Transcriptional Characterization Of The Tumor Immune Microenvironment And Its Prognostic Value For Locally Advanced Lung Adenocarcinoma In A Chinese Population

Objective: We investigated the relationship of the transcriptional tumor immune microenvironment with prognosis of patients with locally advanced lung adenocarcinoma (LUAD).

Materials and methods: A targeted RNA-Seq approach was used to measure the abundance of 395 immune-related transcripts of 24 formalin-fixed paraffin embedded (FFPE) tumor specimens from our institution and transcription data of 85 matched LUAD samples from The Cancer Genome Atlas (TCGA). Gene set variation analysis (GSVA) was used to identify gene sets related to prognosis, and the microenvironment cell-population (MCP)-counter method was used to quantify infiltrated immune cells. Survival analysis with the log rank test was used to determine the relationships of different immune-related transcripts with prognosis. Cox proportional hazards models were also used to identify risk factors associated with poor prognosis.

Results: Among our patients, GSVA and the log rank test demonstrated that enrichment of the antigen processing pathway \( (P = 0.01) \) correlated with a favorable prognosis. MCP-counter and survival analysis demonstrated that greater CD8 T cell infiltration correlated with a favorable prognosis \( (P = 0.05) \), but greater infiltration of neutrophils \( (P = 0.014) \) and NK cells \( (P = 0.015) \) correlated with poor prognosis. Cox hazard analysis showed that greater infiltration of neutrophils was an independent risk factor for poor prognosis. These results were consistent with LUAD data from TCGA.

Conclusion: When integrated with computational bioinformatics methods, targeted RNA-Seq from FFPE specimens provides profiles of the tumor immune microenvironment that have prognostic value for patients with locally advanced LUAD.

Keywords: NSCLC, lung adenocarcinoma, RNA-Seq, tumor immune microenvironment, transcriptomic markers

Introduction

Lung cancer is one of the most common malignancies worldwide, and lung adenocarcinoma (LUAD) accounts for about 60% of all lung cancer cases.\(^1\)\(^-\)\(^4\) The five-year survival rate of LUAD is about 50% for early-stage disease, but only about 20% for locally advanced disease.\(^5\) The presence of lymphatic metastasis is associated with a poor prognosis, a higher rate of local regional relapse, and distant metastasis.\(^6\) However, there is heterogeneity in the prognosis of patients with locally advanced LUAD.\(^7\)\(^-\)\(^8\) Some patients experience relapse or distant metastasis...
Researchers have previously studied the characterization of various types of cells and cell components. \cite{10,11} Researchers have increasingly recognized tumor-infiltrating lymphocytes (TILs) as cells that can profoundly affect the course of tumor progression and the therapeutic response. \cite{12,13,14} Most of these previous studies investigated the composition of TILs using immunohistochemistry (IHC), immunofluorescence, and flow cytometry, but these methods can only provide simultaneous examination of a limited number of immune cells. Thus, little is known about how all these cell components together affect tumor progression and alter the tumor during the course of therapy.

Several recent computational bioinformatic methods quantified immune infiltrates based on transcriptomics data. \cite{15} The ability to simultaneously quantify multiple cell types within a tissue specimen is critical for determining the function of immune infiltrates within a tumor, and may also improve our understanding of the mechanisms underlying different therapeutic responses and allow stratification of patients with different prognoses. \cite{16,17} Previous studies that characterized transcription within the tumor immune microenvironment have used whole transcriptome sequencing data. \cite{15,18,19} However, this method requires many fresh tumor specimens, which are not always available, and the cost is currently too high for routine use in clinical practice.

In the present study, we used targeted RNA-Seq to measure the abundance of 395 immune transcripts from formalin-fixed paraffin embedded (FFPE) LUAD specimens to characterize the transcriptome profile of the tumor immune microenvironment. Although this approach only examined 395 immune-related genes, it provided more deep sequencing, and had superior sensitivity to bulk whole transcriptome sequencing. \cite{20,21} Our purpose was to demonstrate the combined use of computational bioinformatic methods with a targeted RNA-Seq assay to profile the tumor immune microenvironment and to stratify patients with LUAD according to prognosis.

