


# Identification of tumor immune infiltration-associated lncRNAs for improving prognosis and immunotherapy response of patients with non-small cell lung cancer

Jie Sun,<sup>1</sup> Zicheng Zhang,<sup>1</sup> Siqi Bao,<sup>1</sup> Congcong Yan,<sup>1</sup> Ping Hou,<sup>1</sup> Nan Wu,<sup>2</sup> Jianzhong Su,<sup>1</sup> Liangde Xu,<sup>1</sup> Meng Zhou <sup>1</sup>

**To cite:** Sun J, Zhang Z, Bao S, *et al.* Identification of tumor immune infiltration-associated lncRNAs for improving prognosis and immunotherapy response of patients with non-small cell lung cancer. *Journal for ImmunoTherapy of Cancer* 2020;**8**:e000110. doi:10.1136/jitc-2019-000110

► Additional material is published online only. To view please visit the journal online (<http://dx.doi.org/10.1136/jitc-2019-000110>).

JS, ZZ and SB contributed equally.

Accepted 28 December 2019



© Author(s) (or their employer(s)) 2020. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

For numbered affiliations see end of article.

## Correspondence to

Professor Meng Zhou, School of Biomedical Engineering, School of Ophthalmology & Optometry and Eye Hospital, Wenzhou Medical University, Wenzhou, China; [zhoumeng@wmu.edu.cn](mailto:zhoumeng@wmu.edu.cn)

Professor Liangde Xu, School of Biomedical Engineering, School of Ophthalmology & Optometry and Eye Hospital, Wenzhou Medical University, Wenzhou, China; [xuld@eye.ac.cn](mailto:xuld@eye.ac.cn)

## ABSTRACT

**Background** Increasing evidence has demonstrated the functional relevance of long non-coding RNAs (lncRNAs) to immunity regulation and the tumor microenvironment in non-small cell lung cancer (NSCLC). However, tumor immune infiltration-associated lncRNAs and their value in improving clinical outcomes and immunotherapy remain largely unexplored.

**Methods** We developed a computational approach to identify an lncRNA signature (TILSig) as an indicator of immune cell infiltration in patients with NSCLC through integrative analysis for **lncRNA, immune and clinical profiles of 115 immune cell lines, 187 NSCLC cell lines and 1533 patients with NSCLC**. Then the influence of the TILSig on the prognosis and immunotherapy in NSCLC was comprehensively investigated.

**Results** Computational immune and lncRNA profiling analysis identified an lncRNA signature (TILSig) consisting of seven lncRNAs associated with tumor immune infiltration. The TILSig significantly stratified patients into the immune-cold group and immune-hot group in both training and validation cohorts. These immune-hot patients exhibit significantly improved survival outcome and greater immune cell infiltration compared with immune-cold patients. Multivariate analysis revealed that the TILSig is an independent predictive factor after adjusting for other clinical factors. Further analysis accounting for TILSig and immune checkpoint gene revealed that the TILSig has a discriminatory power in patients with similar expression levels of immune checkpoint genes and significantly prolonged survival was observed for patients with low TILSig and low immune checkpoint gene expression implying a better response to immune checkpoint inhibitor (ICI) immunotherapy.

**Conclusions** Our finding demonstrated the importance and value of lncRNAs in evaluating the immune infiltrate of the tumor and highlighted the potential of lncRNA coupled with specific immune checkpoint factors as predictive biomarkers of ICI response to enable a more precise selection of patients.

## INTRODUCTION

Lung cancer is one of the most common cancers diagnosed in men and women

and accounts for one-quarter of all cancer deaths.<sup>1</sup> Non-small cell lung cancer (NSCLC) is the most frequent histological subtype and accounts for approximately 85% of all lung cancers.<sup>2,3</sup> NSCLC has the lowest 5-year relative survival rate with approximately 19% partly because of diagnosed at a distant stage and a paucity of late-stage treatments.<sup>1</sup> Therefore, continued research into molecular biomarkers and novel therapies is critical to predict prognosis and determine the personalized treatment.

The growing research on tumor microenvironment (TME) has indicated that tumor-infiltrating immune cells play a critical role in cancer progression and aggressiveness.<sup>4–6</sup> There is evidence that the microenvironment of NSCLC is rich in different types of immune cells which are associated with clinical outcomes.<sup>7,8</sup> Thus, the quantitative molecular signature of tumor-infiltrating immune cells is increasingly recognized as predictive biomarkers to enable personalized treatment selection and improve patient management. Long non-coding RNAs (lncRNAs), **a type of non-coding RNAs longer than 200 nucleotides in length, have been shown to be involved in a wide range of biological and cellular functions.**<sup>9–12</sup> Recent studies have demonstrated that lncRNAs are emerging as critical regulatory elements in the immune system and play important roles in the development and differentiation of different immune cell lineages.<sup>13–15</sup> A growing body of lncRNAs has been detected and identified in various immune cells.<sup>14,16</sup> For example, Hu *et al* has identified 1524 lncRNAs in T cell populations.<sup>17</sup> Another RNA-seq analysis of human lymphocyte subsets discovered more than 500 expressed lncRNAs and identified

several lymphocyte subset-specific lincRNA signatures.<sup>18</sup> A recent study also highlighted the functional relevance of lincRNAs in cancer immunity regulation and TME which contributes the progression and clinical outcomes of multiple cancers.<sup>19</sup> Whole transcriptome RNA sequencing in diverse immune cell types of patients with melanoma identified 27,625 lincRNAs, some of which are significantly associated with melanoma status.<sup>20</sup> Together these findings demonstrated the roles of lincRNAs in cancer immunology. Furthermore, the correlation between lincRNAs and immune cell infiltrate has also been observed in several cancers,<sup>21, 22</sup> which implied the potential of lincRNAs in evaluating the immune cell infiltrate of tumor.

In this study, we aimed to develop a novel computational frame for identifying tumor-infiltrating immune-related lincRNAs (TILncRNA) by integrative analysis for molecular profiling of purified immune cells, cancer cell lines and bulk cancer tissues and to explore their potential importance as predictive biomarkers for prognosis and immunotherapy in NSCLC tumors.

## MATERIALS AND METHODS

### Patient and tumor cell line cohorts

Clinical information and transcriptional profiles of patients with NSCLC were obtained from the Gene Expression Omnibus (GEO, <http://www.ncbi.nlm.nih.gov/geo>) and The Cancer Genome Atlas (TCGA, <https://portal.gdc.cancer.gov>) according to following selection criteria: (1) have basic clinical information of stage, age, gender, overall survival (OS) and survival status; (2) based on the Affymetrix HG-U133\_Plus 2.0 platform or Illumina-HiSeq platform; and (3) have larger sample size (>200). After filtering, a total of 1533 patients with NSCLC from three data sets were enrolled in this study, including 293 patients in GSE30219 from Rousseaux *et al's* study (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE30219>),<sup>23</sup> 226 patients in GSE31210 from Okayama *et al's* study (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE31210>)<sup>24</sup> and 1014 patients from TCGA (<https://portal.gdc.cancer.gov>). GSE30219 data set was used as a training data set for discovering lincRNA signature, and other two data sets (GSE31210 and TCGA) were used as independent testing data sets for validating the lincRNA signature. The detailed clinical information of three patient sets was shown in table 1. Transcriptional profiles of 187 NSCLC cell lines based on Affymetrix HG-U133\_Plus 2.0 platform were obtained from Cancer Cell Line Encyclopedia project (<https://portals.broadinstitute.org/ccle>).

### Purified immune cell data

Transcriptional profiles of 115 purified cell lines of 19 immune cell types based on Affymetrix HG-U133\_Plus 2.0 platform were obtained from GEO database including GSE13906, GSE23371, GSE25320, GSE27291, GSE27838, GSE28490, GSE28698, GSE28726,

**Table 1** Clinical characteristics of patients with NSCLC in each data set

| Characteristics | Training series |                | Testing series |
|-----------------|-----------------|----------------|----------------|
|                 | GSE30219        | GSE31210       | TCGA           |
| Platform        | HG-U133_Plus_2  | HG-U133_Plus_2 | IlluminaHiSeq  |
| Patients, n     | 293             | 226            | 1014           |
| >60             | 162             | 118            | 721            |
| ≤60             | 130             | 108            | 265            |
| NA              | 1               | 0              | 28             |
| Gender          |                 |                |                |
| Female          | 43              | 121            | 406            |
| Male            | 250             | 105            | 608            |
| Stage           |                 |                |                |
| I               | 152             | 168            | 518            |
| II              | 77              | 58             | 283            |
| III             | 56              | 0              | 168            |
| IV              | 7               | 0              | 33             |
| NA              | 1               | 0              | 12             |
| Survival        |                 |                |                |
| Dead            | 200             | 35             | 284            |
| Alive           | 93              | 191            | 730            |

NA, not applicable; NSCLC, non-small cell lung cancer; TCGA, The Cancer Genome Atlas.

