A fine-grained cell type deconvolution of bulk transcriptome from single-cell RNA-seq data reveals myeloid cells heterogeneity in lung adenocarcinomas

Hao Wu\textsuperscript{1,2†}, Jiale Qin\textsuperscript{3†}, Qiang Zhao\textsuperscript{1,2}, Lu Lu\textsuperscript{1,2}, Dante Neculai\textsuperscript{4} and Chen Li\textsuperscript{1,2*}

1 Department of Human Genetics, and Women’s Hospital, Zhejiang University School of Medicine, Hangzhou, China;
2 Zhejiang Provincial Key Laboratory of Genetic & Developmental Disorders, Zhejiang University School of Medicine, Hangzhou, China;
3 Department of Ultrasound, Women’s Hospital, Zhejiang University School of Medicine, Hangzhou, China.
4 Department of Basic Medical Sciences, Zhejiang University School of Medicine, Hangzhou, China.

†These authors contributed equally to this work.

*Correspondence

Dr. Chen Li, Department of Human Genetics, and Women’s Hospital, Zhejiang University School of Medicine, Hangzhou, China. Tel: +86-571-88981627; Email: chenli2012@zju.edu.cn

Email:

Hao Wu: 21918053@zju.edu.cn
Jiale Qin: qinjiale@zju.edu.cn
Qiang Zhao: 617260696@qq.com
Lu Lu: lulu_2019@zju.edu.cn
Dante Neculai: dneculai@zju.edu.cn
Chen Li: chenli2012@zju.edu.cn
Abstract

**Background:** Tumor infiltrating myeloid cells (TIM) constitute a vital element of the tumor microenvironment. The cell-type heterogeneity of TIM has yet to be fully investigated.

**Methods:** We used a time saving approach to generate a single-cell reference matrix, allowing quantification of cell-type proportions and cell-type-specific gene abundances in bulk RNA-seq data.

**Results:** Two distinct clusters, MSC1 and MSC2 (MSC subtype) were newly identified in lung adenocarcinoma (LUAD) patients, both significantly associated with overall survival and immune blockade therapy responses. Twenty myeloid cell types were detected. Thirteen of these had distinct enrichment patterns between MSC1 and MSC2. LAMP3+ dendritic cells, being a mature and transportable subtype of dendritic cell which may migrate to lymph nodes, were noted as associated with non-responsiveness to immunotargeted therapy. High infiltration level of IFIT3+ neutrophils was strongly related to the response to immune targeted therapy and was seen to activate CD8+ T cells, partly through inflammasome activation. The infiltration levels of TIMP1+ macrophages and S100A8+ neutrophils were both significantly associated with poor prognosis. TIMP1+ macrophages were noted to recruit S100A8+ neutrophils via the CXCL5-CXCR2 axes and promote LUAD progression.

**Conclusion:** Altogether, we performed virtual micro-dissection of the bulk transcriptome at single-cell resolution and provided a promising TIM infiltration landscape that may shed new light on the development of immune therapy.

**Keywords**

Single-cell RNA sequencing, bulk RNA sequencing, lung adenocarcinomas, myeloid cells, tumor microenvironment

**Introduction**

Lung adenocarcinomas (LUAD) account for over 40% of lung cancers and represent its most leading and prevalent histological subtype. Despite an improvement in therapeutic strategies, the rates of objective clinical responses remain low, with only 17.4% lung cancer patients surviving more than five years beyond diagnosis. In this, the dynamic tumor immune microenvironment plays an important role in tumor progression and metastasis [1]. Tumor-infiltrating T lymphocytes are now recognized as the key components of the tumor microenvironment (TME). Therapeutic strategies for targeting these cells are being actively developed and have demonstrated remarkable therapeutic effects [2]. Whilst current
immunotherapies targeting T lymphocytes benefit only few patients [3], it is important to unravel the exact cellular functions of the remaining cell types within the TME that may be involved in tumor progression.

Recently, there has been much focus on cancer immunology, with primary emphasis on myeloid cells as important components in tumor immune evasion [4, 5]. However, the various reports that ascribe macrophages and neutrophils with either pro- or anti-tumor properties, together with acknowledgement that our understanding of tumor infiltrating myeloid cell (TIM) subtypes is quite inadequate [6, 7], leads to a high potential for confusion and/or contradiction within this field. Despite this, some strategies targeting myeloid cells have been developed [8, 9]. However, the limited understanding of clear mechanistic hypotheses had led to difficulties in interpreting clinical outcomes for such approaches [10]. In particular, the complexity of TIM subtypes and the discrepancies between human and mouse models has deeply impeded the implementation of selective myeloid-targeting immunotherapies.

Single-cell RNA sequencing (scRNA-seq) offers an opportunity to dissect the complexity of the TME, enabling the identification of the cell state in a manner independent of any previous knowledge of cellular markers. Single-cell analysis has been applied to reveal the cellular heterogeneity of TIMs, including tumor associated macrophages (TAMs), dendritic cells (DCs) and neutrophils in different cancer types [11, 12]. However, due to high cost and strict requirement for cellular activity, analyses of large patient cohorts have been almost impossible. Computational algorithms, which allow for the estimation of relative cell infiltrate level based on bulk RNA sequencing (RNA-seq) and scRNA-seq data, have now been developed and may compensate for this. CIBERSORTx is a computational framework to accurately infer relative abundance of cell type from bulk RNA-seq according to the signature matrix generated from scRNA-seq by means of a deconvolution algorithm [13]. It has been successfully used and validated for revealing immune cell landscapes in melanomas [14], clear-cell renal cell carcinomas [15] and prostate cancer [16]. Unfortunately, CIBERSORTx does not seem to provide a standard procedure or pipeline describing how to integrate scRNA-seq data to construct a reference matrix, as it just uses all the scRNA-seq data as the input. As such, this procedure requires substantial computational resources and time to handle the huge amount of the data. Although down-sampling can address this problem, scRNA-seq data usually suffers from the problem of extremely high dropout rate, especially for those generated by 10x Genomics. Thus, the reference matrix generated by down-sampling is often unhelpful for such considerations and will result in a poor deconvolution effect.

In this paper, we applied a timesaving approach to create a customized reference matrix for scRNA-seq data for myeloid cells and made a deconvolution of the TCGA LUAD cohort. As a result, we identified for the first time two different enrichment patterns (called MSC subtypes hereafter). A series of analyses, including survival analysis and multivariable Cox
regression analysis with clinical features, revealed MSC subtypes to be robust prognostic factors. We reveal the relationship between MSC subtypes and immune checkpoint blockade (ICB) therapy and identify three TIM subtypes that might contribute to the ICB response. Finally, we explore the heterogeneity of macrophages and detected a functional relationship between macrophage and neutrophil subtypes.

