

# Validation of an Immune-Related Gene Pair Index as a Prognostic Marker of Lung Squamous Cell Carcinoma

**Zihao Wang**

Huazhong University of Science and Technology

**Xuan Xiang**

Huazhong University of Science and Technology

**Xiaoshan Wei**

Huazhong University of Science and Technology

**Linlin Ye**

Huazhong University of Science and Technology

**Yiran Niu**

Huazhong University of Science and Technology

**Wenbei Peng**

Huazhong University of Science and Technology

**Yu Li**

Huazhong University of Science and Technology

**Xu Wang**

Huazhong University of Science and Technology

**Siyu Zhang**

Huazhong University of Science and Technology

**Qiong Zhou** (✉ [zhouqiong@126.com](mailto:zhouqiong@126.com))

Huazhong University of Science and Technology <https://orcid.org/0000-0002-3120-3017>

---

## Primary research

**Keywords:** LUSC, Immune-related genes, Prognosis, Gene pairs, Immune infiltration

**Posted Date:** July 6th, 2020

**DOI:** <https://doi.org/10.21203/rs.3.rs-40047/v1>

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

[Read Full License](#)

# Abstract

**Background:** Lung squamous cell carcinoma (LUSC) is one of the subtypes of non-small-cell lung cancer (NSCLC) and accounts for approximately 20 to 30% of all lung cancers.

**Methods:** In this study, we developed an immune-related gene pair index (IRGPI) for LUSC from 8 public LUSC data sets, including The Cancer Genome Atlas LUSC cohort and Gene Expression Omnibus data sets, and explored the prognostic value of IRGPI for patients with LUSC.

**Results:** IRGPI was constructed by 13 gene pairs consisting of 25 unique immune-related genes from the training cohort. Multivariate cox regression analysis showed that high risk based on IRGPI was an independent risk factor for poor prognosis of patients with LUSC in the training cohort (230 patients; HR=3.40; 95%CI [2.34-4.94];  $p<0.001$ ), the testing cohort (228 patients; HR=2.11; 95%CI [1.48-3.01];  $p<0.001$ ) and the validation cohort (472 patients; HR=1.99; 95%CI [1.5-2.63];  $p<0.001$ ). The infiltrations of naïve B cells, plasma cells, CD8<sup>+</sup> T cells, activated memory CD4<sup>+</sup> T cells, gamma delta ( $\gamma\delta$ ) T cells, M1 macrophages, and activated dendritic cells were lower in the high-risk group, as compared with the low-risk group in the TCGA cohort. The infiltrations of neutrophils, activated mast cells, and monocytes were higher in the high-risk group.

**Conclusions:** IRGPI is a significant prognostic biomarker for predicting overall survival in LUSC patients. Combining clinical features with IRGPI will improve prognostic accuracy.

## Background

With the development of sequencing technology, more and more biomarkers were constructed for the diagnosis and prognosis of patients with lung cancer, including autoantibodies, complement fragments, miRNAs, circulating tumor DNA, DNA methylation, blood protein profiling, and other signatures. For example, Elena et al. reported a novel protein-based prognostic signature for risk stratification in patients with early lung adenocarcinoma [1], Ana et al. constructed a prognostic classifier based on mRNA, microRNA, and DNA methylation biomarkers to identify stage I lung adenocarcinoma patients at high risk of recurrence [2], and Daniel et al. described the diagnostic value of complement fragment C4d for patients with lung cancer [3].

As Guo et al. had reported, because of the instability of quantitative information of gene expression measurements under current technical conditions, it was more reasonable to construct quantitative transcriptional diagnostic signatures based on within-sample relative expression orderings of gene pairs [4]. Gene pairs signatures also demonstrated unique prognostic value in many cancers, such as non-squamous non-small cell lung cancer (NSCLC) [5], gastric cancer [6], and colorectal cancer [7].

Previous studies associated with prognosis of lung squamous cell carcinoma (LUSC) patients mainly concentrated on the alternative mRNA splicing signatures, lncRNA signatures, and other transcriptome signatures [8–13]. As the component of the tumor microenvironment, tumor-infiltrating immune cells

(TIICs) revealed promising value for providing prognostic biomarkers [14–16]. However, few studies focused on the immune-related gene pairs signatures of LUSC patients.

In the present study, we constructed an immune-related gene pair index (IRGPI) for LUSC and explored the prognostic value of IRGPI in multiple data sets. By combining the tumor stage with IRGPI, the prediction accuracy of IRGPI was improved.

## Materials And Methods

### Study design and data process

This retrospective study was conducted by using gene expression data from 8 public LUSC data sets, including 1 RNA-Seq data set and 7 microarray data sets [17–19]. Only patients without neoadjuvant therapy, adjuvant chemotherapy, or other immune-modulating therapy were included. LUSC RNA-Seq FPKM data and clinical information were downloaded from Genomic Data Commons Data Portal (GDC Data Portal) (<https://portal.gdc.cancer.gov/>). 7 GEO data sets (GSE37745, GSE50081, GSE30219, GSE19188, GSE14814, GSE41271, and GSE42127) with clinical information were download from GEO (<https://www.ncbi.nlm.nih.gov/geo/>). All the microarray data sets were merged into 1 data set and then randomly divided into a training cohort and a testing cohort. 472 LUSC patients with complete clinical information from TCGA were included as the independent validation cohort. The information about platforms was shown in **Supplementary Table 1**. For Affymetrix microarrays, gene expression profiles were normalized with the MAS5.0 method (affy, version 1.50.0) [20]. Model-Based Background Correction (MBCB) processed data were downloaded for Illumina datasets.

### Construction of IRGPI as a specific prognostic biomarker for LUSC

The IRGPI was constructed according to a published article [5]. An immune-related genes (IRGs) list consisting of cytokines, cytokine receptors, and genes related to variable immune cells function was constructed by using data downloaded from the ImmPort database (<https://import.niaid.nih.gov>) [21]. Only IRGs which were detected by all platforms were included. To form immune-related gene pairs (IRGPs), the included genes were pairwise compared. For one gene pair, the IRGP score was 0 if the gene expression level of IRG 1 was more than that of IRG 2; otherwise, the IRGP score was 1. Immune-related gene pairs (IRGPs) with constant values (0 or 1) in each data set were excluded considering the probable biases. The log-rank test was used to select the IRGPs which were associated with overall survival in all the cohorts. IRGP with a p-value of less than 0.05 was selected to construct the immune-related gene pair index (IRGPI) in the training cohort. Cox proportional hazards regression model with the least absolute shrinkage and selection operator (LASSO) was used to select the appropriate IRGPs. The optimal model parameter  $\lambda$  was evaluated by tenfold cross-validation at 1 SE (standard error) as previously recommended. The time-dependent receiver operating characteristic (ROC) curve was used to determine the optimal cutoff of IRGPI. The nearest neighbor estimation method was used to estimate the ROC curve. Patients were divided into low-risk and high-risk groups based on the cutoff of IRGPI.