GSE37750, GSE39889, GSE42058, GSE49910, GSE51540, GSE59237, GSE6863 and GSE8059. Marker genes of tumor-infiltrating immune cell types that are representative and are not expressed in cancer cells or in normal tissues were obtained from Charoentong *et al's* study<sup>25</sup> which analyzed 366 microarrays of immune cells collected from 37 studies.

### Preprocessing of transcriptional profiles

All raw data (.cel files) of microarray data sets profiled by Affymetrix HG-U133\_Plus 2.0 platform were downloaded from the GEO database and were processed using the Robust Multi-array Average algorithm with R 'affy' packages from Bioconductor for background correction, quantile normalization, and log2 transformation.<sup>26</sup> lincRNA expression profiles were obtained by reannotating the probes from the Affymetrix HG-U133\_Plus 2.0 microarray data sets. By matching the NetAffx annotation files (HG-U133 Plus 2.0 Annotations, CSV format, release 36, July 12, 2016) of the probe sets and the annotation files of Refseq (release 79) and GENCODE (release 25), we extracted probe sets with Refseq IDs which were labeled as 'NR\_' and annotated with 'long non-coding RNA' in Refseq database and with Ensembl gene IDs which were annotated as 'long non-coding RNA' in GENCODE project. Finally, we obtained 2466 unique lincRNAs corresponding to 3431 probe sets in microarray data sets for further analysis. Sequencing lincRNA expression profiles (IlluminaHiSeq platform) were obtained from TCGA (<https://portal.gdc.cancer.gov>).

## Development of tumor-infiltrating immune-related lncRNA signature

We developed a novel computational frame for identifying tumor-infiltrating immune-related lncRNA signature (TILSig) by integrative immune and lncRNA profiling analysis of purified immune cells, cancer cell lines and bulk cancer tissues as follows (figure 1): (1) For each immune cell line, top 5% expressed lncRNAs were obtained as candidate immune-related lncRNAs. (2) The expression specificity of a candidate immune-related lncRNA with respect to different immune cell types was calculated using tissue specificity index (TSI) developed by Yanai *et al.*<sup>27</sup> as follows:

$$TSI_{lnc} = \frac{\sum_{i=1}^N (1 - x_{lnc,i})}{N-1}$$

where  $N$  is the total number of immune cell types and  $x_{lnc,i}$  is the expression intensity of immune cell  $i$  normalized by the maximal expression of any immune cell types for lncRNA. The TSI ranges from 0 to 1. The lncRNA is a general immune lncRNA when the value is 0, while the lncRNA is one immune cell-specific lncRNA when the value is 1. Those lncRNAs which are universally highly expressed in all immune cell types were defined as immune-related housekeeping lncRNAs (hklncRNA). (3) Those hklncRNAs which are upregulated in immune cell lines and downregulated in NSCLC cell lines were selected as TILncRNAs. (4) A prognostic signature by focusing TILncRNAs (TILSig) was constructed using the linear combination of the expression values of the prognostic TILncRNAs, weighted by their estimated regression coefficients in the multivariate Cox regression analysis.<sup>2,28</sup>

## Statistical analysis

Differentially expressed lncRNAs between immune cell lines and lung cancer cell lines were determined using significance analysis of microarrays method with R packages 'samr'. The prognostic value of NSCLC TILncRNAs was evaluated by univariate Cox proportional hazards regression analysis in the training data set. The Kaplan-Meier survival plots were used to estimate OS and disease-free survival (DFS), and the difference in OS or DFS between the high-risk group and the low-risk group was determined using log-rank tests with R package 'survival'. Univariate and multivariate analyses with Cox proportional hazards regression for OS and DFS were performed on the individual clinical variables with and without TILSig in each data set. HRs and 95% CIs were calculated. The prognostic performance of TILSig was measured using time-dependent receiver operating characteristic (ROC) curves with R packages 'survivalROC'. The predictive performance of TILSig for recurrence status was measured using ROC curves. The Kruskal-Wallis test was used for comparisons among multiple groups. All the statistical analyses were performed in R V.3.1.3 with additional Bioconductor packages

## Gene set enrichment analysis

To evaluate the infiltration of immune cells, gene set enrichment analysis (GSEA) was performed using marker gene set of each immune cell with Bioconductor package 'clusterProfiler',<sup>29</sup> which can identify immune cell types that are over-represented in high-risk group or low-risk group.<sup>25</sup>

## RESULTS

### The landscape of lncRNA expression in human immune cells

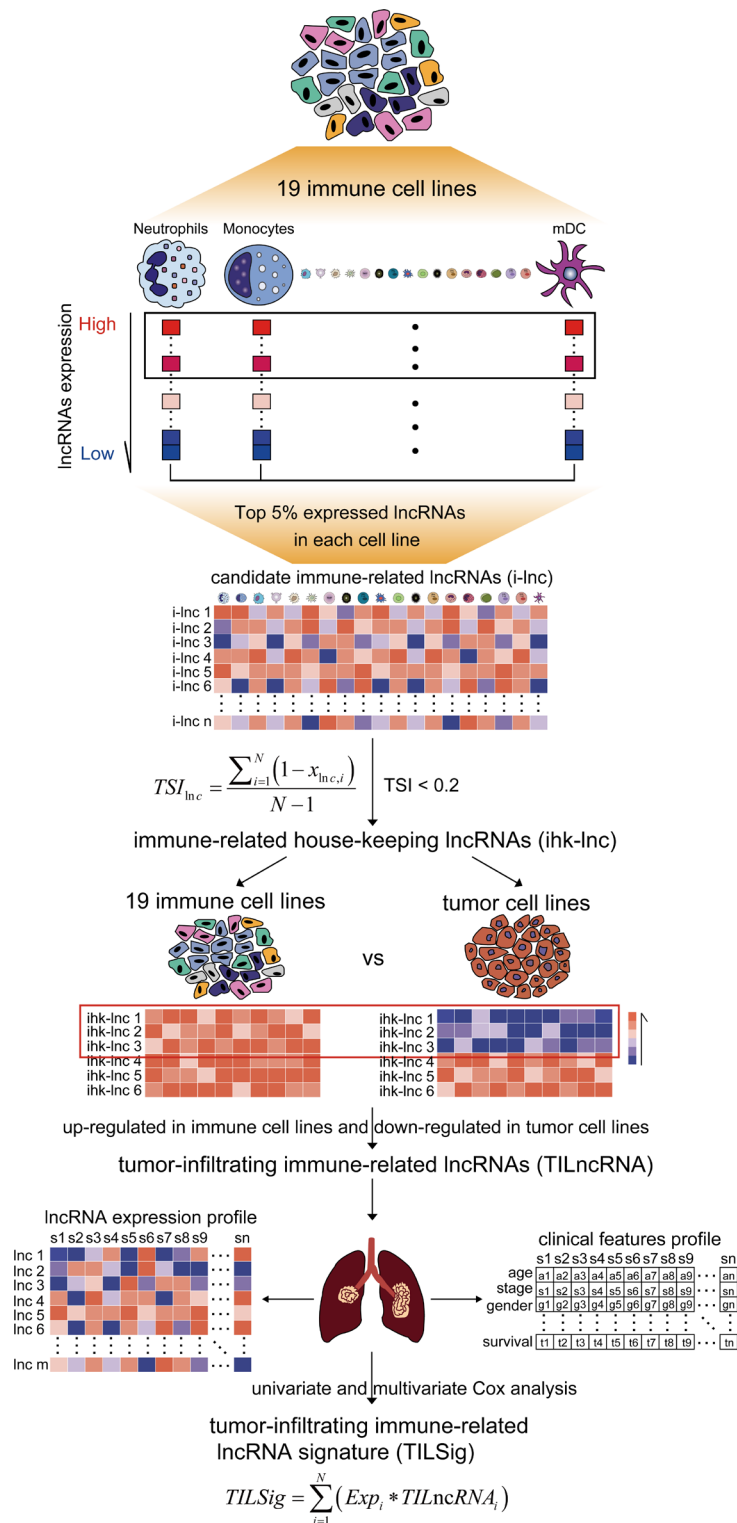
To characterize the expression pattern of lncRNAs in various human immune cells, all lncRNAs for each immune type were first ranked based on their expression levels, and those RNAs whose expression levels were ranked top 5% were considered as candidate immune-related lncRNAs. Of them, 91 lncRNAs were commonly highly expressed in all 19 immune cell types and 117 lncRNAs were highly expressed in exclusively one type of 19 immune cell types (online supplementary table S1 and online supplementary figure S1). Then we calculated the TSI score for each of those 208 candidate immune-related lncRNAs to measure the expression specificity with respect to different immune cell types. Finally, we identified nine immune cell type-specific lncRNAs (icslncRNA) which were expressed in only one type of immune cell and had a high score of cell-type specificity ( $>0.5$ ) (online supplementary table S2 and online supplementary figure S2). Of them, three icslncRNAs (*LOC101926943*, *AP003774.1* and *LINC00996*) were found to be plasmacytoid dendritic cell specific and three icslncRNAs (*LINC01234*, *TLR8-AS1* and *LINC01296*) were significantly higher expressed in mast cells. For natural killer T cells and dendritic cells, we detected two specific icslncRNAs (*LINC00892* and *LINC00515*) and one icslncRNA (*LINC00158*), respectively. In contrast, we also identified 57 hklncRNAs which are highly expressed in all immune cell types and have a lower score of cell-type specificity ( $<0.2$ ), demonstrating their essential for basic immune cellular functions (online supplementary table S3).