**Methods**

**Data source and pre-processing**

The single-cell transcriptome file of five LUAD patients and the validation data for DCs’ distribution were downloaded from the Gene Expression Omnibus (GEO) database under the accession number GSE127465 [17] and GSE131907 [18]. The transcriptome expression profiles and corresponding clinical information for LUAD were retrieved from the Genomic Data Commons Data Portal of The Cancer Genome Atlas (TCGA). Expression data were converted from counts type to transcripts per million (TPM). Three transcriptional microarray expression data (GSE matrix files) for LUAD cohorts (GSE31210 [19], and GSE72094 [20]) were obtained from the GEO. The microarray datasets were log-transformed (on a base 2 scale) and genes were detected with more than one probe retaining its maximum value.

Figure 1 shows a conceptual view of our fine-grained cell type deconvolution and its application to LUAD analysis.

**scRNA-seq data and single-cell data analysis**

The *Seurat* package (version 3.0) was used to perform scRNA-seq analysis [21]. Transcriptomes with more than 300 total counts, less than 10,000 total counts and less than 20% of counts coming from mitochondrial genes, were retained for subsequent analysis. From the remaining cells, gene expression matrices were normalized to the total unique molecular identifier (UMI) counts per cell and were log-transformed (on a base 2 scale). Dimensionality reduction was performed with uniform manifold approximation and projection (UMAP). The marker gene of each cluster was identified using *Seurat*.

**Single cell reference matrix construction**

The most characteristic cells in each cell subtype, rather than all tens of thousands of cell’s data, were selected to create the custom signature matrix. We defined and calculated a so-called cell-type-specific signature score *SigScore* to select the candidate cells from each cell subtype.
\[ \text{SigScore} = \sum_{i=1}^{N} ([\text{Marker}_i] \times k_i) \]

where \([\text{Marker}_i]\) represents the expression level of the marker gene \(i\) and \(k_i\) is the value equal to \(\log_2 FC_{\text{Marker}_i}\) which is calculated using the \text{FindMarkers}() function in \text{Seurat}. \(N\) denotes the number of the markers in each cell type. Note that, \(N\) is variable according to the threshold of \(\log_2 FC\). Specially, 1.25 were selected as the threshold of \(\log_2 FC\) in our paper.

We ranked all cells belonging to a special cell subtype in descending order of their \text{SigScore} and chose the top 50 cells to create the custom signature matrix. In addition, we created another custom signature matrix by randomly selecting 50 cells of each cell type, which would then be used as another matrix for comparison.

**Infiltration estimation of the myeloid cells**

ScRNA-seq data of the top 50 cells was uploaded to CIBERSORTx (http://cibersortx.stanford.edu) to create a customized signature matrix for each myeloid cell subtype. A hundred simulated bulk datasets were created by random sampling of different numbers of each cell types (including non-myeloid cells) and these were used to validate the signature matrix. Since count data were uploaded for creating the signature matrix, count-per-million (CPM) data of the TCGA LUAD cohort within the genes involved in the signature matrix were then generated to estimate the abundance of myeloid cell subtypes in CIBERSORTx.

**MSC subtype identification and prediction model building**

Based on the infiltration level of myeloid cell subtypes, the optimal number of TCGA LUAD cohort clusters were examined using the \text{mclust} package (version 5.4.5) [22]. \textit{K-Means} consensus clustering was conducted in R to determine distinct clusters of MSC subtypes. The least absolute shrinkage and selection operator (LASSO) algorithm was used to reduce the data dimensions and distinguish the most informative genes for predicting the MSC subtype using the \text{glmnet} package (version 3.0-2) [23]. Finally, the MSC score formula was calculated by considering the correlation estimated Cox regression coefficients:

\[ M_{\text{Score}} = \sum_{i=1}^{Z} ([\text{Gene}_i] \times \text{coef}_i) \]

where \(Z\) denotes the gene number determined by LASSO, \([\text{Gene}_i]\) denotes the expression level of gene \(i\), and \(\text{coef}_i\) represents the coefficient of gene \(i\) as determined by LASSO.

**Potential drug targeted gene set selection**

To identify the potentially druggable therapeutic targets for the patients of identified MSC
subtypes, we collected two datasets, the genome-scale CRISPR knockout screens dataset in Project Achilles (https://depmap.org/portal/) and drug-induced gene expression profiles from the Library of Integrated Network-based Cellular Signatures (LINCS; https://commonfund.nih.gov/LINCS/) L1000 dataset. We then performed a two-step analysis to identify candidate drugs. Firstly, we filtered the essential genes for LUAD cell lines based on the Genome-wide CRISPR gene essentiality scores (CERES) from Project Achilles. The genes whose CERES was lower than -0.5 in half of the total LUAD cell lines were retained and then intersected with the up-regulated genes in MSC2 patients. We then ranked the gene expression profiles of LUAD cell lines obtained from LINCS and performed Gene Set Enrichment Analysis (GSEA) using clusterProfiler package based on the above target gene set [24]. Only when the target gene set was significantly enriched in the bottom of ranked gene list, the drug was then considered to have potential.

**Calculation of ligand-receptor interaction**

For the cell-cell interaction analysis, the expression level was normalized according to the total reads count and converted into a TPM-like scale. The expression values were averaged within each cell subtype. We retrieved the ligand-receptor pairs from a systematic research including known ligand–receptor pairs from the existing databases and predicted the ligand-receptor pairs with high confidence [25]. The threshold of 1 TPM was used as the cut-off for ligand-receptor pairs within each cell subtype for further analysis.

**Statistical analysis**

The differentially expressed genes (DEGs) were calculated using the DESeq2 package for R [26]. DEGs satisfying |log2FoldChange| > 1.5 and adjusted P-value < 0.05 criteria were considered statistically significant. Clusterprofiler was used to perform GO function enrichment and KEGG pathway annotation. Within a specific cohort, patients were divided into two groups based on the mean value of $M_{Score}$ in all samples. Survival curves were constructed using the Kaplan-Meier (KM) method and compared using the log-rank test provided in the survival package for R [27]. Multivariate Cox proportional hazard regression modeling was used to verify the prognostic significance for OS. Histologic grade, gender, and age were used as variables. To identify the relationship between clinical state and myeloid cells, we queried the clinical data of the TCGA LUAD cohort. In particular, the TMN stages were categorized to a numeric level. The correlation between the infiltration level of myeloid cells and clinical variables was examined using Pearson’s correlation coefficient (CC), which was considered statistically significant by FDR < 0.05. Area under the curve (AUC) of the receiver operating characteristic (ROC) curve was used to assess the predictive ability of the
predicted signature.

