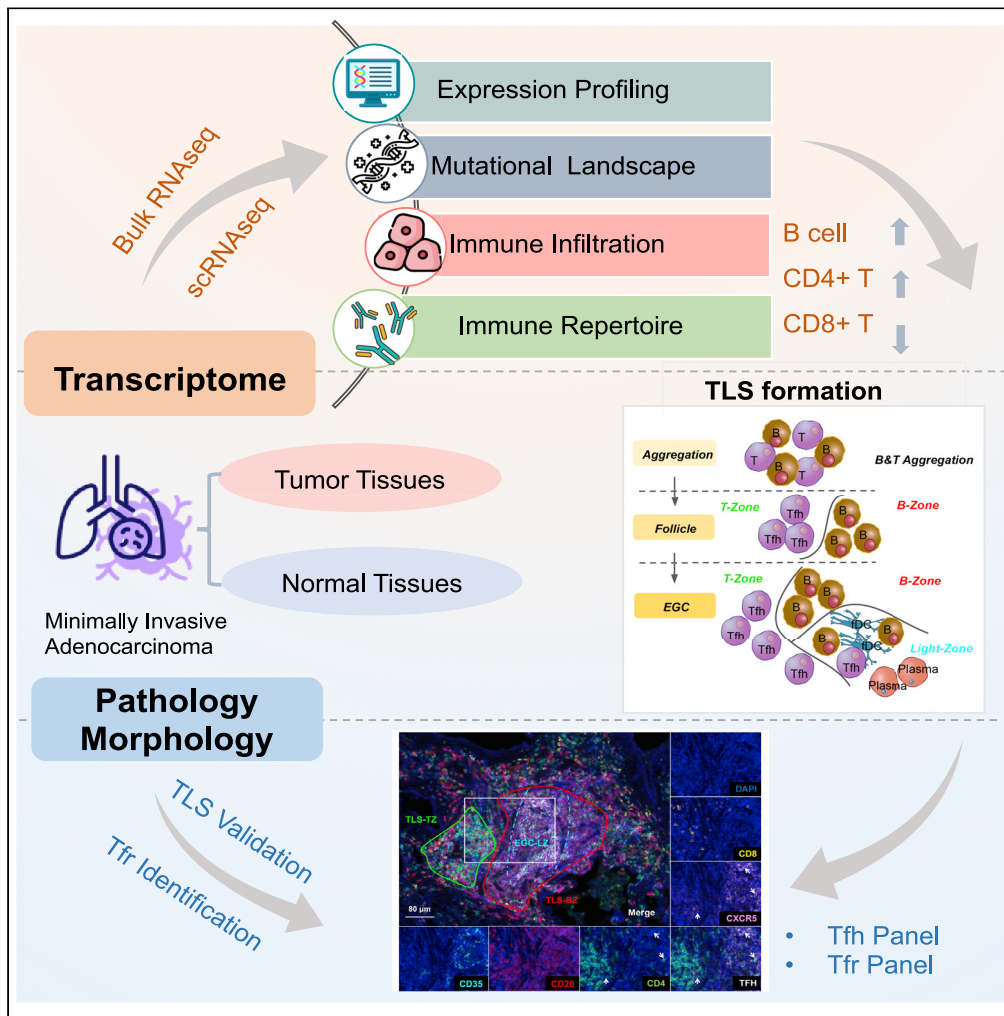


Article

Tertiary lymphoid structure and decreased CD8⁺ T cell infiltration in minimally invasive adenocarcinoma



Jin Wang, Dongbo Jiang, Xiaoqi Zheng, ..., Fan Zhang, Kun Yang, Peng Zhang

fzhang@tongji.edu.cn (F.Z.)
yangkunkun@fmmu.edu.cn (K.Y.)
zhangpeng1121@tongji.edu.cn (P.Z.)

Highlights

Higher infiltration and activation of B and CD4⁺ T cell characterize MIA tumors

MIA tumors are infiltrated with lower CD8⁺ T cells than normal tissues

TLS constituted by B, CD4⁺ T cells, and CD35⁺ FDC is validated in MIA tumors

Decreased CD8⁺ T is associated with Tfr-mediated immunosuppression in MIA tumor



Article

Tertiary lymphoid structure and decreased CD8⁺ T cell infiltration in minimally invasive adenocarcinoma

Jin Wang,^{1,9} Dongbo Jiang,^{2,9} Xiaoqi Zheng,³ Wang Li,⁴ Tian Zhao,⁴ Di Wang,⁵ Huansha Yu,⁵ Dongqing Sun,¹ Ziyi Li,¹ Jian Zhang,⁶ Zhe Zhang,⁷ Likun Hou,⁵ Gening Jiang,⁸ Ke Fei,⁸ Fan Zhang,^{4,*} Kun Yang,^{2,*} and Peng Zhang^{8,10,*}

SUMMARY

Knowledge of the tumor microenvironment (TME) in patients with early lung cancer, especially in comparison with the matched adjacent tissues, remains lacking. To characterize TME of early-stage lung adenocarcinoma, we performed RNA-seq profiling on 58 pairs of minimally invasive adenocarcinoma (MIA) tumors and matched adjacent normal tissues. MIA tumors exhibited an adaptive TME characterized by high CD4⁺ T cell infiltration, high B-cell activation, and low CD8⁺ T cell infiltration. The high expression of markers for B cells, activated CD4⁺ T cells, and follicular helper T (Tfh) cells in bulk MIA samples and three independent single-cell RNA-seq datasets implied tertiary lymphoid structures (TLS) formation. Multiplex immunohistochemistry staining validated TLS formation and revealed an enrichment of follicular regulatory T cells (Tfr) in TLS follicles, which may explain the lower CD8⁺ T cell infiltration and attenuated anti-tumor immunity in MIA. Our study demonstrates how integrating transcriptome and pathology characterize TME and elucidate potential mechanisms of tumor immune evasion.

INTRODUCTION

Lung cancer is the world's second-most common cancer but is by far the leading cause of cancer death. While most early lung cancer cases can be successfully treated, many lung cancers are not diagnosed until symptoms appear, at which time the disease has progressed to a more advanced stage. In recent years, low-dose CT (LDCT) scans have been used for early lung cancer detection on people with a higher risk of lung cancer. Although LDCT has been shown to decrease the rate of lung cancer death, it also has a ~20% false-positive detection rate and exposes people to low-dose radiation (Pinsky, 2014). While these imaging approaches are promising, early detection and treatment of lung cancer might be complemented by molecular approaches that characterize pre-cancerous lesions and their surrounding microenvironment. To better characterize the molecular, cellular, and tissue changes that drive tumor development and progression in its earliest stages, the National Cancer Institute of the United States recently issued new initiatives to construct an atlas of pre-cancer lesions and their surrounding microenvironment. Following The Cancer Genome Atlas (TCGA) project, which successfully profiled thousands of human cancers, this new effort promises to provide better understanding of cancer initiation and bring new strategies for cancer diagnosis and treatment.

Non-small-cell lung cancer (NSCLC) accounts for about 80% of all lung cancer incidences, with a majority belonging to the adenocarcinoma subtype. Early lung adenocarcinoma development is believed to progress from adenocarcinoma *in situ*, to minimally invasive adenocarcinoma (MIA), then to fully invasive adenocarcinoma (Vinayanuwattikun et al., 2016). Despite the difficulty of diagnosing early lung cancer lesions and obtaining samples, there have been a number of genomics studies on early lung cancers in recent years (Vinayanuwattikun et al., 2016; Izumchenko et al., 2015; Lavin et al., 2017; Mascoux et al., 2019). Most of these studies focused on identifying driver mutations potentially responsible for tumor initiation, which have been summarized in a recent review (Inamura, 2018). Besides tumor mutations, there has been growing appreciation for the important roles the tumor microenvironment (TME) plays in cancer

¹Clinical Translational Research Center, Shanghai Pulmonary Hospital, School of Life Sciences and Technology, Tongji University, Shanghai, China

²Department of Immunology, School of Basic Medicine, Air-Force Medical University (Fourth Military Medical University), Xi'an, China

³Department of Mathematics, Shanghai Normal University, Shanghai, China

⁴Department of Thoracic Surgery, Shanghai Pulmonary Hospital, Tongji University School of Life Sciences and Technology, Shanghai, China

⁵Tissue Bank, Department of Pathology, Experimental Animal Center, Shanghai Pulmonary Hospital, Tongji University School of Medicine, Shanghai, China

⁶Academy of Medical Engineering and Translational Medicine, Tianjin University, Tianjin, China

⁷Department of Gynecologic Oncology, Chinese PLA General Hospital, Beijing, China

⁸Department of Thoracic Surgery, Shanghai Pulmonary Hospital, Tongji University School of Medicine, Shanghai, China

⁹These authors contributed equally

¹⁰Lead contact

*Correspondence: f Zhang@tongji.edu.cn (F.Z.), yangkunkun@fmmu.edu.cn (K.Y.), zhangpeng1121@tongji.edu.cn (P.Z.)

<https://doi.org/10.1016/j.isci.2022.103883>



development, progression, and outcome. Changes in cellular and extracellular components of the TME can facilitate tumor behaviors such as unchecked tumor growth and immune evasion (Binnewies et al., 2018). To examine such TME changes, gene expression profiles from RNA-seq are more informative than DNA-based tumor profiling. There have also been multiple studies that use tumor RNA-seq data to infer immune cell populations and sub-types, and to associate immune infiltration with clinical outcomes or driver mutations (Thorsson et al., 2018). However, few such tumor profiling studies include matched adjacent tissues, and often the normal samples are sequenced at much lower coverage. Therefore, knowledge of the tumor immune microenvironment in patients with early lung cancer, especially in comparison with the matched adjacent tissues, remains lacking.

In the TME, B cells and T cells are able to recognize and bind tumor antigens to induce adaptive immune responses (Hennecke and Wiley, 2001; Heesters et al., 2016). How cytotoxic T lymphocytes (CTL) influence immune response to tumor progression has been under intensive investigation in recent years (Tschärke et al., 2015). However, the role of B-T cell interactions in anti-tumor immune response has not been completely understood, especially at the early stage of cancer development. Inflammation is known to be a potential cause and hallmark of early cancer development (Cousens and Werb, 2002), and clusters of B and T lymphocytes can form tertiary lymphoid structure (TLS) at the site of inflammation (Corsiero et al., 2016). Cytokine and chemokine signaling pathway have been shown to promote the recruitment of immune cells within or adjacent to sites of tumors, including T cells, B cells, dendritic cells, etc. The organization and interactions of tumor-infiltrating immune cells have revealed the presence of ectopic lymphoid organization termed TLS in a large number of tumors (Weinstein and Storkus, 2015). The organization of TLS ranges from simple aggregates of B cells and T cells to highly ordered structures, in which B cells undergo affinity maturation and differentiation to memory B cells and plasma cells (Pitzalis et al., 2014). Being the sites for the generation of circulating effector immune cells, TLS exhibits all the characteristics associated with adaptive immune response and constitute a paramount component of the TME, which plays a key role in controlling tumor growth. TLS has been documented in many cancer types, including lung cancer (de Chaisemartin et al., 2011), colon cancer (Bergomas et al., 2011), and breast cancer (Martinet et al., 2011). Not every patient with a specific cancer type develops TLS and the contribution of TLS to the disease varies considerably. There has been increasing research interest in understanding TLS formation and function in human pathology, especially related to clinical implications and potential therapeutic targeting (Pitzalis et al., 2014). A recent review summarized studies on the presence, composition, and functions of TLS in cancers and how TLS is correlated with therapeutic response (Sautès-Fridman et al., 2019). Three recent studies also reported TLS formation to be associated with response to immune checkpoint blockade (Cabrita et al., 2020; Helmink et al., 2020; Bruno, 2020; Petitprez et al., 2020). However, the impact of immune-stimulating or immune-suppressive components of TLS on tumor progression or immune evasion still awaits elucidation. Therefore, investigating TLS formation and its consequences on the tumor immune microenvironment in early stage tumors has the potential to inform disease etiology, precision medicine, and clinical prognosis.

In this study, we sought to characterize the TME of 58 pairs of minimally invasive adenocarcinoma (MIA) tumors and their matched adjacent normal tissues using RNA-seq. These tumors were initially diagnosed with LDCT, then surgically resected and confirmed for MIA by pathologists using H&E staining. To ensure an unbiased comparison between tumor and normal, we sequenced the RNA of both tumor and normal at higher depth than TCGA. This allowed us not only to examine the gene expression changes and infer immune cell infiltration levels but also to compare the clonal expansion patterns of the immune receptor repertoires between the tumors and the matched normal tissues. Furthermore, by combining transcriptome data, H&E staining, and multiplex immunohistochemistry (mIHC) staining on paraffin sections, we confirmed the TLS formation in the MIA TME, which might be associated with immune suppression and tumor initiation. These data and results provide insights into the etiology of early lung adenocarcinoma, which may inform immunotherapy treatment strategies for lung cancers.