**Materials And Methods**

**Patient Enrollment**

This study was approved by the Local Ethics Committee of Xiangya Hospital of Central South University (Changsha, China). All patients provided written informed consent prior to participation. Patients who met the following criteria were enrolled: receipt of surgery as initial treatment, but no preoperative therapy; confirmation of postoperative pathology as locally advanced LUAD (N1–2) with pathological stage of IIIB–IIIA; refusal of postoperative adjuvant therapy; receipt of surgery in our department from December 2014 to December 2016; no other malignant tumors; attainment of the clinical endpoint (local lymph node recurrence or distant metastasis); no previous myocardial infarction, unstable angina, or stroke; no systemic immune disease.

The clinical data of all 24 patients were collected, and FFPE specimens were retrieved from the tissue archive of Xiangya Hospital. Sections from the FFPE blocks (stained with hematoxylin and eosin) were used to confirm the presence of tumor tissue by a qualified pathologist. Only FFPE specimens with 10% or more viable tumor cells were used for RNA extraction.

**RNA Extraction, RNA-Seq Library Preparation And Sequencing**

The truXTRAC™ FFPE RNA Kit (Covaris, Inc., Woburn, MA) was used for extraction of RNA from FFPE sections. Thus, RNA was extracted from FFPE lysates, purified, eluted, and quantified using the Quant-iT RNA HS Assay Kit (Thermo Fisher Scientific, Waltham, MA).

The Ion AmpliSeq™ targeted sequencing system (Thermo Fisher Scientific, Waltham, MA) was used for preparation of RNA-Seq libraries. The assay was designed to quantify the expression of 395 immune-related genes, including 10 housekeeping (HK) genes as controls (Table S1). Following reverse transcription into cDNA, targets of the 395 genes were amplified, and barcode adapters were ligated to the partially digested amplicons. After purification and normalization, equimolar libraries were pooled. Following enrichment and template preparation using the Ion Chef™ system (Thermo Fisher Scientific), 200-bp sequencing was performed on the Ion Torrent S5 system (Thermo Fisher Scientific) to obtain mapped reads (about 2- to 3-million per sample).

**Next Generation Sequencing (NGS), Quality Control And Gene Expression Normalization**

Sequencing data were analyzed using Ion Torrent Suite software version 5.2.0 (Thermo Fisher Scientific) for reference mapping and base calling. Quality control was conducted using acceptance criteria, and gene expression normalization was performed using HK gene background-subtracted reads as a control, as described by...
Finally, the normalized gene expression, presented as normalized reads per million (nRPM), were log₂-transformed.²¹

**Heatmap-Based Hierarchal Clustering And Gene Set Variation Analysis (GSVA)**

Patients were divided into two groups — favorable prognosis or poor prognosis — based on median time of DFS. The “limma” package in R (version 3.4.1) was used to identify differentially expressed genes (DEGs) in the two groups. A heatmap of DEGs was created with z-score normalization for each gene, in which the rows listed genes, and the columns listed patients. The targeted 395 genes are annotated as 36 classes according to signal/function pathway (Table S1). The “GSVA” package in R was then used to analyze and cluster DEGs according to 36 signal/function pathways for the two groups.

**Profiling Of Immune Infiltrates Using MCP-Counter And TCGA LUAD Dataset**

The “MCPcounter” package in R was used for analysis of microenvironment cell populations (MCPs) and quantification of immune cells from transcriptomic data. This method estimates the abundance of different immune cells based on specific molecular markers, as described by Becht et al.¹⁹

Normalized level 3 gene expression data and annotations of LUAD samples from The Cancer Genome Atlas (TCGA) in projects that were not currently embargoed were retrieved from the Genomic Data Commons (GDC) portal. Finally, transcriptional data from 85 patients with locally advanced LUAD (Stage IIb–IIIa, N1–2) and available DFS information were analyzed to explore the relationship of the tumor immune microenvironment with prognosis.