## Validation of the prognostic value of IRGPI

To assess the prognostic value of IRGPI, univariate Cox regression analysis was conducted in the testing and validation cohorts. Age and tumor stage were coded as continuous variables. The tumor stage was coded as I=1, II=2, III=3, IV=4. The risk factors of gender and immune risk are male and high risk based on IRGPI. The prognostic accuracy of IRGPI was assessed by the time-dependent receiver operating characteristic (ROC) curve.

## Immune cells infiltration of low- and high- risk groups based on IRGPI

Tumor-infiltrating immune cells (TIICs) are important components of the tumor microenvironment (TME). The infiltrations of immune cells were assessed by CIBERSORT[22]. CIBERSORT is a computational method that can quantify immune cell fractions by using gene expression profiles (GEPs) from bulk tumor tissue. LUSC RNA-Seq gene expression profiles were downloaded from TCGA and transformed into the appropriate form according to the requirement. A total of 22 immune cell fractions were evaluated via the CIBERSORT website (<https://cibersort.stanford.edu/>). Wilcoxon rank-sum test was used to assess the difference of immune infiltrations between low- and high-risk groups.

## Statistical analysis

All statistical analyses were performed using R version 3.6.1 (<https://www.r-project.org/>). The log-rank test was used to select Immune-related Gene Pairs (IRGPs) associated with overall survival. LASSO Cox proportional hazards regression model was used to select ideal prognostic IRGPs for the construction of IRGPI by glmnet [23]. The time-dependent receiver operating characteristic (ROC) curve was applied to determine the optimal cutoff of IRGPI by survivalROC [24]. All the cohorts were separated into high-risk and low-risk groups by the same cutoff. The discrimination of the model was evaluated by Area Under the Curve (AUC) of the ROC curve. The Kaplan–Meier method was used to estimate survival curves. Multivariate Cox proportional hazards regression model was performed with variates significantly associated with overall survival in the univariate analysis by survival packages. Unless specified otherwise, p-value < 0.05 was considered statistically significant.

# Results

## Characteristics of patients

The analysis process was presented in the flow chart (Fig. 1). A total of 930 lung squamous cell cancer (LUSC) patients from TCGA and GEO data sets were enrolled in this study. 458 LUSC patients from GEO data sets were randomly divided into the training cohort and testing cohort. The demographic and clinical features of patients in the training, testing, and validation cohorts were listed in Table 1. There was no significant difference in the clinical features of the training and testing cohorts.

## Construction of the IRGPI

1065 immune-related genes (IRGs) were detected by the platforms enrolled in this study. 566580 immune-related gene pairs (IRGPs) were constructed and IRGPs with constant values (0 or 1) in each data set were excluded. The log-rank test was used to assess the association between the remaining IRGPs and the overall survival (OS) of patients. To avoid over-fitting of the prediction model, the LASSO regression was performed to screen the top 14 OS-related IRGPs (Figs. 2a, b). The final 13 IRGPs and LASSO coefficients were shown in Table 2. The 13 IRGPs which contained 25 unique immune-related genes were used to construct IRGPI via L1-penalized Cox proportional hazards regression in the training cohort. Based on the time-dependent ROC curve analysis, the optional cutoff for IRGPI was 1.31 (Supplementary Fig. 1). The risk curve and scatterplot were used to demonstrate the IRGPI and vital status of each patient with LUSC. Patients in the high-risk group had higher mortality than patients in the low-risk group (Figs. 2c, d). Survival curves of low- and high-risk groups were estimated by using the Kaplan–Meier method and were compared by using the log-rank test (Fig. 2e). In the training cohort, the AUC of time-dependent ROC curves at 1, 3, and 5 years was 0.744, 0.775, and 0.803, respectively (Fig. 2f).

### **Validation of IRGPI as an independent prognostic factor for LUSC**

The same formula was used to calculate the IRGPI of patients in the testing and validation cohorts. Risk curves and scatterplots were applied to show the IRGPI and vital status of each patient with LUSC in the testing and validation cohorts. Patients with high risk based on IRGPI had higher mortality than patients with low risk in the testing and validation cohorts (Figs. 3a-d). The IRGPI stratified patients with LUSC into different prognostic groups in the testing and training cohorts by using the Kaplan–Meier curves (Figs. 3e, g). In the testing cohort, the AUC of time-dependent ROC curves at 1, 3, and 5 years was 0.601, 0.648, and 0.691, respectively (Fig. 3f). In the validation cohort, the AUC of time-dependent ROC curves at 1, 3, and 5 years was 0.635, 0.698, and 0.677, respectively (Fig. 3h). Multivariate Cox proportional hazards regression model was performed with variates that were significantly associated with overall survival in univariate analysis (Table 3). Multivariate analyses showed that the hazard of death among the high-risk group was 3.4 times that of the low-risk group (HR= 3.40; 95%CI [2.34-4.94];  $p < 0.001$ ) in the training cohort. High risk based on IRGPI was an independent risk factor for poor prognosis of patients with LUSC in the testing cohort (HR=2.11; 95%CI [1.48-3.01];  $p < 0.001$ ) and validation cohort (HR=1.99; 95%CI [1.5-2.63];  $p < 0.001$ ).

### **Immune infiltration related to IRGPI**

CIBERSORT was used to assess the infiltrations of TIICs in the low- and high-risk groups based on IRGPI in the TCGA validation cohort (Fig. 4). The infiltrations of naïve B cells, plasma cells, CD8<sup>+</sup> T cells, activated memory CD4<sup>+</sup> T cells, gamma delta ( $\gamma\delta$ ) T cells, M1 macrophages and activated dendritic cells were lower in the high-risk group, as compared with the low-risk group ( $p=0.002$ ,  $p=0.004$ ,  $p=0.027$ ,  $p=0.024$ ,  $p=0.004$ ,  $p=0.008$ ,  $p=0.001$ , respectively). Patients in the high-risk group had higher proportions of neutrophils, activated mast cells, and monocytes, as compared with patients in the low-risk group ( $p < 0.001$ ;  $p=0.002$ ;  $p=0.043$ , respectively).

## Comparison of biomarkers for LUSC

We summarized current biomarkers for LUSC and compared the biomarkers in Table 4. Li et al. constructed a model with 6 lncRNAs from the TCGA LUSC cohort and the area under the curve (AUC) of the 6-lncRNA signature associated with 3-year survival was 0.672 in the training cohort [11]. Hu et al. constructed a 3-lncRNA signature for LUSC and the AUC of this model associated with 3-year survival was 0.629 in the training cohort [12]. Qi et al. identified 12 miRNAs closely related to the overall survival of patients with LUSC [13]. Yang et al. identified the diagnostic role of miRNA-486-5p in the TCGA LUSC cohort [25]. Li et al. found that 60 genes were statistically related to the overall survival rate in LUSC patients [10]. Shi et al. identified 6 methylation biomarkers for LUSC diagnosis [26]. Li et al. created prognostic predictors based on alternative splicing events for NSCLC patients and the AUC for prognostic predictor was over 0.8 in TCGA LUSC cohort [9]. Choi et al. found that MLL2 mutations predicted poor prognosis in both TP53 mutant and wild-type LUSC [8]. Gao et al. identified a prognostic model contained 5 genes and the AUC of the model for predicting the survival at 1, 3, and 5 years was 0.692, 0.722, and 0.651, respectively [27]. Compared with other prognostic models for LUSC, IRGPI showed a robust ability for predicting the overall survival of patients with LUSC despite using gene expression profiles from different sequencing platforms.