RESULTS

Genes in cytokine-cytokine receptor interactions and mucin glycosylation are overexpressed in MIA

We collected 58 pairs of pathology-confirmed MIA tumors and matched adjacent normal tissues (Table S1). All the MIA tumors were confirmed by pathology using the 2015 WHO classification criteria of lung tumors (Travis et al., 2015). H&E staining confirmed the tumor lepidic growth pattern and the less than 5 mm in

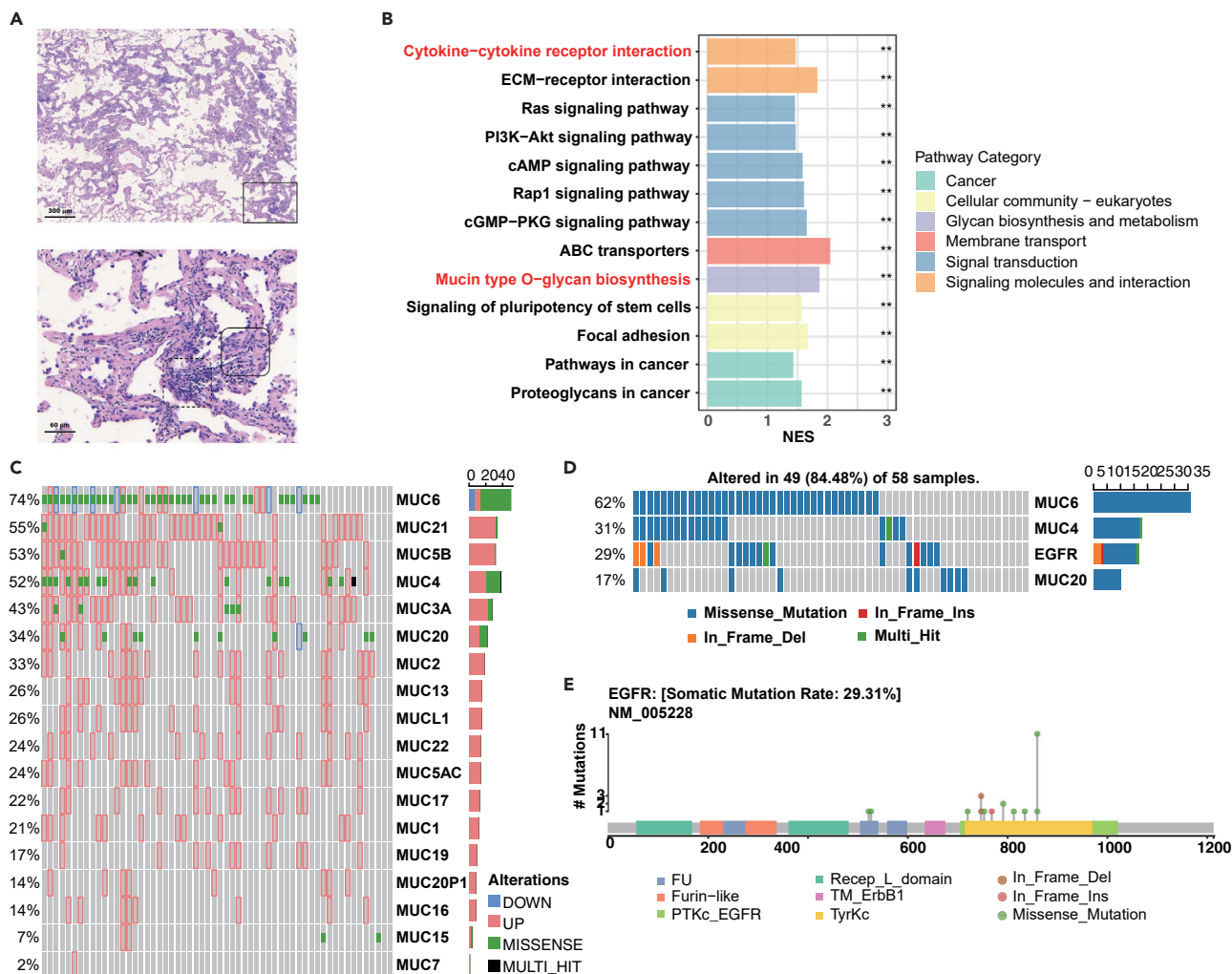


Figure 1. MIA tumors are characterized by highly mutated and overexpressed mucins, cytokine:cytokine-receptor interactions, and EGFR mutations

(A) Representative HE images of MIA. Top at 4× magnification, scale bar 300 μm. Boxed area of top image shown below (20× magnification, scale bar 50 μm). Lymphocyte infiltration (dashed box, bottom) and MIA tumor cells (solid box, bottom) are shown.

(B) Normalized enrichment scores (x axis) of tumor-upregulated KEGG pathways. Colors indicate different KEGG pathway categories. Asterisks denote significance level (***) FDR < 0.001; ** FDR < 0.01; * FDR < 0.05; NS FDR > 0.05).

(C) Alterations of mucin genes in patients with MIA. Each bar represents a patient with MIA. Shown are mutation classifications and expression status. Values on the right indicate the proportion of patients in which the gene was either mutated or had altered expression.

(D) Most commonly mutated genes in patients with MIA. Colors indicate variant type. Shown are the proportion of tumors containing a mutation in the indicated gene (left); total number of mutations in the indicated gene (right).

(E) Schematic of the EGFR gene, showing mutations (stick and ball markers) and protein domains (color blocks)

diameter of invasive components (Figures 1A, S1, and S2). We conducted RNA-seq of all the samples at a minimum depth of 100 million paired-end 150bp fragments, roughly three times the sequencing depth of TCGA samples. We mapped the RNA-seq reads with STAR (Dobin et al., 2013), quantified gene expression with HTSeq (Anders et al., 2015), and called differentially expressed genes between paired tumor and normal samples using edgeR (Robinson et al., 2010). In total, we identified 1,481 upregulated and 444 downregulated genes with over 2-fold change and a false discovery rate (FDR) of less than 0.05 (Table S2). We then applied gene set enrichment analysis (GSEA) (Subramanian et al., 2005) based on Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa and Goto, 2000) to examine altered pathways in MIA tumors (Figure 1B). The most enriched upregulated pathway in MIA is ABC-transporters which have an established tumor-promoting role in regulating cell survival, proliferation, migration, invasion, angiogenesis, and inflammation (Fletcher et al., 2010). In addition, widely studied cancer-associated pathways, such as

Ras signaling and PI3K-AKT signaling (Sanchez-Vega et al., 2018), are also upregulated. Furthermore, MIA tumors exhibit stronger cytokine:cytokine-receptor interactions than adjacent tissues. One particularly overexpressed cytokine, *CXCL13*, encodes a chemokine that has been reported as a major B lymphocyte chemoattractant (Legler et al., 1998) (Figure S3). Another pathway showing significant enrichment in MIA tumors is mucin-type O-glycan biosynthesis. Many genes in the mucin family are overexpressed in MIA, especially *MUC21*, *MUC5B*, and *MUC4* (Figure 1C), a phenomenon also observed in TCGA lung adenocarcinoma (LUAD) (Figure S4). Overexpression of mucin family proteins has been reported to contribute to the carbohydrate-enriched coating of cancer cells, which could promote integrin-mediated tumor growth and survival (Lakshmanan et al., 2015). In addition, overexpression and glycosylation of mucins, especially of *MUC1*, have been reported in multiple cancers and implicated to promote tumor metastasis and drug resistance (Bouillez et al., 2017; Maeda et al., 2018). The overexpression of mucin family members likely accounts for the enrichment of mucin-type O-glycan pathway in MIA tumors.

MIA tumors harbor EGFR and mucin gene mutations

We next sought to detect mutations in MIA tumors by comparing RNA-seq data between tumor and adjacent normal tissue using Varscan2 (Koboldt et al., 2012). Although this is less sensitive than mutation calling from DNA-based approaches such as whole genome or whole exome sequencing, we still managed to find approximately 70 (median) variants per tumor, most of which were missense mutations (Figure S5A). The most enriched type of single nucleotide variants (SNV) is C > T transition (Figure S5B), which is known to be a signature of mismatch repair deficiency resulting in elevated microsatellite instability (MSI) (Harris, 2013). To computationally assess the MSI level of these MIA samples, we applied MSIsensor (Niu et al., 2014) to compare the RNA-seq data between tumor and normal. Although this is also less sensitive than DNA-based approaches, we indeed observed higher MSI score in MIA tumors with higher CT transition rate (Figure S5C).

EGFR was found to be mutated in 29% of the MIA samples (Figure 1D), consistent with the previous report of the prevalent *EGFR* mutations in progressive stages of lung adenocarcinoma (Wang et al., 2018). *EGFR* mutations occur mainly at the mutational hotspot “L858 R/M” (Paez et al., 2004) (Figure 1E) in the kinase domain, in agreement observations in TCGA LUAD (Figure S5D). *EGFR* is known to participate in the signaling of the RTK-RAS pathway (Sanchez-Vega et al., 2018). The gain-of-function *EGFR* mutations primarily contributed to the oncogenic activation of the RTK-RAS pathway. Another important gene in the RTK-RAS pathway, *KRAS*, is mutated in only two tumors (Figure S5E), and as expected, occurs mutually exclusively with *EGFR* mutations (Li et al., 2014).

In addition to RTK-RAS pathway mutations, *MUC6* and *MUC4* genes were also highly mutated in MIA tumors. These two genes, as well as other MUC family members *MUC16*, *MUC5B*, and *MUC17*, are frequently mutated in TCGA LUAD (Figure S4). Most of the MUC family SNVs are missense mutations, and together with their frequent over expression, suggest that tumor-associated mucins might create neoantigens in MIA tumors.

Higher B cell and CD4⁺ T cell infiltration and activation but lower CD8⁺ T cell infiltration in MIA tumors

To decipher immune cell infiltration in the MIA TME, we estimated the immune cell composition from bulk tumor RNA-seq data using ImmuneDeconv (Sturm et al., 2019) and focused on the observations consistently predicted by different algorithms (Table S3). Relative to adjacent normal tissues, MIA samples have significantly higher B cells and CD4⁺ T cells, and significantly lower CD8⁺ T cells and macrophages estimated by TIMER (Li et al., 2016) and CIBERSORT (Newman et al., 2015) (Figures 2A, S6, and S7). Consistent with the observation in MIA, TCGA LUAD tumors also showed higher B cell and lower CD8⁺ T cell infiltration in tumors than the adjacent normal tissues. Interestingly, B cell infiltration gradually decreases toward normal during tumor progression from stage I to stage IV (Figure 2B) in TCGA LUAD, suggesting that increased B cell increase is associated with early LUAD onset. To ensure the consistent immune cell inference, we also performed scRNA-seq analysis from five patients with early-stage LUAD with paired tumor and normal tissues. The higher B cell, higher CD4⁺ T cell, and lower CD8⁺ T cell percentages observed in tumor tissues relative to normal tissues were also observed at single-cell resolution (Table S4; Figure S8).

