

RESEARCH

Open Access



Identification of a cytokine-dominated immunosuppressive class in squamous cell lung carcinoma with implications for immunotherapy resistance

Minglei Yang¹, Chenghao Lin¹, Yanni Wang¹, Kang Chen¹, Haiyue Zhang¹ and Weizhong Li^{1,2,3*} 

Abstract

Background: Immune checkpoint blockade (ICB) therapy has revolutionized the treatment of lung squamous cell carcinoma (LUSC). However, a significant proportion of patients with high tumour PD-L1 expression remain resistant to immune checkpoint inhibitors. To understand the underlying resistance mechanisms, characterization of the immunosuppressive tumour microenvironment and identification of biomarkers to predict resistance in patients are urgently needed.

Methods: Our study retrospectively analysed RNA sequencing data of 624 LUSC samples. We analysed gene expression patterns from tumour microenvironment by unsupervised clustering. We correlated the expression patterns with a set of T cell exhaustion signatures, immunosuppressive cells, clinical characteristics, and immunotherapeutic responses. Internal and external testing datasets were used to validate the presence of exhausted immune status.

Results: Approximately 28 to 36% of LUSC patients were found to exhibit significant enrichments of T cell exhaustion signatures, high fraction of immunosuppressive cells (M2 macrophage and CD4 Treg), co-upregulation of 9 inhibitory checkpoints (*CTLA4*, *PDCD1*, *LAG3*, *BTLA*, *TIGIT*, *HAVCR2*, *IDO1*, *SIGLEC7*, and *VISTA*), and enhanced expression of anti-inflammatory cytokines (e.g. TGF β and CCL18). We defined this immunosuppressive group of patients as exhausted immune class (EIC). Although EIC showed a high density of tumour-infiltrating lymphocytes, these were associated with poor prognosis. EIC had relatively elevated PD-L1 expression, but showed potential resistance to ICB therapy. The signature of 167 genes for EIC prediction was significantly enriched in melanoma patients with ICB therapy resistance. EIC was characterized by a lower chromosomal alteration burden and a unique methylation pattern. We developed a web application (<http://lilab2.sysu.edu.cn/tex> & <http://liwzlab.cn/tex>) for researchers to further investigate potential association of ICB resistance based on our multi-omics analysis data.

Conclusions: We introduced a novel LUSC immunosuppressive class which expressed high PD-L1 but showed potential resistance to ICB therapy. This comprehensive characterization of immunosuppressive tumour microenvironment in LUSC provided new insights for further exploration of resistance mechanisms and optimization of immunotherapy strategies.

*Correspondence: liweizhong@mail.sysu.edu.cn

³ Key Laboratory of Tropical Disease Control of Ministry of Education, Sun Yat-Sen University, Guangzhou 510080, China

Full list of author information is available at the end of the article



© The Author(s) 2022. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Keywords: Immunogenomics, LUSC, T cell exhaustion, Immunosuppressive cytokine, Immune checkpoint blockade resistance, Tumour microenvironment

Background

Lung cancer is the most common cancer and the leading cause of cancer-related death worldwide [1]. Non-small-cell lung cancer (NSCLC) accounts for approximately 85% of lung cancer cases [2]. Besides lung adenocarcinoma, lung squamous carcinoma (LUSC) is the most frequent histologic subtype of NSCLC [3]. However, due to the lack of genetic alterations for which targeted treatments are approved, the patients of LUSC have limited treatment options except chemotherapy [4].

With the recent development of immune checkpoint blockade (ICB) immunotherapy, anti-PD-1/PD-L1 immune checkpoint inhibitors have been approved for the first-line treatment of NSCLC in patients with high tumour programmed death-ligand 1 (PD-L1) expression ($\geq 50\%$) [5]. However, only 45.2% of patients met the screening criteria and benefited from the immune checkpoint blockade therapy [6, 7]. A significant fraction of patients had drug resistance to this immunotherapy. On the other hand, ICB therapy can cause unique toxicity and subsequently immune-related adverse events through enhanced immune responses [8].

ICB therapy aims to reinvigorate dysfunctional or exhausted T cells to eliminate tumours [9]. Tumour microenvironment (TME) consists of cancer cells, immune cells, stromal cells, and cytokines. The TME components dynamically regulated T cell exhaustion [10]. The success of ICB therapy by reinvigorating exhausted T cell relies heavily on the complex interactions between the cancer cells and the components of TME [11]. However, the knowledge of molecular mechanism of resistance to immune checkpoint inhibitors is limited. Therefore, to predict the patients' response or resistance to ICB therapy and to tailor reasonable treatments, characterizing TME molecular features and consequently identifying potential therapeutic markers are greatly required.

The TME of LUSC is highly complex and heterogeneous, but little is known about how TME impacts the efficacy of immunotherapy in LUSC. Virtual microdissection analytical approach based on non-negative matrix factorization (NMF) has enabled effective deconvolution on the gene expression signals derived from tumour cells, inflammatory cells, stromal cells, and cytokines from bulk tumour samples [12, 13]. In this study, we aimed to dissect the RNAseq expression data of 624 human LUSC samples and isolate transcriptomic signals released from immunosuppressive TME through NMF.

Consequently, we identified and validated an exhausted immune class of LUSC with immunosuppressive molecular features and potential ICB resistance. We also associated the exhausted immune class with multi-omics data to investigate the underlying ICB resistance mechanism. Finally, we constructed an interactive web application for researcher to explore immunotherapy resistance based on the multi-omics analysis results.

Methods

LUSC datasets and resources

The gene expression profiles of total 624 LUSC human samples were retrieved from The Cancer Genome Atlas (TCGA) (497 bulk RNA sequencing [RNAseq] datasets) [14] and Gene Expression Omnibus (GEO) (127 microarray datasets) (Additional file 1: Fig. S1). Due to the differences of clinical features and treatment regimens resulted from different tumour stages, we divided the 497 TCGA LUSC patients into the late-stage group (stage IIA to IV, 250 patients) for training and the early-stage group (stage I to II, 247 patients) for internal validation. The relevant data of mutation, copy number variation, methylation, and clinic pathology were obtained from the TCGA Data Portal (<https://tcga-data.nci.nih.gov/tcga/>, June 4th, 2020). The NMF analysis was conducted for the data of protein-coding genes. Other 127 LUSC microarray samples (Affymetrix Human Genome U133 Plus 2.0 Array) of patients from two independent datasets were utilized for external validation. The survival data in two independent datasets, including GSE30219 [15] and GSE37745 [16], were acquired through GEOquery [17] in the R package (<https://www.r-project.org>). The data for clinical outcomes and gene expression profiles of 28 melanoma patient samples (GSE78220) with anti-PD-1 therapy was retrieved from GEO [18]. These samples included 15 responding and 13 non-responding pre-treatment tumours and were profiled by RNAseq. Their response patterns were based on irRECIST [19]. More details for these datasets are listed in Additional file 2: Table S1. The software tools used in this study are summarized in Additional file 2: Table S2.

Identification of exhausted immune class by unsupervised clustering

A virtual microdissection analysis was firstly conducted on the bulk RNAseq-based gene expression profiles of a training cohort of 250 patients using NMF in R [13, 20] (Additional file 1: Fig. S1). The factorization rank r

which defines the number of clusters is the most critical parameter in NMF. When r equalled 4, the highest cophenetic correlation coefficient (Additional file 1: Fig. S2A) was obtained and the TCGA training cohort dataset was effectively decomposed (Fig. 1A) [20]. Therefore, r was set to be 4 in this study. Following the approach in a previous study [12], the immune and the stromal enrichment scores were calculated by the single sample gene set enrichment analysis (ssGSEA) [21, 22], which was wrapped in GSVA [23] to uncover immune and stromal-related expression patterns. When integrating the immune and stromal enrichment scores with the 4 NMF-identified clusters, we observed that Cluster 2 had higher enrichment scores (Fig. 1B) compared with other clusters. Therefore, Cluster 2 herein was referred as the ‘immune-stromal cluster’.

Secondly, we wanted to investigate the abundance of specific immune cells in tumours belonging to the immune-stromal cluster. A number of signatures representing various immune cells (Additional file 2: Table S3) were collected and used to compute the enrichment scores based on the expression profiles by ssGSEA. Then the enrichment scores of immune cells were integrated with the clusters to identify the abundance of specific immune cells in the immune-stromal cluster. In addition, the absolute fraction data of 22 infiltrating immune cells, which was inferred by the CIBERSORT algorithm [24] based on gene expression profiles, was downloaded from the TIMER database [25] (http://timer.cistrome.org/infiltration_estimation_for_tcg_a.csv.gz). And the leukocyte fraction data (TCGA_all_leuk_estimate.masked.20170107.tsv), which was estimated based on DNA methylation in Thorsson’s study [26], was retrieved from <https://gdc.cancer.gov/about-data/publications/panimmune>. The tumour-infiltrating lymphocyte (TIL) percentage, which was evaluated through pathological images of TCGA tumours including LUSC, can be found in the supplementary table (Table S1) in Saltz’s study [27]. The absolute fraction data of 22 infiltrating immune cells, the leukocyte fraction, and the TIL percentages

for TCGA LUSC patients were extracted and then compared between the immune-stromal cluster and the rest clusters to verify the enrichment of lymphocytes in the immune-stromal cluster by Wilcoxon rank-sum test.

Finally, the expression profile analysis of multiple inhibitory receptors was performed, and the gene signatures representing T cell exhaustion were scored by ssGSEA. We observed that the immune-stromal cluster overexpressed multiple inhibitory receptors and had high T cell exhaustion-related signature enrichment scores. Consequently, we defined the patient population within the immune-stromal cluster as the exhausted immune class (EIC) and the rest population as the rest class. We used the method extractFeature [28] wrapped in NMF to extract the relevant genes (named as metagene-specific genes) in order to characterize the EIC expression pattern.

Molecular characterization of exhausted immune class

The analyses by GSEA [22] and ssGSEA were conducted to evaluate the enrichment of molecular pathways and gene expression signatures in the EIC. The data of gene expression signatures was collected from previous studies [29–32] to represent distinct immune cells and exhausted T cells involved in various diseases (e.g. hepatocellular carcinoma, melanoma, and chronic infection). The hallmark gene sets and the KEGG pathway signatures were collated from MSigDB (<https://www.gsea-msigdb.org/gsea/msigdb>). ClueGO [33], a Cytoscape [34] plugin, was applied to generate the functionally grouped network of KEGG pathways to interpret the biological roles of metagene-specific genes. To identify the differentially expressed genes between the EIC and the rest class, the reads-count data of the training cohort was downloaded from TCGA. DESeq2 was used for gene differential expression analysis with false discovery rate (FDR) less than 0.05 and log₂ fold change (log₂FC) greater than 1 [35]. Of the metagene-specific genes, the differentially expressed genes (FDR < 0.05, log₂FC ≥ 1) between the EIC and the rest class were defined as the exhausted

(See figure on next page.)