## Combination of IRGPI and clinical characteristics

Since the IRGPI and tumor stage were independent risk factors for poor prognosis of patients with LUSC in all the cohorts, IRGPI and tumor stage were combined to fit a Cox proportional hazards regression model in the training cohort. A new score immune clinical prognostic index (ICPI) was built ( $ICPI = 0.2709218 * IRGPI + 0.2888875 * stage$ ). The optimal cutoff of ICPI for stratifying patients was determined to be 1.05 based on time-dependent ROC curve analysis in the derivation data set (Supplementary Fig. 2). Compared with IRGPI in continuous form, the continuous form of ICPI improved the prognostic accuracy of overall survival in the testing cohort (C-index=0.63, 95%CI [0.58-0.68] vs C-index= 0.68, 95%CI [0.63-0.73];  $p=0.013$ ) and validation cohort (C-index=0.62, 95%CI [0.58-0.66] vs C-index= 0.66, 95%CI [0.62-0.70];  $p=0.023$ ) (Table 5).

## Discussion

Gene expression profiles had shown enormous potentials in predicting the survival of patients. Many biomarkers from gene signatures were constructed for the evaluation of prognosis in patients with non-squamous NSCLC, such as CCP score [28] and quantitative-PCR-based assay [29]. The cell-cycle progression genes (CCP score) were calculated from the normalized signatures of 31 cell-cycle genes and could identify patient groups of reduced or increased risk of death after surgical resection. The quantitative-PCR-based assay could identify patients with early-stage non-squamous NSCLC at high risk for mortality after surgical resection. For LUSC, gene signatures from tumor microenvironment could also be used to develop clinically useful tests to differentiate patients with different prognoses [30].

Current prognostic biomarkers for LUSC included prognostic alternative mRNA splicing signature, lncRNA signatures, and microRNA signatures [9, 11–13]. However, these studies [9, 11–13] only included LUSC data from the TCGA database. In this study, we built IRGPI based on 13 immune-related gene pairs from the training cohort and explored the prognostic value of IRGPI in the testing and validation cohorts. By combining multi-gene expression, it was possible to construct robust gene expression-based biomarkers to stratify patients with LUSC into low-risk and high-risk groups, even if gene expression profiles were from different sequencing platforms. Results showed that IRGPI could stratify LUSC patients into subgroups with different survival outcomes. Besides, we combined the IRGPI with the tumor stage to build a new score ICPI. Compared with IRGPI, ICPI showed higher prognostic accuracy in the testing and validation cohorts.

More and more evidence had shown the potential role of TIILs in the development, progression, and prognosis of cancers and the response of cancers to therapy [31]. According to the previous study, patients with lower levels of tumor-infiltrating effector T cells demonstrated a relatively worse prognosis [32]. Several studies had reported that tumors could escape immune system surveillance and effective antitumor immunity partly depended on cytotoxic T lymphocytes [33]. In NSCLC, many studies had demonstrated the association between the immune cell populations and prognosis of patients [34–37]. The immunotherapy with immune checkpoint inhibitors against NSCLC had shown remarkable clinical efficacy and the efficacy of immune checkpoint blockade depended on the infiltration of TIICs in the tumor tissues [38–40].

Many agents for NSCLC are not beneficial for the treatment of LUSC because LUSC lacked the specific genetic alterations that agents targeted [41]. In contrast, immune checkpoint inhibitor therapy had shown clinical efficacy in LUSC with prolonged survival [42, 43]. The immunotherapy for LUSC could be improved by prognostic biomarkers related to the tumor immune microenvironment. In this study, we found higher infiltrations of neutrophils and monocytes and lower infiltrations of CD8<sup>+</sup> T cells and CD4<sup>+</sup> T cells in the high-risk group based on IRGPI in the TCGA LUSC cohort. Patients in the high-risk group had lower levels of tumor-infiltrating CD8<sup>+</sup> T cells and displayed a relatively poor prognosis. Both the neutrophil and monocyte lineages had been demonstrated to possess lymphocyte suppressive capabilities [44]. The dysregulation of TIICs might explain the differences of prognosis between low- and high-risk groups based on IRGPI.

The method used to construct the IRGPI was based on the previous study [5]. Li et al. constructed the IRGPI for early-stage non-squamous NSCLC. However, the immune microenvironment differences between squamous and non-squamous NSCLC were significant [45–47]. To construct a robust signature for LUSC, several public LUSC data sets were included in the study. The expression levels of genes in each sample were compared in pairs to generate IRGPI. This gene pair-based approach was based on the relative gene expression of each sample and did not need normalization. There were also some limitations in our study. Although batch effects were reduced by the pairwise comparison of immune-related genes, some of them were unavoidable due to the combination of different platforms [48].

Meanwhile, sampling bias caused by intratumor genetic heterogeneity was inevitable for gene expression-based IRGPI.

## Conclusions

The immune-related gene pair index (IRGPI) is a prospective prognostic biomarker for LUSC, especially for early-stage LUSC. Further studies are still needed to validate its accuracy for assessing prognosis and to explore its clinical utility in the individualized management of LUSC.

## Abbreviations

ROC: Receiver Operating Characteristics; AUC: area under the curve; C-index: concordance index; TCGA: The Cancer Genome Atlas; GEO: gene expression omnibus; HR: hazard ratio; LASSO: least absolute shrinkage and selection operator; LUAD: lung adenocarcinoma; LUSC: lung squamous cell carcinoma; NSCLC: non-small cell lung cancer; OS: overall survival; IRGPI: immune-related gene pair index.

## Declarations

### Ethics approval and consent to participate

This study was approved by Ethics Committees of Tongji Medical College, Huazhong University of Science and Technology (2020-0120).

### Consent for publication

Not applicable.

### Availability of data and materials

All datasets enrolled in this study are available in the GEO (<https://www.ncbi.nlm.nih.gov/geo/>) and TCGA (<https://portal.gdc.cancer.gov/>) databases.

### Competing interests

The authors declare that they have no competing interests.

### Funding

This work was supported in part by grants from National Natural Science Foundation of China (No. 81973990, No. 81900096, No.81770090).

### Authors' contributions

QZ conceived the idea, designed and supervised the study, had full access to all data and took responsibility for the integrity of the data. YL, LLY, WBP, SYZ and YRN recorded and sorted the data. XSW and XW analyzed the data. ZHW and XX interpreted the data and wrote the manuscript. All authors contributed significantly to this work and agreed to be accountable for the work. All authors read and approved the final manuscript.

## Acknowledgements

We would like to thank GuoZiXueShengXin for the bioinformatic skills.