### Identification of NSCLC TILncRNAs

To identify NSCLC TILncRNAs, we only focused on these 57 hklncRNAs which are universally highly expressed in all immune cell types and are critical for the maintenance of basic immune cellular functions. We performed differential expression analysis of 57 hklncRNAs between immune cell lines and NSCLC cell lines, and identified 17 hklncRNAs which are upregulated in immune cell lines and downregulated in NSCLC cell lines demonstrating their expression specificity to immune cells rather than tumor cells. These 17 hklncRNAs were considered as NSCLC TILncRNAs.

### Development of TILSig

We performed univariate Cox proportional hazards regression analysis to investigate the association between expressions of 17 TILncRNAs and patients' OS time in

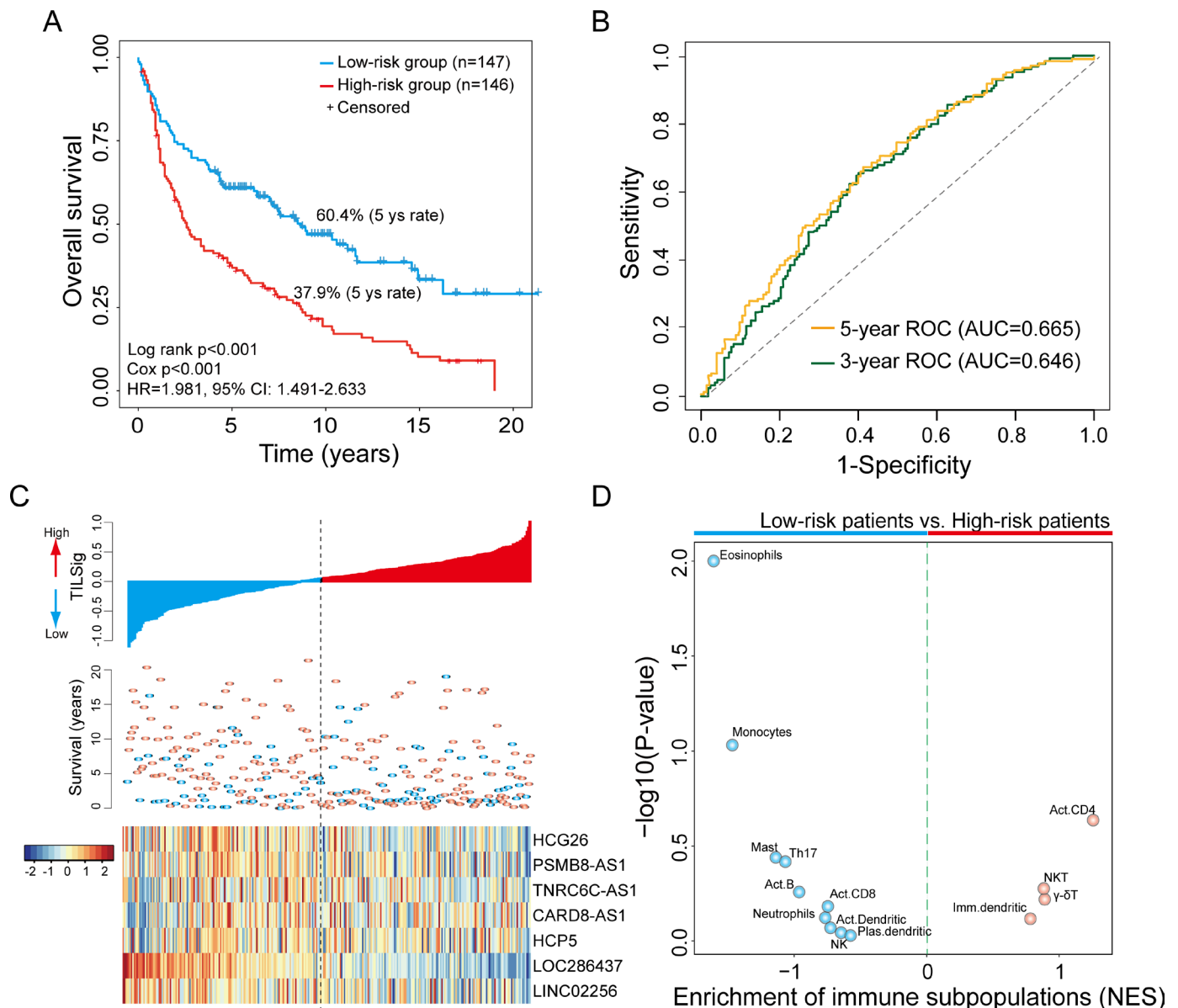


**Figure 1** Strategy for identifying tumor-infiltrating immune-related lncRNA signature (TILSig) in this study. Top 5% expressed long non-coding RNAs (lncRNAs) were obtained as candidate immune-related lncRNAs for each immune cell line. The specificity of expression of a candidate immune-related lncRNA with respect to different immune cell types was calculated using tissue specificity index (TSI). Those housekeeping lncRNAs (hk-lncRNAs) which are upregulated in immune cell lines and downregulated in non-small cell lung cancer (NSCLC) cell lines were selected as tumor-infiltrating immune-related lncRNAs. A prognostic signature by focusing tumor-infiltrating immune-related lncRNAs (TILSig) was constructed using the linear combination of the expression values of the prognostic tumor-infiltrating immune-related lncRNAs, weighted by their estimated regression coefficients in the multivariate Cox regression analysis.



the training data set. A set of seven TILncRNAs (*HCG26*, *PSMB8-AS1*, *TNRC6C-AS1*, *CARD8-AS1*, *HCP5*, *LOC286437* and *LINC02256*) was significantly correlated with NSCLC patients' OS and therefore was considered as prognostic immune lncRNAs. Then we develop an lncRNA-based prognostic signature indicative of immune infiltration (TILSig) using the expression of seven TILncRNAs weighted by the multivariate Cox regression coefficient as follows:  $TILSig = (-0.1323 \times \text{expression value of } HCG26) + (-0.2323 \times \text{expression value of } PSMB8-AS1) + (0.0009 \times \text{expression value of } TNRC6C-AS1) + (0.2472 \times \text{expression value of } CARD8-AS1) + (0.1190 \times \text{expression value of } HCP5) + (-0.3019 \times \text{expression value of } LOC286437) + (-0.0572 \times \text{expression value of } LINC02256)$ .

