcinoma based on immune-infiltrating Treg-related genes: from bench to bedside.** *Translational lung cancer research* 2021, **10**(1):167-182.
7. Yang T, Hao L, Cui R, Liu H, Chen J, An J, Qi S, Li Z: **Identification of an immune prognostic 11-gene signature for lung adenocarcinoma.** *PeerJ* 2021, **9**:e10749.
8. Geng W, Lv Z, Fan J, Xu J, Mao K, Yin Z, Qing W, Jin Y: **Identification of the Prognostic Significance of Somatic Mutation-Derived LncRNA Signatures of Genomic Instability in Lung Adenocarcinoma.** *Frontiers in cell and developmental biology* 2021, **9**:657667.
9. Yang J, Xu T, Gomez DR, Yuan X, Nguyen QN, Jeter M, Song Y, Komaki R, Hu Y, Hahn SM *et al*: **Nomograms incorporating genetic variants in BMP/Smad4/Hamp pathway to predict disease outcomes after definitive radiotherapy for non-small cell lung cancer.** *Cancer medicine* 2018, **7**(6):2247-2255.
10. Zhang J, Luo L, Dong J, Liu M, Zhai D, Huang D, Ling L, Jia X, Luo K, Zheng G: **A prognostic 11-DNA methylation signature for lung squamous cell carcinoma.** *Journal of thoracic disease* 2020, **12**(5):2569-2582.
11. Ma X, Cheng J, Zhao P, Li L, Tao K, Chen H: **DNA methylation profiling to predict recurrence risk in stage Iota lung adenocarcinoma: Development and validation of a nomogram to clinical management.** *Journal of cellular and molecular medicine* 2020, **24**(13):7576-7589.
12. Zhang M, Sun L, Ru Y, Zhang S, Miao J, Guo P, Lv J, Guo F, Liu B: **A risk score system based on DNA methylation levels and a nomogram survival model for lung squamous cell carcinoma.** *International journal of molecular medicine* 2020, **46**(1):252-264.
13. Dong X, Zhang R, He J, Lai L, Alolga RN, Shen S, Zhu Y, You D, Lin L, Chen C *et al*: **Trans-omics biomarker model improves prognostic prediction accuracy for early-stage lung adenocarcinoma.** *Aging* 2019, **11**(16):6312-6335.
14. Christofi T, Zaravinos A: **RNA editing in the forefront of epitranscriptomics and human health.** *Journal of translational medicine* 2019, **17**(1):319.
15. Bazak L, Haviv A, Barak M, Jacob-Hirsch J, Deng P, Zhang R, Isaacs FJ, Rechavi G, Li JB, Eisenberg E *et al*: **A-to-I RNA editing occurs at over a hundred million genomic sites, located in a majority of human genes.** *Genome research* 2014, **24**(3):365-376.
16. Peng X, Xu X, Wang Y, Hawke DH, Yu S, Han L, Zhou Z, Mojumdar K, Jeong KJ, Labrie M *et al*: **A-to-I RNA Editing Contributes to Proteomic Diversity in Cancer.** *Cancer cell* 2018, **33**(5):817-828 e817.

17. Han L, Diao L, Yu S, Xu X, Li J, Zhang R, Yang Y, Werner HMJ, Eterovic AK, Yuan Y *et al*: **The Genomic Landscape and Clinical Relevance of A-to-I RNA Editing in Human Cancers.** *Cancer cell* 2015, **28**(4):515-528.
18. Gumireddy K, Li A, Kossenkova AV, Sakurai M, Yan J, Li Y, Xu H, Wang J, Zhang PJ, Zhang L *et al*: **The mRNA-edited form of GABRA3 suppresses GABRA3-mediated Akt activation and breast cancer metastasis.** *Nature communications* 2016, **7**:10715.
19. Shigeyasu K, Okugawa Y, Toden S, Miyoshi J, Toiyama Y, Nagasaka T, Takahashi N, Kusunoki M, Takayama T, Yamada Y *et al*: **AZIN1 RNA editing confers cancer stemness and enhances oncogenic potential in colorectal cancer.** *JCI insight* 2018, **3**(12).
20. Xu X, Wang Y, Mojumdar K, Zhou Z, Jeong KJ, Mangala LS, Yu S, Tsang YH, Rodriguez-Aguayo C, Lu Y *et al*: **A-to-I-edited miRNA-379-5p inhibits cancer cell proliferation through CD97-induced apoptosis.** *The Journal of clinical investigation* 2019, **129**(12):5343-5356.
21. Shen P, Yang T, Chen Q, Yuan H, Wu P, Cai B, Meng L, Huang X, Liu J, Zhang Y *et al*: **CircNEIL3 regulatory loop promotes pancreatic ductal adenocarcinoma progression via miRNA sponging and A-to-I RNA-editing.** *Molecular cancer* 2021, **20**(1):51.
22. Han J, An O, Hong H, Chan THM, Song Y, Shen H, Tang SJ, Lin JS, Ng VHE, Tay DJT *et al*: **Suppression of adenosine-to-inosine (A-to-I) RNA editome by death associated protein 3 (DAP3) promotes cancer progression.** *Science advances* 2020, **6**(25):eaba5136.
23. Fu L, Qin YR, Ming XY, Zuo XB, Diao YW, Zhang LY, Ai J, Liu BL, Huang TX, Cao TT *et al*: **RNA editing of SLC22A3 drives early tumor invasion and metastasis in familial esophageal cancer.** *Proceedings of the National Academy of Sciences of the United States of America* 2017, **114**(23):E4631-E4640.
24. Kurkowiak M, Arcimowicz L, Chrusciel E, Urban-Wojciuk Z, Papak I, Keegan L, O'Connell M, Kowalski J, Hupp T, Marek-Trzonkowska N: **The effects of RNA editing in cancer tissue at different stages in carcinogenesis.** *RNA biology* 2021:1-16.
25. Song Y, An O, Ren X, Chan THM, Tay DJT, Tang SJ, Han J, Hong H, Ng VHE, Ke X *et al*: **RNA editing mediates the functional switch of COPA in a novel mechanism of hepatocarcinogenesis.** *Journal of hepatology* 2021, **74**(1):135-147.
26. Ramirez-Moya J, Baker AR, Slack FJ, Santisteban P: **ADAR1-mediated RNA editing is a novel oncogenic process in thyroid cancer and regulates miR-200 activity.** *Oncogene* 2020, **39**(18):3738-3753.
27. Wood S, Willbanks A, Cheng JX: **The Role of RNA Modifications and RNA-modifying Proteins in Cancer Therapy and Drug Resistance.** *Current cancer drug targets* 2021.
28. Tibshirani R: **The lasso method for variable selection in the Cox model.** *Statistics in medicine* 1997, **16**(4):385-395.
29. Steyerberg EW, Vergouwe Y: **Towards better clinical prediction models: seven steps for development and an ABCD for validation.** *European heart journal* 2014, **35**(29):1925-1931.
30. Vickers AJ, Van Calster B, Steyerberg EW: **Net benefit approaches to the evaluation of prediction models, molecular markers, and diagnostic tests.** *Bmj* 2016, **352**:i6.