Fig. 1 The identification and molecular characterization of EIC. **A** The heatmap of gene expression clusters for 250 late-stage (IIA-IV) LUSC samples by unsupervised NMF illustrates 4 distinct expression patterns. **B** Stromal and immune enrichment analysis defined the cluster 2 of four expression patterns as an immune-stromal cluster. High and low gene enrichment scores are delineated in red and grey, respectively. **C** The enrichment scores of gene signatures identified the immune cells for the immune-stromal and other clusters. **D** The comparison of the absolute fractions of TME cells inferred by CIBERSORT between two classes. **E,F** Box plots show the differences of leukocyte fraction and TIL percentage between two classes. **G** Box plots show different expression levels of multiple inhibitory receptors in the immune-stromal cluster compared to the other clusters. **H** The consensus-clustered heatmap of 250 LUSC samples defined the immune-stromal cluster as exhausted immune class (EIC). High and low gene enrichment scores are represented in red and grey, respectively. **I** GSEA analysis indicated the EIC showed significant enrichments of hallmark gene sets and KEGG pathways related to cytokine, T cell receptor, epithelial mesenchymal transition, and apoptosis. **J** The functionally grouped network of KEGG pathways by ClueGO/CluePedia for the interpretation of metagene-specific genes’ biological roles. Colourless and colour nodes represent metagene-specific genes and KEGG pathway terms, respectively. Node colours represent distinct functional groups. Node size represents the significance of KEGG pathways. The more significant KEGG pathways are, the larger highlighted nodes. All statistical differences of two classes were compared by Wilcoxon rank-sum test; *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$; ****, $P < 0.0001$

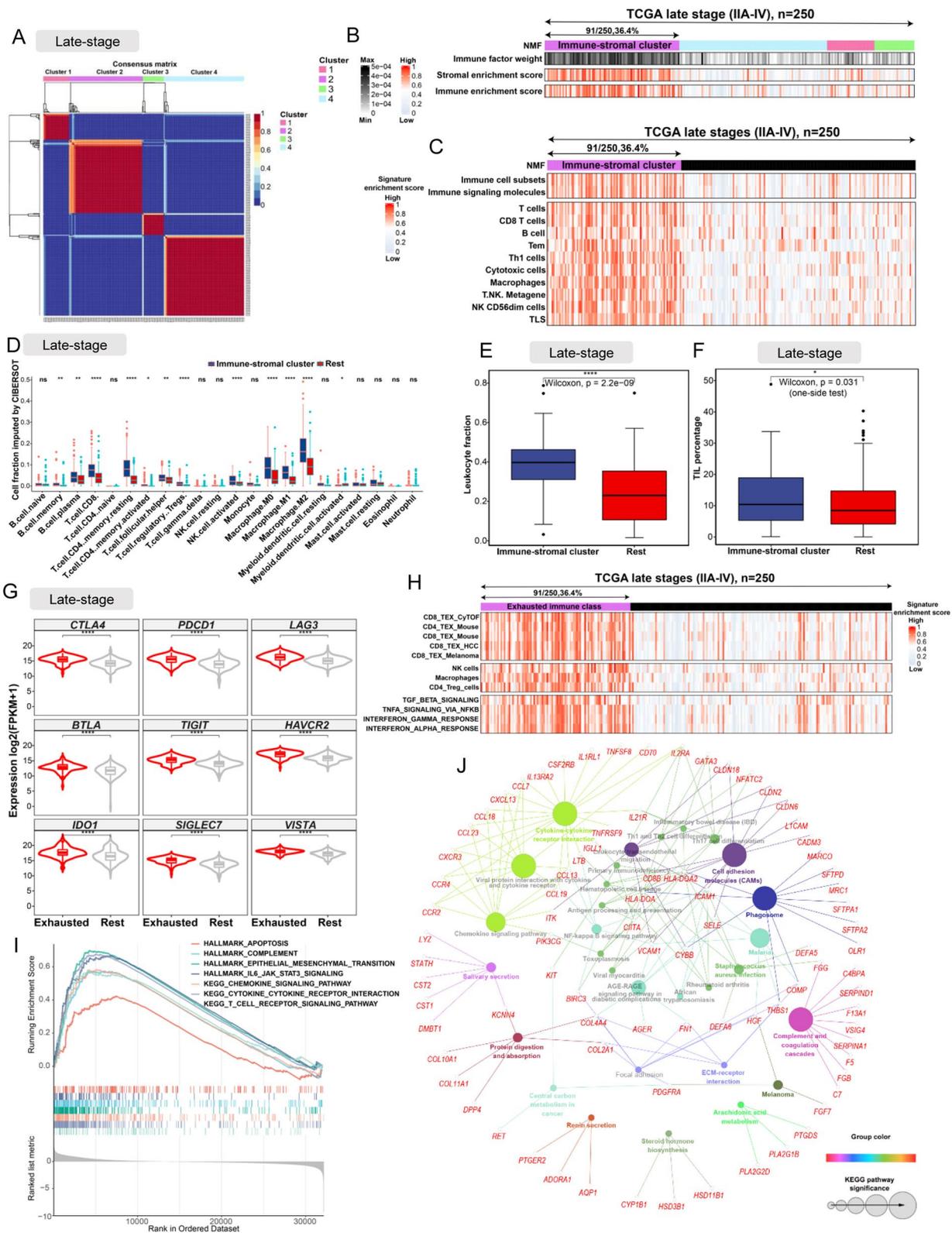


Fig. 1 (See legend on previous page.)

immune classifier genes for identifying the EIC in the training cohort from TCGA. GSEA was applied to identify activated pathways and hallmark gene sets enriched in the EIC.

Internal validation of EIC in early-stage TCGA LUSC

To confirm the presence of exhausted immune status in early-stage LUSC, we performed NMF and ssGSEA analysis on the bulk RNAseq-based expression profiles of 247 early-stage samples by using the same approach described above. Similarly, we also obtained 4 clusters for the early-stage LUSC cohort. When integrating the enrichment scores of early-stage samples for the signatures used in the late-stage LUSC cohort, we also observed that Cluster 2 had higher immune cell, stromal, and TEX-related signature enrichment scores. Therefore, Cluster2 was identified as EIC for the early-stage LUSC. The proportion of immune cells, leukocyte fraction, TIL percentage, and the expression of multiple inhibitory receptors were compared between the EIC and the rest class in these samples. GSEA was conducted to identify the enrichment of hallmarks and KEGG pathways. We used ssGSEA to calculate the enrichment score of 167 exhausted immune classifier genes obtained in the training stage and defined it as the EIC score. Receiver operating characteristic (ROC) analysis was utilized to evaluate the predictive capacity of the EIC score.

Prediction of ICB therapy for EIC

The tumour immune dysfunction and exclusion (TIDE) algorithm [36] was used to predict potential ICB therapy response. We also retrieved the public data of melanoma tumour expressions and the clinical outcomes of patients treated with anti-PD-1 [37] to validate the association between ICB therapy resistance and 167 exhausted immune classifier genes. Specifically, the EIC scores of melanoma tumour samples were calculated by ssGSEA using the 167 exhausted immune classifier genes and compared between the responders and the non-responders.

Analysis of genomic mutation, chromosomal alteration, and methylation profile for EIC

Maftools [38] was used to visualize and analyse somatic mutations and the total number of somatic mutations was counted. The statistical information of neoantigens for LUSC patients was obtained from Rooney's study [39]. The copy number data generated by GISTIC2.0 [40] for TCGA LUSC samples was retrieved from the cBioPortal for Cancer Genomics database (<http://www.cbioportal.org>), followed by the data extraction for the copy number alterations of cytobands and focal genes. Specifically, the amplification and deletion data for cytobands of

each LUSC sample was extracted from the file 'all_lesions.conf_99.txt', and the amplification and deletion data for focal genes was extracted from the file 'all_thresholded_by_genes.txt'. Both files were generated by GISTIC2.0 and available from <http://www.cbioportal.org>. Then we assessed the difference in somatic mutations, the number of neoantigens, and the copy number alterations between the EIC and the rest class.

The genomic methylation data for 367 LUSC patient samples of all tumour stages was obtained from TCGA. The 367 patients were a subset of 497 patients used in RNAseq data analysis. The methylation CpGs' Beta-value data of different samples was merged into a Beta-value matrix in which columns corresponded to samples and rows to CpGs. To correct the probe design bias in the Illumina Infinium 450K DNA methylation data, the Beta-value matrix was normalized by the Beta-mixture quantile normalization method using ChAMP in R [41]. We then performed a linear model using 'limma' in R to identify the significant CpG sites (FDR adjusted p -value < 0.05) that were differentially methylated between the EIC and the rest class identified by transcriptome analysis ($\Delta\text{Beta} > 0.2$). We only selected differentially methylated CpG sites located in the promoter regions of differentially expressed genes in the EIC compared with the rest class. The selected CpG sites were used to generate and cluster supervised heatmaps based on the Euclidean distances aggregated by Complete-linkage. Finally, the correlations between the methylation level of promoter regions and the corresponding gene expression were computed.

Validation of EIC in independent datasets

The robustness of exhausted immune classifier genes was evaluated by an unsupervised Random Forest procedure [42]. Based on the expression value of these genes, we used the randomForest R package to perform an unsupervised learning on the training cohort with the parameter 'ntree = 500', and then visualized the samples distribution through the function 'MDSplot' wrapped in this package. The ability of exhausted immune classifier genes to predict exhausted immune status was validated in two additional datasets by using NMF. The gene signatures related to T cell exhaustion (Additional file 2: Table S3) were applied to validate and characterize the EIC captured by the classifier in each dataset.

Highly visual interactive web application

Based on the analysis data in this study, we used 'shiny' in R to build an interactive web application (<http://lilab2.sysu.edu.cn/tex> & <http://liwzlab.cn/tex>) for other researchers to explore potential mechanisms of immunotherapy resistance at multi-omics level. The

web application has included several immunotherapy resistance-related analysis modules, such as exhausted immune classifier module, signature expression module, somatic mutation module, clinical prognosis module, microRNA module, and methylation module. The source codes generated to build the web application are available at the GitHub repository (https://github.com/Lilab-SYSU/LUSC_Tex) [43].

Protein expression analysis

To investigate protein expression alterations in EIC, the protein expression data of 319 LUSC patients was downloaded from TCGA. This data included 487 proteins profiled by reverse-phase protein arrays (RPPAs). The 319 LUSC patients were a subset of 497 patients for the training and internal validating cohort used in the bulk RNAseq data analysis. Then the protein expression was compared between the EIC and the rest class. The significantly upregulated proteins in EIC were identified. Moreover, we explored these protein expression profiles of LUSC on the immunohistochemically stained images via the pathology section of the Human Protein Atlas (HPA; <https://www.proteinatlas.org/>) [44]. Specifically, we searched the genes in HPA and found 12 LUSC patient samples which had the quantity information of the corresponding protein expressions in immunohistochemically stained images. Typical images were then selected to demonstrate the IDO protein expression level.

Statistical analysis

All statistical discrete analyses were performed in SPSS (version 19.0, IBM) and R (version 3.5.1, <http://www.r-project.org>). We correlated the EIC and the rest class with TIL percentage, copy number alteration, mutation number, and neoantigen number by the Wilcoxon rank-sum test for continuous data. Kaplan–Meier estimate and log-rank testing were used to perform the survival analysis for overall survival (OS) and progression-free survival (PFS). We incorporated all clinicopathological variables into Cox model to identify the combination of variables. Two-tailed or one-tailed P -values < 0.05 were considered statistically significant. Pearson correlation was applied to measure the strength of the linear relationship between two variables.

Results

Identification and characterization of a novel exhausted immune class in late-stage LUSC

We performed a NMF analysis on the bulk RNAseq-based gene expression profiles of 250 late-stage LUSC samples in the training cohort and isolated the transcriptomic signals related to TEX in TME. The dataset of training cohort was effectively separated into four

expression clusters (Fig. 1A). LUSC patients in cluster 2 possessed both high immune and stromal enrichment scores calculated by ssGSEA with bulk RNAseq-based gene expression profiles [12], indicating the significant enrichment of immune cell and stromal component gene expression signatures. Subsequently, we herein named this group as the immune-stromal cluster (Fig. 1B). The immune-stromal cluster also showed the significant enrichment of the signatures of immune cells (Fig. 1C), including immune cell subsets, T cells, B cells, macrophages, tertiary lymphoid structures (TLS), Tem, th1 cell, cytotoxic cells, and T.NK. metagene (all, $P < 0.001$). To further verify the enrichment of these immune cells in the immune-stromal cluster, the absolute proportions of immune cells, which were imputed by CIBERSORT using bulk RNAseq data, were compared between the immune-stromal and the rest clusters. Consistent with our enrichment analysis by ssGSEA, the immune-stromal cluster had higher proportions of CD8 T cells, macrophages, and B cells than the rest clusters (Fig. 1D). Moreover, the leukocyte fraction estimated by DNA methylation [26] was significantly higher in the immune-stromal cluster (median fraction 0.40) than that in the rest clusters (median fraction 0.23) (Fig. 1E). Similarly, the tumour-infiltrating lymphocytes (TILs) evaluated through the pathological images [27] were significantly higher in the immune-stromal cluster (median TILs percentage 10.35) than that in the rest clusters (median TILs percentage 8.25) (Fig. 1F).