When TILSig was applied to the training data set, 293 patients were classified into the high-risk group (n=146) and low-risk group (n=147) using the median value (0.0429) of TILSig as risk cut-off. As shown in figure 2A, patients in the low-risk group had significantly longer OS time than those in the high-risk group (HR=1.981, 95% CI 1.491 to 2.633, log-rank p<0.001) (figure 2A). The 5-year survival rate of the low-risk group was 60.4% which is significantly higher than that of the high-risk group (37.9%). The area under the curve (AUC) of the TILSig



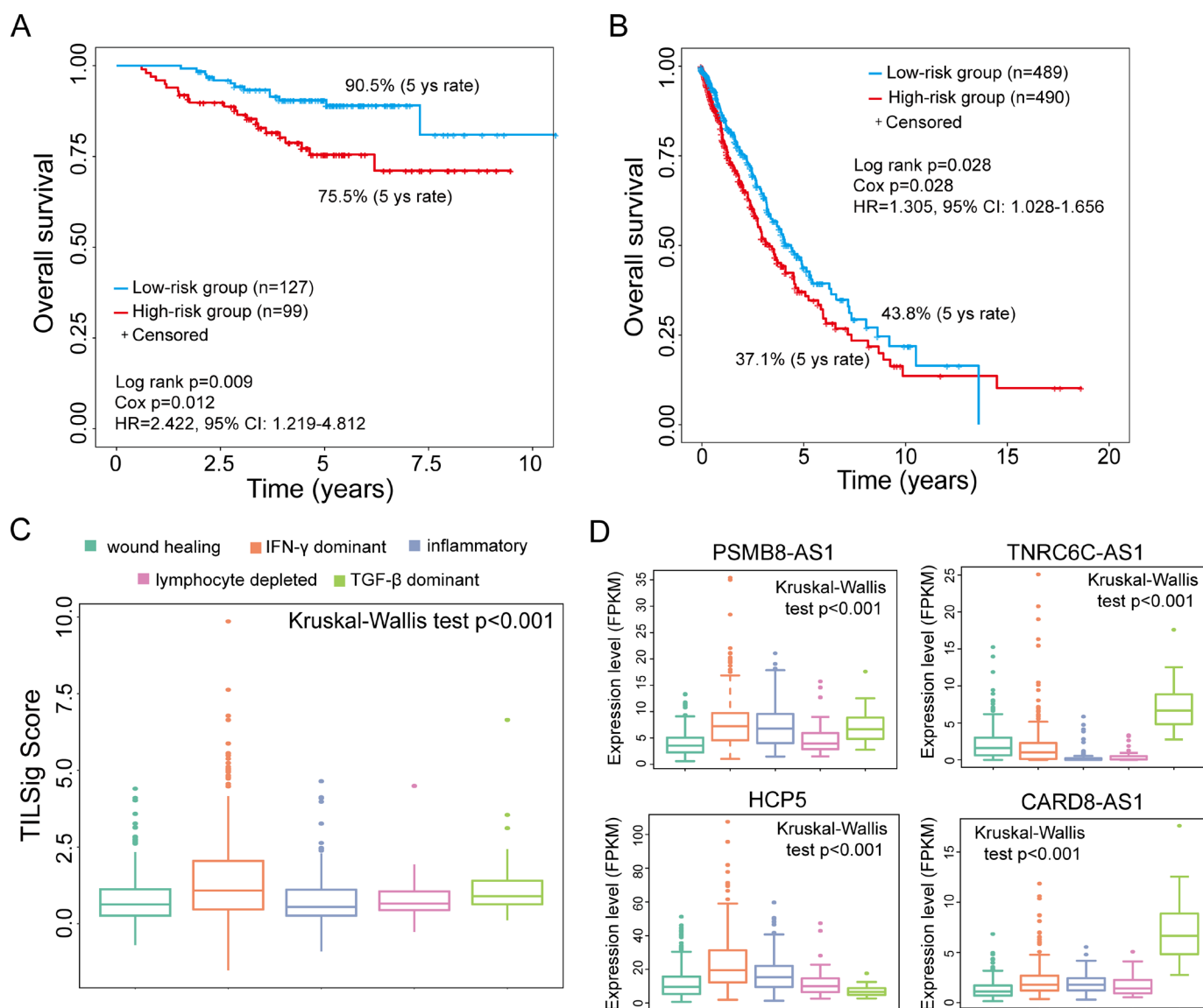
**Figure 2** The tumor-infiltrating immune-related lncRNA signature (TILSig) is associated with outcome and immune cell infiltrates. (A) Kaplan-Meier survival curves of overall survival between patients with a higher score of TILSig and with the lower score of TILSig. (B) Time-dependent receiver operating characteristic (ROC) curve at 5 and 3 years of overall survival (OS). (C) The distribution of TILSig, patients' survival status and long non-coding RNA (lncRNA) expression pattern. (D) Volcano plots for the enrichment of immune cell types for tumors with high TILSig and low TILSig calculated based on the NES score from the gene set enrichment analysis (GSEA). AUC, area under the curve; NES, normalized enrichment scores; NK, natural killer; NKT, natural killer T cell.

was 0.665 and 0.646 at 5 and 3 years of OS (figure 2B). The distribution of TILSig and expression pattern of lncRNAs in the TILSig was revealed in figure 2C. To further examine whether the TILSig is highly reflective of immune infiltrates, we estimated the infiltration of 19 immune subpopulations for patients of high-risk and low-risk groups using single-sample GSEA analysis as previously described.<sup>25</sup> We found that patient risk groups stratified by TILSig showed distinct immune infiltrate patterns. As shown in figure 2D, patients in the low-risk group were enriched with 10 immune subpopulations, while only four immune subpopulations were enriched in patients with high risk. These results suggested that the higher score of TILSig corresponded to less immune cell infiltration and poor outcome, the lower score of TILSig

corresponded to greater immune cell infiltration and better outcome.

### Confirmation of the TILSig in two independent data sets with different platform

To evaluate the robustness of the TILSig, we test its prognostic power using lncRNA expression and clinical data from an independent cohort of 226 patients profiled by microarray platform. The TILSig could stratify patients into low risk (n=127) and high risk (n=99) with significantly different OS (HR=2.422, 95% CI 1.219 to 4.812, log-rank p=0.009) using the same risk score derived from the training data set (figure 3A). The 5-year survival rate of the low-risk group is 90.5%, whereas the corresponding rate in the high-risk group is 75.5%.



**Figure 3** Validation of the tumor-infiltrating immune-related lncRNA signature (TILSig) in independent cohorts. Kaplan-Meier survival curves of overall survival between patients with a higher score of TILSig and with the lower score of TILSig in the GSE31210 data set (A) and The Cancer Genome Atlas (TCGA) data set (B). (C) Distribution and comparison of the TILSig among five immune subtypes. (D) Expression pattern of four tumor-infiltrating immune-related lncRNAs (TILncRNAs) among five immune subtypes. FPKM, fragments per kilobase of exon model per million reads mapped; IFN- $\gamma$ , interferon gamma; TGF- $\beta$ , transforming growth factor beta.

The prognostic value of the TILSig was further tested using another completely independent TCGA cohort of 979 patients profiled by the IlluminaHiSeq platform. Due to different platforms, only four of seven lncRNAs in the TILSig were covered by the IlluminaHiSeq platform. Therefore, the TILSig only based on four lncRNAs (*PSMB8-AS1*, *TNRC6C-AS1*, *HCP5* and *CARD8-AS1*) without re-estimating parameters was applied to TCGA data set. The median risk score cut-off point obtained from TCGA data set classified 979 patients into the high-risk group (n=490) and low-risk group (n=489). Survival analysis demonstrated that the high and low-risk groups are significantly different in OS (HR=1.305, 95% CI 1.028 to 1.656, log-rank p=0.028) (figure 3B). The 5-year survival rate of the low-risk group is 43.8%, whereas the corresponding rate in the high-risk group is 37.1%. We further examined the distribution of the TILSig among five immune subtypes reported by a recent study<sup>30</sup> and found that there is significant difference in TILSig among five immune subtypes (Kruskal-Wallis test p<0.001) (figure 3C), suggesting that the TILSig is closely associated with tumor immune microenvironment. Furthermore, further analysis revealed that expression levels of four TILncRNAs in the TILSig are also significantly different among five immune subtypes (figure 3D). As shown in figure 3D, lncRNAs *TNRC6C-AS1* and *CARD8-AS1* tended to be overexpressed in transforming growth factor beta (TGF- $\beta$ ) dominant immune subtype than other four immune subtypes, while lncRNA *HCP5* has higher expression levels in interferon gamma (IFN- $\gamma$ ) dominant compared with other four immune subtypes. lncRNA *PSMB8-AS1* tended to be overexpressed in three of five immune subtypes, including IFN- $\gamma$  dominant,

inflammatory and TGF- $\beta$  dominant. We further examine the association of the TILSig with OS in other 17 solid cancers using univariate Cox analysis. **Pan-cancer analysis of the TILSig revealed significant association between the TILSig and OS in other two cancers: colon adenocarcinoma (COAD) and kidney renal clear cell carcinoma (KIRC) (online supplementary figure S3A).** The TILSig could stratify patients into two groups with significantly different OS in COAD and KIRC (log-rank p=0.019 for COAD and p=0.0006 for KIRC) (online supplementary figure S3B).