Next, we characterized the immune repertoires in these samples to investigate the activation and clonal expansion of tumor-infiltrating B and T cells. We ran TRUST, a computational algorithm we developed

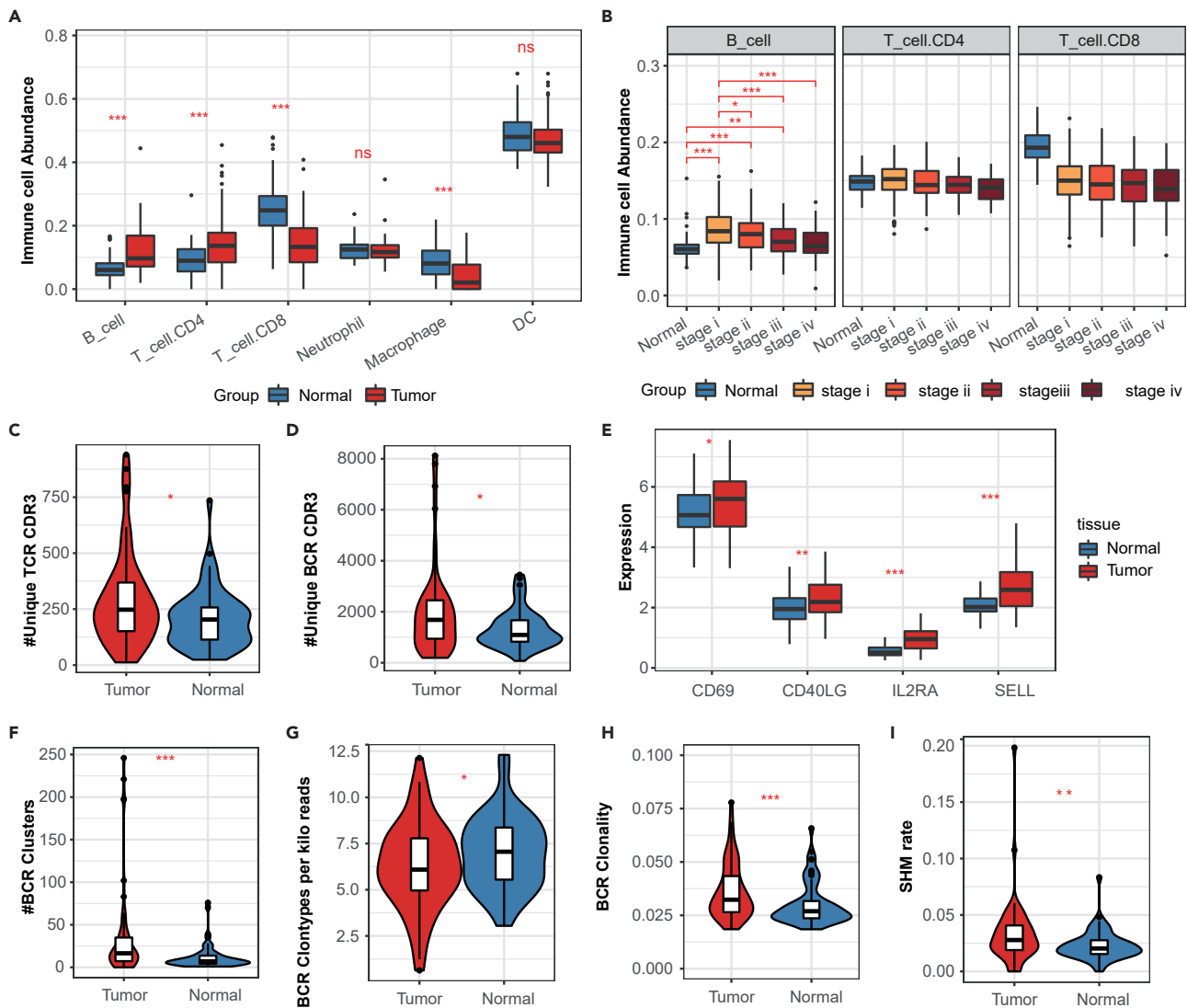


Figure 2. Higher B and CD4⁺ T cell infiltration and activation but lower CD8⁺ T cell infiltration characterize MIA tumors

(A) Comparison of immune components between tumor (red) and adjacent normal tissues (blue) in MIA. Immune cell abundance values (y axis) were generated from TIMER. Significance levels are indicated.

(B) Immune infiltration among TCGA lung adenocarcinoma (LUAD) samples separated by clinical stage (colors).

(C and D) Comparison of fraction of TCR reads (C) and fraction of BCR reads (D) between MIA tumor and adjacent normal tissues.

(E) Comparison of gene expressions on CD4⁺ T cell activation markers (CD69, CD40LG, IL2RA, and CD62) between MIA tumor and adjacent normal tissues.

(F–I) Comparison of the number of BCR clusters (F), BCR clonotypes per kilo reads (G), BCR clonality (H), and somatic hyper-mutation rate of BCR (I) between MIA tumor and adjacent normal tissues. Asterisks denote significance on difference between MIA tumors and adjacent normal tissues. p values are calculated using Mann-Whitney U test; ***p < 0.001; **p < 0.01; *p < 0.05; ns p > 0.05

(Li et al., 2017; Hu et al., 2019), to assemble the T cell and B cell receptor (TCR and BCR) repertoires from RNA-seq data. Significantly higher fractions of raw reads were mapped to TCR and BCR in MIA tumors than normal tissues (Figures S9A and S9B). Consistent with this observation, TRUST assembled significantly higher number TCR and BCR unique clonotypes of complementary determining region 3 (CDR3) sequences in MIA tumors (Figures 2C and 2D). Because TRUST has better power to call the abundant and clonal-expanded TCRs and BCRs, the fact that TRUST assembled more TCR/BCR clonotypes already suggest that T cells and B cells have clonal expansion in the MIA tumors. Based on the CD4⁺ and CD8⁺ T cell infiltration trend in MIA vs adjacent normal, we expect most of the T cell activation and clonal expansion to occur in CD4⁺ T cells. Indeed, MIA tumors have significantly higher expression of T cell activation markers CD69, CD40LG, CD25 (IL2RA), and CD62L (SELL) (Hosono et al., 2003; Ivetic et al., 2019) (Figure 2E), which

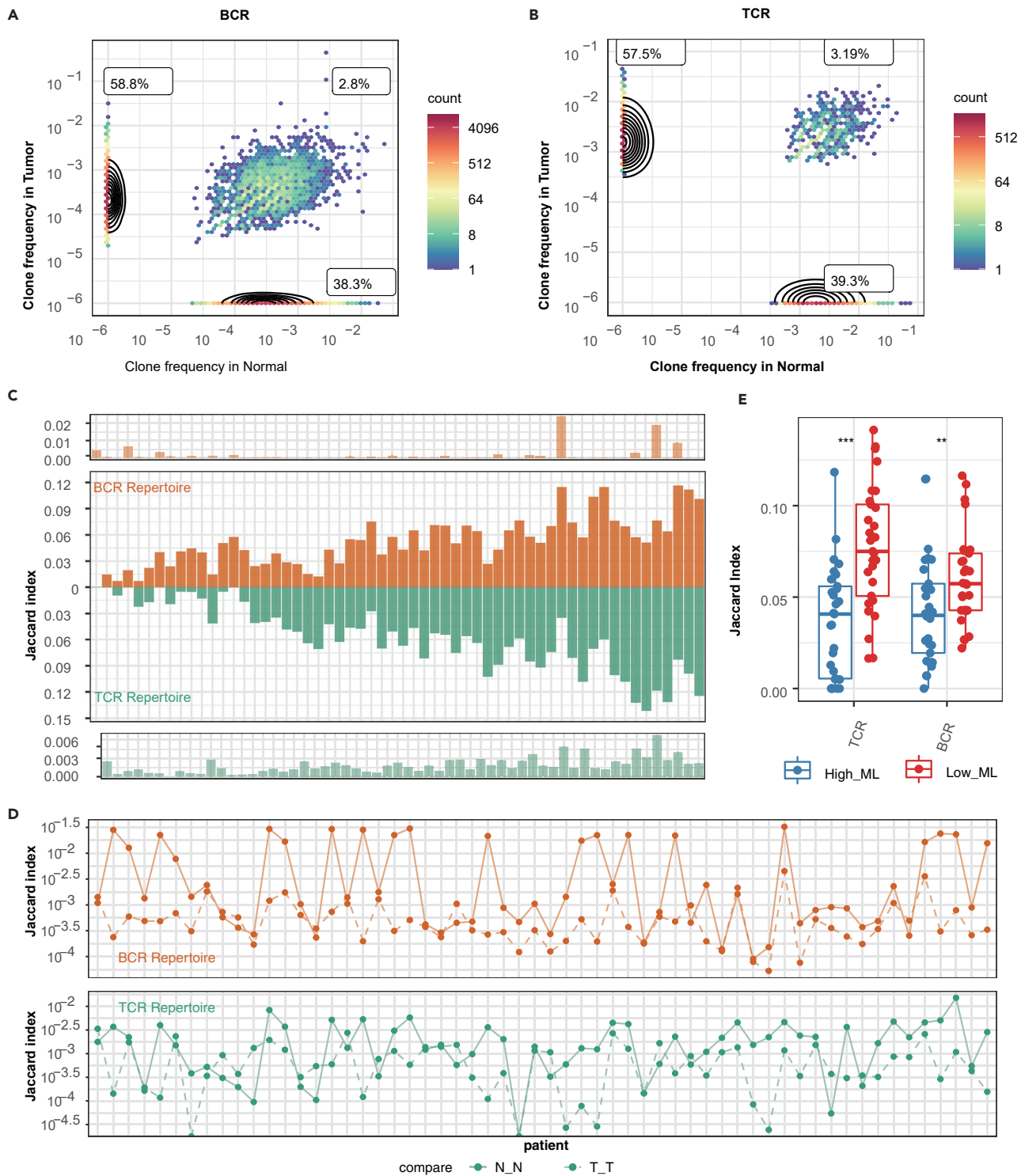


Figure 3. Tumor-infiltrating B cells and T cells are activated by tumor antigens in MIA

(A) Normalized BCR clonotype frequency on patient level. Each dot represents a BCR clonotype and the normalized clonotype abundance in tumor and adjacent normal tissue. The three percentage values correspond to the overall percentage of tumor-specific clonotypes (58.6% in BCR), normal-specific clonotypes (38.6% in BCR), and shared (2.86% in BCR) clonotypes.

(B) Normalized TCR clonotype frequency on patient level.

Figure 3. Continued

(C) Jaccard similarity of the BCR/TCR clonotypes between tumor and adjacent normal tissues. Shown are shared clonotypes within each patient (dark brown for BCR, dark green for TCR) and average shared clonotypes between tumors and unmatched normal samples (light brown for BCR, light green for TCR). (D) Jaccard similarity of BCR and TCR repertoire among tumors (dashed line) and normal tissues (solid line) shown in top (BCR) and bottom (TCR) panel, respectively.

(E) Jaccard similarity in groups with high (blue) and low (red) mutation load. p values are calculated using Mann-Whitney U test; ***p < 0.001; **p < 0.01; *p < 0.05; ns p > 0.05

are mainly contributed by CD4⁺ T cells (Figure S9C). However, because TCR diversity and clonality calculation could be confounded and because we couldn't distinguish TCRs from CD4⁺ vs CD8⁺ T cells, we turned to BCR to further evaluate immune cell clonal expansion. For B cells, we not only observed higher total number of BCR CDR3 clonotypes (Figure 2D) and BCR clusters (Figure 2F) but also lower BCR diversity (Figure 2G) and higher BCR clonality (Figure 2H) in MIA than adjacent normal, which strongly support B cell clonal expansion. In addition, B cells, upon antigen recognition and clonal expansion, undergo somatic hyper-mutation (SHM) to evolve BCRs with better antigen-binding affinity, and class-switch recombination (CSR) to transition to antibody-producing plasma cells. Indeed, we also observed higher BCR SHM rate (Figure 2I), higher CSR (Figure S9D) in MIA than adjacent normal, further supporting higher activated and clonal-expanded B cells.

To check whether the activated tumor-infiltrating B and T cells could generate more diverse clonotypes for binding to different antigens in the TME, we compared the clonotypes detected between tumors and adjacent normal tissues. Majorities of the BCR (58.8%, Figure 3A) or TCR (57.5%) clonotypes were uniquely detected in tumors (Figure 3B), suggesting more B cell and T cell infiltration and clonal expansion in MIA tumors. We measured the repertoire similarity between each pair of MIA and adjacent tissue using Jaccard index (Thomas et al., 2014), and found within-individual repertoire similarity to be much higher than the across-patient repertoire similarity (Figure 3C). Interestingly, the TCR and BCR similarities between tumor and normal (Figure 3C) are correlated, suggesting that T cells and B cells have coordinated clonal expansion in the MIA tumors. In addition, repertoire similarities among normal samples (N-N) are higher than that among MIA samples (T-T) from differential patients (Figure 3D), suggesting that tumor-specific antigens stimulate the expansion of different TCR and BCR clones in different tumors. Furthermore, tumors with higher mutation burden, as inferred from VarScan2, have significantly lower BCR (p value < 0.01) and TCR (p value < 0.001) repertoire similarity between tumor and adjacent tissues (Figure 3E). Collectively, these observations suggest that TRUST-inferred immune receptor repertoires represent activated and clonal-expanded B cells and potentially CD4⁺ T cells for recognizing and binding to tumor antigens in MIA.