To explore TEX in LUSC, we carried out an expression profile analysis of multiple inhibitory receptors, such as *CTLA4*, *PDCD1* (known as *PD-1*), *LAG3*, *BTLA*, *TIGIT*, *HAVCR2* (known as TIM-3), *IDO1*, *SIGLEC7*, and *VISTA*. These inhibitory receptors were significantly upregulated in the tumour samples within the immune-stromal cluster (fold-change > 2 , Benjamini-Hochberg false discovery rate [FDR] $< 1.1 \times 10^{-7}$) (Fig. 1G). Meanwhile, patient samples within the immune-stromal cluster showed a significant enrichment of multiple gene sets for identifying TEX (Fig. 1H), such as CD8 TEX revealed by mass cytometry profiling (CyTOF) (CD8_TEX_CyTOF, $P < 0.001$) [29], human gene sets homologous to CD4 TEX (CD4_TEX_Mouse, $P < 0.001$) and CD8 TEX (CD8_TEX_Mouse, $P < .001$) in mice with chronic viral infection [30], CD8 TEX in hepatocellular carcinoma (HCC) (CD8_TEX_HCC, $P < .001$) [45], and CD8 TEX in melanoma patients (CD8_TEX_Melanoma, $P < 0.001$) [32].

Based on the above expression analysis of inhibitory receptors and the enrichment scores of TEX signal gene sets, we identified a new population subgroup belonging to the immune-stromal cluster, accounting for 36.4% of the training cohort (91/250), referred herein as the

exhausted immune class (EIC) (Fig. 1H). The rest sub-population group in the training cohort was defined as the rest class. To further validate the T cell exhaustion of EIC, we collected 3 gene signatures which characterize three immunosuppressive cell types (NK cells, macrophages, and CD4_{Treg} cells) involved in TEX regulation, and 4 hallmark gene sets of cytokines that promote TEX [31] (Fig. 1H). The high enrichment scores of the above gene sets verified the immunosuppressive TME of the EIC. The above results demonstrated that we successfully identified an exhausted immune class, which showed a significant enrichment of gene expression signatures of exhausted T cells and other immunosuppressive cells in TME.

Cytokine enrichment is a specific molecular feature of EIC

To characterize molecular features of EIC, gene set enrichment analysis (GSEA) based on the gene expression profile of the training cohort identified 48 KEGG pathways (Additional file 1: Fig. S2B) and 17 hallmark gene sets (Additional file 1: Fig. S2C) with enrichment in the EIC. Particularly, cytokine-related pathways and hallmarks, such as cytokine and cytokine receptor interaction pathway, chemokine signalling pathway, and complement hallmark (Fig. 1 I), were significantly enriched in the EIC. Previous studies suggested that severely exhausted T cells may suffer from apoptosis [46, 47], and our analysis showed that apoptosis hallmark was significantly enriched in the EIC. These implied severe T cell exhaustion in the EIC of late-stage LUSC.

In total, 253 metagene-specific genes representing the EIC expression pattern were extracted by NMF and grouped into a network of KEGG pathways by ClueGO (Fig. 1 J) to further reveal their molecular functions. Most of these genes were also associated with cytokine and its complements, which mediate immune response and inflammatory response. Transcriptomics comparison between the EIC and the rest class identified 167 significantly different genes (FDR < 0.01 and fold-change > 2), which were a part of the metagene-specific genes aforementioned. These genes were defined as exhausted immune classifier genes that can be used to confirm the exhausted immune status in LUSC, including 165 upregulated genes related to immunosuppression, such as chemokine and chemokine receptors (*CCL13*, *CCL18*, *CCR4*, *CCR2*, *CXCR3*, *CCL7*, and *CCL19*), interleukin receptor molecules (*IL2RA*, *IL21R*, and *IL17REL*), tumour necrosis factor-related molecules (*TNFSF8* and *TNFRSF9*), complement-related molecules (*C7* and *C4BPA*), and *WNT6* (FDR < 0.001) (Additional file 2: Table S4). The above characterization of molecular functions indicated that cytokine-related expression signals

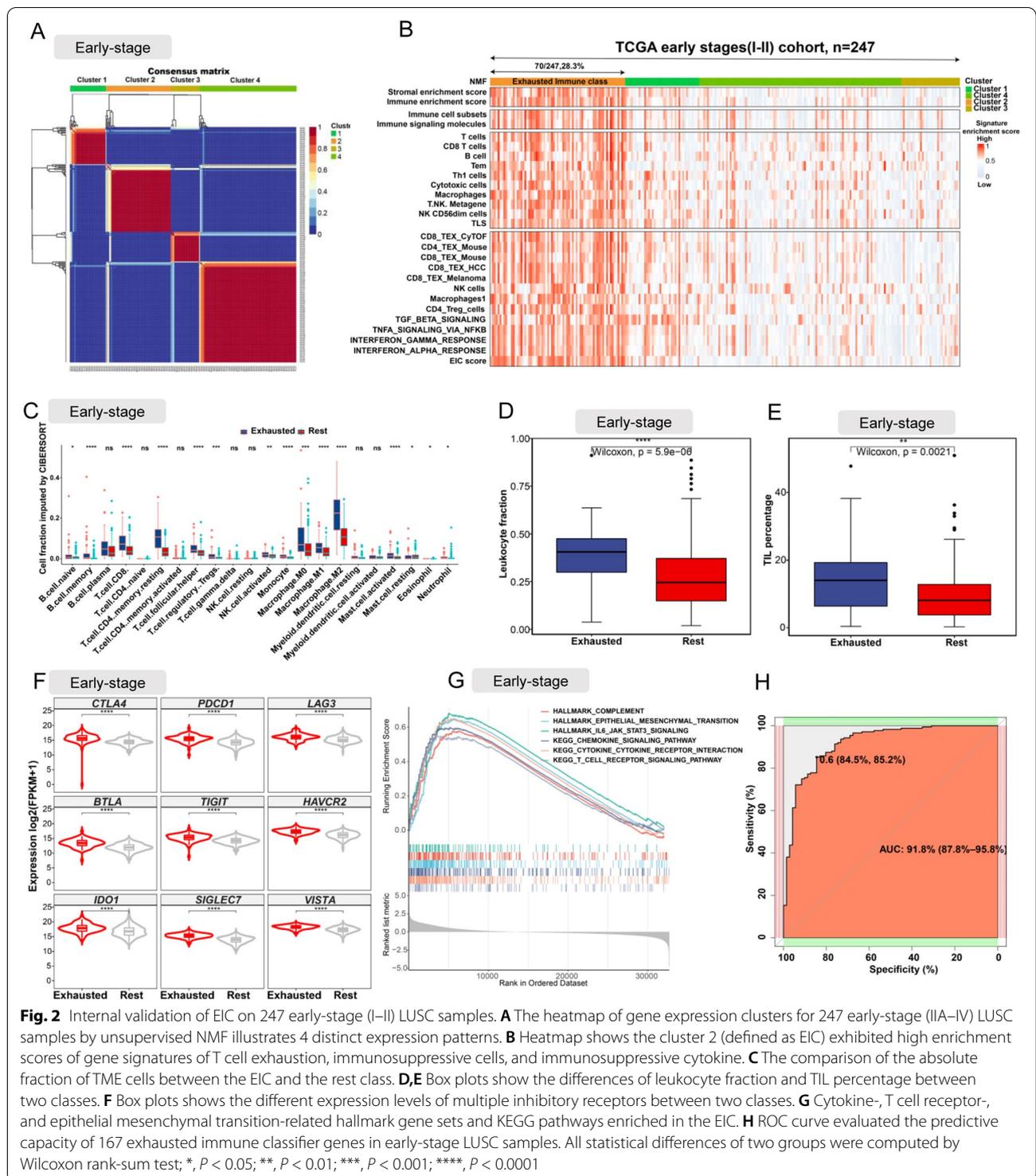
played a dominant role in the exhausted immune status of LUSC.

Internal validation of EIC in early-stage TCGA LUSC

To verify whether exhausted immune class existed in early-stage LUSC, we performed NMF on the bulk RNAseq-based gene expression profiles of the additional 247 early-stage TCGA LUSC samples and subsequently obtained 4 clusters (Fig. 2A). When integrating the clusters with signature enrichment scores calculated by ssGSEA with bulk RNAseq-based gene expression profiles, we observed that stromal and immune enrichment scores of Cluster 2 were higher than those of other clusters (Fig. 2B). Moreover, Cluster 2 also showed a significant enrichment of gene expression signatures of immune cells (CD8 T cells, T cells, macrophages, Tem, th1 cell, cytotoxic cells, NK cell, and exhausted CD8 T cells [all, $P < 0.001$]) and other cytokine hallmarks related to TEX (Fig. 2B). Therefore, Cluster 2 was defined as the Exhausted Immune class in the early-stage TCGA LUSC samples. Consistent with the EIC in the late-stage LUSC, the samples in the EIC of early-stage LUSC had higher proportions of immune cells (CD8 T cells, macrophage, and NK cells) imputed by CIBERSORT (Fig. 2C), higher leukocyte fraction (Fig. 2D), and higher tumour-infiltrating lymphocytes percentage (Fig. 2E) than the rest samples. And co-upregulation of multiple inhibitory receptors was also observed in the EIC of early-stage LUSC (Fig. 2F). GSEA analysis showed significant enrichments of cytokine pathways and complement hallmarks in the EIC of early-stage LUSC. However, we did not observe enrichment of apoptosis hallmark in the EIC of early-stage LUSC, suggesting that the EIC of early-stage LUSC had a lower level of TEX than the EIC of late-stage LUSC. The ROC curve with an AUC of 0.918 (Fig. 2H) showed that our 167 exhausted immune classifier genes had good performance in predicting EIC.

EIC had poorer prognosis in late-stage LUSC

We investigated the prognostic value of exhausted immune status by correlating the classes with clinicopathologic variables. Previous studies suggested that the high densities of TILs correlated with favourable prognoses such as longer progression-free survival (PFS) or improved overall survival (OS) [48, 49]. In our study, the higher percentage of TILs was observed in the EIC compared with the rest class in both late-stage and early-stage LUSC (Figs. 1 F and 2E). However, in the late-stage LUSC, Kaplan–Meier estimates demonstrated that patients within the EIC had significantly poorer OS ($P < 0.001$; Fig. 3A) and PFS ($P < 0.01$; Fig. 3D) than the rest class. Multivariate survival analysis using the Cox regression model indicated that the EIC was retained as



an independent prognostic factor for OS ($P < 0.001$) in the late-stage LUSC (Fig. 3G). With regard to the early-stage LUSC, there was no difference of both OS and PFS between the EIC and the rest class (Fig. 3 B,E). Finally, we

investigated prognostic value in all stage LUSC patients. As expected, patients within the EIC exhibited worse OS and PFS than within the rest class (Fig. 3 C,F). These results verified that although abundant T cells existed