31. Kim JJ, Lee YA, Su D, Lee J, Park SJ, Kim B, Jane Lee JH, Liu X, Kim SS, Bae MA *et al*: **A Near-Infrared Probe Tracks and Treats Lung Tumor Initiating Cells by Targeting HMOX2**. *Journal of the American Chemical Society* 2019, **141**(37):14673-14686.
32. Liu KK, Qiu WR, Naveen Raj E, Liu HF, Huang HS, Lin YW, Chang CJ, Chen TH, Chen C, Chang HC *et al*: **Ubiquitin-coated nanodiamonds bind to autophagy receptors for entry into the selective autophagy pathway**. *Autophagy* 2017, **13**(1):187-200.
33. Newman AC, Scholefield CL, Kemp AJ, Newman M, McIver EG, Kamal A, Wilkinson S: **TBK1 kinase addiction in lung cancer cells is mediated via autophagy of Tax1bp1/Ndp52 and non-canonical NF-kappaB signalling**. *PloS one* 2012, **7**(11):e50672.
34. Ghosh D, Ulasov IV, Chen L, Harkins LE, Wallenborg K, Hothi P, Rostad S, Hood L, Cobbs CS: **TGFbeta-Responsive HMOX1 Expression Is Associated with Stemness and Invasion in Glioblastoma Multiforme**. *Stem cells* 2016, **34**(9):2276-2289.
35. Wang Z, Wang X: **miR-122-5p promotes aggression and epithelial-mesenchymal transition in triple-negative breast cancer by suppressing charged multivesicular body protein 3 through mitogen-activated protein kinase signaling**. *Journal of cellular physiology* 2020, **235**(3):2825-2835.
36. Eisenberg E: **Proteome Diversification by RNA Editing**. *Methods in molecular biology* 2021, **2181**:229-251.

Figures

Figure 1

Identification of an A-to-I RNA editing risk biosignature for OS of LUSC patients. A. The workflow of construction of a nomogram based on OS-related A-to-I sites and clinicopathological features. B. Scatter plot of P values in $-\log_{10}$ scale from the univariate Cox model analysis on transcriptome-wide A-to-I editing sites from TCGA. C. LASSO coefficient profiles of the A-to-I editing sites after removing OS-related sites with low editing level (<5%).

Figure 2

Association of the A-to-I risk biosignature with OS survival and clinical stages of LUSC patients. A-D. Distribution of the risk score based on the seven A-to-I sites, survival status, and editing level of each A-to-I site in the training set (A) and validation set (C), and Kaplan–Meier plot to visualize the survival probabilities for risk score in the training set (B) and validation set (D). P values were calculated by the log-rank test. E. Performance of the A-to-I biosignature on LUSC OS in different sub-groups with regard to

clinicopathological features. F-I. Correlation between the A-to-I signature and clinicopathological features including T stages (F), N stages (G), clinical stages (H), and age at diagnosis (I). P values were calculated by the ANOVA test.

Figure 3

Developed prognostic nomogram based on the A-to-I risk biosignature and clinicopathological features. A. The nomogram for predicting probabilities of patients with 1-, 3- and 5-year OS; B-C. Calibration curves

of the nomogram in the training group (B) and validation group (C). D-E. Decision curves of the nomogram in the training group (D) and validation group (E) at 1 year.

Figure 4

Editing levels of each A-to-I and their correlations with corresponding genes. A-E. The difference in editing level of each A-to-I site between tumor tissues and normal tissues of LUSC. P values were calculated by

the student's t test. F-J. The correlation between editing level of each A-to-I site and expression of corresponding gene. P values were assessed by the Pearson correlation test.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [FigureS1.jpeg](#)
- [FigureS2.jpg](#)