## References

1. Martínez-Terroba E, Behrens C, de Miguel FJ, Agorreta J, Monsó E, Millares L, Sainz C, Mesa-Guzman M, Pérez-Gracia JL, Lozano MD *et al*: **A novel protein-based prognostic signature improves risk stratification to guide clinical management in early-stage lung adenocarcinoma patients.** *The Journal of pathology* 2018, **245**(4):421-432.
2. Robles AI, Arai E, Mathé EA, Okayama H, Schetter AJ, Brown D, Petersen D, Bowman ED, Noro R, Welsh JA *et al*: **An Integrated Prognostic Classifier for Stage I Lung Adenocarcinoma Based on mRNA, microRNA, and DNA Methylation Biomarkers.** *Journal of thoracic oncology : official publication of the International Association for the Study of Lung Cancer* 2015, **10**(7):1037-1048.
3. Ajona D, Pajares MJ, Corrales L, Perez-Gracia JL, Agorreta J, Lozano MD, Torre W, Massion PP, de-Torres JP, Jantus-Lewintre E *et al*: **Investigation of complement activation product c4d as a diagnostic and prognostic biomarker for lung cancer.** *Journal of the National Cancer Institute* 2013, **105**(18):1385-1393.
4. Guan Q, Yan H, Chen Y, Zheng B, Cai H, He J, Song K, Guo Y, Ao L, Liu H *et al*: **Quantitative or qualitative transcriptional diagnostic signatures? A case study for colorectal cancer.** *BMC genomics* 2018, **19**(1):99.
5. Li B, Cui Y, Diehn M, Li R: **Development and Validation of an Individualized Immune Prognostic Signature in Early-Stage Nonsquamous Non-Small Cell Lung Cancer.** *JAMA oncology* 2017, **3**(11):1529-1537.
6. Peng PL, Zhou XY, Yi GD, Chen PF, Wang F, Dong WG: **Identification of a novel gene pairs signature in the prognosis of gastric cancer.** *Cancer medicine* 2018, **7**(2):344-350.
7. Wu J, Zhao Y, Zhang J, Wu Q, Wang W: **Development and validation of an immune-related gene pairs signature in colorectal cancer.** *Oncoimmunology* 2019, **8**(7):1596715.
8. Choi M, Kadara H, Zhang J, Parra ER, Rodriguez-Canales J, Gaffney SG, Zhao Z, Behrens C, Fujimoto J, Chow C *et al*: **Mutation profiles in early-stage lung squamous cell carcinoma with clinical follow-up and correlation with markers of immune function.** *Annals of oncology : official journal of the European Society for Medical Oncology* 2017, **28**(1):83-89.

9. Li Y, Sun N, Lu Z, Sun S, Huang J, Chen Z, He J: **Prognostic alternative mRNA splicing signature in non-small cell lung cancer.** *Cancer letters* 2017, **393**:40-51.
10. Li Y, Gu J, Xu F, Zhu Q, Ge D, Lu C: **Transcriptomic and functional network features of lung squamous cell carcinoma through integrative analysis of GEO and TCGA data.** *Scientific reports* 2018, **8**(1):15834.
11. Li R, Yang YE, Jin J, Zhang MY, Liu X, Liu XX, Yin YH, Qu YQ: **Identification of lncRNA biomarkers in lung squamous cell carcinoma using comprehensive analysis of lncRNA mediated ceRNA network.** *Artificial cells, nanomedicine, and biotechnology* 2019, **47**(1):3246-3258.
12. Hu J, Xu L, Shou T, Chen Q: **Systematic analysis identifies three-lncRNA signature as a potentially prognostic biomarker for lung squamous cell carcinoma using bioinformatics strategy.** *Translational lung cancer research* 2019, **8**(5):614-635.
13. Qi L, Gao C, Feng F, Zhang T, Yao Y, Wang X, Liu C, Li J, Li J, Sun C: **MicroRNAs associated with lung squamous cell carcinoma: New prognostic biomarkers and therapeutic targets.** *Journal of cellular biochemistry* 2019, **120**(11):18956-18966.
14. Coussens LM, Werb Z: **Inflammation and cancer.** *Nature* 2002, **420**(6917):860-867.
15. Zhang L, Conejo-Garcia JR, Katsaros D, Gimotty PA, Massobrio M, Regnani G, Makrigiannakis A, Gray H, Schlienger K, Liebman MN *et al*: **Intratumoral T cells, recurrence, and survival in epithelial ovarian cancer.** *The New England journal of medicine* 2003, **348**(3):203-213.
16. Angell H, Galon J: **From the immune contexture to the Immunoscore: the role of prognostic and predictive immune markers in cancer.** *Current opinion in immunology* 2013, **25**(2):261-267.
17. **Comprehensive genomic characterization of squamous cell lung cancers.** *Nature* 2012, **489**(7417):519-525.
18. Botling J, Edlund K, Lohr M, Hellwig B, Holmberg L, Lambe M, Berglund A, Ekman S, Bergqvist M, Ponten F *et al*: **Biomarker discovery in non-small cell lung cancer: integrating gene expression profiling, meta-analysis, and tissue microarray validation.** *Clinical cancer research : an official journal of the American Association for Cancer Research* 2013, **19**(1):194-204.
19. Sato M, Larsen JE, Lee W, Sun H, Shames DS, Dalvi MP, Ramirez RD, Tang H, DiMaio JM, Gao B *et al*: **Human lung epithelial cells progressed to malignancy through specific oncogenic manipulations.** *Molecular cancer research : MCR* 2013, **11**(6):638-650.
20. Gautier L, Cope L, Bolstad BM, Irizarry RA: **affy—analysis of Affymetrix GeneChip data at the probe level.** *Bioinformatics (Oxford, England)* 2004, **20**(3):307-315.
21. Bhattacharya S, Andorf S, Gomes L, Dunn P, Schaefer H, Pontius J, Berger P, Desborough V, Smith T, Campbell J *et al*: **ImmPort: disseminating data to the public for the future of immunology.** *Immunologic research* 2014, **58**(2-3):234-239.
22. Chen B, Khodadoust MS, Liu CL, Newman AM, Alizadeh AA: **Profiling Tumor Infiltrating Immune Cells with CIBERSORT.** *Methods in molecular biology (Clifton, NJ)* 2018, **1711**:243-259.
23. Simon N, Friedman J, Hastie T, Tibshirani R: **Regularization Paths for Cox's Proportional Hazards Model via Coordinate Descent.** *Journal of statistical software* 2011, **39**(5):1-13.