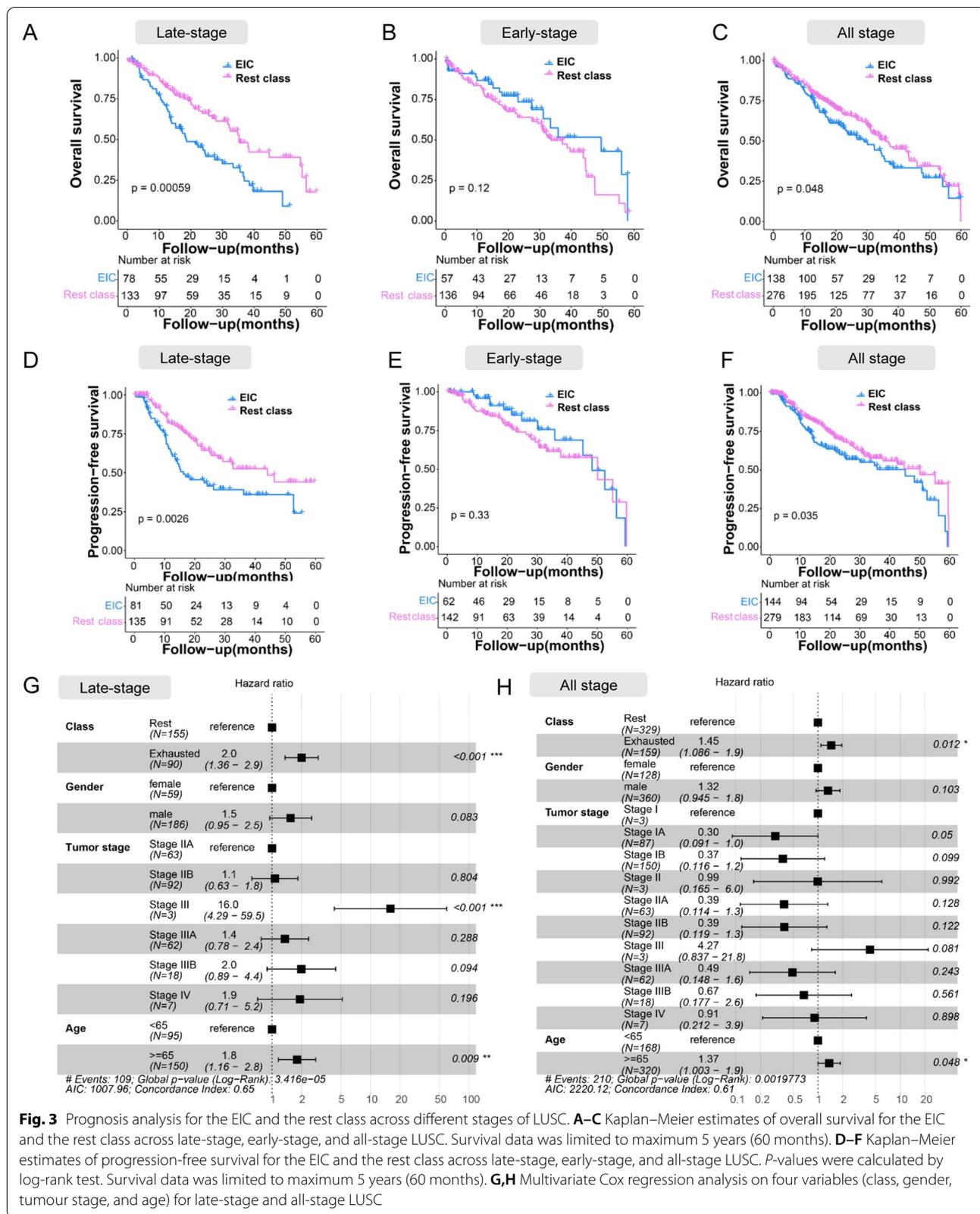


Fig. 3 Prognosis analysis for the EIC and the rest class across different stages of LUSC. **A–C** Kaplan–Meier estimates of overall survival for the EIC and the rest class across late-stage, early-stage, and all-stage LUSC. Survival data was limited to maximum 5 years (60 months). **D–F** Kaplan–Meier estimates of progression-free survival for the EIC and the rest class across late-stage, early-stage, and all-stage LUSC. *P*-values were calculated by log-rank test. Survival data was limited to maximum 5 years (60 months). **G,H** Multivariate Cox regression analysis on four variables (class, gender, tumour stage, and age) for late-stage and all-stage LUSC

in the EIC, the majority of T cells were in the immunosuppressive status and lost the effector function to control tumour progress, resulting in poorer prognosis. The survival results also verified that the EIC in the late-stage LUSC showed more severe T cell exhaustion than in the early-stage LUSC.

EIC is associated with immunotherapy resistance

To investigate the response of patients within the EIC to ICB therapy, we compared PD-L1 expression between the EIC and the rest class and found that the EIC had a higher expression level of PD-L1 than the rest class in both early-stage and late-stage LUSC (Fig. 4 A,B). We also used the tumour immune dysfunction and exclusion (TIDE) algorithm [36] to predict ICB therapy response and observed that the EIC had higher TIDE prediction scores than the rest class in both early-stage and late-stage LUSC (Fig. 4 A,B). A higher TIDE prediction score is usually associated with worse ICB response. Our result suggested that although patients within the EIC had high *PD-L1* expression, they were possibly resistant to ICB therapy.

To further validate the immunotherapy resistance of EIC, we calculated the enrichment score of 167 exhausted immune classifier genes on the RNAseq data of 28 metastatic melanoma patients (GEO: GSE78220) treated with anti-PD-1 [18] and 497 LUSC patients of all tumour stages from TCGA (Project ID: TCGA LUSC, dbGaP Study Accession: phs000178) [14]. We found the melanoma patients who did not respond to ICB therapy had higher enrichment scores than those who respond to ICB therapy (Fig. 4C). As expected, the patients within the EIC also showed higher enrichment scores than the rest class, further evidencing the resistance of EIC (Fig. 4D). We observed a higher expression level of *TGFB1* in the EIC than in the rest class (Fig. 4E), and this is consistent with Mariathasan's study which found cytokine *TGF β* (encoded by *TGFB1*) suppressed anti-tumour immunotherapy [50]. These all suggested the ICB therapy resistance of EIC.

EIC has distinctive methylation patterns

De novo DNA methylation programs can promote T cell exhaustion, and blocking these programs can enhance exhausted T cell rejuvenation, aiding tumour control by immune checkpoint blockade [51]. To explore epigenetic alteration related to deregulated genes in the EIC,

a whole-genome methylation profiling analysis on the TCGA cohort of all tumour stages was conducted and found that 216 CpG sites located in 162 immune-related gene promoter regions were differentially methylated in the EIC compared to the rest class (FDR<0.05) (Fig. 5 A,B and Additional file 2: Table S5). A total of 111 of 162 genes had significant correlations between methylation and gene expression (Additional file 2: Table S6). This indicated that the EIC showed specific methylation profiles, and most of the 162 genes were significantly associated with their promoter methylations.

In particular, multiple genes were regulated by their promoter methylations and involved in the *TGF- β* signaling pathway, which plays an important role in immune evasion [52] and immunotherapy resistance [50]. For example, *ARTN* is a ligand of transforming growth factor-beta (*TGF-beta*) superfamily proteins and binds various *TGF-beta* receptors, leading to the recruitment and activation of *SMAD* family transcription factors [53, 54]. We observed that *ARTN* exhibited a higher methylation level in its promoter region and a lower gene expression level in the EIC compared to the rest class (Fig. 5 C–E). On the contrary, the transcription factor *SMAD7*, which can be activated by *TGF- β* to attenuate or restrain immune cell activation [55], showed a lower methylation level and a higher expression in the EIC (Fig. 5 C–E). In addition, plasma membrane-associated class I myosin (*MYO1G*), C-C Motif Chemokine Receptor 4 (*CCR4*), and interferon regulatory factor 7 (*IRF7*) were also overexpressed and lowly methylated in the EIC (Fig. 5 C–E). *MYO1G* is abundant in T and B lymphocytes and mast cells [56], and *CCR4* is a novel-specific molecular target for immunotherapy in Hodgkin lymphoma, able to regulate the cell trafficking of various leukocyte types [57].

EIC shows no difference in tumour mutational burden or number of neoantigens, but has a lower burden of copy number alterations

Genomic mutations of tumours had a strong association with immunotherapy outcome [58, 59]. We depicted a landscape of commonly mutated genes between the EIC and the rest class in both late-stage and early-stage LUSC. There was no significant difference of individual gene mutation between the EIC and the rest class (Fig 6 A,D). We also linked the burden of somatic mutations, and mutated neoantigens with the exhausted immune status in LUSC. We did not observe significant changes

(See figure on next page.)

Fig. 4 Prediction of ICB therapy resistance. **A,B** Patients in the EIC showed a relatively higher expression level of PD-L1 and higher TIDE prediction score for ICB therapy. **C** Metastatic melanoma patients with no response to anti-PD-1 therapy had higher enrichment scores of 167 exhausted immune classifier genes compared to patients with response. **D** The EIC showed higher enrichment score than the rest class in all tumour stage TCGA LUSC. **E** Box plot shows higher expression of *TGFB1* in the EIC than rest class

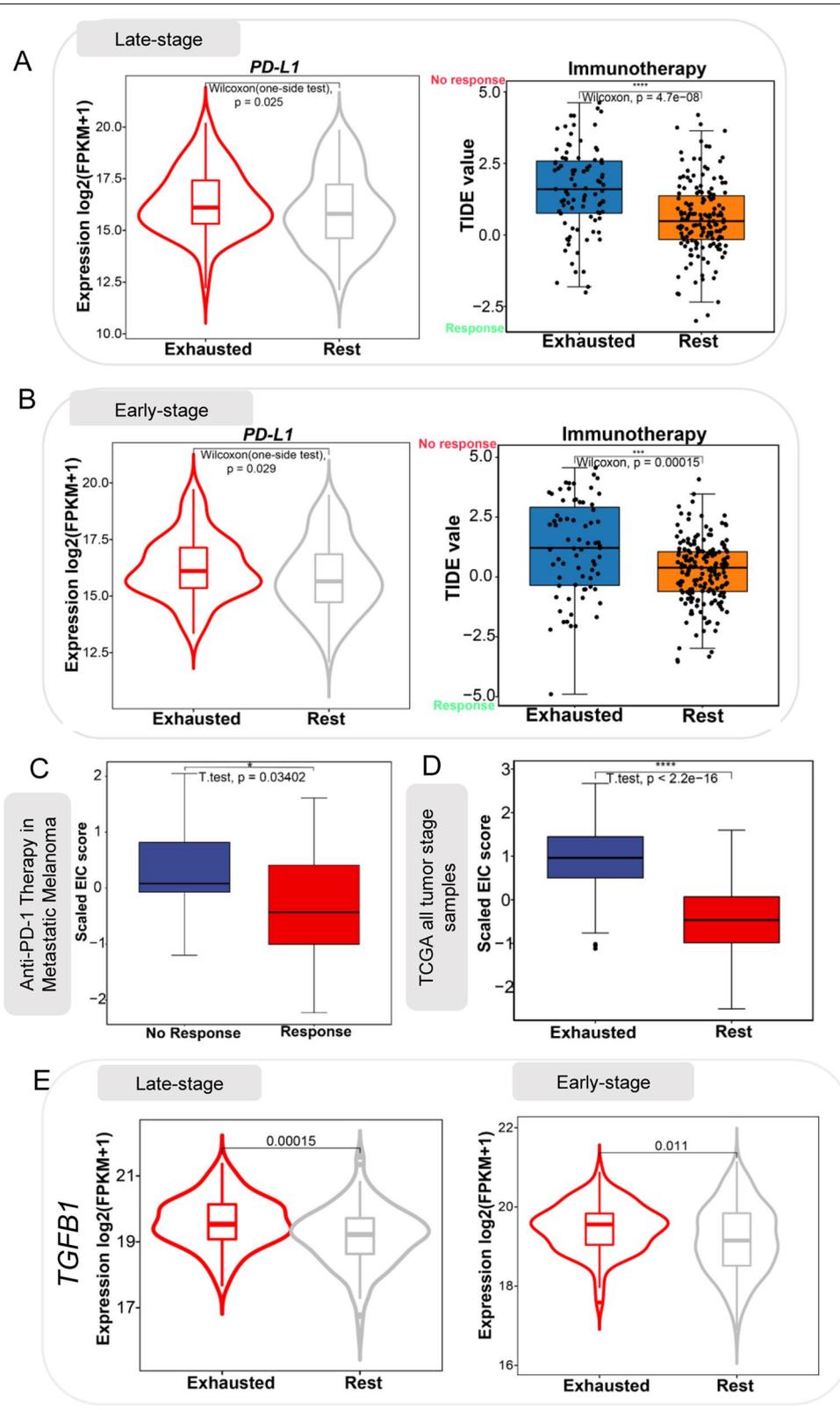


Fig. 4 (See legend on previous page.)