**Results**

**Construction and validation of SigScore-based reference matrix**

We downloaded the scRNA-seq data from Zilionis et al [17] in which the authors demonstrated the major aspects of the lung tumor immune microenvironment. Twenty subtypes of myeloid cells were identified using Seurat (Figure S1A). The corresponding scRNA-seq data of the top cells with the highest SigScore to CIBERSORTx were then uploaded and the underlying reference matrix obtained.

We examined the SigScore-based reference matrix compared to a randomly selected reference matrix within 100 simulated datasets. Our reference matrix showed better performance (Figure 2A).

**Identification of two distinct myeloid cell infiltration subtypes in LUAD patients**

We applied the SigScore-based reference matrix to investigate the fractions of infiltrated myeloid cells in the TCGA LUAD dataset (Figure S1B). Among the total samples, 485 tumor samples were eligible for CIBERSORTx under $P$-value < 0.05 and CC > 0.5. We performed K-means clustering with the optimal number ($k = 2$) and identified two distinct myeloid cell infiltration clusters, namely MSC1 and MSC2, according to the contexts of the myeloid cells (Figures S1C and S1D). Thirteen of 20 had distinct enrichment patterns between MSC1 and MSC2 ($\mid \log2\text{FoldChange}\mid > 1.25$, $P$-value < 0.05) (Figure 2B). MSC2 was significantly associated with a shorter OS compared with MSC1 (Figure 2C, $P$-value = 0.00074). GO analysis suggested that the MCS1 subtype, with its favorable outcome, has a stronger immune response ability including both innate and adaptive immunity. Conversely, the MCS2 subtype was significantly associated with the cell cycle and negative regulation of cell apoptosis, in line with unfavorable outcomes (Figure 2D).

We then queried the distribution of distinct TIM cell types within tumors and adjacent tissues. The statistical result is showed in Table S1. Some of those cell types are strongly related with OS. A high fraction of LAMP3+DC3, S100A8+N3, PI3+N4, TIMP1+M3 and TUBB+Mcy1 were identified as poor prognostic factors, while high fraction of CLEC9A+DC2 were identified as a protective (Figure S2).

We then observed that SELENOP+M4, the most enriched cell type in MSC1, was the subtype of macrophages which highly expressed SELENOP and TM4SF1 mRNAs. TUBB+Mcy1, the most enriched cell type in MSC2, was the subtype of the monocytes which highly expressed TUBB. The infiltration levels of two subtypes between the tumor and paired
adjacent tissues of all patients, MSC1 patients and MSC2 patients were further analyzed, respectively. SELENOP+M4 showed preferential enrichment in normal tissues. While considering the MSC subtypes, we found a significant lower infiltration level of SELENOP+M4 in MCS2, while TUBB+Mcy1 showed the opposite trends (Figure 2E). The infiltration levels of other cell subtypes in three states are shown in Figure S3.

**Association of predicted MSC subtype with prognostic impact in TCGA and two independent LUAD cohorts**

To predict the MSC subtype for bulk RNA-seq or microarray data, we determined the most informative genes and constructed a resulting 14-gene signature (i.e. $M_{Score}$) (Table S2). The TCGA LUAD cohort were randomly divided into a training set (n = 292) and a test set (n = 193). The LASSO Cox regression model with 20-fold cross validation was performed to train the model in the training set. We then assessed the model performance in the test set. According to the $M_{Score}$, LUAD patients were well classified into MSC1 and MSC2 subtypes. The AUC of the ROC curve achieved 0.91 and 0.89 in the training and test sets respectively, indicating a strong prediction ability to stratify the patients (Figure 3A).

The prediction capability of 14-gene signature was executed to further examine the prognostic significance in two independent LUAD microarray cohorts (GEO72094 and GEO31210). Similarly, a poor prognostic impact of MSC1 compared with MSC2 was found in both cohorts ($P$-value < 0.0001 and $P$-value = 0.0012, respectively) (Figures 3B and 3C). We also observed that nearly twice the relapse rate had occurred in MSC2 compared with MSC1 in the GSE31210 dataset. This suggested that the myeloid cell distribution might be associated with cancer relapse (Figure 3D).

To determine whether MSC subtype is an independent predictor of the prognosis, we performed multivariate analysis. In the TCGA LUAD and GSE72094 cohorts, MSC2 strongly predicted a shorter OS compared with MSC1, independent of known risk factors (HR: 1.765, 95% CI: 1.172-2.567, $P$-value = 0.0065 for TCGA; HR: 2.680, 95% CI: 1.789-4.020, $P$-value < 0.001 for GSE72094), including age, gender and cancer stage (Figure 3F and Figure S4A).

**MSC subtype significantly associated with tumor stages of LUAD**

We analyzed the distribution of $M_{Score}$ in different tumor and TNM stages. The higher $M_{Score}$ was associated with a higher tumor stage (Figure 3E). We also observed a similar positive correlation in both T, M and N stages, suggesting the distribution of myeloid cells were potentially related to the clinical stage (Figure S4B-D). To further investigate whether the relative presence of these myeloid cell subtypes was associated with tumor progression, we calculated the correlation between each myeloid cell subtype and cancer stage. Five cell
types, TIMP1+M3, S100A8+N3, TUBB+Mcyl, LAMP3+DC3, TCL1A+pDC, all of which were enriched in MSC2, were positively associated with tumor progression (FDR < 0.05). Four cell types, SELENOP+M4, CXCL9+M2, CD1C+DC1, CLEC9A+DC2, were negatively associated with tumor progression (FDR < 0.05) (Figure 3E).

**MSC Subtype associated with immunotherapy response**

Myeloid cells have been reported to be associated with ICB [28]. In the TCGA LUAD cohort, we introduced a tumor immune dysfunction and exclusion (TIDE) algorithm to explore the relationship between the MSC subtype and ICB response [29]. TIDE is a computational framework for predicting the clinical response to ICB in patients. A low TIDE prediction score indicates that the patients would potentially exhibit a greater immune therapy response. We observed the TIDE score as significantly lower in MSC1, suggesting the MSC1 group is more likely to respond to ICB therapy (Wilcoxon rank-sum test, \( P\text{-value} < 0.05 \)) (Figure 4A). This association was verified in two independent cohorts using univariate analysis (Wilcoxon rank-sum test, \( P\text{-value} < 0.05 \)) (Figure 4B). In the other independent cohort (i.e. GSE126044), which includes the RNA-seq data and response states of 16 patients before anti-programmed cell death protein 1 (PD-1) treatment, we compared the TIDE and \( M_{Score} \) to evaluate the prediction performance of ICB response. All five responders of 16 patients were clearly identified as MSC1 (Figure 4C). \( M_{Score} \) showed the best predictive power (AUC = 0.891) as compared to TIDE and PD-1 (Figure 4D).