#### Independence of the TILSig from other clinical factors

To examine whether the prognostic performance of the TILSig is independent of other clinical factors, we used multivariate Cox analysis to test the performance of the TILSig after adjusted by other clinical factors, including age, gender and stage. In multivariate analysis, the HRs of high TILSig versus low TILSig for OS were 1.584 (p=0.002; 95% CI 1.184 to 2.118) in the training set, 2.322 (p=0.019; 95% CI 1.147 to 4.703) in the GSE31210 test set and 1.338 (p=0.017; 95% CI 1.054 to 1.698) in the TCGA test set (table 2), respectively, showing that TILSig is still significantly correlated with unfavorable OS in the training set and other two independent test sets after adjusting for other clinical factors. The results of the multivariable analysis thus indicated that the prognostic performance of TILSig is independent of other clinical factors for OS prediction.

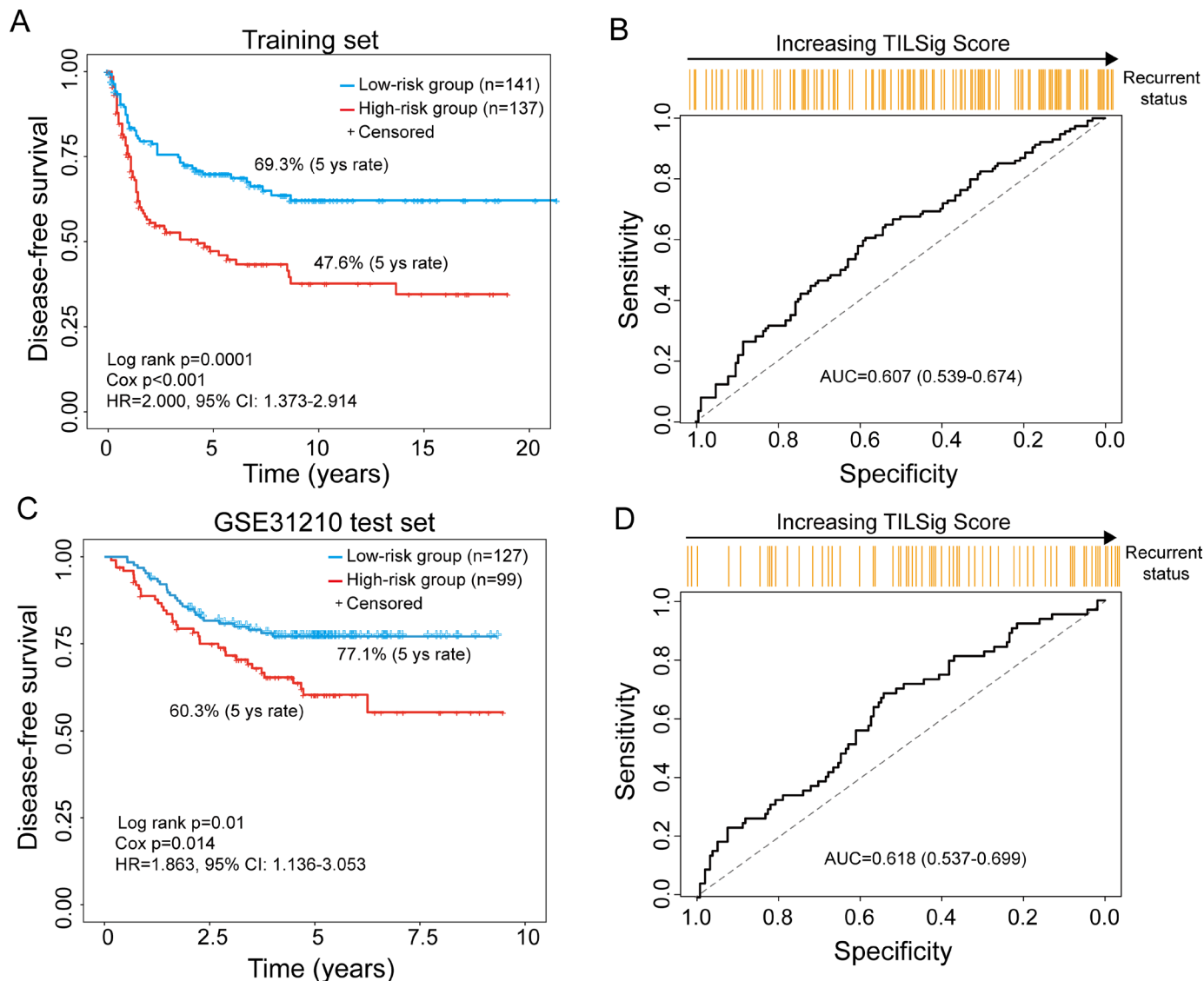
#### The TILSig is associated with tumor recurrence

To examine whether the TILSig is also related to tumor recurrence, we first used DFS time as recurrence

**Table 2** Univariate and multivariate Cox regression analysis of overall survival in each data set

| Variables              | Univariate analysis |                |         | Multivariate analysis |                |         |
|------------------------|---------------------|----------------|---------|-----------------------|----------------|---------|
|                        | HR                  | 95% CI of HR   | P value | HR                    | 95% CI of HR   | P value |
| GSE30219               |                     |                |         |                       |                |         |
| TILSig (high/low)      | 1.981               | 1.491 to 2.633 | <0.001  | 1.584                 | 1.184 to 2.118 | 0.002   |
| Age                    | 1.038               | 1.024 to 1.053 | <0.001  | 1.037                 | 1.022 to 1.052 | <0.001  |
| Gender (male/female)   | 1.701               | 1.082 to 2.674 | 0.022   | 1.398                 | 0.886 to 2.206 | 0.150   |
| Stage (III/IV vs I/II) | 3.036               | 2.234 to 4.127 | <0.001  | 2.810                 | 2.059 to 3.835 | <0.001  |
| GSE31210               |                     |                |         |                       |                |         |
| TILSig (high/low)      | 2.422               | 1.219 to 4.812 | 0.012   | 2.322                 | 1.147 to 4.703 | 0.019   |
| Age                    | 1.025               | 0.978 to 1.075 | 0.306   | 1.035                 | 0.988 to 1.085 | 0.147   |
| Gender (male/female)   | 1.519               | 0.780 to 2.955 | 0.219   | 1.100                 | 0.549 to 2.201 | 0.789   |
| Stage (II /I)          | 4.232               | 2.175 to 8.236 | <0.001  | 4.290                 | 2.164 to 8.502 | <0.001  |
| TCGA                   |                     |                |         |                       |                |         |
| TILSig (high/low)      | 1.305               | 1.028 to 1.656 | 0.028   | 1.338                 | 1.054 to 1.698 | 0.017   |
| Age                    | 1.017               | 1.003 to 1.030 | 0.014   | 1.020                 | 1.006 to 1.034 | 0.004   |
| Gender (male/female)   | 0.995               | 0.780 to 1.270 | 0.970   | 0.966                 | 0.756 to 1.234 | 0.782   |
| Stage (III/IV vs I/II) | 1.901               | 1.476 to 2.45  | <0.001  | 2.008                 | 1.555 to 2.592 | <0.001  |

TCGA, The Cancer Genome Atlas; TILSig, tumor-infiltrating immune-related lncRNA signature.



**Figure 4** The tumor-infiltrating immune-related lncRNA signature (TILSig) is associated with tumor recurrence. Kaplan-Meier survival curves of disease-free survival between patients with a higher score of TILSig and with the lower score of TILSig in the training data set (A) and GSE31210 data set (C). Receiver operating characteristic (ROC) curve of the TILSig for predicting tumor recurrence in the training data set (B) and GSE31210 data set (D). AUC, area under the curve.

measures to investigate the effectiveness of the TILSig for recurrence risk prediction in the training set and independent GSE31210 test set. As shown in figure 4A,C, survival analyses demonstrated that the DFS distribution from the high-risk group was significantly different from that of the low-risk group (log-rank  $p=0.0001$  for the training set and  $p=0.01$  for GSE31210 test set). In the high-risk group, the proportions of the disease-free patient were 47.6% and 60.3% at 5 years in the training and GSE31210 sets, respectively, which were also significantly lower than those in the low-risk group (69.3% and 77.1%) (figure 4A,C). Next, we used TILSig to predict patients' recurrence status, which achieved better predictive performance with AUC of 0.607 in the training set and 0.618 in the GSE31210 test set (figure 4B,D). Moreover, increasing TILSig score is associated with a greater probability of

disease recurrence. We also performed multivariate Cox analysis of DFS for each data set and found that the TILSig still maintained an independent correlation with DFS ( $p=0.024$ , HR=1.570, 95% CI 1.062 to 2.320 for training set and  $p=0.034$ , HR=1.728, 95% CI 1.043 to 2.863) (table 3).