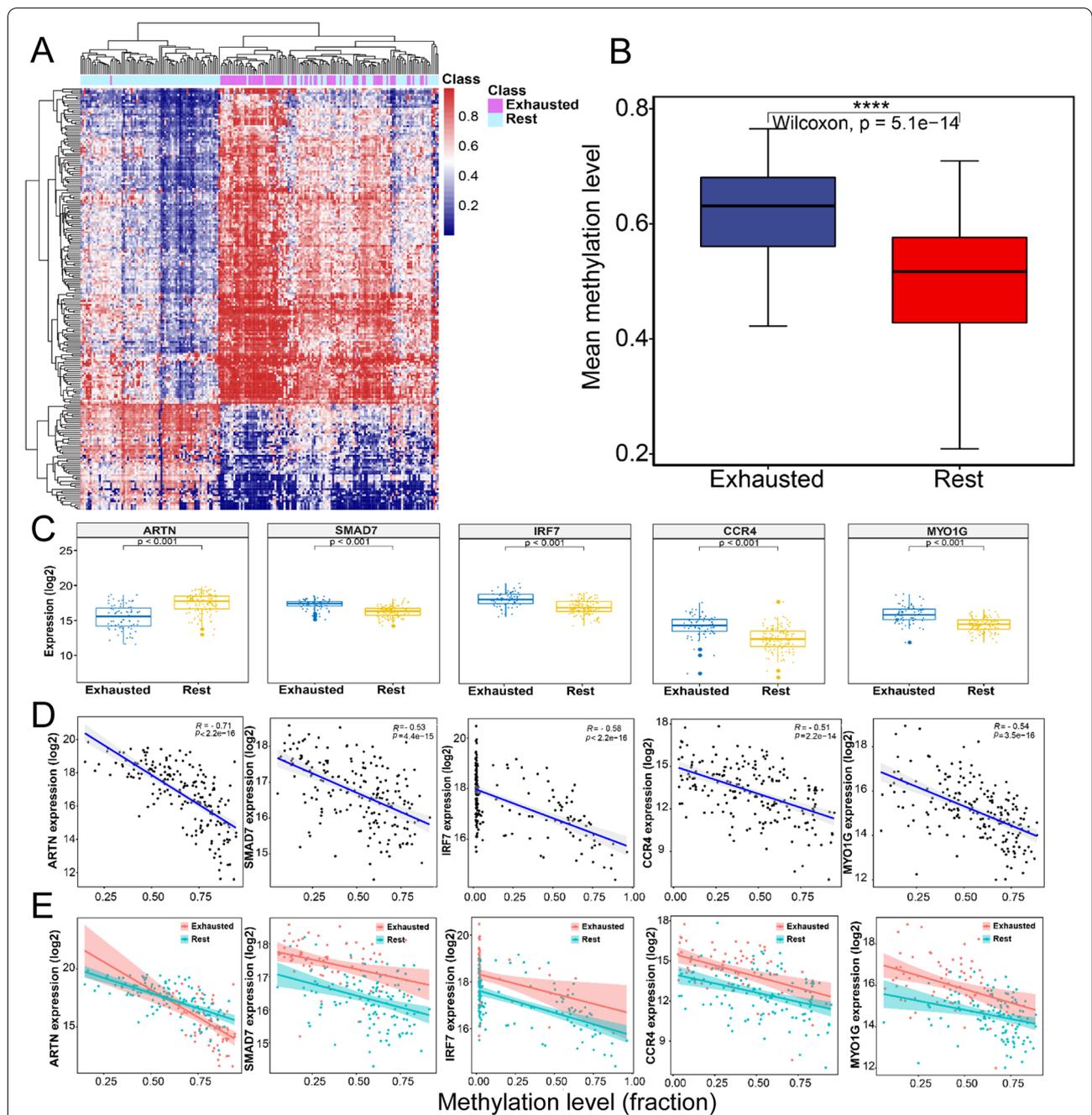


Fig. 5 Distinctive methylation signatures characterized the EIC of LUSC. **A** Hierarchical clustering heatmap of 216 CpG methylation values located within 162 immunosuppression-related gene promoters show significant difference between the EIC and the rest class (FDR < 0.05, diff > 0.2). **B** Boxplot displays the mean methylation levels of 216 CpG sites within 162 immune exhaustion-related gene promoters for 2 classes. Wilcoxon rank-sum test ($p < 0.0001$). Exhausted vs Rest: $p = 5.1 \times 10^{-15}$. **C–E** Correlations between expression and promoter methylation levels for deregulated genes in the EIC to the rest class. **C** *ARTN* exhibited significantly lower expression in the EIC, while *SMAD7*, *IRF7*, *CCR4*, and *MYO1G* were synergistically overexpressed ($P < 0.001$, Wilcoxon rank-sum test). **D** The expressions of *ARTN*, *SMAD7*, *IRF7*, *CCR4*, and *MYO1G* negatively correlated with their promoter methylation level for the whole training cohorts. **E** *SMAD7*, *IRF7*, *CCR4*, and *MYO1G* had lower promoter methylation levels mirroring higher expression levels in the EIC, whereas *ARTN* had an opposite status. Red dots in the plotting represent the members of EIC, and blue dots represent the members of the rest class

of the burden of tumour mutations or neoantigens between two classes in both late-stage (Fig. 6 B,C) and early-stage (Fig. 6 E,F) LUSC. These results indicated that somatic mutations and relevant neoantigens showed no significant association with immunosuppressive TME. Recent studies on molecular characteristics of cancer patients treated with immune checkpoint inhibitors demonstrated that tumour mutational burden (TMB) and gene expression profile-based biomarkers, such as IFN γ -6-related and T cell-inflamed gene expression profiles, had a low correlation and thus were independently predictive of response [60–63]. These studies also indicated that these biomarkers combined with TMB could improve the prediction of response. In our study, the EIC score was calculated by ssGSEA based on gene expression profiles, and the EIC patients showed no difference in TMB compared with the rest class patients. Therefore, we investigated the correlation between the EIC score and TMB. As Fig. 6G displays, there was no significant correlation between the EIC score and TMB across the LUSC cohorts of late-stage ($R=-0.017$, $P=0.79$), early-stage ($R=-0.053$, $P=0.41$), and all-stage ($R=-0.039$, $P=0.39$), indicating that the EIC score was independently predictive of immunotherapy resistance. Both our result and the literature evidences supported that the combination of the EIC score and TMB may improve the prediction of response to immunotherapy.

On the other hand, we observed the samples of EIC showed lower burden of copy number alterations in cytobands. Specifically, the EIC had a lower number of cytoband amplifications than the rest class in both late-stage (Fig. 6 H) and early-stage LUSC (Fig. 6 L), and EIC had a lower number of cytoband deletions too (Fig. 6 I,M). With regard to the amplification (Fig. 6 J,N) or deletion (Fig. 6 K,O) of driver genes, we found these genes also had lower alteration frequency in the EIC than the rest class.

External validation of the novel exhausted immune class across two independent datasets

To validate the presence of exhausted immune status in the training cohort of 250 LUSC samples, we performed unsupervised random forest clustering by using

167 exhausted immune classifier genes. As displayed in the MDS plot (Additional file 1: Fig. S3A), most of the patients were successfully separated into two clusters, which were consistent with the classifications for the EIC and the rest class. The ability of exhausted immune classifier genes to predict immune exhaustion status was estimated in two additionally independent testing datasets ($n=127$ LUSC samples). Similar to the training cohort, about 30–35% of LUSC samples were successfully identified as EIC across the testing datasets. Based on the gene expression profiles, the molecular characterization of the testing datasets also verified that the EIC showed both high immune and stroma enrichment scores.

Taking GSE30219 ($n=61$ LUSC samples) as an example, 19 (31%) LUSC samples were predicted as the EIC, which had remarkable enrichment of immune exhaustion signatures (CD8_TEX_CytoF, CD4_TEX_Mouse, CD8_TEX_Mouse, CD8_TEX_HCC and CD8_TEX_Melanoma; all, $P < 0.001$), TGF- β signalling pathway promoting TEX ($P < 0.001$), IFN signatures (INTERFERON_GAMA_RESPONSE and INTERFERON_ALPHA_RESPONSE; both, $P < 0.001$), and hallmark genes regulated by NF- κ B in response to TNF (TNFA_SIGNALING_VIA_NFKB, $P < 0.001$) (Additional file 1: Fig. S3B). Furthermore, Kaplan–Meier survival analysis showed that the immune exhaustion status was also correlated with poor prognosis ($P=0.013$) (Additional file 1: Fig. S3D). In another example of GSE37745 ($n=66$ LUSC), 23 (34%) patients also showed the enrichment of T cell exhaustion signatures and exhausted molecular features (Additional file 1: Fig. S3C). Kaplan–Meier survival analysis on the 66 patients based on immune exhaustion status indicated that 23 of the 66 patients tended poor prognosis (Additional file 1: Fig. S3E).

The correlation between the clinical outcomes and the immune exhaustion status suggested that exhausted immune cells cannot control the progression of the tumour, leading to deteriorated survival. The prediction of ICB response by TIDE for both GSE30219 and GSE37745 also suggested that the EIC showed potential resistance to ICB therapy (Additional file 1: Fig. S3F&G).

(See figure on next page.)

Fig. 6 Association of EIC with somatic mutations, neoantigens, and copy number alteration. **A, D** The landscape of most frequently mutated genes between the EIC and the rest class in late-stage and early-stage LUSC, respectively. **B, E** Box plots show the number of mutations between two classes in late-stage and early-stage LUSC, respectively. **C, F** Box plots show the number of neoantigens between two classes in late-stage and early-stage LUSC, respectively. **G** Pearson correlation analysis between scaled the EIC score and the number of mutations across LUSC cohorts of late-stage, early-stage, and all-stage, respectively. **H, L** Box plots show significant difference of amplification burden of cytoband between two classes in late-stage and early-stage LUSC, respectively. **I, M** Box plots show significant difference of deletion burden of cytoband between two classes in late-stage and early-stage LUSC, respectively. For late-stage (**J**) and early-stage (**N**) LUSC, the frequency of patients with amplification of driver genes in two classes. For late-stage (**K**) and early-stage (**O**) LUSC, the frequency of patients with deletion of driver genes in two classes. All statistical significances of two classes were computed by Wilcoxon rank-sum test; *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$; ****, $P < 0.0001$