**Identification of potential drug for MSC2 patients**

Since MSC2 patients seem more unlikely to respond to ICB therapy and show a worse survival state compared with MSC1 patients, we next focused on identifying the potential drugs for patients of the MSC2 subtype. The upregulated DEGs in MSC2 patients are supposed to be the therapeutic target. Note that MSC subtypes are classified according to the marker genes of myeloid cells, leading us to query whether the DEGs represent the diversity of tumor microenvironment. Using a hypergeometric test, we found the myeloid cell markers were almost irrelevant with DEGs between MSC1 and MSC2 patients (Figure S5A). That suggests that the DEGs might reflect differences in tumor cell state.

We screened the data of lung adenocarcinoma cell lines in Project Achilles (https://depmap.org/portal/). The genes with a CERES lower than -0.5 in half of the lung adenocarcinoma cell lines were retained. We then intersected the DEGs which were upregulated in MSC2 patients with survival-related genes and obtained 29 genes which represent the targets for ICB therapy in MSC2 patients (Figures S5B-C).

Eight of 29 genes were related to the cell cycle, suggesting the tumor cells in MSC2
patients might be in a relatively strong state of cellular proliferation. The remaining genes were involved in a wide range of cancer-related pathways, such as DNA replication, Ras signaling, and mTOR signaling (Figure S5C). Twenty-eight of 29 genes have been reported to be related with LUAD [30-33]. However, TOPBP1 interacting checkpoint and replication regulator (TICRR), had only been previously reported to be important in DNA replication [34]. In our study, TICRR seems to be a promising therapeutic target for LUAD, especially in MSC2 patients.

Since the potential drug target gene set was obtained, we employed LINCS L1000 dataset to identify potential drugs. We focused on four LUAD cell lines, DV90, SKLU1, NCIH2073, and NCIH596. Testing was conducted of 361 drugs on the four cell lines, with a total of 1498 gene expression profiles extracted. After computing the robust z-scores for each profile relative to control, we ranked the gene based on the expression levels and performed GSEA analysis. We totally identified 129 drugs showed potential inhibition effect for at least a particular cell line (FDR < 0.05, Figure 4E). To obtain more reliable drugs, we selected the drugs which showed significant suppression effects on all four cell lines and discovered 10 drug candidates (Figure 4E).

**Identification of new myeloid cell subtypes related with ICB response**

To further understand which myeloid cell subtype contributed to ICB response, we further investigated the fractions of infiltrated myeloid cells in GSE126044 dataset by applying CIBERSORTx. We observed IFIT3+N5 and PPARG+M7 were enriched in responders, while LAMP3+DC3 was the only subtype enriched in Non-responders (Figure 5A). CD1C+DC1 highly expressed CD1C, FCER1A, and CLEC10A, corresponding to conventional cDC2, while CLEC9A+DC2 highly expressed CLEC9A, BATF3 and CADM1, corresponding to conventional cDC1 [35, 36] (Figure 4B).

According to existing knowledge, there is no LAMP3+DC3 counterpart in the classic DC subsets, we thus compared transcript profiles among three DC subtypes. We observed that LAMP3+DC3 highly expressed an “activated” DC signature in line with the previous study [12] and showed a higher migration ability, according a gene signature derived from mouse tissue-migratory cDCs (Figure 5B) [37]. We then compared the ratio of LAMP3+ DCs/Total DCs in GSE131907 dataset [18]. The result showed LAMP3+DCs were enriched in the lymph nodes in LUAD patients, which further demonstrates its migration ability (Figure 5C). Interestingly, GO analysis showed LAMP3+DC3 was associated with negative regulation of the immune system (Figure S6A), which seems to explain its enrichment in Non-responders. We also note that CD274 (PD-L1) was highly expressed in LAMP3+DC3 (Figure S6B). Altogether, LAMP3+DC3 might play an important role in immunosuppression, especially in
T cell dysfunction, albeit more mature [12].

The degree of cytotoxic T cell infiltration (CTL) has been reported to influence ICB effectiveness and been used as a parameter of TIDE [29, 38]. We used the average expression of PRF1, GZMA, GZMB, CD8A and CD8B to estimate the CTL levels and examined the correlation between $M_{\text{score}}$ and CTL levels. Interestingly, a significant but moderate negative correlation was found ($r=-0.26$, P-value=$9e^{-9}$), suggesting myeloid cells might affect the CTL level (Figure 5D). It could be considered that the genes with high correlation with CTL levels not only exist as biomarkers but also as a clue of potential useful cell subtype. Notably, the unique marker gene of IFIT3+N5 was enriched in the gene set which highly correlated with CTL level (Figure 5E). Nine of the top 20 highly correlated genes were unique markers of IFIT3+N5. IFIT3+N5 was found to be involved in the viral defense response, response to INF-gamma, and in the type I interferon signaling pathway (Figure 5F). KEGG pathway analysis showed IFIT3+N5 was strongly associated with the NOD-like receptor signaling pathway and NF-kappa B signaling pathway (Figure 5F). We noted that three guanylate-binding family proteins (GBP5, GBP4 and GBP1) in the top 20 correlated genes, had been identified to be linked to the inflammasome activation [39, 40], and were also highly correlated with CTL level (0.876, 0.818, and 0.806). Altogether, IFIT3+N5 might play a role in activate the CD8+ cytotoxic cell response in LUAD, which act, in part, through inflammasome activation.

**TIMP1+M3 macrophages recruit S100A8+ neutrophils via CXCL5-CXCR2 axes to promote LUAD progression**

We used predefined ligand-receptor pairs [25] to examine the interactions between the myeloid cells. In terms of cell communications, we observed that macrophages showed higher proportions than other cell types (51% in Ligand; 45% in Receptor), suggesting macrophages may act as a hub for other myeloid cells (Figure 6A).

The macrophage subtypes included both pro-inflammatory (M1), anti-inflammatory (M2) and a mixed phenotype, suggesting the M1 and M2 types may be underestimate of tumor-associated macrophage complexity (Figure 6B). CCL18+M1, CXCL9+M2 and TIMP1+M3 were enriched in tumor tissues and were regarded as TAMs, whereas SELENOP+M4, MMP7+M5, CHIT1+M6 and PPARG+M7 were enriched in normal tissues and considered as resident tissue macrophages (RTMs) (Figure S7A).