#### Potential of TILSig as an indicator of immunotherapy response in patients with NSCLC

The previous study has revealed that immune infiltration can be modulated by the immune checkpoint inhibitor (ICI) genes. To further investigate the complex cross-talk between immune infiltration and ICI genes, we first compared the expression pattern of ICI genes (PD1, PD-L1 and CTLA-4) between different patient groups stratified by the TILSig. As shown in figure 5A, patients with high TILSig tend to express high ICI genes compared with



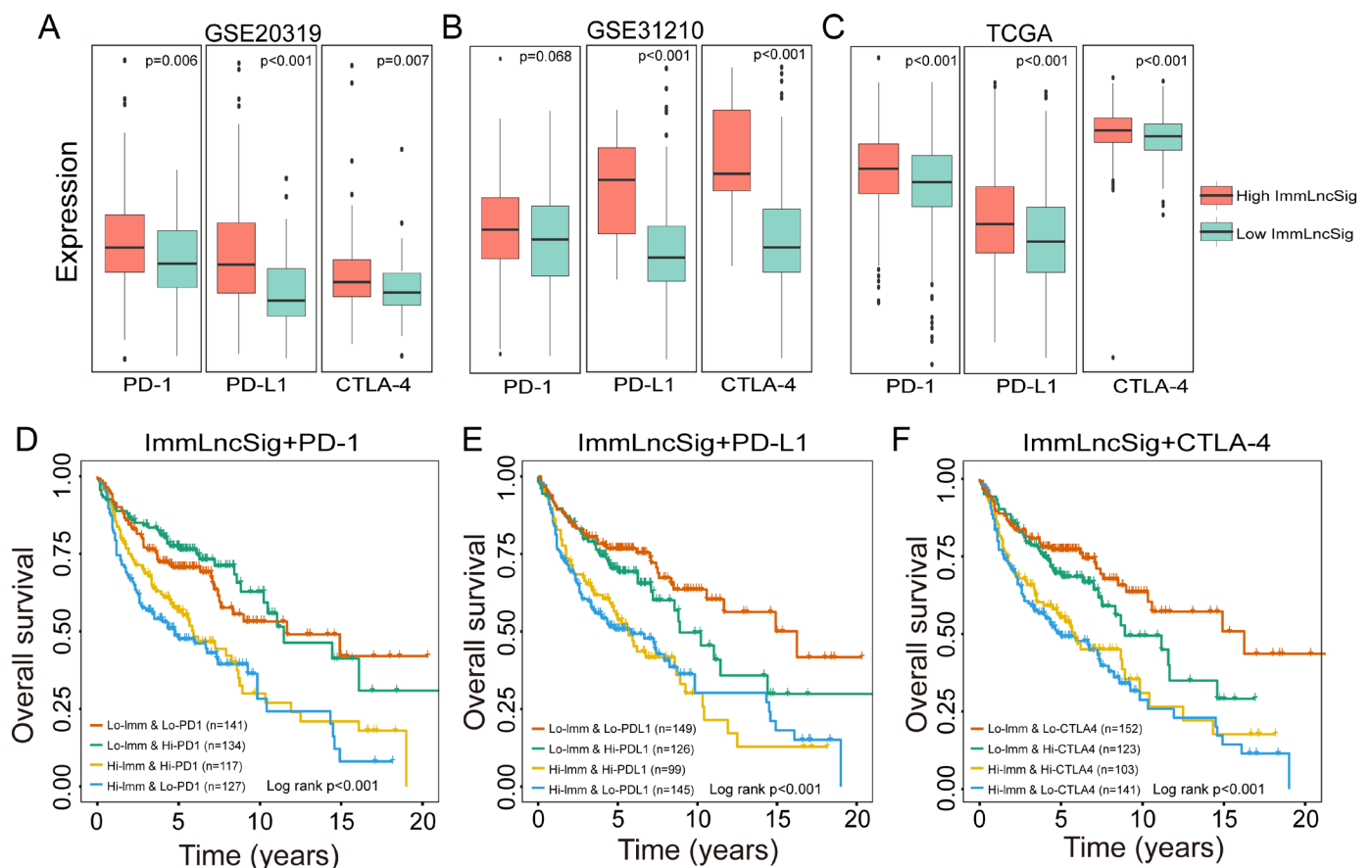
**Table 3** Univariate and multivariate regression analysis of disease-free survival in each data set

| Variables              | Univariate analysis |                |         | Multivariate analysis |                |         |
|------------------------|---------------------|----------------|---------|-----------------------|----------------|---------|
|                        | HR                  | 95% CI of HR   | P value | HR                    | 95% CI of HR   | P value |
| GSE30219               |                     |                |         |                       |                |         |
| TILSig (high/low)      | 2.000               | 1.373 to 2.914 | <0.001  | 1.570                 | 1.062 to 2.320 | 0.024   |
| Age                    | 1.021               | 1.004 to 1.038 | 0.017   | 1.019                 | 1.000 to 1.037 | 0.045   |
| Gender (male/female)   | 1.355               | 0.787 to 2.333 | 0.273   | 1.216                 | 0.704 to 2.100 | 0.484   |
| Stage (III/IV vs I/II) | 3.959               | 2.685 to 5.837 | <0.001  | 3.583                 | 2.411 to 5.326 | <0.001  |
| GSE31210               |                     |                |         |                       |                |         |
| TILSig (high/low)      | 1.863               | 1.136 to 3.053 | 0.014   | 1.728                 | 1.043 to 2.863 | 0.034   |
| Age                    | 1.034               | 0.997 to 1.072 | 0.074   | 1.039                 | 1.003 to 1.077 | 0.0342  |
| Gender (male/female)   | 1.271               | 0.778 to 2.075 | 0.338   | 1.012                 | 0.609 to 1.682 | 0.964   |
| Stage (II /I)          | 3.164               | 1.921 to 5.213 | <0.001  | 3.280                 | 1.968 to 5.466 | <0.001  |

TILSig, tumor-infiltrating immune-related lncRNA signature.

those with low TILSig in the GSE20319 data set (Mann-Whitney U test  $p=0.006$ ,  $p<0.001$ ,  $p=0.007$  for PD1, PD-L1 and CTLA-4, respectively; [figure 5A](#)). Similar trends also were observed both in GSE31210 (Mann-Whitney U test  $p=0.068$ ,  $p<0.001$ ,  $p<0.001$  for PD1, PD-L1 and CTLA-4,

respectively; [figure 5B](#)) and TCGA data sets (Mann-Whitney U test  $p<0.001$ ,  $p<0.001$ ,  $p<0.001$  for PD1, PD-L1 and CTLA-4, respectively; [figure 5C](#)). This trend is concordant with previous observations linking higher expression of immune checkpoint genes to poor outcomes. We



**Figure 5** Impact of immune checkpoint gene expression and tumor-infiltrating immune-related lncRNA signature (TILSig) on clinical outcome. Comparison of the expression pattern of immune checkpoint genes (PD1, PD-L1, and CTLA-4) between patients with a higher score of TILSig and with lower score of TILSig in the training data set (A), GSE31210 data set (B) and TCGA data set (C). Kaplan-Meier survival curves of overall survival among four patient groups stratified by the TILSig and PD-1 (D), PD-L1 (E) and CTLA-4 (F). TCGA, The Cancer Genome Atlas.

further examined whether the immune infiltration has an impact on clinical outcomes in patients with similar expression levels of immune checkpoint genes. Survival distribution of four patient groups stratified by the TILSig and high/low immune checkpoint gene expression was compared. As shown in [figure 5D](#), patients with low TILSig and high PD-1 have significantly better survival than those with high TILSig and high PD-1 (log-rank  $p < 0.001$ ), and patients with low TILSig and low PD-1 also are associated with prolonged survival relative to those with high TILSig and low PD-1 (log-rank  $p < 0.001$ ) ([figure 5D](#)). A similar analysis was repeated using TILSig and PD-L1 or CTLA-4, and revealed that patients stratified by TILSig and PD-L1 or CTLA-4 exhibited survival patterns similar to those of PD-1 ([figure 5E,F](#)). When stratifying patients with low TILSig based on immune checkpoint gene expression, we also observed that immune checkpoint gene expressions are associated with noticeable survival differences in patients with low TILSig. However, survival differences were not present for patients with high TILSig when stratifying patients with TILSig-based immune checkpoint gene expression. Furthermore, patients with low TILSig and low immune checkpoint gene expression tend to have significantly better survival than other three patient groups. These results indicated that the TILSig may be a potential predictive biomarker of treatment response to ICI immunotherapy.