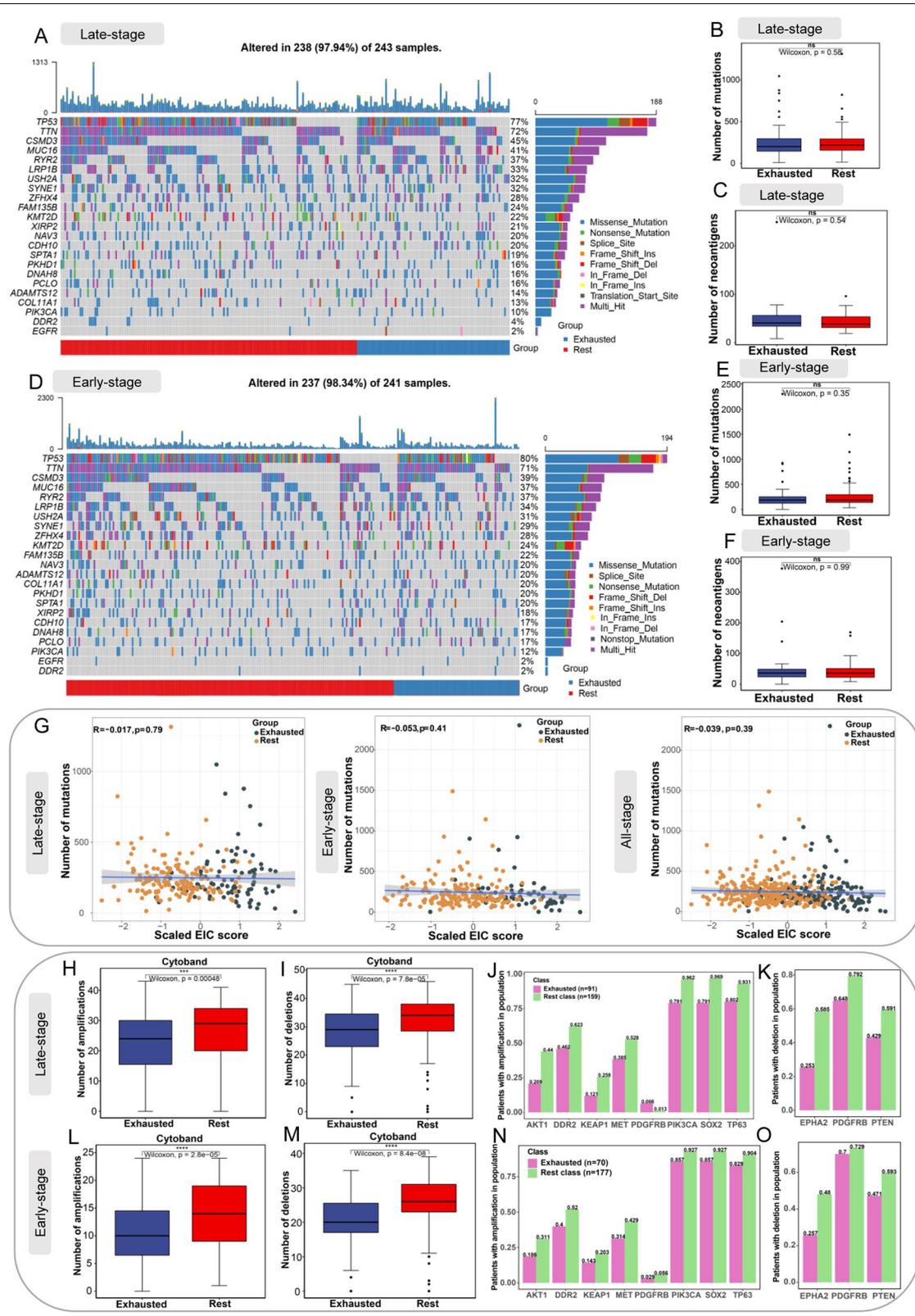


Fig. 6 (See legend on previous page.)

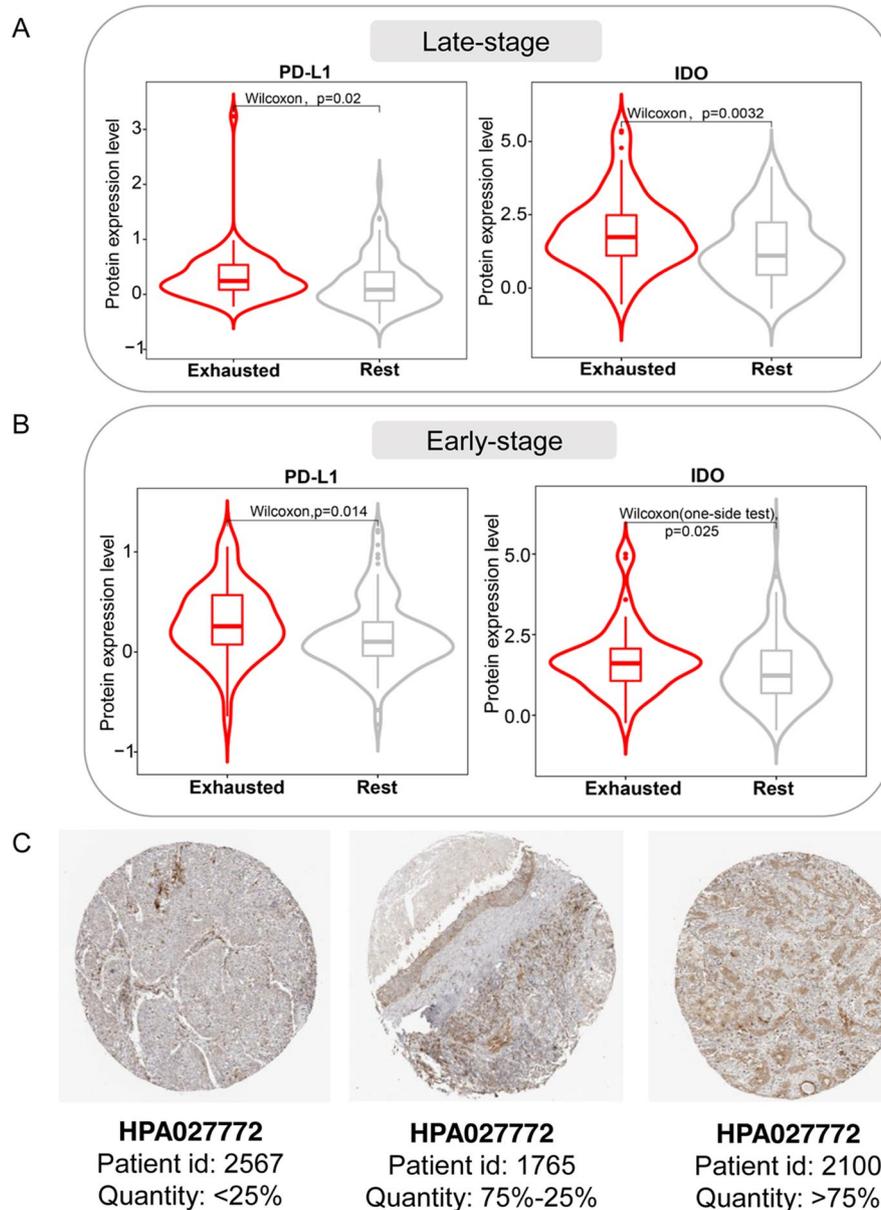


Fig. 7 IDO protein expression analysis in EIC. **A, B** Boxplots showing protein expression difference between the EIC and the rest class in late-stage and early-stage LUSC, respectively. Wilcoxon rank-sum test was utilized to perform comparison analysis. **C** The immunohistochemically stained tissue images show different expression levels of IDO protein across three LUSC patients from HPA

IDO protein expression is higher in EIC

We explored the difference in protein expressions between the EIC and the rest class using the RPPA data of LUSC. As expected, EIC showed significantly higher PD-L1 and IDO (Indoleamine-2,3-dioxygenase encoded by IDO1) protein expressions than the rest class in both late-stage (Fig. 7A) and early-stage (Fig. 7B) LUSC, consistent with the transcriptomic expression analysis aforementioned (Fig. 1G and Fig. 2 F). As well known,

the evaluation of PD-L1 protein expression by immunohistochemistry was applied to select NSCLC patients to receive anti-PD-1 inhibitor treatment. Therefore, we investigated the IDO protein expression of different LUSC patients by immunohistochemically stained tissue images from HPA. We observed that LUSC patients showed different IDO immunohistochemical expression quantities. For example, Fig. 7 C shows three LUSC patients had different IDO expression quantities of less

than 25%, between 25 and 75%, and more than 75%, respectively. Additionally, high IDO activity was reported to be associated with primary resistance to immunotherapy in NSCLC [64]. These evidences indicated different IDO expressions of LUSC patients at the levels of transcriptomics, proteomics, and stained tissues, suggesting that IDO immunohistochemical expression may be a potential biomarker of immunotherapy resistance of LUSC with high PD-L1 expression.

Discussion

Immunotherapy, especially ICB therapy, has revolutionarily transformed the treatment of LUSC and remarkably improved the overall survival of advanced-stage patients [5, 6]. However, more than half of the patients with high PD-L1 expression showed resistance to immune checkpoint inhibitors [6, 65]. Additionally, immune-related adverse events are frequent. Current understanding of the mechanisms of ICB therapy resistance remains limited. Immunosuppressive TME, which consists of tumour cells, immune cells, and other stromal components, may play a vital role in ICB resistance [65]. Thus, characterizing the molecular features of immunosuppressive TME is fundamental for identifying LUSC patients with ICB resistance and thus optimizing immunotherapy strategies.

In our study, we used NMF to deconvolute the gene expression signals derived from exhausted T cells, immune cells, and stromal components in TME of LUSC; then, we successfully identified a novel immunosuppressive class of LUSC (~30% of 624), herein defined as the EIC. Consistent with the exhausted immune class observed in head and neck squamous cell carcinoma [66] and hepatocellular carcinoma [67], our EIC had both high immune and stromal enrichment scores, which suggested the presence of abundant immune cells and stromal components. As expected, the EIC had specifically molecular features, including high immune cell infiltration, co-upregulation of multiple inhibitory receptors, enhanced immunosuppressive cytokine expression, and elevated PD-L1 expression. Of these immune cells, the M2 subtype of tumour-associated macrophages and CD4 Treg cells, as immunosuppressive cells, played a vital role in immune evasion and impacted ICB therapy [68, 69]. The EIC broadly existed in different tumour stages, but there was a potential difference at the T cell exhaustion level between early-stage (stages I–II) and late-stage (stages IIA to IV) LUSC. A previous study suggested that severely exhausted T cell may undergo apoptosis [70]. The apoptosis hallmark gene set was enriched in the EIC of the late stage, but not in the EIC of the early stage, suggesting that the EIC of the late stage had a higher T cell exhaustion level.

Previous observations indicated that high densities of TILs correlated with favourable prognosis [71]. Our EIC also had a high percentage of TILs. However, the EIC showed both poor overall survival and progression-free survival in late-stage LUSC; on the other hand, the EIC showed no significant difference in early-stage LUSC. This suggested that although there was a high T cell infiltration in TME, most T cells were in exhausted status and lacking of effector function to control tumour progress and to prolong survival time of the EIC patients. This also validated the EIC showed more severe immunosuppression in the late stage than in the early stage.

Clinical outcome analysis further confirmed that the EIC of LUSC had immunosuppressive TME. To understand the impact of immunosuppressive TME on ICB therapy, we used the TIDE algorithm to predict response to ICB therapy and observed that the EIC had higher TIDE prediction scores for resistance to ICB therapy [36]. Then 167 exhausted immune classifier genes for identifying EIC in LUSC were also significantly enriched in melanoma patients who showed no response to anti-PD-1 therapy. This further evidenced EIC's potentiality of resistance to ICB therapy. A previous study indicated cytokine TGF β promoted tumour immune evasion and resistance to ICB therapy [50], and the elevated expression level of *TGFBI* in our EIC also verified the potential resistance of the EIC patients.

Blocking de novo DNA methylation program may reactivate the effector function of exhausted T cell and improve the effectiveness of ICB therapy [51]. Our investigation of the epigenetic alterations indicated that the EIC exhibited a unique methylation pattern. Similar to a previous study in HCC [67], some exhaustion-related genes with differentially methylated promoters were involved in the TGF- β signal pathway. In particular, *ARTN* and *SMAD7* were differentially expressed in the EIC compared to the rest class, and their gene expressions were significantly correlated with their promoter methylation. *ARTN* encoded a secreted ligand of the TGF- β protein superfamily and its high expression was associated with the progression of NSCLC [72]. On the contrary, *ARTN* in our EIC had a lower expression level and a higher promoter methylation level, which were correlated with poor PFS. Additionally, a previous study suggested that the overexpression of *SMAD7* may suppress tumour progression by antagonizing TGF- β [73], while the high expression and low methylation of *SMAD7* in our EIC were associated with poor prognosis for LUSC. These suggested that there was a potential mechanism for regulating immunosuppressive TME by DNA methylation.