We then focused on the subtypes enriched in either MSC1 or MSC2 types. TIMP1+M3 was the unique subtype enriched in MSC2, suggesting its potential ability to promote tumor progression. CXCL9+M2, the subtype also enriched in tumor tissues, showed an opposite enrichment pattern. By further comparing the RNA expression profiles between two TAM
subtypes (i.e. TIMP1+M3 and CXCL9+M2), we observed that CXCL9+M2 exhibited a high expression of C1Q family genes, apolipoprotein family genes and antigen presentation related genes (Figure 6C). In contrast, TIMP1+M3 showed specific expression of TIMP1, VCAN, and CXCL5 (Figure 6C). GO analysis revealed a strong enrichment of complement activation, immune response, and antigen processing and presentation pathway in CXCL9+M2, while positive regulation of angiogenesis and neutrophil chemotaxis showed significant enrichment in TIMP1+M3 (Figure 6D). Meanwhile, we found TIMP1+M3 showed a functional relationship with neutrophil, as indicated by GO annotation. By the cell-cell interaction analysis, TIMP1+M3 was predicted to interact with neutrophils (IFITM2+N2 and S100A8+N3) via CXCL5-CXCR2 axes, suggesting TIMP3+M3 attracted the neutrophil via the chemokine (Figure 6E). S100A8+N3, which was attracted by TIMP1+M3 via the CXCL5-CXCR2 axes, was identified as a risk factor and showed moderate positive correlation with TIMP3+M3, whilst its correlation with IFITM2+N2 was negative (Figures S2 and S7B). We also observed the higher value CXCL5 × CXCR2 was associated with an unfavorable OS, whereas no such association was observed for either CXCL5 or CXCR2 (Figure 6F). Interestingly, TIMP1+M3 and S100A8+N3 were positively correlated with N stage, suggesting their important role in lymph node metastasis (Figure S4C).

We also investigated other macrophage subtypes. SELENOP+M4 showed high expressions of CCL4L2, CCL3L3, CCL3, and CCL4. GO analysis revealed leukocyte chemotaxis and positive regulation of cytokine production pathway as enriched in SELENOP+M4 (Figure S7C). SELENOP+M4 also highly expressed antigen presentation and T cell activation related genes, suggesting it might play an important anti-cancer role and involvement in immune activation (Figure 6B). PPARG+M7 was identified as resident alveolar macrophage with high expressions of PPARG, FABP4, INHBA and ALDH2. We noted that PPARG+M7 was enriched in ICB responders, suggesting it may also play a key role in regulating anti-tumor immunity as a tissue-specific macrophage.

**Discussion**

We calculated the infiltration of TIM subtypes in the TCGA LUAD cohort by combining the scRNA-seq and deconvolution algorithm. According to intra-tumor TIM heterogeneity, patients were stratified into two groups, MSC1 and MSC2. We found that MSC subtypes were strongly associated with OS and ICB responses. Specific TIM subtypes showed particular functions in tumor progression. In discussing this work, we focused on the results that have promising applications and those which are closely related to clinical treatment.

We proved the MSC subtypes represent the states which either may or may not respond to ICB therapy in multiple datasets. In particular, the MSC subtype was useful for the
estimation of differences in TIMs infiltration states. We validated the effectiveness of MSC subtype by comparing it with the TIDE Score, which mainly considered the function of cytotoxic T cells as predictive of ICB response. The TIDE Score shows significant differences within two MSC subtypes in TCGA and in the two GEO datasets. Since both lymphocytes and myeloid cells have been reported to be related with ICB response [41], we believe that MSC subtypes and TIDE Score reflected different aspects of ICB responses of patients and that MSC subtypes could be jointly used with TIDE score to achieve a better estimations in a clinical context.

In our paper, three TIM subtypes were identified as ICB response related, including LAMP3+DC3, IFIT3+N5, and PPARG+M7. LAMP3+DC3 was enriched in ICB non-responders and was identified as a more mature DC subtype. Compared with other two DC subtypes, CD1C+DC1 and CLEC9A+DC2, the negative regulation of immune system process pathways were enriched in LAMP3+DC3. CCR7 is necessary for the migration of tumor-infiltrating DCs into tumor-draining lymph nodes [42]. LAMP3+DC3 highly expressed CCR7 and showed the strongest migration ability, indicting LAMP3+DC3 might migrate to the lymph node and suppress immune activation. Similar DCs were identified in the single cell study of hepatocellular carcinomas and were described to be related with T cell dysfunction by interacting with T lymphocytes [12]. Zhang et al. also suggested that LAMP3+DCs in tumors might originate from cDC1 and cDC2. Thus, LAMP3+DCs may not only be a predictive factor of ICB response, but also could be considered as a new target for immunotherapy.

As multiple studies have demonstrated that tumor infiltrating neutrophils are related to cytotoxic T cells in various ways [43, 44], we confirmed that our identified IFIT3+N5, a subtype of neutrophil, was enriched in ICB responders and showed positive correlation with CTL. Functional annotation indicated that IFIT3+N5 might activate CD8+ T lymphocytes, partly via inflammasome activation. According to a previous study, IFIT3+N5 corresponds to a group of mature neutrophils which are expanded in virus infected tissues [45]. Considering that neutrophils might be converted into different phenotypes, either anti-tumoral or pro-tumoral, the clarification of how the precursor cells are changed to IFIT3+N5 will be an important consideration for future studies.

It has been reported that CXCR2+ neutrophils are recruited by CXCL5 in tumor tissues to promote tumor progression in liver and non-small cell lung cancers [46, 47]. However, most of these studies used cell lines, tissue sections and mouse models, which made it difficult to identify the specific cell subtypes involved. In this article, we clearly identified that TIMP3+M3 recruited S100A8+N3 via CXCL5-CXCR2 axes. TIMP3+M3 and S100A8+N3 were both identified as protumoral cell types and related with lymph node metastasis, suggesting those cells might promote the tumor progression in synergy. CXCR2
and CXCR4 were seen as required when neutrophils egress from the bone marrow and are retained in the lungs [48]. Here, we noticed a repulsive expression pattern between CXCR2 and CXCR4 in the neutrophils (Figure S6D). Thus, a blockade of CXCR2 might lead to decreasing infiltration of S100A8+N3, which might partly explain the high performance of CXCR2 antagonists [49].

**Conclusions**

We generated the landscape of myeloid cells in LUAD and stratified the patients into two infiltrating patterns (MSC1, MSC2). We observed a significant relationship between TIM infiltrating pattern and OS and ICB responses, and validated this finding in two external independent cohorts. We identified special myeloid subtypes related with tumor progression and ICB response, leading to new insights into the function of TIMs in cancer. These findings could assist scientists in understanding the complexity of TIMs and help optimize related immunotherapy strategies.