## DISCUSSION

The immune infiltrates in the TME are increasingly recognized to be associated with patient prognosis in NSCLC and other cancers.<sup>5 31 32</sup> For example, the immunoscore was proposed based on the mean of four density percentiles including two markers (CD3+ and CD8+) and two regions (tumor and invasive margin regions), and has been internationally validated as a risk assessment tool in colon cancer,<sup>33</sup> highlighting the potential importance of evaluating the immune infiltrate of tumor in guiding decision-making in the clinic. Therefore, quantitative evaluation of the tumor immune infiltrates in the TME is a major challenge. Considerable efforts have been devoted to quantifying the composition of immune cell infiltration. However, the traditional immunohistochemistry immunoscore approach remains suboptimal because of lacking a consistent standard and limited number of biomarkers assessed simultaneously.<sup>34</sup> Recently, molecular profiling-based approaches and signatures have been used to infer immune infiltration, such as mRNAs,<sup>31</sup> microRNA<sup>35</sup> and DNA methylation.<sup>36</sup> lncRNA, a recently identified key players of genome regulatory network, has also been found to play critical roles in the development and activation of immune cells, and may, therefore, serve as a specific molecular marker for tumor-infiltrating immune cells.<sup>14 15</sup> However, genome-wide screening of tumor immune infiltration-associated lncRNA and their value in evaluating immune infiltrate of tumor and clinical outcome has barely been explored.

In this study, we first reannotated the expression profiles of 19 immune cell types and uncovered previously uncharacterized immune-related lncRNAs including nine iclncRNAs and 57 hklncRNAs. Gene Ontology (GO) enrichment analysis suggested that mRNAs coexpressed with hklncRNAs were significantly enriched in various immune-related biological processes (online supplementary figure S4). Based on the assumption that lncRNAs that have universal expression across different immune cell types may be essential for basic immune cellular functions, we developed a novel computational frame by integrative analysis of molecular profiling of purified immune cells, cancer cell lines, and bulk cancer tissues to identify TILncRNAs for evaluating immune cell infiltration in TME. By applying this computational frame to NSCLC cell line and patient data sets, we identified the lncRNA signature, named TILSig, as an indicator of immune cell infiltration in patients with NSCLC. This computational frame for identifying TILSig may be referenced for other cancer types. The ability of the TILSig to evaluate immune cell infiltration is based on highly differential expression of hklncRNAs in immune cells versus lung cancer cell lines. In particular, some components of the TILSig have recently been associated with immunological functions and cancer prognosis. lncRNA *HCG26* is located in the human leucocyte antigen (HLA) class I region which is a crucial determinant of the adaptive immune response, and recent study also linked *HCG26* with bone metastasis of NSCLC.<sup>37</sup> *CARD8-AS1* has recently been shown to be involved in the regulation of immune response and epithelial-mesenchymal transition.<sup>38</sup> *HCP5*, a hybrid HLA class I endogenous retroviral gene, is known for its functional roles in adaptive and innate immune responses, and newly is recognized as SMAD3-responsive lncRNA which can promote lung adenocarcinoma metastasis via miR-203/SNAI axis.<sup>39</sup> These published experimental efforts provide further evidence supporting the association between the TILSig and immune responses.

To further understand the prognostic role of the TILSig, we then correlate the TILSig with patient survival, resulting in significant risk stratification of survival in patients with NSCLC ([figure 2A](#)). Using the single-sample GSEA analysis to measure the infiltration of 19 immune subpopulations within NSCLC tumors of different risk groups stratified by the TILSig, respectively, we found significantly different immune contexture between different risk groups classified by the TILSig. The high-risk group defined by the TILSig seems to be an immune-cold patient group with less immune cell infiltration and low-risk group defined by the TILSig is an immune-hot patient group with greater immune cell infiltration. This observation is concordant with previous findings that there are two distinct immunophenotypes in patients with NSCLC.<sup>40</sup> These results showed that the TILSig is highly reflective of immune cell infiltrates resulting in a better prognosis prediction of patients with NSCLC. To confirm that these associations were not specific to one data set, we applied the TILSig to other two additional

larger NSCLC patient data sets from GEO and TCGA. The TILSig is reproducible across two independent patient data sets with microarray or RNA-seq platforms and was capable of distinguishing patients with a different outcome in terms of survival and recurrence.

Over the past several years, checkpoint inhibitor-based immunotherapies have revealed substantial advancement in clinical care for many cancer types, including NSCLC.<sup>41–42</sup> An early assessment for ICI response by predictive biomarkers is crucial for the selection of patients who are most likely to benefit from ICI. The expression of immune checkpoint genes such as PD-L1 is currently available biomarkers in clinical practice, but not a sufficient independent predictor of ICI response.<sup>43–44</sup> Patients with NSCLC still have heterogeneous response rates to immunotherapy despite high tumor PD-L1/PD-1 levels and only a small subset of patients with NSCLC are responders for immune checkpoint blockade therapy.<sup>45</sup> The immune infiltrates in the TME are increasingly recognized to be associated with immunotherapy response. By comparing the survival distributions of patients stratified by the TILSig and immune checkpoint gene expression, we found that the TILSig has a discriminatory power in patients with similar expression levels of immune checkpoint genes, and the prognosis of patients with similar TILSig could be affected by immune checkpoint gene expression. These findings suggest that the complex interplay between immune infiltrate and immune checkpoint gene in the tumor immune microenvironment has an impact on patient survival in NSCLC which is in line with previous study in immunomodulatory interactions between immune cell infiltration and checkpoint gene expression.<sup>31</sup> Furthermore, significantly prolonged survival was observed for patients with low TILSig and low immune checkpoint gene expression, implying that these tumors with low TILSig may be associated with better response to ICI therapy.

## CONCLUSIONS

We performed a genome-wide screening by integrative analysis of molecular profiling of immune cells, cancer cell lines and bulk cancer tissues to identify lncRNAs associated with tumor immune cell infiltration and highlighted the importance and value of lncRNAs in evaluating the immune infiltrate of the tumor. Furthermore, for the first time, our study identified and validated an lncRNA signature (TILSig) that is based on seven TILncRNAs, as an indicator of immune cell infiltration in the TME and has independent prognostic significance for patients with NSCLC. Lastly, we characterized complex interplay between immune infiltrate and immune checkpoint gene in patient's outcome, and suggested the potential of TILSig coupled with specific immune checkpoint factors as predictive biomarkers of ICI response to enable a more precise selection of patients who will benefit from checkpoint inhibitor immunotherapy which need further

validating with continuously releasing immunotherapy data sets.

## Author affiliations

<sup>1</sup>School of Biomedical Engineering, School of Ophthalmology & Optometry and Eye Hospital, Wenzhou Medical University, Wenzhou, China

<sup>2</sup>Department of Orthopedic Surgery, Peking Union Medical College Hospital, Peking Union Medical College and Chinese Academy of Medical Sciences, Beijing, China

**Contributors** MZ and LX designed the study. JS and MZ developed the methodology. ZZ, SB, CY, PH and NW performed the data analysis. MZ, JS and JZS drafted the manuscript. All authors read and approved the final manuscript.

**Funding** This study was supported by the National Natural Science Foundation of China (grant numbers 61973240) and the Scientific Research Foundation for Talents of Wenzhou Medical University (QTJ18024 and QTJ18023).

**Disclaimer** The funders had no roles in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests** None declared.

**Patient consent for publication** Not required.

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Data availability statement** Data are available in a public, open access repository. The data sets analyzed during the current study are available in the Gene Expression Omnibus (GEO, <https://www.ncbi.nlm.nih.gov/geo/>) and The Cancer Genome Atlas (TCGA) network (<https://cancergenome.nih.gov/>) and the Cancer Cell Line Encyclopedia (CCLE) project (<https://portals.broadinstitute.org/ccle>).

**Open access** This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See <http://creativecommons.org/licenses/by-nc/4.0/>.