24. Saha-Chaudhuri P, Heagerty PJ: **Non-parametric estimation of a time-dependent predictive accuracy curve.** *Biostatistics (Oxford, England)* 2013, **14**(1):42-59.
25. Yang S, Sui J, Liu T, Wu W, Xu S, Yin L, Pu Y, Zhang X, Zhang Y, Shen B *et al*: **Expression of miR-486-5p and its significance in lung squamous cell carcinoma.** *Journal of cellular biochemistry* 2019, **120**(8):13912-13923.
26. Shi YX, Wang Y, Li X, Zhang W, Zhou HH, Yin JY, Liu ZQ: **Genome-wide DNA methylation profiling reveals novel epigenetic signatures in squamous cell lung cancer.** *BMC genomics* 2017, **18**(1):901.
27. Gao M, Kong W, Huang Z, Xie Z: **Identification of Key Genes Related to Lung Squamous Cell Carcinoma Using Bioinformatics Analysis.** *International journal of molecular sciences* 2020, **21**(8).
28. Wistuba, II, Behrens C, Lombardi F, Wagner S, Fujimoto J, Raso MG, Spaggiari L, Galetta D, Riley R, Hughes E *et al*: **Validation of a proliferation-based expression signature as prognostic marker in early stage lung adenocarcinoma.** *Clinical cancer research : an official journal of the American Association for Cancer Research* 2013, **19**(22):6261-6271.
29. Kratz JR, He J, Van Den Eeden SK, Zhu ZH, Gao W, Pham PT, Mulvihill MS, Ziaei F, Zhang H, Su B *et al*: **A practical molecular assay to predict survival in resected non-squamous, non-small-cell lung cancer: development and international validation studies.** *Lancet (London, England)* 2012, **379**(9818):823-832.
30. Gandara DR, Hammerman PS, Sos ML, Lara PN, Jr., Hirsch FR: **Squamous cell lung cancer: from tumor genomics to cancer therapeutics.** *Clinical cancer research : an official journal of the American Association for Cancer Research* 2015, **21**(10):2236-2243.
31. Fridman WH, Pagès F, Sautès-Fridman C, Galon J: **The immune contexture in human tumours: impact on clinical outcome.** *Nature reviews Cancer* 2012, **12**(4):298-306.
32. Schalper KA, Brown J, Carvajal-Hausdorf D, McLaughlin J, Velcheti V, Syrigos KN, Herbst RS, Rimm DL: **Objective measurement and clinical significance of TILs in non-small cell lung cancer.** *Journal of the National Cancer Institute* 2015, **107**(3).
33. Pardoll DM: **The blockade of immune checkpoints in cancer immunotherapy.** *Nature reviews Cancer* 2012, **12**(4):252-264.
34. Hiraoka K, Miyamoto M, Cho Y, Suzuoki M, Oshikiri T, Nakakubo Y, Itoh T, Ohbuchi T, Kondo S, Katoh H: **Concurrent infiltration by CD8+ T cells and CD4+ T cells is a favourable prognostic factor in non-small-cell lung carcinoma.** *British journal of cancer* 2006, **94**(2):275-280.
35. Al-Shibli KI, Donnem T, Al-Saad S, Persson M, Bremnes RM, Busund LT: **Prognostic effect of epithelial and stromal lymphocyte infiltration in non-small cell lung cancer.** *Clinical cancer research : an official journal of the American Association for Cancer Research* 2008, **14**(16):5220-5227.
36. Kawai O, Ishii G, Kubota K, Murata Y, Naito Y, Mizuno T, Aokage K, Saijo N, Nishiwaki Y, Gemma A *et al*: **Predominant infiltration of macrophages and CD8(+) T Cells in cancer nests is a significant predictor of survival in stage IV nonsmall cell lung cancer.** *Cancer* 2008, **113**(6):1387-1395.
37. Platonova S, Cherfils-Vicini J, Damotte D, Crozet L, Vieillard V, Validire P, André P, Dieu-Nosjean MC, Alifano M, Régnard JF *et al*: **Profound coordinated alterations of intratumoral NK cell phenotype and**

- function in lung carcinoma.** *Cancer research* 2011, **71**(16):5412-5422.
38. Ansell SM, Lesokhin AM, Borrello I, Halwani A, Scott EC, Gutierrez M, Schuster SJ, Millenson MM, Cattrly D, Freeman GJ *et al*: **PD-1 blockade with nivolumab in relapsed or refractory Hodgkin's lymphoma.** *The New England journal of medicine* 2015, **372**(4):311-319.
39. Herbst RS, Soria JC, Kowanetz M, Fine GD, Hamid O, Gordon MS, Sosman JA, McDermott DF, Powderly JD, Gettinger SN *et al*: **Predictive correlates of response to the anti-PD-L1 antibody MPDL3280A in cancer patients.** *Nature* 2014, **515**(7528):563-567.
40. Topalian SL, Hodi FS, Brahmer JR, Gettinger SN, Smith DC, McDermott DF, Powderly JD, Carvajal RD, Sosman JA, Atkins MB *et al*: **Safety, activity, and immune correlates of anti-PD-1 antibody in cancer.** *The New England journal of medicine* 2012, **366**(26):2443-2454.
41. Jiang J, Huang L, Liang X, Zhou X, Huang R, Chu Z, Zhan Q: **Gefitinib versus docetaxel in previously treated advanced non-small-cell lung cancer: a meta-analysis of randomized controlled trials.** *Acta oncologica (Stockholm, Sweden)* 2011, **50**(4):582-588.
42. Brahmer J, Reckamp KL, Baas P, Crinò L, Eberhardt WE, Poddubskaya E, Antonia S, Pluzanski A, Vokes EE, Holgado E *et al*: **Nivolumab versus Docetaxel in Advanced Squamous-Cell Non-Small-Cell Lung Cancer.** *The New England journal of medicine* 2015, **373**(2):123-135.
43. Horn L, Spigel DR, Vokes EE, Holgado E, Ready N, Steins M, Poddubskaya E, Borghaei H, Felip E, Paz-Ares L *et al*: **Nivolumab Versus Docetaxel in Previously Treated Patients With Advanced Non-Small-Cell Lung Cancer: Two-Year Outcomes From Two Randomized, Open-Label, Phase III Trials (CheckMate 017 and CheckMate 057).** *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 2017, **35**(35):3924-3933.
44. Gabilovich DI, Nagaraj S: **Myeloid-derived suppressor cells as regulators of the immune system.** *Nature reviews Immunology* 2009, **9**(3):162-174.
45. Meng X, Gao Y, Yang L, Jing H, Teng F, Huang Z, Xing L: **Immune Microenvironment Differences Between Squamous and Non-squamous Non-small-cell Lung Cancer and Their Influence on the Prognosis.** *Clinical lung cancer* 2019, **20**(1):48-58.
46. Kargl J, Busch SE, Yang GH, Kim KH, Hanke ML, Metz HE, Hubbard JJ, Lee SM, Madtes DK, McIntosh MW *et al*: **Neutrophils dominate the immune cell composition in non-small cell lung cancer.** *Nature communications* 2017, **8**:14381.
47. Faruki H, Mayhew GM, Serody JS, Hayes DN, Perou CM, Lai-Goldman M: **Lung Adenocarcinoma and Squamous Cell Carcinoma Gene Expression Subtypes Demonstrate Significant Differences in Tumor Immune Landscape.** *Journal of thoracic oncology : official publication of the International Association for the Study of Lung Cancer* 2017, **12**(6):943-953.
48. Leek JT, Scharpf RB, Bravo HC, Simcha D, Langmead B, Johnson WE, Geman D, Baggerly K, Irizarry RA: **Tackling the widespread and critical impact of batch effects in high-throughput data.** *Nature reviews Genetics* 2010, **11**(10):733-739.