**Statistical Analysis**

All statistical tests were performed using R (version 3.4.1). The Wilcoxon rank-sum test was used for hierarchical data, and the chi-square test or Fisher’s exact test for binary data. To analyze the impacts of gene set enrichment score and immune infiltrates on clinical outcomes, Kaplan-Meier curves were used with the log rank test to analyze DFS, in which the median score was the cutoff value. Cox proportional hazards models for DFS were also used to identify the association of prognosis with infiltration by different types of immune cells. A p-value of 0.05 or less was regarded as significant.

**Results**

**Correlation Of Clinical Data And Prognosis**

We analyzed transcriptional data of 395 immune-related genes from FFPE specimens of 24 patients from our institution who had locally advanced LUAD (Table S2). We divided these patients into two groups — favorable prognosis or poor prognosis — using the median time (16 months) of DFS as the cutoff (Table 1). These two groups had no significant differences in any major clinical parameters. Survival analysis with the log rank test (data not shown) showed no significant correlation of clinical outcomes with major clinical parameters, including sex (P = 0.47), age (P = 0.30), T stage (P = 0.13), N stage (P = 0.35), or degree of differentiation (P = 0.44).

**Analysis Of DEGs And Function Analysis Using Gene Set Variation Analysis (GSVA)**

A comparison of the two groups indicated the presence of 23 DEGs (filter criteria: fold change ≥2, P < 0.05; Figure 1A and B). Nineteen genes had significantly higher expression in patients with favorable prognoses, and 4

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Favorable Prognosis (n=12)</th>
<th>Poor Prognosis (n=12)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≥60</td>
<td>4</td>
<td>8</td>
<td>&gt;0.05</td>
</tr>
<tr>
<td>&lt;60</td>
<td>8</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>8</td>
<td>4</td>
<td>&gt;0.05</td>
</tr>
<tr>
<td>Female</td>
<td>4</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>T Stage</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T1</td>
<td>6</td>
<td>6</td>
<td>&gt;0.05</td>
</tr>
<tr>
<td>T2</td>
<td>4</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>T3</td>
<td>2</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>N Stage</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N1</td>
<td>3</td>
<td>1</td>
<td>&gt;0.05</td>
</tr>
<tr>
<td>N2</td>
<td>9</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>Differentiation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High</td>
<td>3</td>
<td>4</td>
<td>&gt;0.05</td>
</tr>
<tr>
<td>Medium</td>
<td>7</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>2</td>
<td>6</td>
<td></td>
</tr>
</tbody>
</table>
genes had significantly higher expression in patients with poor prognoses (Figure 1C).

Gene set variation analysis (GSVA) demonstrated the two groups had significant differences in 5 pathways ($P_{adj} < 0.05$): antigen processing, innate immune response, lymphocyte activation, myeloid markers, and helper T cells (Figure 2). Furthermore, survival analysis with the log rank test demonstrated that greater enrichment of gene sets that function in the antigen processing pathway ($P = 0.01$) and lymphocyte activation pathway ($P = 0.043$) correlated with a favorable prognosis (Figure S1).

**Profiles Of Tumor Immune Infiltrates And Impact On Clinical Outcome: Xiangya Hospital LUAD Data**

Analysis of data from our institution demonstrated that greater infiltration of CD8 T cells correlated with longer DFS (log-rank test: $P = 0.05$), but greater infiltration of neutrophils ($P = 0.015$) and NK cells ($P = 0.015$) correlated with shorter DFS (Figure 3). However, there were no significant associations of DFS time with infiltration of cytotoxic lymphocytes, endothelial cells, cells with monocytic lineage, or T cells. Univariate and multivariate Cox hazard analysis also showed that infiltration of neutrophils was an independent risk factor for poor prognosis (adjusted hazard ratio [aHR] = 1.67, 95% confidence interval [CI] = 1.09 to 2.55, $P = 0.019$; Table 2).

**Profiles Of Tumor Immune Infiltrates And Impact On Clinical Outcome: TCGA LUAD Data**

We also used LUAD data from TCGA with MCP-counter estimates to examine 85 different patients with LUAD (Stage IIb–IIIa, N1–2, with available DFS data). A Cox proportional hazard ratio analysis indicated greater infiltration of neutrophils was associated with shorter DFS ($P = 0.01$), but the presence of more endothelial cells was associated with slower
disease progression \((P = 0.005; \text{Figure } 4)\). Next, we determined the correlation between immune cell infiltration and overall survival (OS) in the same cohort (Figure 5). The results indicated that infiltration of NK cells \((P < 0.001)\) and neutrophils \((P = 0.055)\) negatively correlated with OS; on the other hand, cytotoxic lymphocytes \((P = 0.008)\) and endothelial cells \((P = 0.014)\) positively correlated with OS. These observations were largely in accordance with the data from our institution.