**List of abbreviations**

AUC: Area under the curve; CERES: Genome-wide CRISPR gene essentiality scores; CTL: Cytotoxic T cell infiltration; DC: Dendritic cell; DEG: Differentially expressed genes; ICB: Immune blockade therapy; LASSO: Least absolute shrinkage and selection operator; LN: Lymph node; LUAD: Lung adenocarcinoma; OS: Overall survival; RTM: Resident tissue macrophages; scRNA-seq: Single-cell RNA sequencing; TAM: Tumor associated macrophage; TIM: Tumor infiltrating myeloid cell; TME: Tumor microenvironment; TPM: Transcripts per million; UMAP: Uniform manifold approximation and projection; UMI: Unique molecular identifier.

**Declarations**

**Ethics approval and consent to participate**

Not applicable

**Consent for publication**

Not applicable
Availability of data and materials

The datasets analyzed during the current study are available in the GEO repository (accession numbers GSE127465, GSE131907, GSE31210, and GSE72094), and Genomic Data Commons Data Portal of The Cancer Genome Atlas (TCGA) LUAD. Supplementary Information files are also available upon reasonable request from the corresponding author.

Competing interests

The authors declare no conflict of interest.

Funding

This work was supported by the grant from Zhejiang Provincial Natural Science Foundation of China (LY17C060003), Chinese National Natural Science Foundation (81601515), and Zhejiang Provincial Key Projects of Technology Research (WKJ-ZJ-2033).

Authors' contributions

HW and CL conceived the basic idea. HW, JQ and QZ further analyzed scRNA-seq and bulk RNA-seq data of LUAD. HW, QZ and LL evaluated the results and validated the performance of the new approach. HW wrote the draft of the manuscript. JQ and DN revised the whole contents. CL supervised the whole study. All authors read and approved of the final manuscript.

Acknowledgments

We would like to thank Chris Wood for improvement of English writing of the manuscript. The authors are grateful to the research groups for providing the clinical information and high-throughput data in this study. We are deeply thankful for the computation time provided by the Super Computer System, Human Genome Center, Institute of Medical Science, University of Tokyo, Japan.

Reference

1. Maynard A, McCoach CE, Rotow JK, Harris L, Haderk F, Kerr DL, et al. Therapy-Induced Evolution of Human Lung Cancer Revealed by Single-Cell RNA Sequencing. Cell. 2020;182:1232-1251 e1222.

2. Sharma P, Allison JP. The future of immune checkpoint therapy. Science. 2015;348:56-61.
3. Awad RM, De Vlaeminck Y, Maebe J, Goyvaerts C, Breckpot K. Turn Back the TiMe: Targeting Tumor Infiltrating Myeloid Cells to Revert Cancer Progression. Front Immunol. 2018;9:1977.
4. Engblom C, Pfirschke C, Pittet MJ. The role of myeloid cells in cancer therapies. Nat Rev Cancer. 2016;16:447-462.
5. DeNardo DG, Ruffell B. Macrophages as regulators of tumour immunity and immunotherapy. Nat Rev Immunol. 2019;19:369-382.
6. Ginhoux F, Schultzje JL, Murray PJ, Ochando J, Biswas SK. New insights into the multidimensional concept of macrophage ontogeny, activation and function. Nat Immunol. 2016;17:34-40.
7. Coffelt SB, Wellenstein MD, de Visser KE. Neutrophils in cancer: neutral no more. Nat Rev Cancer. 2016;16:431-446.
8. Papadopoulos KP, Gluck L, Martin LP, Olszanski AJ, Tolcher AW, Ngarmchamnanrith G, Rasmussen E, Amore BM, Ngarmchamnanrith G, Rasmussen E, Amore BM, Nagorsen D, Hill JS, Stephenson J, Jr. First-in-Human Study of AMG 820, a Monoclonal Anti-Colony-Stimulating Factor 1 Receptor Antibody, in Patients with Advanced Solid Tumors. Clin Cancer Res. 2017;23:5703-5710.
9. Vonderheide RH. CD40 Agonist Antibodies in Cancer Immunotherapy. Annu Rev Med. 2020;71:47-58.
10. Zhang L, Li Z, Skrzypczynska KM, Fang Q, Zhang W, O'Brien SA, et al. Single-Cell Analyses Inform Mechanisms of Myeloid-Targeted Therapies in Colon Cancer. Cell. 2020;181:442-459 e429.
11. Azizi E, Carr AJ, Plitas G, Cornish AE, Konopacki C, Prabhakaran S, et al. Single-Cell Map of Diverse Immune Phenotypes in the Breast Tumor Microenvironment. Cell. 2018;174:1293-1308 e1236.
12. Zhang Q, He Y, Luo N, Patel SJ, Han Y, Gao R, et al. Landscape and Dynamics of Single Immune Cells in Hepatocellular Carcinoma. Cell. 2019;179:829-845 e820.
13. Newman AM, Steen CB, Liu CL, Gentles AJ, Chaudhuri AA, Scherer F, et al. Determining cell type abundance and expression from bulk tissues with digital cytometry. Nat Biotechnol. 2019;37:773-782.
14. Huang L, Chen H, Xu Y, Chen J, Liu Z, Xu Q. Correlation of tumor-infiltrating immune cells of melanoma with overall survival by immunogenomic analysis. Cancer Med. 2020;9:8444-8456.
15. Hu J, Chen Z, Bao L, Zhou L, Hou Y, Liu L, Xiong M, Zhang Y, Wang B, Tao Z, Chen K. Single-Cell Transcriptome Analysis Reveals Intratumoral Heterogeneity in ccRCC, which Results in Different Clinical Outcomes. Mol Ther. 2020;28:1658-1672.
16. Jairath NK, Farha MW, Srivinasan S, Jairath R, Green MD, Hess RT, et al. Tumor Immune Microenvironment Clusters in Localized Prostate Adenocarcinoma: Prognostic Impact of Macrophage Enriched/Plasma Cell Non-Enriched Subtypes. J Clin Med. 2020;9.
17. Zilionis R, Engblom C, Pfirschke C, Savova V, Zemmour D, Saatioglou HD, et al. Single-Cell Transcriptomics of Human and Mouse Lung Cancers Reveals Conserved Myeloid Populations across Individuals and Species. Immunity. 2019;50:1317-1334 e1310.
18. Kim N, Kim HK, Lee K, Hong Y, Cho JH, Choi JW, et al. Single-cell RNA sequencing demonstrates the molecular and cellular reprogramming of metastatic lung adenocarcinoma. Nat Commun. 2020;11:2285.
19. Yamauchi M, Yamaguchi R, Nakata A, Kohno T, Nagasaki M, Shimamura T, et al. Epidermal growth factor receptor tyrosine kinase defines critical prognostic genes of stage I lung
adenocarcinoma. PLoS One. 2012;7:e43923.