## Tables

Table 1. Clinical characteristics of patients enrolled in this study

Variable	Training cohort	Testing cohort	P value*	Validation cohort
No. of patients	230	228	-	472
Age, median(IQR)	68(62-73)	65(60-71)	0.104	68(62-73)
Gender				
Female(%)	71(30.9)	74(32.5)	0.715	122(25.8)
Male(%)	159(69.1)	154(67.5)	-	350(74.2)
Stage				
I(%)	125(54.4)	120(52.7)	0.713	230(48.7)
II(%)	58(25.2)	60(26.3)	0.788	155(32.8)
III(%)	46(20)	47(20.6)	0.870	81(17.2)
IV(%)	1(0.4)	1(0.4)	1	6(1.3)
Overall survival, median(IQR)	1482(670-2275)	1398(612-2363)	0.420	696(366-1335)
No. of death(%)	119(51.7)	130(57.0)	0.061	199(42.2)
*Wilcoxon rank-sum test, Chi-square test ( $\chi^2$ test), Fisher's exact test, and log-rank test were used to compare the clinical features of the training and testing cohorts.				

Table 2. Immune-related genes (IRGs) used to construct IRGPI.

<b>IRG1</b>	<b>IRG2</b>	<b>IRGP</b>	<b>Lasso Coefficients</b>
ACVR1B	TNFRSF10C	ACVR1B - TNFRSF10C	0.566744196
ALB	VIPR1	ALB - VIPR1	0.393997022
BMP2	HLA-DPA1	BMP2 - HLA-DPA1	-0.174188484
CCL27	LEFTY1	CCL27 - LEFTY1	-0.568081528
CXCL12	TNFRSF4	CXCL12 - TNFRSF4	0.533501309
CYSLTR2	LHB	CYSLTR2 - LHB	-0.542300411
DEFB126	ESM1	DEFB126 - ESM1	0.342268191
DUOX2	SEMA3C	DUOX2 - SEMA3C	0.194021162
ECD	NUDT6	ECD - NUDT6	-0.252974437
EREG	GALR3	EREG - GALR3	-0.355428588
NPR1	PDGFA	NPR1 - PDGFA	-0.752795486
NTS	OPRD1	NTS - OPRD1	-0.473873924
NTS	TPT1	NTS - TPT1	0.529699366

Table 3. Univariate and multivariate analyses of IRGPI and clinical characteristics in the training, testing and validation data sets.

Cohorts	Variables <sup>1</sup>	Univariate			Multivariate		
		Harzard ratio	95%CI	P Value	Harzard ratio	95%CI	P Value
<b>Training</b>	risk	3.26	2.26 - 4.71	<0.001	3.40	2.34-4.94	<0.001
	gender	1.54	0.99 - 2.38	0.053	-	-	-
	age	1.04	1.01 - 1.06	<0.001	1.05	1.02-1.07	<0.001
	stage	1.45	1.16 - 1.81	<0.001	1.52	1.21-1.91	<0.001
<b>Testing</b>	risk	2.32	1.64 - 3.28	<0.001	2.11	1.48-3.01	<0.001
	gender	0.95	0.64 - 1.39	0.780	-	-	-
	age	1.01	0.99 - 1.03	<0.001	1.01	0.99-1.03	0.361
	stage	1.9	1.53 - 2.37	<0.001	1.82	1.46-2.27	<0.001
<b>Validation</b>	risk	1.92	1.45 - 2.54	<0.001	1.99	1.5-2.63	<0.001
	gender	1.24	0.88 - 1.72	0.213	-	-	-
	age	1.01	1.00 - 1.03	<0.001	1.02	1-1.04	0.035
	stage	1.32	1.11 - 1.57	<0.001	1.35	1.14-1.6	<0.001

<sup>1</sup>Age and tumor stage were coded as continuous variables. The tumor stage was coded as I=1, II=2, III=3, IV=4. The risk factors of gender and immune risk are male and high risk based on IRGPI.

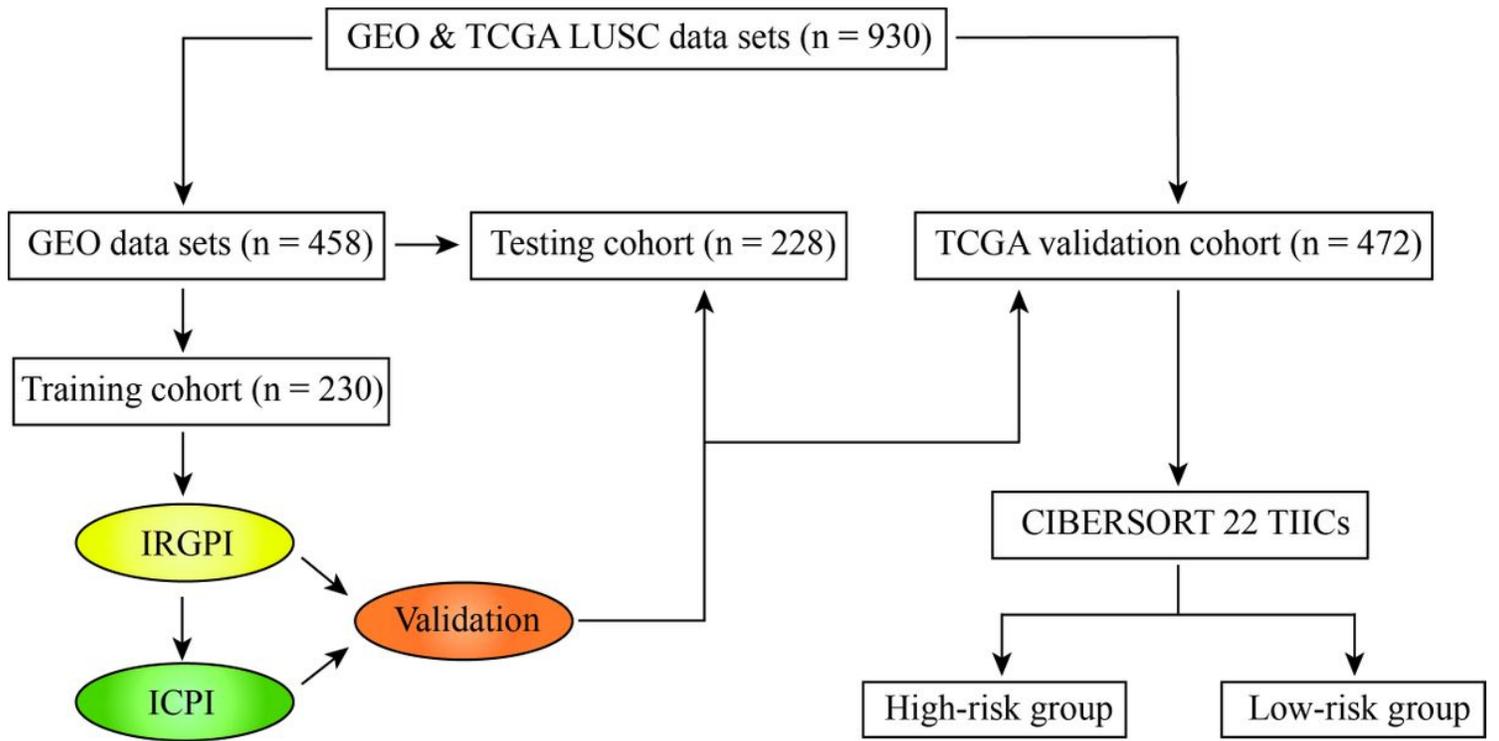
Table 4. Comparison of different biomarkers for LUSC.