**Discussion**

The targeted RNA-Seq assay used in this study is more sensitive than bulk whole transcriptome data \(^{20,21}\) and requires very little starting material. Thus, RNA-Seq can be used with FFPE specimens, even though formalin can cause RNA degradation. Paluch et al. \(^{20}\) first developed this assay in 2017 to examine a cohort of 13 patients with ovarian cancer. They found that targeted RNA-Seq data from FFPE specimens were highly correlated with identical samples that were freshly frozen from the same tumors. These researchers also found that the expression data from FFPE specimens measured by RNA-Seq were highly concordant with data measured by quantitative RT-PCR and IHC of the same specimens. \(^{20}\) Intriguingly, a recent study used the targeted RNA-Seq approach (Thermo Fisher Scientific) on FFPE samples to explore the tumor microenvironment and identify potential responders to immune therapy. \(^{22}\)
In 2016, Becht et al. first described MCP-counter, a method used to quantify tumor-infiltrating immune cells based on a stringent set of marker genes. These researchers performed an in vitro RNA mixture validation experiment and found that MCP-counter score had a strong correlation with the known proportions of RNAs in these mixtures. Moreover, their in vivo experiments indicated that MCP-counter score had a high correlation with IHC data used to determine the densities of different cells. Thus, our combined use of the targeted RNA-Seq assay and the MCP-counter method is a valid approach for comparison of tumor samples based on tumor infiltrated immune cells and transcriptome data.

Our results indicated that greater infiltration of CD8+ T cells correlated with better prognosis in patients with LUAD. Previous studies also reported that greater infiltration of CD8+ T cells into tumors was associated with slower progression in melanoma, ovarian cancer, colorectal cancer, and breast cancer. Furthermore, greater CD8+ T cell infiltration of tumors can predict response to standard chemotherapy and to immune checkpoint blockade therapy, such as anti-CTLA-4 therapy or anti-PD-1 therapy. Therefore, it is crucial to measure the extent of CD8+ T cell infiltration of solid tumors.

Our GSVA analysis also demonstrated that enrichment of the antigen processing pathway ($P = 0.01$), and possibly

<table>
<thead>
<tr>
<th>Cell Type</th>
<th>Univariate Cox Hazard Analysis</th>
<th>Multivariate Cox Hazard Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HR</td>
<td>95% CI</td>
</tr>
<tr>
<td>T cells</td>
<td>1</td>
<td>0.68–1.48</td>
</tr>
<tr>
<td>CD8 T cells</td>
<td>0.78</td>
<td>0.48–1.25</td>
</tr>
<tr>
<td>Cytotoxic lymphocytes</td>
<td>1.12</td>
<td>0.72–1.76</td>
</tr>
<tr>
<td>NK cells</td>
<td>1.51</td>
<td>0.96–2.39</td>
</tr>
<tr>
<td>B lineage</td>
<td>1</td>
<td>0.63–1.6</td>
</tr>
<tr>
<td>Monocytic lineage</td>
<td>1.01</td>
<td>0.67–1.51</td>
</tr>
<tr>
<td>Neutrophils</td>
<td>1.72</td>
<td>1.12–2.65</td>
</tr>
<tr>
<td>Endothelial cells</td>
<td>0.91</td>
<td>0.61–1.38</td>
</tr>
</tbody>
</table>

Abbreviations: HR, hazard ratio; CI95, 95% confidence interval; NA, non-available.