Mucin mutation and overexpression is associated with CD4⁺ T cell-mediated B cell activation in MIA tumors

Because B cells could be activated by direct antigen stimulation and interaction with CD4⁺ T cells, we further checked how B cell activation was induced in MIA. It has been reported that MUC1-glycosylated peptides can induce MHCII-restricted B-T cell interactions (Cai et al., 2016). In addition, a synthetic peptide from the MUC1 tandem repeat region was shown to be presented by MHCII molecules which resulted in the initiation of T helper cell responses (Cai et al., 2013). In our MIA samples, we observed a positive association between mucin level and MHCII level (Figure 4A), raising the possibility that higher mucin expression in cancer cells might be associated with higher MHCII presentation. In addition, CD40 and its ligand CD40L are well known to mediate the interaction between B cells and CD4⁺ T cells (Schoenberger et al., 1998). We observed significant positive correlation between CD40 and CD40L in MIA tumors (Figure 4B), which suggests CD4⁺ T cell-mediated B cell activation. The average level of CD40 and CD40L is associated with mucin expression (Figure 4C). Furthermore, we observed MIA tumors with overexpressed mucins have higher levels of IgG and IgA isotypes, suggesting a higher level of B cell activation and transition to plasma cells (Figure 4D). MIA tumors with frequently mutated mucins (MUC6, MUC4, and MUC20) also have significantly higher B cell infiltration (Figure S10A), suggesting that mucin mutations act as cancer antigens to trigger B cell response. Taken together, these results suggest interactions between the mutation or overexpression of mucins, MHCII presentation to CD4⁺ T cells, and B cell activation and plasma cell transition in MIA tumors.

To further investigate B cell activation among heterogeneous tumors, we characterized the overall immune status among the MIA samples using five signatures derived from the TCGA pan-cancer study (Thorsson et al., 2018). These signatures include "IFN-g" (representing IFN-g response), "Lymphocyte" (representing overall

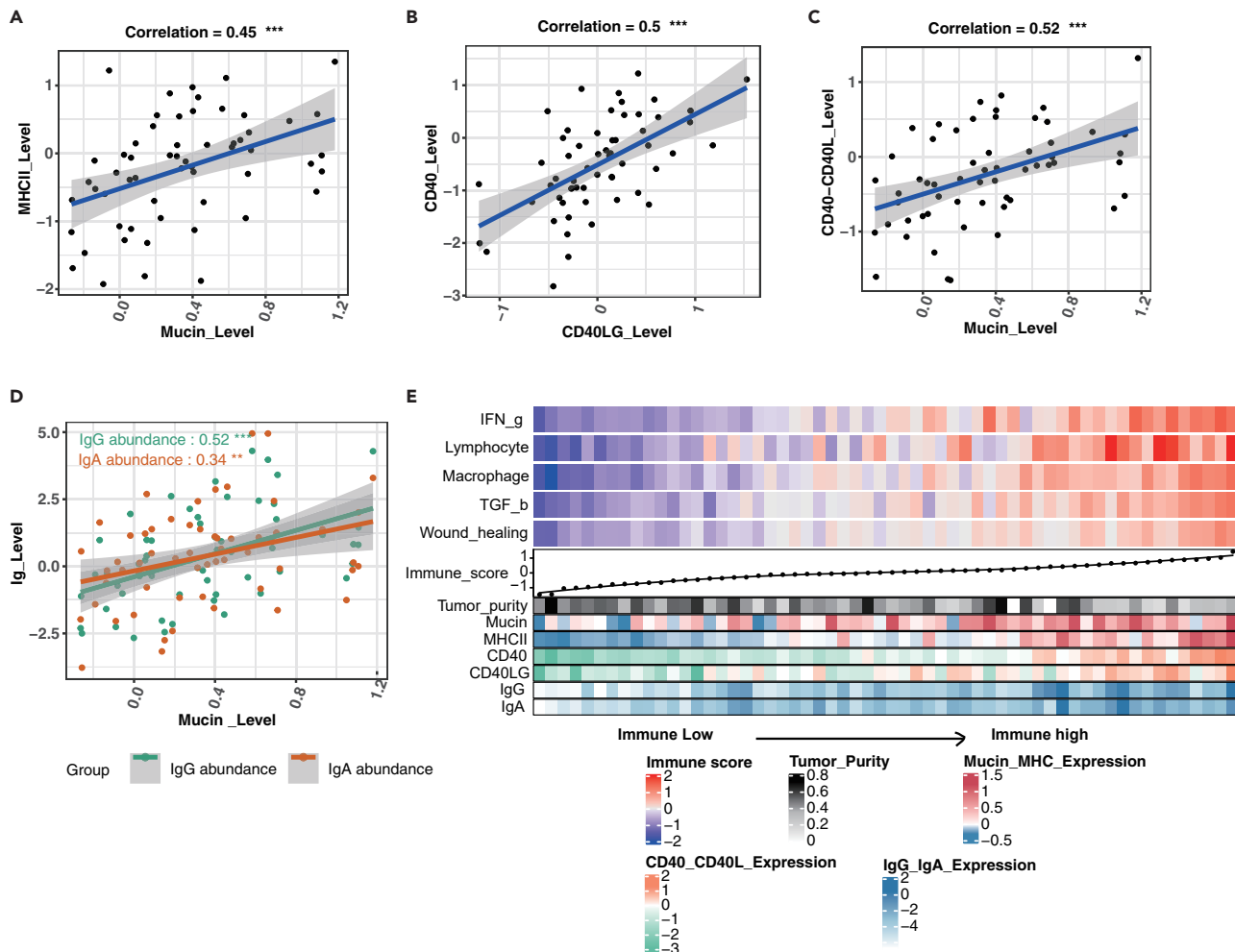


Figure 4. Mucin mutation and overexpression is associated with CD4⁺ T cell-mediated B cell activation in MIA tumors.

(A) The correlation between mucins and MHCII expression.

(B) The correlation between CD40 and CD40LG expression.

(C) The correlation between mucins and averaged level of CD40 and CD40LG.

(D) The correlation between mucin expression with IgG (green) and IgA (brown) expression in patients with MIA.

(E) MIA tumors (x axis) ranked by immune-related signature score (heatmap colors). The top bar shows the average immune signature score for each patient. The bottom bars show the tumor purity, expression level of mucins, MHCII molecules, CD40, CD40LG, IgG, and IgA. p values are calculated using Spearman correlation test; ***p < 0.001; **p < 0.01; *p < 0.05; ns p > 0.05

lymphocyte infiltration by B and T cells), “Macrophage” (representing the activation of macrophages/monocytes), “TGF-β” (representing TGF-β response), and “Wound healing” (representing core serum response). These five signatures seem to have correlated signals in the MIA tumors, so we used the average of the five to rank the MIA tumors from “immune-low” to “immune-high” (Figure 4E, STAR Methods). The “immune-high” tumors are associated with increased expression of MHCII (Figure S10B) and mucins (Figure S10C). In addition, the “immune-high” tumors are also associated with increased expression of CD40 (Figure S10D) and CD40L (Figure S10E), suggesting stronger interactions between B cells and CD4⁺ T cells. Furthermore, “immune-high” tumors are associated with higher B cell signals, such as IgG and IgA expression (Figures S10F and S10G) and unique BCR clonotypes (Figure S10H). Collectively, our observations suggest that mucin expression is associated with CD4⁺ T cell-mediated B cell activation in MIA adaptive immunity.

Formation of tertiary lymphoid structure (TLS) in MIA tumors

Previous literature (Corsiero et al., 2016) reported that B cell and CD4⁺ T cell aggregation could lead to the formation of tertiary lymphoid structures (TLS). TLS starts from scattered B and T cell aggregation, progress to

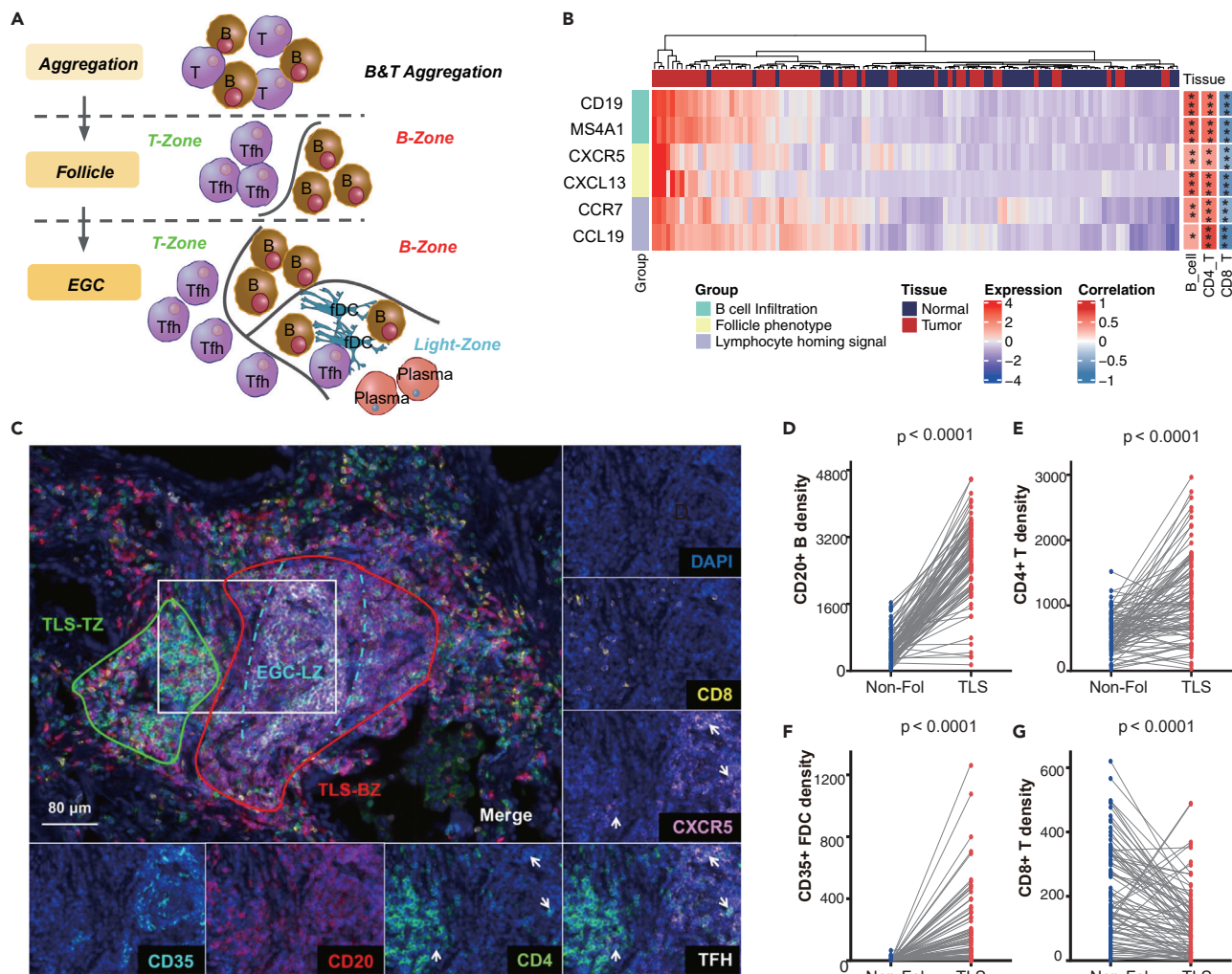


Figure 5. TLS formation in MIA tumors

(A) The schema of TLS formation. The first stage is B & T cell aggregation. The second stage is follicle formation with formation with follicular dendritic cell in light-zone.

(B) Hierarchical clustering of Z score normalized gene expression of B cell infiltration markers (CD19 and CD20), follicle phenotype markers (CXCR5 and CXCL13), and lymphocyte homing signal markers (CCR7 and CCL19). Color of the top bar represents tumor (red) or adjacent normal (blue) tissue. Correlations between expression levels from these six TLS markers and immune cell abundance corrected by tumor purity are shown in the right bar. p Values are calculated using Spearman correlation; ***p < 0.001; **p < 0.01; *p < 0.05; ns p > 0.05.

(C) Typical observation field of MIA six-plex TLS panel staining. The merged-channel multispectral image was captured after 5-cycle TSA staining and DAPI counterstain, revealing a classic lymphoid structure residing in MIA stromal tissue. Single channels are shown for the area bounded by the white rectangle. Fake colors are assigned as DAPI: blue, CD8: yellow, CXCR5: pink, CD4: green, CD20: red, CD35: cyan. Tfh cells are highlighted with white arrows (lower right corner images).