Biomarkers	Signature	Datasets	AUC of derivation cohort	AUC of validation cohort	PMID
A model with 6 lncRNAs	LncRNA	TCGA	0.629	-	31364871
A model with 3 lncRNAs	LncRNA	TCGA	0.672	-	31737498
12 microRNAs	MicroRNA	TCGA	-	-	31241205
MiR-486-5p	MicroRNA	TCGA	0.908	-	30963622
60 Genes	Gene	TCGA GEO	-	-	30367091
Methylation biomarkers	Gene	TCGA	-	>0.9	29169318
Alternative mRNA splicing events	mRNA	TCGA	>0.8	-	28223168
Mutated genes	Mutation	-	-	-	28177435
A model contained 5 genes	Gene	TCGA	-	0.692/0.722/0.651	32340320

Table 5. Comparison of IRGPI and ICPI in the training, testing and validation cohorts.

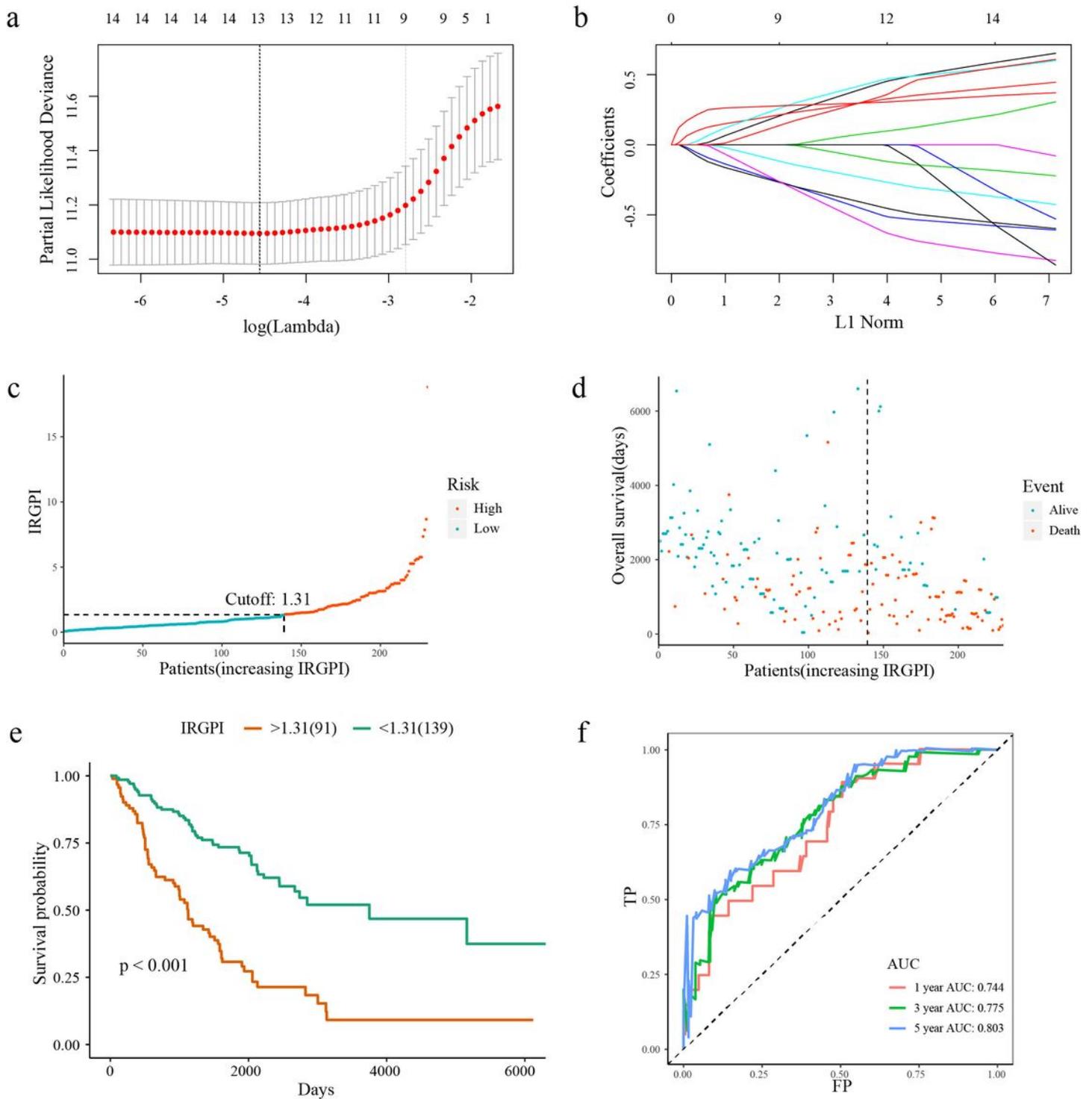
Cohorts	IRGPI		ICPI		P Value
	C-index	95%CI	C-index	95%CI	
<b>Training</b>	0.73	0.69-0.77	0.70	0.65-0.75	0.056
<b>Testing</b>	0.63	0.58-0.68	0.68	0.63-0.73	0.013
<b>Validation</b>	0.62	0.58-0.66	0.66	0.62-0.70	0.023