20. Schabath MB, Welsh EA, Fulp WJ, Chen L, Teer JK, Thompson ZJ, et al. Differential association of STK11 and TP53 with KRAS mutation-associated gene expression, proliferation and immune surveillance in lung adenocarcinoma. Oncogene. 2016;35:3209-3216.

21. Butler A, Hoffman P, Smibert P, Papalexi E, Satija R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat Biotechnol. 2018;36:411-420.

22. Scrucca L, Fop M, Murphy TB, Raftery AE. mclust 5: Clustering, Classification and Density Estimation Using Gaussian Finite Mixture Models. J r. 2016;8:289-317.

23. Tibshirani R. The lasso method for variable selection in the Cox model. Stat Med. 1997;16:385-395.

24. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. Omics. 2012;16:284-287.

25. Ramilowski JA, Goldberg T, Harshbarger J, Kloppmann E, Lizio M, Satagopam VP, Itoh M, Kawaji H, Carninci P, Rost B, Forrest AR. A draft network of ligand-receptor-mediated multicellular signalling in human. Nat Commun. 2015;6:7866.

26. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15:550.

27. Therneau TM, Grambsch PM: The Cox model. In Modeling survival data: extending the Cox model. Springer; 2000: 39-77

28. Bao X, Shi R, Zhao T, Wang Y, Anastasov N, Rosemann M, Fang W. Integrated analysis of single-cell RNA-seq and bulk RNA-seq unravels tumour heterogeneity plus M2-like tumour-associated macrophage infiltration and aggressiveness in TNBC. Cancer Immunol Immunother. 2020.

29. Jiang P, Gu S, Pan D, Fu J, Sahu A, Hu X, et al. Signatures of T cell dysfunction and exclusion predict cancer immunotheraph response. Nat Med. 2018;24:1550-1558.

30. Liu WT, Wang Y, Zhang J, Ye F, Huang XH, Li B, He QY. A novel strategy of integrated microarray analysis identifies CENPA, CDK1 and CDC20 as a cluster of diagnostic biomarkers in lung adenocarcinoma. Cancer Lett. 2018;425:43-53.

31. Niemira M, Collin F, Szalkowska A, Bielska A, Chwialkowska K, Reszec J, Niklinski J, Kwasniewski M, Kretowski A. Molecular Signature of Subtypes of Non-Small-Cell Lung Cancer by Large-Scale Transcriptional Profiling: Identification of Key Modules and Genes by Weighted Gene Co-Expression Network Analysis (WGCNA). Cancers (Basel). 2019;12.

32. Yu X, Zhang Y, Wu B, Kurie JM, Pertsemlidis A. The miR-195 Axis Regulates Chemoresistance through TUBB and Lung Cancer Progression through BIRC5. Mol Ther Oncolytics. 2019;14:288-298.

33. Corson TW, Zhu CQ, Lau SK, Shepherd FA, Tsao MS, Gallie BL. KIF14 messenger RNA expression is independently prognostic for outcome in lung cancer. Clin Cancer Res. 2007;13:3229-3234.

34. Sansam CL, Cruz NM, Danielian PS, Amsterdam A, Lau ML, Hopkins N, Lees JA. A vertebrate gene, ticrr, is an essential checkpoint and replication regulator. Genes Dev. 2010;24:183-194.

35. Villani AC, Satija R, Reynolds G, Sarkizova S, Shekhar K, Fletcher J, et al. Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes, and progenitors. Science. 2017;356.

36. Patente TA, Pinho MP, Oliveira AA, Evangelista GCM, Bergami-Santos PC, Barbuto JAM.
Human Dendritic Cells: Their Heterogeneity and Clinical Application Potential in Cancer Immunotherapy. Front Immunol. 2018;9:3176.

37. Miller JC, Brown BD, Shay T, Gautier EL, Jojic V, Cohain A, et al. Deciphering the transcriptional network of the dendritic cell lineage. Nat Immunol. 2012;13:888-899.

38. Van Allen EM, Miao D, Schilling B, Shukla SA, Blank C, Zimmer L, et al. Genomic correlates of response to CTLA-4 blockade in metastatic melanoma. Science. 2015;350:207-211.

39. Meunier E, Wallet P, Dreier RF, Costanzo S, Anton L, Ruhl S, et al. Guanylate-binding proteins promote activation of the AIM2 inflammasome during infection with Francisella novicida. Nat Immunol. 2015;16:476-484.

40. Shenoy AR, Wellington DA, Kumar P, Kassa H, Booth CJ, Cresswell P, MacMicking JD. GBP5 promotes NLRP3 inflammasome assembly and immunity in mammals. Science. 2012;336:481-485.

41. Gebhardt C, Sevko A, Jiang H, Lichtenberger R, Reith M, Tarnanidis K, et al. Myeloid Cells and Related Chronic Inflammatory Factors as Novel Predictive Markers in Melanoma Treatment with Ipilimumab. Clin Cancer Res. 2015;21:5453-5459.

42. Roberts EW, Broz ML, Binnewies M, Headley MB, Nelson AE, Wolf DM, Kaisho T, Bogunovic D, Bhardwaj N, Krummel MF. Critical Role for CD103(+)/CD141(+) Dendritic Cells Bearing CCR7 for Tumor Antigen Trafficking and Priming of T Cell Immunity in Melanoma. Cancer Cell. 2016;30:324-336.

43. Singel KL, Emmons TR, Khan ANH, Mayor PC, Shen S, Wong JT, et al. Mature neutrophils suppress T cell immunity in ovarian cancer microenvironment. JCI Insight. 2019;4.

44. Kandasamy M, Bay BH, Lee YK, Mahendran R. Lactobacilli secreting a tumor antigen and IL15 activates neutrophils and dendritic cells and generates cytotoxic T lymphocytes against cancer cells. Cell Immunol. 2011;271:89-96.

45. Xie X, Shi Q, Wu P, Zhang X, Kambara H, Su J, et al. Single-cell transcriptome profiling reveals neutrophil heterogeneity in ovarian cancer microenvironment. Nat Immunol. 2020;21:1119-1133.

46. Zhou SL, Dai Z, Zhou ZJ, Wang XY, Yang GH, Wang Z, Huang XW, Fan J, Zhou J. Overexpression of CXCL5 mediates neutrophil infiltration and indicates poor prognosis for hepatocellular carcinoma. Hepatology. 2012;56:2242-2254.

47. Lee HO, Hong Y, Etlioglu HE, Cho YB, Pomella V, Van den Bosch B, et al. Lineage-dependent gene expression programs influence the immune landscape of colorectal cancer. Nat Genet. 2020;52:594-603.

48. Yamada M, Kubo H, Kobayashi S, Ishizawa K, He M, Suzuki T, et al. The increase in surface CXCR4 expression on lung extravascular neutrophils and its effects on neutrophils during endotoxin-induced lung injury. Cell Mol Immunol. 2011;8:305-314.