(D–G) Paired comparisons of CD20⁺ B cell density (D), CD4⁺ T cell density (E), CD35⁺ follicle dendritic cell density (F), and CD8⁺ T cell density (G) between non-follicle (Non-Fol) and TLS areas

follicles with separate T-zones and B-zones, and finally matures into a highly organized structure, including ectopic germinal centers (EGC) with follicular dendritic cells (FDC) in the light-zone for better B cell activation (Sautès-Fridman et al., 2019) (Figure 5A). In the EGC of mature TLS, B cells activates somatic hypermutations, isotype class-switches, and antibody production (Pitzalis et al., 2014). So far, our analysis revealed higher B cell and CD4⁺ T cell infiltration, lower CD8⁺ T cell infiltration (Figure 2A), and higher CD4⁺ T cell-mediated B cell activation (Figures 3 and 4) in MIA tumors. This motivated us to examine the expression of typical TLS markers for evidence of TLS formation in MIA. CD19 and MS4A1 (CD20), which are typical B cell markers (Ginaldi et al., 1998), were highly expressed in MIA tumors. The ligand-receptor pair CXCL13 and CXCR5, which are conventional indicators of follicles (Havenar-Daughton et al., 2016), were also significantly higher expressed in MIA tumors. Furthermore,

the ligand-receptor pair CCL19 and CCR7, which are important for promoting T cell and B cell homing to lymphoid structures (Kobayashi et al., 2017), were also highly expressed in tumors (Figure 5B). These six TLS markers were all positively correlated with TIMER-estimated B cell and CD4⁺ T cell abundance but negatively correlated with CD8⁺ T cell abundance (Figure 5B), suggesting that TLS formation provides an unfavorable microenvironment for CD8⁺ T cells.

Besides the six TLS markers we evaluated above, another 12-chemokine gene signature has also been proposed to predict the presence of TLS (Messina et al., 2012). We observed highly correlated expression signals between our six TLS markers and the 12-chemokine signatures (Figure S11A). As a result, the two approaches give concordant TLS predictions in MIA tumors, and found TLS signals to be associated with immune-high tumors with stronger CD4⁺ T cell-mediated B cell activation (Figures S11B and S11C). To further confirm the presence of aggregated B cells and CD4⁺ T cells in non-small-cell lung cancer (NSCLC), we also examined two recently published NSCLC single-cell RNA-seq datasets with tumors and adjacent normal tissues (Lambrechts et al., 2018; Wu et al., 2020b). Indeed, NSCLC tumors have higher level of MS4A1⁺ B cells, CD4⁺CXCR5⁺ follicular helper T cells (Tfh) than adjacent normal tissues (Figures S8J, S8K, S12A, S12B, S13A, and S13B). In addition, we used a computational algorithm CellPhoneDB (Efremova et al., 2020) infer cell-cell communication from combined expression of multi-subunit ligand-receptor complexes to confirm the interaction between CD40LG in CD4⁺ T cells with CD40 in B cells (Figure S13C). Collectively, our MIA RNA-seq data and NSCLC scRNA-seq data support TLS formation in lung adenocarcinoma (Figure 4).

With the transcriptome evidences of TLS formation in MIA tumors, we next sought to confirm TLS formation from pathology imaging. Firstly, we examined the H&E pathology slides and observed lymphocytes aggregation (Figure S14), which support TLS formation in MIA tumors. Then, we used a six-marker multiplex immunohistochemistry (mIHC) to stain for TLS in 22 MIA tumors with high Tfh phenotype, which included the markers CD20 for B cells, CD4 and CXCR5 for follicular helper T cells (Tfh), CD8 for CD8⁺ T cells, and CD35 for follicular dendritic cells (FDC) (Figure 5C). mIHC confirmed the formation of TLS in MIA tumors, although different samples showed varying level of TLS maturation (Figure S15). In the samples with mature EGC, we observed TLS B cell zone (TLS-BZ) as marked by CD20, TLS T cell zone (TLS-TZ) as marked by CD4, and EGC light zone (EGC-LZ) as marked by CD20, CD4, CXCR5, and CD35 (Figures 5C, S16, and S17). Furthermore, we segmented each observation field into different areas (adenocarcinoma tissue, non-Follicle, TLS T zone, TLS B zone, and EGC light zone) to quantify and compare the immune cell density between TLS follicles and non-follicle areas. Here, TLS follicle area includes TLS-TZ, TLS-BZ, and EGC-LZ, and non-follicle area includes scattered T cells and B cells outside of TLS. In MIA tumors, CD20⁺ B cells (Figure 5D, p value < 0.0001), CD4⁺ T cells (Figure 5E, p value < 0.0001), and CD35⁺ FDC (Figure 5F, p value < 0.0001) were significantly higher in TLS than in non-follicular area. In contrast, CD8⁺ T cell showed significantly lower density (p value < 0.0001) in TLS follicles than in non-follicle areas (Figure 5G).

Treg recruitment as potential mechanism of decreased CD8⁺ T cell infiltration in tertiary lymphoid structures

Our transcriptome and mIHC data indicated that MIA tumors had lower CD8⁺ T cell infiltration (Figure 2A) and lower CD8⁺ T cell density in TLS (Figure 5G), respectively. In addition, MIA tumors have lower PD-L1 (CD274) expression than adjacent normal, suggesting PD-L1 independent mechanisms for CD8⁺ T cell suppression (Figure S18). This prompted us to explore the potential mechanism of CD8⁺ T cell decrease in TLS and MIA. We examined immune suppressive cells, such as Tregs, M2 macrophages, neutrophils, monocytes, and fibroblasts (Cassetta and Pollard, 2018) for possible cause of CD8⁺ T cell decrease. Most suppressive cells were lower in MIA tumors than adjacent normal, with the exception of Treg cells (Figure 6A), suggesting a Treg-mediated CD8⁺ T cell decrease. From three independent scRNA-seq datasets, we also observed CD4⁺CXCR5⁺FOXP3⁺ follicular T regulatory (Tfr) cells to be enriched in NSCLC tumors (Figure S8K, S12B, and S13B). These data suggest that Treg cells might be responsible to the decreased CD8⁺ T cell infiltration in the region of lymphocyte aggregations (Figure 4). To quantify the degree of immune suppression over immune stimulation, we measured the ratio of follicular regulatory T cells (Tfr) to follicular helper T cells (Tfh) in TLS by the normalized expression of three pairs of phenotype markers for Tfr and Tfh cells (FOXP3 over BCL6, PDCD1 over ICOS, and IL2RA over IL7R) (Xu et al., 2017), respectively. For each pair of markers, the ratio of Tfr/Tfh was negatively associated with CD8⁺ T cell infiltration (Figure 6B), suggesting that Tfr in TLS might have negative effects on CD8⁺ T cell infiltration in local inflammation.

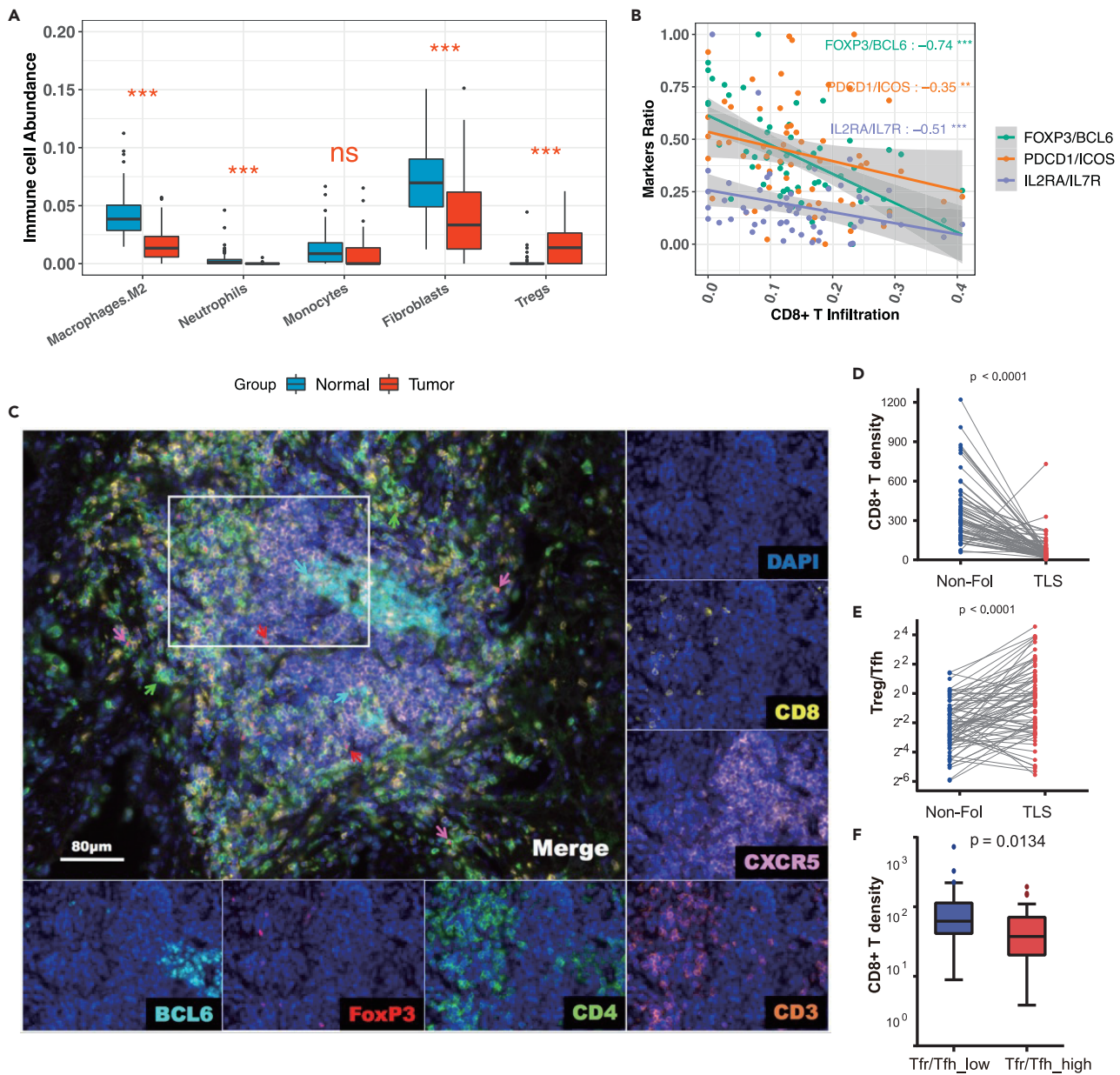


Figure 6. Treg recruitment as potential mechanism of decreased CD8⁺ T cell infiltration in tertiary lymphoid structures

(A) Comparison of immune suppressive cell infiltration estimated from XCELL between MIA tumor and adjacent normal tissues. p values are calculated using Mann-Whitney U test; ***p < 0.001; **p < 0.01; *p < 0.05; ns p > 0.05.

(B) Plot of Tfr/Tfh marker ratio vs CD8⁺ T cell infiltration. FoxP3, PDCD1, and IL2RA are markers of Tfr cells, while BCL6, ICOS, and IL7R are markers of Tfh cells. See text. p values are calculated using Spearman correlation test; ***p < 0.001; **p < 0.01; *p < 0.05; ns p > 0.05.

(C) Typical observation field of MIA 7-plex panel staining for Tfr and Tfh. Merged multispectral image was captured after 6-cycle TSA staining and DAPI counterstain and reveals a mature TLS architecture. Single channels are shown for the area bounded by the white rectangle. Fake colors are assigned as DAPI: blue, CD8: yellow, CXCR5: pink, CD4: green, FOXP3: red, BCL6: cyan and CD3: orange. Sub-populations of CD4⁺ T cells are highlighted by arrows (Tfr: red, Treg: purple, Th: green and Tfh: cyan).

(D) Paired comparison of CD8⁺ T cell density between non-follicle (Non-Fol) and TLS areas.

(E) Paired comparison of ratio of regulatory T cells to helper T cells between TLS and non-follicle areas.

(F) Comparison of CD8⁺ T cell density between groups with high and low Tfr/Tfh ratio

To further quantify the Tfr and Tfh in TLS, we used seven-marker mIHC staining on sequential slides of 22 MIA samples with high TLS expression markers. Specifically, we used CD3, CD4, CXCR5, and FOXP3 to detect Tfr cells and CD3, CD4, CXCR5, and BCL-6 to detect Tfh cells (Figure 6C), respectively. Within the 69 available observation fields from the 22 MIA tumors, we found lower CD8⁺ T cell density in TLS follicles than non-follicular areas (Figure 6D, $p < 0.0001$), which was consistent with our observations in the six-plex mIHC panel (Figure 5E). In addition, we observed a higher ratio of regulatory T cells to helper T cells in TLS follicles than in non-follicular areas (Figure 6E, $p = 0.0032$), indicating stronger immune suppression associated with Tfr cells in TLS. To determine how the balance between immune suppression from Tfr and immune stimulation from Tfh influence CD8⁺ T cell infiltration in TLS, we split all the observation fields into two groups based on the Tfr/Tfh ratio. While higher Tfr/Tfh ratio was associated with lower CD8⁺ T cell density in TLS follicles (Figure 6F), no such difference was found in non-follicular areas (Figure S19). These results suggest TLS formation to be associated with decreased CD8⁺ T cell infiltration through Tfr-mediated immune suppression.