## ORCID iD

Meng Zhou <http://orcid.org/0000-0001-9987-9024>

## REFERENCES

- 1 Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. *CA Cancer J Clin* 2019;69:7–34.
- 2 Zhou M, Guo M, He D, et al. A potential signature of eight long non-coding RNAs predicts survival in patients with non-small cell lung cancer. *J Transl Med* 2015;13:231.
- 3 Herbst RS, Morgensztern D, Boshoff C. The biology and management of non-small cell lung cancer. *Nature* 2018;553:446–54.
- 4 Bense RD, Sotiriou C, Piccart-Gebhart MJ, et al. Relevance of tumor-infiltrating immune cell composition and functionality for disease outcome in breast cancer. *J Natl Cancer Inst* 2017;109:djw192.
- 5 Barnes TA, Amir E. Erratum: hype or hope: the prognostic value of infiltrating immune cells in cancer. *Br J Cancer* 2018;118:e5.
- 6 Jochims C, Schlom J. Tumor-Infiltrating immune cells and prognosis: the potential link between conventional cancer therapy and immunity. *Exp Biol Med* 2011;236:567–79.
- 7 Zheng X, Hu Y, Yao C. The paradoxical role of tumor-infiltrating immune cells in lung cancer. *IRDR* 2017;6:234–41.
- 8 Stankovic B, Bjørhovde HAK, Skarshaug R, et al. Immune cell composition in human non-small cell lung cancer. *Front Immunol* 2018;9:3101.
- 9 Yao R-W, Wang Y, Chen L-L. Cellular functions of long noncoding RNAs. *Nat Cell Biol* 2019;21:542–51.
- 10 Moran VA, Perera RJ, Khalil AM. Emerging functional and mechanistic paradigms of mammalian long non-coding RNAs. *Nucleic Acids Res* 2012;40:6391–400.
- 11 Zhou M, Zhao H, Wang X, et al. Analysis of long noncoding RNAs highlights region-specific altered expression patterns and diagnostic roles in Alzheimer's disease. *Brief Bioinform* 2019;20:598–608.
- 12 Bao S, Zhao H, Yuan J, et al. Computational identification of mutator-derived lncRNA signatures of genome instability for improving the clinical outcome of cancers: a case study in breast cancer. *Brief Bioinform* 2019;68.
- 13 Turner M, Galloway A, Vigorito E. Noncoding RNA and its associated proteins as regulatory elements of the immune system. *Nat Immunol* 2014;15:484–91.

- 14 Chen YG, Satpathy AT, Chang HY. Gene regulation in the immune system by long noncoding RNAs. *Nat Immunol* 2017;18:962–72.
- 15 Atianand MK, Caffrey DR, Fitzgerald KA. Immunobiology of long noncoding RNAs. *Annu Rev Immunol* 2017;35:177–98.
- 16 Elling R, Chan J, Fitzgerald KA. Emerging role of long noncoding RNAs as regulators of innate immune cell development and inflammatory gene expression. *Eur J Immunol* 2016;46:504–12.
- 17 Hu G, Tang Q, Sharma S, et al. Expression and regulation of intergenic long noncoding RNAs during T cell development and differentiation. *Nat Immunol* 2013;14:1190–8.
- 18 Ranzani V, Rossetti G, Panzeri I, et al. The long intergenic noncoding RNA landscape of human lymphocytes highlights the regulation of T cell differentiation by linc-MAF-4. *Nat Immunol* 2015;16:318–25.
- 19 Denaro N, Merlano MC, CJMo LN. Long noncoding RNA S as regulators of cancer immunity 2019;13:61–73.
- 20 Wang L, Felts SJ, Van Keulen VP, et al. Integrative genome-wide analysis of long noncoding RNAs in diverse immune cell types of melanoma patients. *Cancer Res* 2018;78:4411–23.
- 21 Sage A, Ng K, Marshall E, et al. MA24. 06 long non-coding RNA expression patterns delineate infiltrating immune cells in the lung tumour microenvironment 2018;13:S443–4.
- 22 Xu M, Xu X, Pan B, et al. Lncrna SATB2-AS1 inhibits tumor metastasis and affects the tumor immune cell microenvironment in colorectal cancer by regulating SATB2. *Mol Cancer* 2019;18:135.
- 23 Rousseaux S, Debernardi A, Jacquiou B, et al. Ectopic activation of germline and placental genes identifies aggressive metastasis-prone lung cancers. *Sci Transl Med* 2013;5:186ra66.
- 24 Okayama H, Kohno T, Ishii Y, et al. Identification of genes upregulated in ALK-positive and EGFR/KRAS/ALK-negative lung adenocarcinomas. *Cancer Res* 2012;72:100–11.
- 25 Charoentong P, Finotello F, Angelova M, et al. Pan-Cancer Immunogenomic analyses reveal Genotype-Immunophenotype relationships and predictors of response to checkpoint blockade. *Cell Rep* 2017;18:248–62.
- 26 Irizarry RA et al. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* 2003;4:249–64.
- 27 Yanai I, Benjamin H, Shmoish M, et al. Genome-Wide midrange transcription profiles reveal expression level relationships in human tissue specification. *Bioinformatics* 2005;21:650–9.
- 28 Zhou M, Zhao H, Wang Z, et al. Identification and validation of potential prognostic lncRNA biomarkers for predicting survival in patients with multiple myeloma. *J Exp Clin Cancer Res* 2015;34:102.
- 29 Yu G, Wang L-G, Han Y, et al. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* 2012;16:284–7.
- 30 Thorsson V, Gibbs DL, Brown SD, et al. The immune landscape of cancer. *Immunity* 2018;48:812–30.
- 31 Varn FS, Tafe LJ, Amos CI, et al. Computational immune profiling in lung adenocarcinoma reveals reproducible prognostic associations with implications for immunotherapy. *Oncoimmunology* 2018;7:e1431084.
- 32 Bonanno L, Pavan A, Dieci MV, et al. The role of immune microenvironment in small-cell lung cancer: distribution of PD-L1 expression and prognostic role of FOXP3-positive tumour infiltrating lymphocytes. *Eur J Cancer* 2018;101:191–200.
- 33 Pagès F, Mlecnik B, Marliot F, et al. International validation of the consensus immunoscore for the classification of colon cancer: a prognostic and accuracy study. *The Lancet* 2018;391:2128–39.
- 34 Hao Y, Yan M, Heath BR, et al. Fast and robust deconvolution of tumor infiltrating lymphocyte from expression profiles using least trimmed squares. *PLoS Comput Biol* 2019;15:e1006976.
- 35 Korsunsky I, Parameswaran J, Shapira I, et al. Two microRNA signatures for malignancy and immune infiltration predict overall survival in advanced epithelial ovarian cancer. *J Investig Med* 2017;65:1068–76.
- 36 Jeschke J, Bizet M, Desmedt C, et al. Dna methylation-based immune response signature improves patient diagnosis in multiple cancers. *J Clin Invest* 2017;127:3090–102.
- 37 Dat le T, Matsuo T, Yoshimaru T, et al. Identification of genes potentially involved in bone metastasis by genome-wide gene expression profile analysis of non-small cell lung cancer in mice. *Int J Oncol* 2012;40:1455–69.
- 38 Lin X, Jiang T, Bai J, et al. Characterization of transcriptome transition associates long noncoding RNAs with glioma progression. *Mol Ther Nucleic Acids* 2018;13:620–32.
- 39 Kulski JK. Long noncoding RNA HCP5, a hybrid HLA class I endogenous retroviral gene: structure, expression, and disease associations. *Cells* 2019;8:480.
- 40 Lizotte PH, Ivanova EV, Awad MM, et al. Multiparametric profiling of non-small-cell lung cancers reveals distinct immunophenotypes. *JCI Insight* 2016;1:e89014.
- 41 Havel JJ, Chowell D, Chan TA. The evolving landscape of biomarkers for checkpoint inhibitor immunotherapy. *Nat Rev Cancer* 2019;19:133–50.
- 42 Doroshov DB, Sanmamed MF, Hastings K, et al. Immunotherapy in Non-Small cell lung cancer: facts and hopes. *Clin Cancer Res* 2019;25:4592–602.
- 43 Lupo A, Alifano M, Wislez M, et al. Biomarkers predictive of PD1/PD-L1 immunotherapy in non-small cell lung cancer]. *Rev Pneumol Clin* 2018;74:339–50.
- 44 Gridelli C, Ardizzoni A, Barberis M, et al. Predictive biomarkers of immunotherapy for non-small cell lung cancer: results from an experts panel meeting of the Italian association of thoracic oncology. *Transl Lung Cancer Res* 2017;6:373–86.
- 45 Pu X, Wu L, Su D, et al. Immunotherapy for non-small cell lung cancers: biomarkers for predicting responses and strategies to overcome resistance. *BMC Cancer* 2018;18:1082.