## Figures



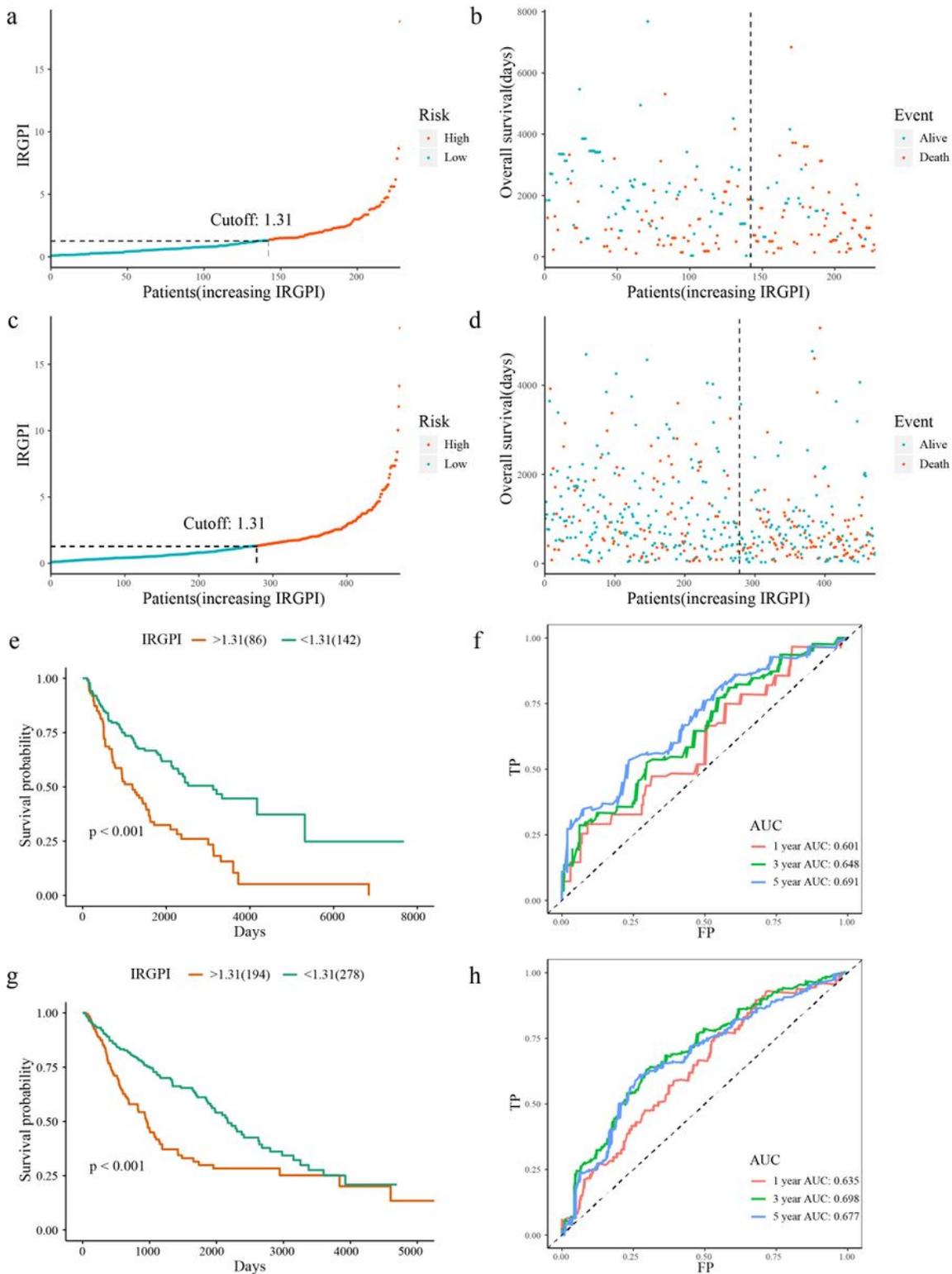
**Figure 1**

Flow chart of the research.



**Figure 2**

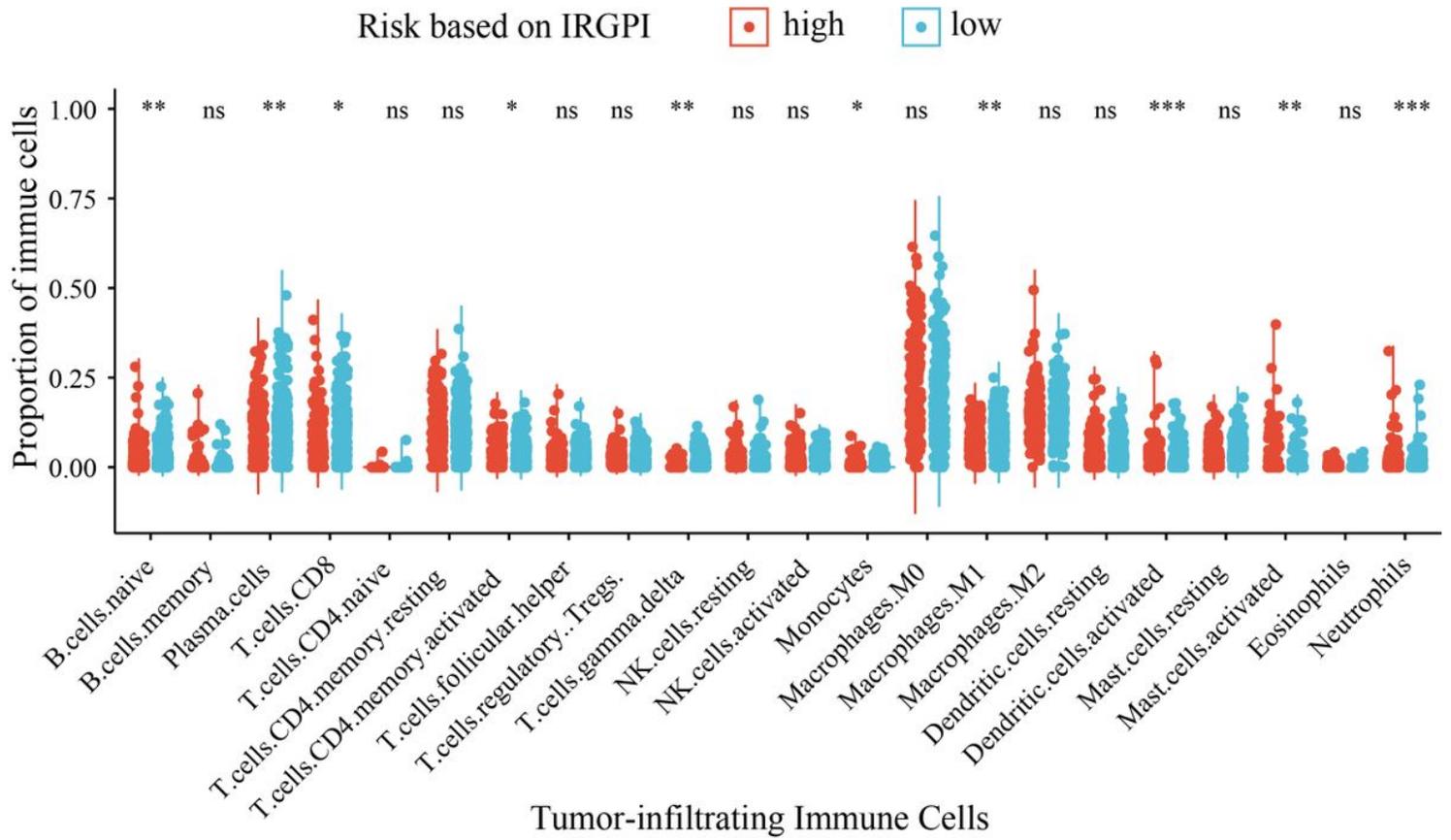
Construction of the IRGPI. (a); (b) The penalty parameter was estimated by 10-fold cross-validation in the training data set at 1 SE beyond the minimum partial likelihood deviance; (c); (d) Risk curve of IRGPI and scatterplot of vital status of each patient in the training cohort; (e) The survival curves of low- and high-risk groups. (f) The ROC curves and AUC of IRGPI in the training cohort.



**Figure 3**

Patients in the testing and validation cohorts were divided into low- and high-risk groups based on IRGPI. (a); (b) Risk curve of IRGPI and scatterplot of vital status of each patient in the testing cohort; (c); (d) Risk curve of IRGPI and scatterplot of vital status of each patient in the validation cohort; (e) The survival curves of low- and high-risk groups in the testing cohort; (f) The ROC curves and AUC of IRGPI in the

testing cohort; (g) The survival curves of low- and high-risk groups in the validation cohort; (h) The ROC curves and AUC of IRGPI in the validation cohort.



**Figure 4**

The infiltrations of 22 tumor-infiltrating immune cells (TIICs) between low-and high-risk groups based on IRGPI were compared by Wilcoxon rank-sum test.

## Supplementary Files

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

- [Supplement.pdf](#)