DISCUSSION

This study is not a randomized controlled clinical trial but a research study mainly investigating tumorigenesis mechanism and immune evasion in MIA to provide insights into the pathophysiology of early lung cancer development. We characterized the tumor microenvironment (TME) of minimally invasive adenocarcinomas (MIA) by integrating transcriptome data (molecular profiling, mutation analysis, immune cell estimation, and immune repertoire inference) and pathology staining (H&E staining and multiplex immune-histochemistry). We observed varying levels of TLS formation, from the initial T cell and B cell aggregates, to follicle formation, to the mature ectopic germinal centers (EGC), in the MIA TME. Furthermore, we propose follicular Treg (Tfr) cells in the TLS potentially as an attenuated CD8⁺ T cell infiltration mechanism in MIA tumors.

Our analysis revealed mucins in MIA samples to be highly mutated, overexpressed, and highly glycosylated. Glycosylated tumor-associated mucin peptides have been reported to be bound by MHCII molecules and presented by B cells, leading to the activation of CD4⁺ T cells which in turn activates B cells through CD40⁻CD40L interactions. Our study identified higher tumor-infiltrating B cells and CD4⁺ T cells in both MIA and TCGA lung adenocarcinoma. Moreover, MIA tumors exhibited strong signals of B cell activation, somatic hyper-mutation, and clonal expansion. Our observation of the positive associations of mucins with MHCII, B cell activation, and CD40⁻CD40L interactions potentially suggested mucin overexpression as a potential cause of CD4⁺ T cell-mediated B cell activation in MIA tumors. In addition, this CD4⁺ T cell-mediated B cell activation seemed to be associated with “immune-high” tumors.

Our observation of tumor-infiltrating B and CD4⁺ T cells suggested lymphocyte aggregation and TLS formation in the MIA TME. We therefore examined additional transcriptome and pathology evidences of TLS formation in MIA tumors. From MIA transcriptome, we observed upregulation of six key TLS indicators of B cell infiltration (CD19 and CD20), follicle phenotype (CXCR5 and CXCL13), and lymphocyte homing (CCR7 and CCL19). The results were corroborated using another 12-chemokine TLS signature (Messina et al., 2012). From pathology, we not only observed lymphocyte aggregation in the H&E slides but also confirmed the co-localization of B cells, Tfh, and FDC in TLS using six-panel mIHC. We observed higher density of CD20⁺ B cells, CD4⁺ T cells, and CD35⁺ FDC, but lower density of CD8⁺ T cells in TLS follicles than in non-follicular areas. This motivated us to investigate whether cells previously reported to suppress CD8⁺ T cell infiltration, such as Tregs, M2 macrophage, neutrophils, monocytes, or fibroblasts (Cassetta and Pollard, 2018), inhibit CD8⁺ T cell infiltration in MIA. Further examination of bulk RNA-seq on our MIA tumors implicated Tregs in decreased CD8⁺ T cell infiltration in MIA TLS. This hypothesis is supported by previous autoimmune studies in human and mouse (Sage and Sharpe, 2015; Sage et al., 2016; Wing et al., 2018; Viguier et al., 2004). Using mIHC to evaluate the ratio between immune suppression (Tfr) and immune stimulation (Tfh), we found a negative association between the Tfr/Tfh ratio and CD8⁺ T cell infiltration, further supporting Tfr-mediated CD8⁺ T cell decrease in MIA tumors. Besides, Li et al. has confirmed the co-culturing of Tfr cells and CD8⁺ T cells *in vitro* resulted in CD8⁺ T cell suppression through TGF- β regulation (Li et al., 2019). Taken together, these results show that our findings are consistent from a biological, computational, and experimental perspective.

TLS is recognized as a heterogeneous collection of immune cells, the composition of which differs based on the cancer-specific immune contextures. Controversy exists regarding the impact of TLS on clinical prognosis. It has also been reported that TLSs are often but not always associated with a positive prognosis

(Binnewies et al., 2018). For example, TLS was reported to be associated with poor prognosis in human hepatocellular carcinoma (Finkin et al., 2015) but good prognosis in colorectal cancer and breast cancer (di Caro et al., 2014; Gu-Trantien et al., 2013). In NSCLC cohorts, TLS in the tumors was observed to be associated with favorable prognosis (Dieu-Nosjean et al., 2008; Remark et al., 2016; Germain et al., 2014), dependent on high CD8⁺ T cell and mature dendritic cell infiltration in the tumors, although we observed a negative association between TLS with CD8⁺ T cell infiltration in MIA tumors. Intriguingly, recent studies also reported TLS to be associated with better response to immune checkpoint inhibitors in melanoma and sarcoma (Helmink et al., 2020; Cabrita et al., 2020; Petitprez et al., 2020). These studies suggest that the discrepancies of TLS on clinical prognosis may be dependent on the cancer type and clinical contexts. Our observation that TLS formation could be associated with Tfr recruitment, CD8⁺ T cell attenuation, and immune suppression have potentially important implications in cancer immunotherapy: while tumors with TLS and a predominant Tfh recruitment might directly respond to anti-PD1 treatment, those with TLS and a predominant Tfr recruitment might benefit from initial anti-CTLA4 treatment to deplete the Tregs before responding to anti-PD1 treatment. Novel therapeutic interventions to shift TLS from Tfr to Tfh recruitment could also hold promise for cancer treatment. Current clinical trials are testing TLS-related cytokine and chemokine administrations and targeting TLS-related pathways as therapeutic approaches in autoimmune disease (Pitzalis et al., 2014). However, it remains uncertain whether the therapeutics associated with B cell infiltration and activation will modulate the function of TLS to change anti-tumor immunity, and if so how long such effects will last.

Limitations of the study

Currently, our analysis only revealed associations without having functional validation of the causal effects, although another study reported similar findings in early lung cancers (Dejima et al., 2020). It is challenging to perform the co-culturing of Tfr cells and CD8⁺ T cells *in vitro* due to the difficulty of extracting enough Tfr cells. It has been reported that the decreased CD8⁺ T cells might be the potential reason of insensitive anti-PD1/PD-L1-based treatment in patients with LUAD with the synchronous ground-glass nodules (Wu et al., 2020a). Additional research and clinical studies are required to fully elucidate the mechanism of CD8⁺ T cell decrease and explore better immunotherapy strategies in cancer.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
 - Lead contact
 - Material availability
 - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
 - Human sample collection
- METHOD DETAILS
 - Bulk RNA sequencing
 - Gene profile processing of bulk RNA-Seq data
 - TCGA data downloading
 - Mutation calling from paired RNA-Seq data
 - Normalization for altered mucin expression
 - Immune infiltration estimation from RNA-seq data
 - The immune signature score for classifying “immune-low” and “immune-high” MIA tumors
 - Definitions of BCR repertoire metrics
 - Tumor purity evaluation
 - Single-cell RNA-seq data analysis
 - Immunohistochemistry
 - Multiplex immunohistochemistry staining for lymphocyte infiltrating in MIA
 - Multispectral image analysis for tumor microenvironment of MIA
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.isci.2022.103883>.

ACKNOWLEDGMENTS

The authors acknowledge Dr. X. Shirley Liu in Dana-Farber Cancer Institute, Dr. Arlene Sharpe in Harvard Medical School, Sherlock Hu in Dana-Farber Cancer Institute, and Stan Gu in Dana-Farber Cancer Institute for the helpful discussion and advice, to Drs. Jia Liang and Rongjian Su in Jinzhou Medical University, Mr. Wei Liu and fellows of the SKLBCE-PerkinElmer Center of Excellence, Beijing, China for their assistance on the miHC experiments. This work was supported by the National Natural Science Foundation of China (Grant No. 82125001, 81872290, 82073154, 81972172, 81772763, and 81772465), the Shanghai Science and Technology Committee (No. 19XD1423200, 18140903900).

AUTHOR CONTRIBUTIONS

WJ carried out the computational analysis and study design, participated in the methodology development, and drafted the manuscript. JDB carried out the experimental analysis and participated in method writing. ZXQ participated in data pre-processing. ZF, LW, and ZT carried out the data sequencing. WD and YHS collected samples and supported materials. HLK, JGN, and FK acquired and managed patients and provided facilities. ZZ, ZJ, SDQ, and LZY participated part of analysis and gave useful suggestions. ZP and YK participated in the design of the study. All authors read and approved the final manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: October 5, 2021

Revised: December 3, 2021

Accepted: February 2, 2022

Published: March 18, 2022

REFERENCES

- Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq—a python framework to work with high-throughput sequencing data. *Bioinformatics* 31, 166–169.
- Bergomas, F., Grizzi, F., Doni, A., Pesce, S., Laghi, L., Allavena, P., Mantovani, A., and Marchesi, F. (2011). Tertiary intratumoral lymphoid tissue in colo-rectal cancer. *Cancers* 4, 1–10.
- Binnewies, M., Roberts, E.W., Kersten, K., Chan, V., Fearon, D.F., Merad, M., Coussens, L.M., Gabrilovich, D.I., Ostrand-Rosenberg, S., Hedrick, C.C., et al. (2018). Understanding the tumor immune microenvironment (TIME) for effective therapy. *Nat. Med.* 24, 541–550.
- Bouillez, A., Adeegbe, D., Jin, C., Hu, X., Tagde, A., Alam, M., Rajabi, H., Wong, K.K., and Kufe, D. (2017). MUC1-C promotes the suppressive immune microenvironment in non-small cell lung cancer. *Oncoimmunology* 6, e1338998.
- Bruno, T.C. (2020). New predictors for immunotherapy responses sharpen our view of the tumour microenvironment. *Nature* 577, 474–476.
- Cabrita, R., Lauss, M., Sanna, A., Donia, M., Skaarup Larsen, M., Mitra, S., Johansson, I., Phung, B., Harbst, K., Vallon-Christersson, J., et al. (2020). Tertiary lymphoid structures improve immunotherapy and survival in melanoma. *Nature* 577, 561–565.
- Cai, H., Chen, M.S., Sun, Z.Y., Zhao, Y.F., Kunz, H., and Li, Y.M. (2013). Self-adsjuvanting synthetic antitumor vaccines from MUC1 glycopeptides conjugated to T-cell epitopes from tetanus toxoid. *Angew. Chem. Int. Ed. Engl.* 52, 6106–6110.
- Cai, H., Degliangeli, F., Palitzsch, B., Gerlitzki, B., Kunz, H., Schmitt, E., Fiamengo, R., and Westerlind, U. (2016). Glycopeptide-functionalized gold nanoparticles for antibody induction against the tumor associated mucin-1 glycoprotein. *Bioorg. Med. Chem.* 24, 1132–1135.
- Cassetta, L., and Pollard, J.W. (2018). Targeting macrophages: therapeutic approaches in cancer. *Nat. Rev. Drug Discov.* 17, 887–904.
- Corsiero, E., Nerviani, A., Bombardieri, M., and Pitzalis, C. (2016). Ectopic lymphoid structures: powerhouse of autoimmunity. *Front. Immunol.* 7, 430.
- Coussens, L.M., and Werb, Z. (2002). Inflammation and cancer. *Nature* 420, 860–867.
- de Chaisemartin, L., Goc, J., Damotte, D., Validire, P., Magdeleinat, P., Alifano, M., Cremer, I., Fridman, W.H., Sautes-Fridman, C., and Dieu-Nosjean, M.C. (2011). Characterization of chemokines and adhesion molecules associated with T cell presence in tertiary lymphoid structures in human lung cancer. *Cancer Res.* 71, 6391–6399.
- Dejima, H., Hu, X., Chen, R., Zhang, J., Junya, F., Parra, E.C., Haymaker, C., Duose, D.Y., Solis, S.L.M., Su, D., et al. (2020). Immune evolution from preneoplasia to invasive lung adenocarcinomas and underlying molecular features. *Nat. Commun.* 12, 2722.
- di Caro, G., Bergomas, F., Grizzi, F., Doni, A., Bianchi, P., Malesci, A., Laghi, L., Allavena, P., Mantovani, A., and Marchesi, F. (2014). Occurrence of tertiary lymphoid tissue is associated with T-cell infiltration and predicts better prognosis in early-stage colorectal cancers. *Clin. Cancer Res.* 20, 2147–2158.
- Dieu-Nosjean, M.C., Antoine, M., Danel, C., Heudes, D., Wislez, M., Poulot, V., Rabbe, N., Laurans, L., Tartour, E., de Chaisemartin, L., et al. (2008). Long-term survival for patients with non-small-cell lung cancer with intratumoral lymphoid structures. *J. Clin. Oncol.* 26, 4410–4417.
- Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15–21.
- Efremova, M., Vento-Tormo, M., Teichmann, S.A., and Vento-Tormo, R. (2020). CellPhoneDB: inferring cell-cell communication from combined expression of multi-subunit ligand-receptor complexes. *Nat. Protoc.* 15, 1484–1506.
- Finkin, S., Yuan, D., Stein, I., Taniguchi, K., Weber, A., Unger, K., Browning, J.L., Goossens, N., Nakagawa, S., Gunasekaran, G., et al. (2015). Ectopic lymphoid structures function as microniches for tumor progenitor cells in hepatocellular carcinoma. *Nat. Immunol.* 16, 1235–1244.
- Fletcher, J.I., Haber, M., Henderson, M.J., and Norris, M.D. (2010). ABC transporters in cancer: more than just drug efflux pumps. *Nat. Rev. Cancer* 10, 147–156.
- Germain, C., Gnjatich, S., Tamzalit, F., Knockaert, S., Remark, R., Goc, J., Lepelley, A., Becht, E., Katsahian, S., Bizouard, G., et al. (2014). Presence