A previous study suggested that high tumour mutation burden and neoantigen load were related to the

response to ICB therapy [74]. Interestingly, in our study, neither mutational burden nor neoantigen load showed a significant association with the EIC which has high lymphocyte infiltration. The EIC scores for predicting EIC were computed based on gene expression profiles; therefore, we reviewed recent studies on the association between TMB and gene expression profile-based biomarkers for predicting response to immunotherapy. Cristescu's study on pan-tumour genomic biomarkers for predicting clinical response to PD-1 checkpoint blockade indicated that TMB and T cell-inflamed gene expression profiles were independently predictive of response and showed weak correlation [61]. KEYNOTE-028, a clinical trial to evaluate the associations between the biomarkers (e.g. TMB, PD-L1, and T cell-inflamed gene expression profile) and the clinical efficacy of pembrolizumab across 20 cancers, demonstrated that the correlations of TMB with gene expression profile and PD-L1 were low [62]. In addition, recent studies on the immunotherapy of melanoma exhibited that TMB and IFN γ -related gene expression show no significant correlation and both are able to predict response independently [60, 63]. Moreover, TMB did not correlate with the proportion of CD8 T cells estimated by CIBERSORT using gene expression profiles [60]. These studies also demonstrated that the combination of gene expression profile-based biomarkers and TMB could improve the prediction of response to ICB therapy [60, 61]. Consistently, our EIC score had no significant correlation with TMB, suggesting that the combination of the EIC score and TMB may improve the prediction of response to immunotherapy. Specifically, the patients who had high EIC scores and low TMB may have the higher likelihood of immunotherapy resistance, but this needs to be validated in patients treated with immune checkpoint inhibitors in clinical trials. Additionally, the genomic differences (mutation burden and neoantigen) of tumour between the EIC and the rest class in LUSC were also similar to those in other tumours such as head and neck squamous cell carcinoma [66] and hepatocellular carcinoma [67]. These evidences suggested that tumour-intrinsic mutations may not impact immunosuppressive microenvironment and the EIC score is independently predictive for potential immunotherapy resistance. However, we found that the EIC had a lower chromosomal alteration burden and a lower frequency of copy number alteration for the driver genes (e.g., *AKT1*, *DDR2*, *KEAP1*), suggesting that copy number alteration may play an important role in regulating immunosuppressive microenvironment.

The robustness of EIC was successfully verified in three validating datasets including 374 LUSC samples. The EIC in the validation datasets also showed potential resistance to ICB therapy and tended to poor prognosis, verifying

its predictive value. However, this finding needs further validation on LUSC patients treated with ICB therapy.

Understanding the molecular features of immunosuppressive TME is critical for finding successful solutions of TEX reversion and immunotherapy. Comprehensively multidimensional data analysis unfolded a complex immunosuppressive network that may be dominated by cytokines in TME for LUSC. In our study, tumours within the EIC overexpressed PD-1 and CTLA-4, but did not have a higher tumour mutation burden than those within the rest class. However, recent clinical trials indicated that patients benefitting from anti-PD-1 antibody plus anti-CTLA-4 antibody were associated with a high tumour mutational burden. Therefore, patients within the EIC may not be able to respond to the immunotherapy with the two antibodies. Our comparison between the EIC and the rest class indicated that signal alterations related to TGF- β occurred in both transcriptome and epigenome. Thus, patients within the EIC may benefit from TGF- β inhibition plus ICB. In this regard, a phase 1b/2 clinical trial testing the novel TGF- β inhibitor, galunisertib, in combination with nivolumab in advanced refractory solid tumours and in recurrent or refractory NSCLC, or in hepatocellular carcinoma, is currently ongoing (NCT02423343). The rest patients who was lacking of the characteristics of immunosuppressive TME may be selected by immunohistochemical assay to determine the tumour-expressed PD-L1 status for eligibility.

Protein expression analysis indicated that PD-L1 and IDO were significantly increased in EIC, and this was consistent with the high transcriptomic expressions of two corresponding genes. IDO is an immune regulatory enzyme which suppresses T cell response. A recent study demonstrated that high IDO activity is associated with primary resistance to immunotherapy in NSCLC [64]. In our study, a high IDO protein expression level illustrated the potential immunotherapy resistance of EIC. Volaric's study based on immunohistochemistry also demonstrated that IDO is a targetable mechanism of immune resistance frequently coexpressed with PD-L1 [75]. In addition, through checking immunohistochemically stained tissue images from HPA, we found that LUSC patients had different IDO protein expressions. Overall, these evidences suggested that IDO immunohistochemistry may be a biomarker for identifying patients with potential immunotherapy resistance.

Conclusions

In conclusion, we identified an immunosuppressive class accounting for approximately 30% LUSC patients, which had elevated PD-L1 expression but showed potential resistance to ICB therapy and uniquely

immunosuppressive molecular features of TME. Our findings provide new insights for understanding the molecular mechanism of ICB therapy resistance and tailoring appropriate immunotherapy strategies for patients with different molecular characteristics.

Abbreviations

BTLA: Band T lymphocyte attenuator; *CCR4*: C-C Motif Chemokine Receptor 4; *CTLA-4*: Cytotoxic T lymphocyte antigen-4; *EIC*: Exhausted immune class; *GEO*: Gene Expression Omnibus; *GSEA*: Gene set enrichment analysis; *HCC*: Hepatocellular carcinoma; *HPA*: Human Protein Atlas; *ICB*: Immune checkpoint blockade; *IDO*: Indoleamine 2,3-dioxygenase; *IFN γ* : Interferon gamma; *IRF7*: Interferon regulatory factor 7; *LAG3*: Lymphocyte activation gene 3; *LUAD*: Lung adenocarcinoma; *LUSC*: Lung squamous cell carcinoma; *MYO1G*: Plasma membrane-associated class I myosin; *NMF*: Non-negative matrix factorization; *NK*: Natural killer; *NSCLC*: Non-small-cell lung cancer; *OS*: Overall survival; *PD-1*: Programmed death 1; *PD-L1*: Programmed Cell Death 1 Ligand 1; *PFS*: Progression-free survival; *RNAseq*: RNA sequencing; *ROC*: Receiver operating characteristic; *RPPAs*: Reverse-phase protein arrays; *ssGSEA*: Single sample gene set enrichment analysis; *TCGA*: The Cancer Genome Atlas; *TIDE*: The tumour immune dysfunction and exclusion; *TEX*: T cell exhaustion; *TGF- β* : Transforming growth factor-beta; *TIGIT*: T cell immunoreceptor with Ig and ITIM domains; *TILs*: Tumour-infiltrating lymphocytes; *TIM-3*: T cell immunoglobulin and mucin domain protein; *TKIs*: Tyrosine kinase inhibitor; *TLS*: Tertiary lymphoid structure; *TMB*: Tumour mutational burden; *TME*: Tumour microenvironment.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13073-022-01079-x>.

Additional file 1: Figure S1. Analysis flowchart of this study. **Figure S2.** Identification and characterization of the exhausted immune expression pattern. **Figure S3.** The external validation of the Exhausted Immune class.

Additional file 2: Table S1. Publicly available data sets used in this study. **Table S2.** Tools and software used in this study. **Table S3.** Publicly available gene signatures used in the study. **Table S4.** Exhausted Immune classifier (167 genes) with significantly different expression between the Exhausted Immune class and the rest class. **Table S5.** Differently methylation comparison of the 216 CpG located in the promoters of 162 exhaustion related genes. **Table S6.** Correlation between gene expression and methylation in the 162 exhaustion related genes with differentially methylated promoter.

Authors' contributions

M.Y. designed the study, performed the major data analysis, and drafted the manuscript; C.L. collected a part of data, performed a part of data analysis, and help to generated figures. Y.W. performed a part of data analysis. K.C. helped to generate figures. H.Z. assisted project management and reviewed and revised the manuscript. W.L. provided funding source, designed, oversaw, and supervised the project and edited, reviewed, and finalized the paper. All authors read and approved the final manuscript.

Funding

This work was supported by the National Key R&D Program of China (2021YFF1200903, 2021YFF1200900, 2016YFC0901604 & 2018YFC0910401), the Natural Science Foundation of Guangdong Province (2021A1515012108), the Guangdong Project (2017GC010608), and the Support Scheme of Guangzhou for Leading Talents in Innovation and Entrepreneurship (2020007).

Availability of data and materials

All data analysed during this study were retrieved from public repositories, which have been listed in the Additional files. The source codes generated to build the web application are available at the GitHub repository (https://github.com/Lilab-SYSU/LUSC_Tex) [43].

Declarations

Ethics approval and consent to participate

This study only uses publicly available data. The research performed conformed to the principles of the Helsinki Declaration.

Consent for publication

Not applicable.

Competing interests

The authors declare that they do not have any competing interests.

Author details

¹Zhongshan School of Medicine, Sun Yat-sen University, Guangzhou 510080, China. ²Center for Precision Medicine, Sun Yat-sen University, Guangzhou 510080, China. ³Key Laboratory of Tropical Disease Control of Ministry of Education, Sun Yat-Sen University, Guangzhou 510080, China.

Received: 25 November 2021 Accepted: 24 June 2022

Published online: 08 July 2022

References

- Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2018;68(6):394–424.
- Duma N, Santana-Davila R, Molina JR. Non-small cell lung cancer: epidemiology, screening, diagnosis, and treatment. *Mayo Clin Proc*. 2019;94(8):1623–40.
- Youlten DR, Cramb SM, Baade PD. The International Epidemiology of Lung Cancer: geographical distribution and secular trends. *J Thorac Oncol*. 2008;3(8):819–31.
- Morgensztern D, Campo MJ, Dahlberg SE, Doebele RC, Garon E, Gerber DE, et al. Molecularly targeted therapies in non-small-cell lung cancer annual update 2014. *J Thorac Oncol*. 2015;10(1 Suppl 1):S1–63.
- Socinski MA, Obasaju C, Gandara D, Hirsch FR, Biondi P, Bunn PA Jr, et al. Current and emergent therapy options for advanced squamous cell lung cancer. *J Thorac Oncol*. 2018;13(2):165–83.
- Garon EB, Rizvi NA, Hui R, Leighl N, Balmanoukian AS, Eder JP, et al. Pembrolizumab for the treatment of non-small-cell lung cancer. *N Engl J Med*. 2015;372(21):2018–28.
- Thompson JA. New NCCN Guidelines: recognition and management of immunotherapy-related toxicity. *J Natl Compr Canc Netw*. 2018;16(5):594–6.
- Kennedy LB, Salama AKS. A review of cancer immunotherapy toxicity. *CA Cancer J Clin*. 2020;70(2):86–104.
- Speiser DE, Ho PC, Verdeil G. Regulatory circuits of T cell function in cancer. *Nat Rev Immunol*. 2016;16(10):599–611.
- Jiang Y, Li Y, Zhu B. T-cell exhaustion in the tumor microenvironment. *Cell Death Dis*. 2015;6:e1792.
- Murciano-Goroff YR, Warner AB, Wolchok JD. The future of cancer immunotherapy: microenvironment-targeting combinations. *Cell Res*. 2020;30(6):507–19.
- Yoshihara K, Shahmoradgoli M, Martinez E, Vegesna R, Kim H, Torres-Garcia W, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat Commun*. 2013;4:2612.
- Moffitt RA, Marayati R, Flate EL, Volmar KE, Loeza SG, Hoadley KA, et al. Virtual microdissection identifies distinct tumor- and stroma-specific subtypes of pancreatic ductal adenocarcinoma. *Nat Genet*. 2015;47(10):1168–78.
- Cancer Genome Atlas Research N. Comprehensive genomic characterization of squamous cell lung cancers. *Nature*. 2012;489(7417):519–25.
- Rousseaux S, Debernardi A, Jacquiou B, Vitte AL, Vesin A, Nagy-Mignotte H, et al. Ectopic activation of germline and placental genes identifies aggressive metastasis-prone lung cancers. *Sci Transl Med*. 2013;5(186):186ra66.
- Botling J, Edlund K, Lohr M, Hellwig B, Holmberg L, Lambe M, et al. Biomarker discovery in non-small cell lung cancer: integrating gene