49. Steele CW, Karim SA, Leach JDG, Bailey P, Upstill-Goddard R, Rishi L, et al. CXCR2 Inhibition Profoundly Suppresses Metastases and Augments Immunotherapy in Pancreatic Ductal Adenocarcinoma. Cancer Cell. 2016;29:832-845.

Figure Legends
Figure 1. Conceptual view of study design.

Figure 2. TIM infiltration landscape of myeloid cells in LUAD. A. Performance comparison of two reference matrixes. B. Different infiltration levels of myeloid cells between MSC1 and MSC2 patients. For brevity, myeloid cells were named as marker gene plus short name. M, N, DC, Mcyl, Mono, pDC, MonoDC represent macrophages, neutrophils, dendritic cells, monocytes highly expressing cell cycle related gene, monocytes, plasmacytoid DCs, and one cell type showing the signatures of both monocytes and DCs, respectively. C. KM plot for MSC subtypes in TCGA LUAD cohort. D. GO functional annotation of MSC1 and MSC2. E. Infiltration levels of two myeloid cell subtypes (SELENOP+M4 and TUBB+Mcyl) in tumor and tumor-adjacent tissues.

Figure 3. Clinical traits of MSC subtypes. A. ROC curve of MSC subtype classifier in the training data and the test data. Training set, red; Test set, blue. B-C. KM plots for MSC types in GSE72094 (B) and in GSE31210 (C). MSC1, yellow; MSC2, blue. D. Rate of relapse in MSC1 and MSC2 patients in GSE31210. Patients with relapse (n = 28) and non-relapse (n = 105) in MSC1 type; Patients with relapse (n = 36) and non-relapse (n = 57) in MSC2 type. E. Correlation of tumor stages and the infiltration of myeloid cells in TCGA LUAD cohort. Left: MSC score in different tumor stages (stage I, n = 257; stage II, n = 115; stage III, n= 81; stage IV, n =24); Right: correlation between tumor stages and myeloid cell types. The size of the circle represents log10 (FDR), the dark blue circle indicates a significant correlation (FDR < 0.05). F. Multivariable Cox proportional hazard regression analysis in TCGA LUAD cohort.

Figure 4. Exploration of treatment for two MSC type. A-B. TIDE score in different MSC subtypes for TCGA LUAD and two independent cohorts. C. Rate of ICB response in MSC1 and MSC2 patients in GSE126044. Patients with response (n = 5) and nonresponse (n = 2) in MSC1 type; Patients with relapse (n = 0) and non-relapse (n = 9) in MSC2 type. D. ROC curve of ICB response prediction. $M_{Score}$, red; Expression of PD-1, black; TIDE score, blue. E. Enrichment score for each drug in four cell lines. On the left most column, red box indicates the drug has therapeutic potential in total four cell lines, which are listed on the right-side.
Figure 5. Characteristics of the myeloid cell subtypes related to ICB response. A. Differences of myeloid cell infiltration levels between ICB responders and non-responders. B. Transcriptome traits of dendritic cells. Activation and migration signatures are shown on the right. C. Ratio of LAMP3+DCs of total DCs. mLN, nLN, nLung, and tLung denote metastatic, normal lymph nodes, normal lung tissues, and lung tumor tissues, respectively. D. Correlation between $M_{\text{Score}}$ and CTL level. E. Correlation of myeloid cell signature and CTL level. Marker genes were ranked based on the Pearson’s correlation coefficient. Left, the marker genes of IFIT3+N5 are denoted in red. Top 10 enriched marker gene sets of myeloid cells as identified by GSEA analysis. F. GO and KEGG functional annotation of IFIT3+N5.

Figure 6. Tumor-associated macrophages in tumor microenvironment. A. Cell communication in LUAD immune microenvironment. The ligand (left) is displayed separately from the receptor (right). B. Functional gene expressions in each macrophage subtypes. Tissue and MSC subtype enrichment state is shown as annotations. C. Differentially expressed genes between CXCL9+M2 and TIMP3+M3 subtypes. Top 20 genes in each macrophage subtype were labeled. D. GO functional annotation of CXCL9+M2 and TIMP3+M3. E. Circos plot for predicted interactions mediated by CXCL5-CXCR2. F. KM plots for TCGA LUAD cohort stratified by the expression levels of CXCL5, CXCR2 and the product of CXCL5 and CXCR2.

List of Supporting Information:

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Figure S1. Construction of MSC subtypes. A. UMAP plot for myeloid cells in LUAD patients. B. Infiltrating fractions of myeloid immune cells in 485 LUAD patients. X-axis denotes the samples. Y-axis denotes the infiltrating fraction. C. Dot plot for principal component analysis of TCGA LUAD patients. The MSC subtype is labeled. D. Bayesian Information Criterion (BIC) score of different models and different cluster number.

Figure S2. KM plots for TCGA LUAD cohort stratified by the infiltrating levels of myeloid cell types.

Figure S3. Paired box plot for the infiltrating levels of myeloid cell types in tumor and tumor-adjacent tissues.

Figure S4. Clinical traits of MSC subtypes. A. Multivariable Cox proportional hazard regression analysis in GSE72094 dataset. B-D. Correlation between the myeloid cell type infiltration levels. B. T stage; C. N stage; D. M stage in TCGA LUAD cohort. The red line
indicates a significant correlation with the threshold of FDR < 0.05.

**Figure S5.** Potential drug targeted gene set of MSC2 patients. **A.** Overlap of DEGs between MSC1 and MSC2 patients and myeloid cell markers. **B.** Overlap of upregulated genes in MSC2 patients and survival-related genes in LUAD cell lines. **C.** Heatmap of functional annotation for therapeutic target of MSC2 patients. The red box indicates the gene is related to the pathway showed at the right-side.

**Figure S6.** Transcriptomic characteristics of dendritic cell subtypes. **A.** Dot plot for GO functional annotation of dendritic cell subtypes. **B.** Violin plot for CD274 (PD-L1) expression level in dendritic cell subtypes.

**Figure S7.** Macrophage-related transcription characteristics. **A.** Heatmap for tissue prevalence estimated by macrophage subtype infiltrating level. **B.** Scatter plot for the correlation of TIMP3+M3 with S100A8+N3 (left) and IFITM2+N2 (right) respectively. **C.** GO functional annotation of SELENOP+M4. **D.** The expression levels of CXCR2 and CXCR4 in neutrophils.

**Table S1.** Infiltration levels of myeloid cell types in tumor and paired normal tissues of TCGA LUAD cohort.

**Table S2.** Coefficients of 14-gene signature.