- of B cells in tertiary lymphoid structures is associated with a protective immunity in patients with lung cancer. *Am. J. Respir. Crit. Care Med.* 189, 832–844.
- Ginaldi, L., de Martinis, M., Matutes, E., Farahat, N., Morilla, R., and Catovsky, D. (1998). Levels of expression of CD19 and CD20 in chronic B cell leukaemias. *J. Clin. Pathol.* 51, 364–369.
- Gu, Z., Eils, R., and Schlesner, M. (2016). Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* 32, 2847–2849.
- Gu-Trantien, C., Loi, S., Garaud, S., Equeter, C., Libin, M., de Wind, A., Ravoet, M., le Buanec, H., Sibille, C., Manfouo-Foutsop, G., et al. (2013). CD4(+) follicular helper T cell infiltration predicts breast cancer survival. *J. Clin. Invest.* 123, 2873–2892.
- Harris, R.S. (2013). Cancer mutation signatures, DNA damage mechanisms, and potential clinical implications. *Genome Med.* 5, 87.
- Havenar-Daughton, C., Lindqvist, M., Heit, A., Wu, J.E., Reiss, S.M., Kendrick, K., Belanger, S., Kasturi, S.P., Landais, E., Akondy, R.S., et al. (2016). CXCL13 is a plasma biomarker of germinal center activity. *Proc. Natl. Acad. Sci. U S A* 113, 2702–2707.
- He, D., Wang, D., Lu, P., Yang, N., Xue, Z., Zhu, X., Zhang, P., and Fan, G. (2020). Single-cell RNA sequencing reveals heterogeneous tumor and immune cell populations in early-stage lung adenocarcinomas harboring EGFR mutations. *Oncogene* 40, 355–368.
- Heesters, B.A., van der Poel, C.E., Das, A., and Carroll, M.C. (2016). Antigen presentation to B cells. *Trends Immunol.* 37, 844–854.
- Helmink, B.A., Reddy, S.M., Gao, J., Zhang, S., Basar, R., Thakur, R., Yizhak, K., Sade-Feldman, M., Blando, J., Han, G., et al. (2020). B cells and tertiary lymphoid structures promote immunotherapy response. *Nature* 577, 549–555.
- Hennecke, J., and Wiley, D.C. (2001). T cell receptor-MHC interactions up close. *Cell* 104, 1–4.
- Hosono, M., de Boer, O.J., van der Wal, A.C., van der Loos, C.M., Teeling, P., Piek, J.J., Ueda, M., and Becker, A.E. (2003). Increased expression of T cell activation markers (CD25, CD26, CD40L and CD69) in atherectomy specimens of patients with unstable angina and acute myocardial infarction. *Atherosclerosis* 168, 73–80.
- Hu, X., Zhang, J., Wang, J., Fu, J., Li, T., Zheng, X., Wang, B., Gu, S., Jiang, P., Fan, J., et al. (2019). Landscape of B cell immunity and related immune evasion in human cancers. *Nat. Genet.* 51, 560–567.
- Inamura, K. (2018). Clinicopathological characteristics and mutations driving development of early lung adenocarcinoma: tumor initiation and progression. *Int. J. Mol. Sci.* 19, 1259.
- Ivetic, A., Hoskins Green, H.L., and Hart, S.J. (2019). L-Selectin: a major regulator of leukocyte adhesion, migration and signaling. *Front. Immunol.* 10, 1068.
- Izumchenko, E., Chang, X., Brait, M., Fertig, E., Kagohara, L.T., Bedi, A., Marchionni, L., Agrawal, N., Ravi, R., Jones, S., et al. (2015). Targeted sequencing reveals clonal genetic changes in the progression of early lung neoplasms and paired circulating DNA. *Nat. Commun.* 6, 8258.
- Kanehisa, M., and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 28, 27–30.
- Kobayashi, D., Endo, M., Ochi, H., Hojo, H., Miyasaka, M., and Hayasaka, H. (2017). Regulation of CCR7-dependent cell migration through CCR7 homodimer formation. *Sci. Rep.* 7, 8536.
- Koboldt, D.C., Zhang, Q., Larson, D.E., Shen, D., McEellan, M.D., Lin, L., Miller, C.A., Mardis, E.R., Ding, L., and Wilson, R.K. (2012). VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res.* 22, 568–576.
- Lakshmanan, I., Ponnusamy, M.P., Macha, M.A., Haridas, D., Majhi, P.D., Kaur, S., Jain, M., Batra, S.K., and Ganti, A.K. (2015). Mucins in lung cancer: diagnostic, prognostic, and therapeutic implications. *J. Thorac. Oncol.* 10, 19–27.
- Lambrechts, D., Wauters, E., Boeckx, B., Aibar, S., Nittner, D., Burton, O., Bassez, A., Decaluwe, H., Pircher, A., Van Den Eynde, K., et al. (2018). Phenotype molding of stromal cells in the lung tumor microenvironment. *Nat. Med.* 24, 1277–1289.
- Lavin, Y., Kobayashi, S., Leader, A., Amir, E.D., Elefant, N., Bigenwald, C., Remark, R., Sweeney, R., Becker, C.D., Levine, J.H., et al. (2017). Innate immune landscape in early lung adenocarcinoma by paired single-cell analyses. *Cell* 169, 750–765 e17.
- Legler, D.F., Loetscher, M., Roos, R.S., Clark-Lewis, I., Baggiolini, M., and Moser, B. (1998). B cell-attracting chemokine 1, a human CXCL chemokine expressed in lymphoid tissues, selectively attracts B lymphocytes via BLR1/CXCR5. *J. Exp. Med.* 187, 655–660.
- Li, B., Li, T., Wang, B., Dou, R., Zhang, J., Liu, J.S., and Liu, X.S. (2017). Ultrasensitive detection of TCR hypervariable-region sequences in solid-tissue RNA-seq data. *Nat. Genet.* 49, 482–483.
- Li, B., Severson, E., Pignon, J.C., Zhao, H., Li, T., Novak, J., Jiang, P., Shen, H., Aster, J.C., Rodig, S., et al. (2016). Comprehensive analyses of tumor immunity: implications for cancer immunotherapy. *Genome Biol.* 17, 174.
- Li, L., Ma, Y., and Xu, Y. (2019). Follicular regulatory T cells infiltrated the ovarian carcinoma and resulted in CD8 T cell dysfunction dependent on IL-10 pathway. *Int. Immunopharmacol.* 68, 81–87.
- Li, S., Li, L., Zhu, Y., Huang, C., Qin, Y., Liu, H., Ren-Heidenreich, L., Shi, B., Ren, H., Chu, X., et al. (2014). Coexistence of EGFR with KRAS, or BRAF, or PIK3CA somatic mutations in lung cancer: a comprehensive mutation profiling from 5125 Chinese cohorts. *Br. J. Cancer* 110, 2812–2820.
- Looney, T.J., Lee, J.Y., Roskin, K.M., Hoh, R.A., King, J., Glanville, J., Liu, Y., Pham, T.D., Dekker, C.L., Davis, M.M., and Boyd, S.D. (2016). Human B-cell isotype switching origins of IgE. *J. Allergy Clin. Immunol.* 137, 579–586 e7.
- Maeda, T., Hiraki, M., Jin, C., Rajabi, H., Tagde, A., Alam, M., Bouillez, A., Hu, X., Suzuki, Y., Miyoi, M., et al. (2018). MUC1-C induces PD-L1 and immune evasion in triple-negative breast cancer. *Cancer Res.* 78, 205–215.
- Martinet, L., Garrido, I., Filleron, T., le Guellec, S., Bellard, E., Fournie, J.J., Rochaix, P., and Girard, J.P. (2011). Human solid tumors contain high endothelial venules: association with T- and B-lymphocyte infiltration and favorable prognosis in breast cancer. *Cancer Res.* 71, 5678–5687.
- Mascaux, C., Angelova, M., Vasaturo, A., Beane, J., Hijazi, K., Anthoine, G., Buttard, B., Rothe, F., Willard-Gallo, K., Haller, A., et al. (2019). Immune evasion before tumour invasion in early lung squamous carcinogenesis. *Nature* 571, 570–575.
- Mayakonda, A., Lin, D.C., Assenov, Y., Plass, C., and Koeffler, H.P. (2018). Maftools: efficient and comprehensive analysis of somatic variants in cancer. *Genome Res.* 28, 1747–1756.
- McLaren, W., Gil, L., Hunt, S.E., Riat, H.S., Ritchie, G.R., Thormann, A., Flicek, P., and Cunningham, F. (2016). The ensembl variant effect predictor. *Genome Biol.* 17, 122.
- Messina, J.L., Fenstermacher, D.A., Eschrich, S., Qu, X., Berglund, A.E., Lloyd, M.C., Schell, M.J., Sondak, V.K., Weber, J.S., and Mule, J.J. (2012). 12-Chemokine gene signature identifies lymph node-like structures in melanoma: potential for patient selection for immunotherapy? *Sci. Rep.* 2, 765.
- Newman, A.M., Liu, C.L., Green, M.R., Gentles, A.J., Feng, W., Xu, Y., Hoang, C.D., Diehn, M., and Alizadeh, A.A. (2015). Robust enumeration of cell subsets from tissue expression profiles. *Nat. Methods* 12, 453–457.
- Niu, B., Ye, K., Zhang, Q., Lu, C., Xie, M., McEellan, M.D., Wendt, M.C., and Ding, L. (2014). MSIsensor: microsatellite instability detection using paired tumor-normal sequence data. *Bioinformatics* 30, 1015–1016.
- Paez, J.G., Janne, P.A., Lee, J.C., Tracy, S., Greulich, H., Gabriel, S., Herman, P., Kaye, F.J., Lindeman, N., Boggon, T.J., et al. (2004). EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. *Science* 304, 1497–1500.
- Petitprez, F., de Reynies, A., Keung, E.Z., Chen, T.W., Sun, C.M., Calderaro, J., Jeng, Y.M., Hsiao, L.P., Lacroix, L., Bougouin, A., et al. (2020). B cells are associated with survival and immunotherapy response in sarcoma. *Nature* 577, 556–560.
- Pinsky, P.F. (2014). Assessing the benefits and harms of low-dose computed tomography screening for lung cancer. *Lung Cancer Manag.* 3, 491–498.
- Pitzalis, C., Jones, G.W., Bombardieri, M., and Jones, S.A. (2014). Ectopic lymphoid-like structures in infection, cancer and autoimmunity. *Nat. Rev. Immunol.* 14, 447–462.
- Racle, J., de Jonge, K., Baumgaertner, P., Speiser, D.E., and Gfeller, D. (2017). Simultaneous enumeration of cancer and immune cell types