Figure 4 Forest plot showing multivariate Cox regression analysis of the effect of different tumor immunological infiltrates on DFS (TCGA LUAD data, n = 85). *P<0.05; **P<0.01.
the lymphocyte activation pathway ($P = 0.08$), correlated with better prognosis (Figure S1). The antigen processing pathway prepares antigens by converting them into MHC-I bound peptides for presentation to T lymphocytes. Activation of cytotoxic T lymphocytes is an antigen-specific process that requires an interaction of the TCR-CD3 complex with a processed tumor antigen-derived peptide bound to an MHC class I molecule, as well as CD8 and CD28. These two pathways interact, and both are critical for an effective immune response against tumors. Only the activated CD8+ T lymphocytes can recognize and kill tumor cells that harbor neoantigens. Thus, the higher enrichment scores in the antigen processing pathway and lymphocyte activation pathway that we observed indicate a stronger adaptive immune response, protective T cell mounting, and therefore a more favorable prognosis.

We found similarities of our LUAD patients (Figure 3) with LUAD patients in TCGA in terms of NK cell infiltration (Figure 5), in that abundant NK cell infiltration correlated with poor prognosis. A study by Gordon et al showed that continuous exposure of NK cells to tumors reduced their cytotoxic activity, and that excessive proliferation of NK cells contributed to their dysfunction. Several mechanisms may influence NK cell activity within tumor tissues. Platonova et al showed that intratumoral NK cells have strongly down-regulated activating receptors, and this impairs their cytotoxic effects. Additionally, Carrega et al found that non-small cell lung cancer (NSCLC) samples had more CD56hiCD16-NK cells than normal lung tissues, leading to expression of inhibitory killer cell immunoglobulin-like receptors and suppression of the activity of NK cells within tumor tissue. Furthermore, these researchers found that NK cells were mainly retained in the stroma and had little direct contact with NSCLC cells.

Our results showed that stronger neutrophil infiltration into the tumor was associated with a shorter DFS ($P = 0.015$), and our Cox proportional hazard analysis showed that greater tumor infiltration by neutrophils was an independent risk factor for poor prognosis. The results were similar in the TCGA LUAD cohort (Figures 4 and 5). Previous studies also found that greater infiltration of neutrophils into tumors predicts poor prognosis in other solid tumors, such as lung cancer, breast cancer, gastric cancer, and esophageal cancer. There is also evidence that the presence of neutrophils can directly enhance tumor growth. In particular, a major neutrophil protein, neutrophil elastase (NE), can enter cancer cells via clathrin-coated pit-mediated endocytosis, leading to degradation of the insulin receptor substrate 1 (IRS-1), which generally binds to phosphatidylinositol 3-kinase (PI-3K). Following degradation of IRS-1, tumor cells have increased availability of PI-3K as an energy source, leading to enhanced proliferation. Another possibility is that neutrophils...
suppress anti-tumor immune responses by suppressing cytotoxic T-cell function and proliferation. Under such circumstances, these neutrophils are regarded as granulocytic-myeloid-derived suppressor cells, which function in immunosuppression.\(^{42}\)

One of the drawbacks of the present study is that the sample size was small, and we only analyzed patients from a single institution. Studies with larger samples and from multiple institutions are needed to verify our results. Second, we only examined the abundance of 395 immune-related transcripts, so this could not be considered a thorough transcriptional immune profile of LUAD. A comprehensive profile of the whole transcriptome for analysis of immune infiltrations of fresh frozen tissues would be a valuable step forward.

In conclusion, our study of patients with LUAD indicated that greater tumor infiltration by CD8 T cells correlated with a favorable prognosis, but greater tumor infiltration by neutrophils or NK cells correlated with a poor prognosis. Thus, our results demonstrate that targeted RNA-Seq of immune-related genes from FFPE specimens, when integrated with computational bioinformatic methods, can effectively profile the tumor immune microenvironment. This method thus provides a fundamental understanding of the versatile role of immune infiltrates within tumors, allows stratification of patients with different prognoses, and is especially useful when bulk tumor tissues are unavailable.

**Ethical Approval**

This study has been performed in accordance with the ethical standards laid down in the 1964 Declaration of Helsinki and its later amendments.

**Acknowledgments**

The present study was funded by the Human Provincial Science and Technology Department (2017J2345) and the National Natural Science Foundation of China (No. 81673516).

**Disclosure**

The authors report no conflicts of interest in this work.

**References**