- expression profiling, meta-analysis, and tissue microarray validation. *Clin Cancer Res.* 2013;19(1):194–204.
17. Davis S, Meltzer PS. GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor. *Bioinformatics.* 2007;23(14):1846–7.
 18. Hugo W, Zaretsky JM, Sun L, Song C, Moreno BH, Hu-Lieskovan S, et al. Genomic and transcriptomic features of response to anti-PD-1 therapy in metastatic melanoma. *Cell.* 2016;165(1):35–44.
 19. Wolchok JD, Hoos A, O'Day S, Weber JS, Hamid O, Lebbé C, et al. Guidelines for the evaluation of immune therapy activity in solid tumors: immune-related response criteria. *Clin Cancer Res.* 2009;15(23):7412–20.
 20. Brunet JP, Tamayo P, Golub TR, Mesirov JP. Metagenes and molecular pattern discovery using matrix factorization. *P Natl Acad Sci USA.* 2004;101(12):4164–9.
 21. Barbie DA, Tamayo P, Boehm JS, Kim SY, Moody SE, Dunn IF, et al. Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. *Nature.* 2009;462(7269):108–U22.
 22. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A.* 2005;102(43):15545–50.
 23. Hanzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-Seq data. *BMC Bioinformatics.* 2013;14(1):7.
 24. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, et al. Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods.* 2015;12(5):453–7.
 25. Li T, Fu J, Zeng Z, Cohen D, Li J, Chen Q, et al. TIMER2.0 for analysis of tumor-infiltrating immune cells. *Nucleic Acids Res.* 2020;48(W1):W509–W14.
 26. Thorsson V, Gibbs DL, Brown SD, Wolf D, Bortone DS, Ou Yang TH, et al. The immune landscape of cancer. *Immunity.* 2018;48(4):812–30 e14.
 27. Saltz J, Gupta R, Hou L, Kurc T, Singh P, Nguyen V, et al. Spatial organization and molecular correlation of tumor-infiltrating lymphocytes using deep learning on pathology images. *Cell Rep.* 2018;23(1):181–93 e7.
 28. Kim H, Park H. Sparse non-negative matrix factorizations via alternating non-negativity-constrained least squares for microarray data analysis. *Bioinformatics.* 2007;23(12):1495–502.
 29. Bengsch B, Ohtani T, Khan O, Setty M, Manne S, O'Brien S, et al. Epigenomic-guided mass cytometry profiling reveals disease-specific features of exhausted CD8 T cells. *Immunity.* 2018;48(5):1029–45 e5.
 30. Crawford A, Angelosanto JM, Kao C, Doering TA, Odorizzi PM, Barnett BE, et al. Molecular and transcriptional basis of CD4(+) T cell dysfunction during chronic infection. *Immunity.* 2014;40(2):289–302.
 31. McLane LM, Abdel-Hakeem MS, Wherry EJ. CD8 T cell exhaustion during chronic viral infection and cancer. *Annu Rev Immunol.* 2019;37:457–95.
 32. Sade-Feldman M, Yizhak K, Bjorgaard SL, Ray JP, de Boer CG, Jenkins RV, et al. Defining T cell states associated with response to checkpoint immunotherapy in melanoma. *Cell.* 2019;176(1–2):404.
 33. Bindea G, Mlecnik B, Hackl H, Charoentong P, Tosolini M, Kirilovsky A, et al. ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics.* 2009;25(8):1091–3.
 34. Otasek D, Morris JH, Boucas J, Pico AR, Demchak B. Cytoscape automation: empowering workflow-based network analysis. *Genome Biol.* 2019;20(1):185.
 35. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 2014;15(12):550.
 36. Jiang P, Gu S, Pan D, Fu J, Sahu A, Hu X, et al. Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response. *Nat Med.* 2018;24(10):1550–8.
 37. Hugo W, Zaretsky JM, Sun L, Song C, Moreno BH, Hu-Lieskovan S, et al. Genomic and transcriptomic features of response to anti-PD-1 therapy in metastatic melanoma. *Cell.* 2017;168(3):542.
 38. Mayakonda A, Lin DC, Assenov Y, Plass C, Koeffler HP. Maftools: efficient and comprehensive analysis of somatic variants in cancer. *Genome Res.* 2018;28(11):1747–56.
 39. Rooney MS, Shukla SA, Wu CJ, Getz G, Hacohen N. Molecular and genetic properties of tumors associated with local immune cytolytic activity. *Cell.* 2015;160(1–2):48–61.
 40. Mermel CH, Schumacher SE, Hill B, Meyerson ML, Beroukhim R, Getz G. GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol.* 2011;12(4):R41.
 41. Teschendorff AE, Marabita F, Lechner M, Bartlett T, Tegner J, Gomez-Cabrero D, et al. A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data. *Bioinformatics.* 2013;29(2):189–96.
 42. Breiman L. Random forests. *Mach Learn.* 2001;45(1):5–32.
 43. Minglei Y. LUSC_Tex. Github. 2022. https://github.com/Lilab-SYSU/LUSC_Tex.
 44. Uhlén M, Fagerberg L, Hallström BM, Lindskog C, Oksvold P, Mardinoglu A, et al. Proteomics. Tissue-based map of the human proteome. *Science.* 2015;347(6220):1260419.
 45. Zheng C, Zheng L, Yoo JK, Guo H, Zhang Y, Guo X, et al. Landscape of infiltrating T cells in liver cancer revealed by single-cell sequencing. *Cell.* 2017;169(7):1342–56 e16.
 46. Moskopidis D, Lechner F, Pircher H, Zinkernagel RM. Virus persistence in acutely infected immunocompetent mice by exhaustion of antiviral cytotoxic effector T cells. *Nature.* 1993;362(6422):758–61.
 47. Barathan M, Mohamed R, Yong YK, Kannan M, Vadivelu J, Saeidi A, et al. Viral persistence and chronicity in hepatitis C virus infection: role of T-cell apoptosis, Senescence and Exhaustion. *Cells.* 2018;7(10):165.
 48. Mlecnik B, Bindea G, Pagès F, Galon J. Tumor immunosurveillance in human cancers. *Cancer Metastasis Rev.* 2011;30(1):5–12.
 49. Fridman WH, Pages F, Sautes-Fridman C, Galon J. The immune contexture in human tumours: impact on clinical outcome. *Nat Rev Cancer.* 2012;12(4):298–306.
 50. Mariathasan S, Turley SJ, Nickles D, Castiglioni A, Yuen K, Wang Y, et al. TGFβ attenuates tumour response to PD-L1 blockade by contributing to exclusion of T cells. *Nature.* 2018;554(7693):544–8.
 51. Ghoneim HE, Fan Y, Moustaki A, Abdelsamed HA, Dash P, Dogra P, et al. De novo epigenetic programs inhibit PD-1 blockade-mediated T cell rejuvenation. *Cell.* 2017;170(1):142–57 e19.
 52. Batlle E, Massagué J. Transforming growth factor-beta signaling in immunity and cancer. *Immunity.* 2019;50(4):924–40.
 53. Lin LF, Doherty DH, Lile JD, Bektesh S, Collins F. GDNF: a glial cell line-derived neurotrophic factor for midbrain dopaminergic neurons. *Science.* 1993;260(5111):1130–2.
 54. Airaksinen MS, Saarma M. The GDNF family: signalling, biological functions and therapeutic value. *Nat Rev Neurosci.* 2002;3(5):383–94.
 55. Derynck R, Zhang YE. Smad-dependent and Smad-independent pathways in TGF-beta family signalling. *Nature.* 2003;425(6958):577–84.
 56. Pierce RA, Field ED, Mutis T, Golovina TN, Von Kap-Herr C, Wilke M, et al. The HA-2 minor histocompatibility antigen is derived from a diallelic gene encoding a novel human class I myosin protein. *J Immunol.* 2001;167(6):3223–30.
 57. Ishida T, Ishii T, Inagaki A, Yano H, Kusumoto S, Ri M, et al. The CCR4 as a novel-specific molecular target for immunotherapy in Hodgkin lymphoma. *Leukemia.* 2006;20(12):2162–8.
 58. Boussiotis VA. Somatic mutations and immunotherapy outcome with CTLA-4 blockade in melanoma. *N Engl J Med.* 2014;371(23):2230–2.
 59. Chan TA, Yarchoan M, Jaffee E, Swanton C, Quezada SA, Stenzinger A, et al. Development of tumor mutation burden as an immunotherapy biomarker: utility for the oncology clinic. *Ann Oncol.* 2019;30(1):44–56.
 60. Newell F, Pires da Silva I, Johansson PA, Menzies AM, Wilmott JS, Addala V, et al. Multiomic profiling of checkpoint inhibitor-treated melanoma: identifying predictors of response and resistance, and markers of biological discordance. *Cancer Cell.* 2022;40(1):88–102.e7.
 61. Cristescu R, Mogg R, Ayers M, Albright A, Murphy E, Yearley J, et al. Pan-tumor genomic biomarkers for PD-1 checkpoint blockade-based immunotherapy. *Science.* 2018;362(6411):eaar3593.
 62. Ott PA, Bang YJ, Piha-Paul SA, Razak ARA, Bannouna J, Soria JC, et al. T-cell-inflamed gene expression profile, programmed death ligand 1 expression, and tumor mutational burden predict efficacy in patients treated with pembrolizumab across 20 cancers: KEYNOTE-028. *J Clin Oncol.* 2019;37(4):318–27.
 63. Rozeman EA, Hoefsmit EP, Reijers ILM, Saw RPM, Versluis JM, Krijgsman O, et al. Survival and biomarker analyses from the OpACIN-neo and OpACIN neoadjuvant immunotherapy trials in stage III melanoma. *Nat Med.* 2021;27(2):256–63.
 64. Kocher F, Amann A, Zimmer K, Geisler S, Fuchs D, Pichler R, et al. High indoleamine-2,3-dioxygenase 1 (IDO) activity is linked to primary resistance to immunotherapy in non-small cell lung cancer (NSCLC). *Transl Lung Cancer Res.* 2021;10(1):304–13.

65. Horvath L, Thienpont B, Zhao L, Wolf D, Pircher A. Overcoming immunotherapy resistance in non-small cell lung cancer (NSCLC) - novel approaches and future outlook. *Mol Cancer*. 2020;19(1):141.
66. Chen YP, Wang YQ, Lv JW, Li YQ, Chua MLK, Le QT, et al. Identification and validation of novel microenvironment-based immune molecular subgroups of head and neck squamous cell carcinoma: implications for immunotherapy. *Ann Oncol*. 2019;30(1):68–75.
67. Sia D, Jiao Y, Martinez-Quetglas I, Kuchuk O, Villacorta-Martin C, Castro de Moura M, et al. Identification of an immune-specific class of hepatocellular carcinoma, based on molecular features. *Gastroenterology*. 2017;153(3):812–26.
68. Tay RE, Richardson EK, Toh HC. Revisiting the role of CD4(+) T cells in cancer immunotherapy-new insights into old paradigms. *Cancer Gene Ther*. 2021;28(1-2):5–17.
69. Duan Z, Luo Y. Targeting macrophages in cancer immunotherapy. *Signal Transduct Target Ther*. 2021;6(1):127.
70. Kahan SM, Wherry EJ, Zajac AJ. T cell exhaustion during persistent viral infections. *Virology*. 2015;479-480:180–93.
71. Mlecnik B, Bindea G, Pages F, Galon J. Tumor immunosurveillance in human cancers. *Cancer Metast Rev*. 2011;30(1):5–12.
72. Tang JZ, Kong XJ, Kang J, Fielder GC, Steiner M, Perry JK, et al. Artemin-stimulated progression of human non-small cell lung carcinoma is mediated by BCL2. *Mol Cancer Ther*. 2010;9(6):1697–708.
73. Yan X, Liu Z, Chen Y. Regulation of TGF-beta signaling by Smad7. *Acta Biochim Biophys Sin (Shanghai)*. 2009;41(4):263–72.
74. Bai R, Lv Z, Xu D, Cui J. Predictive biomarkers for cancer immunotherapy with immune checkpoint inhibitors. *Biomark Res*. 2020;8:34.
75. Volaric A, Gentzler R, Hall R, Mehaffey JH, Stelow EB, Bullock TN, et al. Indoleamine-2,3-dioxygenase in non-small cell lung cancer: a targetable mechanism of immune resistance frequently coexpressed with PD-L1. *The American journal of surgical pathology*. 2018;42(9):1216–23.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