- from bulk tumor gene expression data. *Elife* 6, e26476.
- Remark, R., Lupo, A., Alifano, M., Biton, J., Ouakrim, H., Stefani, A., Cremer, I., Goc, J., Regnard, J.F., Dieu-Nosjean, M.C., and Damotte, D. (2016). Immune contexture and histological response after neoadjuvant chemotherapy predict clinical outcome of lung cancer patients. *Oncoimmunology* 5, e1255394.
- Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139–140.
- Sage, P.T., Ron-Harel, N., Juneja, V.R., Sen, D.R., Maleri, S., Sunngak, W., Kuchroo, V.K., Haining, W.N., Chevrier, N., Haigis, M., and Sharpe, A.H. (2016). Suppression by TFR cells leads to durable and selective inhibition of B cell effector function. *Nat. Immunol.* 17, 1436–1446.
- Sage, P.T., and Sharpe, A.H. (2015). T follicular regulatory cells in the regulation of B cell responses. *Trends Immunol.* 36, 410–418.
- Sanchez-Vega, F., Mina, M., Armenia, J., Chatila, W.K., Luna, A., La, K.C., Dimitriadou, S., Liu, D.L., Kantheti, H.S., Saghaforia, S., et al. (2018). Oncogenic signaling pathways in the cancer genome atlas. *Cell* 173, 321–337 e10.
- Sautès-Fridman, C., Petitprez, F., Calderaro, J., and Fridman, W.H. (2019). Tertiary lymphoid structures in the era of cancer immunotherapy. *Nat. Rev. Cancer* 19, 307–325.
- Schoenberger, S.P., Toes, R.E., van der Voort, E.I., Offringa, R., and Melief, C.J. (1998). T-cell help for cytotoxic T lymphocytes is mediated by CD40-CD40L interactions. *Nature* 393, 480–483.
- Sturm, G., Finotello, F., Petitprez, F., Zhang, J.D., Baumbach, J., Fridman, W.H., List, M., and Anechik, T. (2019). Comprehensive evaluation of transcriptome-based cell-type quantification methods for immuno-oncology. *Bioinformatics* 35, i436–i445.
- Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., and Mesirov, J.P. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U S A* 102, 15545–15550.
- Sun, D., Wang, J., Han, Y., Dong, X., Ge, J., Zheng, R., Shi, X., Wang, B., Li, Z., Ren, P., et al. (2020). TISCH: a comprehensive web resource enabling interactive single-cell transcriptome visualization of tumor microenvironment. *Nucleic Acids Res.* 49, D1420–D1430.
- Thomas, N., Best, K., Cinelli, M., Reich-Zeliger, S., Gal, H., Shifrut, E., Madi, A., Friedman, N., Shawe-Taylor, J., and Chain, B. (2014). Tracking global changes induced in the CD4 T-cell receptor repertoire by immunization with a complex antigen using short stretches of CDR3 protein sequence. *Bioinformatics* 30, 3181–3188.
- Thorsson, V., Gibbs, D.L., Brown, S.D., Wolf, D., Bortone, D.S., Ou Yang, T.H., Porta-Pardo, E., Gao, G.F., Plaisier, C.L., Eddy, J.A., et al. (2018). The immune landscape of cancer. *Immunity* 48, 812–830.e14.
- Travis, W.D., Brambilla, E., Nicholson, A.G., Yatabe, Y., Austin, J.H.M., Beasley, M.B., Chirieac, L.R., Dacic, S., Duhig, E., Flieder, D.B., et al. (2015). The 2015 world health organization classification of lung tumors: impact of genetic, clinical and radiologic advances since the 2004 classification. *J. Thorac. Oncol.* 10, 1243–1260.
- Tscharke, D.C., Croft, N.P., Doherty, P.C., and la Gruta, N.L. (2015). Sizing up the key determinants of the CD8+ T cell response. *Nat. Rev. Immunol.* 15, 705–716.
- Viguier, M., Lemaître, F., Verola, O., Cho, M.S., Gorocho, G., Dubertret, L., Bachelez, H., Kourilsky, P., and Ferradini, L. (2004). Foxp3 expressing CD4+CD25(high) regulatory T cells are overrepresented in human metastatic melanoma lymph nodes and inhibit the function of infiltrating T cells. *J. Immunol.* 173, 1444–1453.
- Vinayanuwattikun, C., le Calvez-Kelm, F., Abedi-Ardekani, B., Zaridze, D., Mukeria, A., Voegelé, C., Vallee, M., Purnomosari, D., Forey, N., Durand, G., et al. (2016). Elucidating genomic characteristics of lung cancer progression from in situ to invasive adenocarcinoma. *Sci. Rep.* 6, 31628.
- Wang, C., Yin, R., Dai, J., Gu, Y., Cui, S., Ma, H., Zhang, Z., Huang, J., Qin, N., Jiang, T., et al. (2018). Whole-genome sequencing reveals genomic signatures associated with the inflammatory microenvironments in Chinese NSCLC patients. *Nat. Commun.* 9, 2054.
- Wang, C., Sun, D., Huang, X., Wan, C., Li, Z., Han, Y., Qin, Q., Fan, J., Qiu, X., Xie, Y., et al. (2020). MAESTRO: model-based analyses of single-cell transcriptome and regulome. *Genome Biol.* 21, 198.
- Weinstein, A.M., and Storkus, W.J. (2015). Therapeutic lymphoid organogenesis in the tumor microenvironment. *Adv. Cancer Res.* 128, 197–233.
- Wing, J.B., Tegguc, M., and Sakaguchi, S. (2018). Control of germinal center responses by T-follicular regulatory cells. *Front. Immunol.* 9, 1910.
- Wu, F., Li, W., Zhao, W., Zhou, F., Xie, H., Shi, J., Yu, G., Fan, J., Jiang, T., and Zhou, C. (2020a). Synchronous ground-glass nodules showed limited response to anti-PD-1/PD-L1 therapy in patients with advanced lung adenocarcinoma. *Clin. Transl. Med.* 10, e149.
- Wu, T.D., Madireddi, S., de Almeida, P.E., Bancheureau, R., Chen, Y.J., Chitre, A.S., Chiang, E.Y., Iftikhar, H., O’Gorman, W.E., Au-Yeung, A., et al. (2020b). Peripheral T cell expansion predicts tumour infiltration and clinical response. *Nature* 579, 274–278.
- Xu, B., Wang, S., Zhou, M., Huang, Y., Fu, R., Guo, C., Chen, J., Zhao, J., Gaskin, F., Fu, S.M., and Yang, N. (2017). The ratio of circulating follicular T helper cell to follicular T regulatory cell is correlated with disease activity in systemic lupus erythematosus. *Clin. Immunol.* 183, 46–53.
- Zheng, G.X., Terry, J.M., Belgrader, P., Ryvkin, P., Bent, Z.W., Wilson, R., Ziraldo, S.B., Wheeler, T.D., McDermott, G.P., Zhu, J., et al. (2017). Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* 8, 14049.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Rabbit anti-CD4 monoclonal antibody (EP204)	CST	Cat#: 48274S
Rabbit anti-CD8a monoclonal antibody (D8A8Y)	CST	Cat#: 85336S
Rabbit anti-CR1 monoclonal antibody (EP197)	ZSbio	Cat#: ZA-0638
Rabbit anti-CXCR5 monoclonal antibody (D6L3C)	CST	Cat#: 72172S
Mouse anti-CD20 monoclonal antibody (L26)	Abcam	Cat#: ab9475
Rabbit anti-FOXP3 monoclonal antibody (D2W8E)	CST	Cat#: 98377S
Mouse anti-CD3e monoclonal antibody (OTI3E10)	ZSbio	Cat#: TA506064
Rabbit anti-BCL6 monoclonal antibody (1E6B1)	Proteintech	Cat#: 66340-1-Ig
Chemicals, peptides, and recombinant proteins		
EDTA Antigen Retrieval Solution,50x	Solarbio	Cat: C1034
DAPI solution(ready-to-use)	Solarbio	Cat: C0065
Software and algorithms		
STAR	Dobin et al. (2013)	https://code.google.com/archive/p/rna-star/
Samtools	Open resource	http://samtools.sourceforge.net/
Conda	Open resource	https://docs.conda.io/en/latest/
Varscan2	Koboldt et al. (2012)	http://varscan.sourceforge.net/
VEP	McLaren et al. (2016)	https://github.com/Ensembl/ensembl-vep
Maftools	Bioconductor	https://github.com/PoisonAlien/maftools
HTSeq	Anders et al. (2015)	http://www-huber.embl.de/HTSeq
edgeR	Robinson et al. (2010)	http://bioconductor.org
GSEA	Subramanian et al. (2005)	http://bioconductor.org
TIMER	Li et al. (2017)	http://cistrome.org/TIMER/
CIBERSORT	Newman et al. (2015)	https://cibersort.stanford.edu/
EPIC	Racle et al. (2017)	http://epic.gfellerlab.org
Xcell	xCell Website	https://xcell.ucsf.edu/
CRI-iAtlas	Open resource	https://www.cri-iatlas.org/
TRUST3	Hu et al. (2019)	https://bitbucket.org/liulab/trust/
CellRanger	10x Genomics	https://bioinformatics.uconn.edu/single-cell-rna-sequencing-cell-ranger-2/
MAESTRO	Wang et al. (2020)	http://github.com/liulab-dfci/MAESTRO
Seurat	Open resource	https://satijalab.org/seurat/
TISCH	Sun et al. (2020)	http://tisch.comp-genomics.org/
CellPhoneDB	Efremova et al. (2020)	https://github.com/Teichlab/cellphonedb
R software	Open resource	http://www.r-project.org/
RStudio	Open resource	https://www.rstudio.com/
inForm 2.4.1 Automated Image Analysis Software	PerkinElmer	NA
Other		
OPAL 7-COLOR MANUAL IHC KIT 50 SLIDES [kit]	PerkinElmer	Cat#: NEL811001KT
MIA RNAseq samples	This paper	PRJCA003185
Independent single-cell RNAseq samples	This paper	GSE139555, E-MTAB-6149, PMID 33144684

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Peng Zhang (pengzhang1121@tongji.edu.cn).

Material availability

This study did not generate new unique reagents.

Data and code availability

MIA RNA-seq data has been deposited at Genome Sequence Archive for Human (GSA-Human) of National Genomics Data Center (NGDC: PRJCA003185, <https://ngdc.cncb.ac.cn/gsa-human/>) and are publicly available as of the date of publication. Accession numbers are listed in the [key resource table](#). All the other data could be found at the [key resource table](#). Code is available in the github repository (<https://github.com/jwang0611/MIA-RNAseq>). Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Human sample collection

All 116 samples were collected from 58 MIA patients at Pulmonary Hospital, Shanghai, China. All specimens were obtained from patients with appropriate consent from the relevant institutional review board. The clinical information of these patients was shown in [Table S1](#). These MIA samples were all obtained through surgical resection and pathologically diagnosed as stage IA in 2016.

METHOD DETAILS

Bulk RNA sequencing

The RNA extraction procedure was compiled with standard protocols. DNA and RNA were collected from samples using AllPrep DNA/RNA Mini Kit (Qiagen, 80204, Germany). The cDNAs were synthesized and used to construct RNA libraries using NEBNext Ultra RNA Library Prep Kit (NEB, E7530, USA) and NEBNext Multiplex Oligos for Illumina (NEB, E7335, USA). The libraries were sequenced on the Illumina HiSeq2500 platform via a 2x150bp paired-end sequencing at Annoroad Gene Technology (Beijing, China). For each sample, more than 100M fragments were sequenced.

Gene profile processing of bulk RNA-Seq data

The transcriptome data were processed by the standard RNA-Seq pipeline. We performed sequence mapping by STAR alignment ([Dobin et al., 2013](#)), reads quantifying by HT-Seq Count ([Anders et al., 2015](#)), and gene expression measurement by TPM converting from read count. Differential expression analysis on paired samples was performed using edgeR ([Robinson et al., 2010](#)). The reference fasta file and annotation gtf file was downloaded from Ensembl website on version 79 of hg38. Gene set enrichment analysis was performed using GSEA ([Subramanian et al., 2005](#)). Pathway categories are referred from KEGG website (<https://www.genome.jp/kegg/pathway.html>)

TCGA data downloading

For TCGA LUAD dataset, we downloaded gene expression data and mutation data from firebrowse (<http://firebrowse.org>).

Mutation calling from paired RNA-Seq data

Mutation detection from paired RNA-Seq data was analyzed by Varscan2. Somatic variants from Varscan2 ([Koboldt et al., 2012](#)) were annotated by Variant Effect Predictor (VEP) ([McLaren et al., 2016](#)) and vcf2maf. Variants with low quality were removed based on the criteria which GDC pipeline recommends. Mutation visualization was implemented by Maftools ([Mayakonda et al., 2018](#)) and Complexheatmap ([Gu et al., 2016](#)).

Normalization for altered mucin expression

Z-score normalization of mRNA expression for patients was performed. One over-expressed or down-expressed gene was expected to have a difference of at least two standard deviations from the mean of expression in the reference samples.

Immune infiltration estimation from RNA-seq data

To ensure robust abundance estimation, we applied ImmuneDeconv of TIMER2.0 to conduct deconvolution (9) of immune cell components from TPM expression data and only focused on the observations consistently predicted by different algorithms. Each algorithm estimates one general cell type or specific cell sub-types (Table S3). We obtained consistent trends for the comparison of major immune cell types (B, CD4⁺ T, and CD8⁺ T) between tumor and normal tissues from the different algorithms and visualized the estimation predicted by the common methods, TIMER and CIBERSORT.

The immune signature score for classifying “immune-low” and “immune-high” MIA tumors

From the pan-cancer paper (Thorsson et al., 2018), the immune signature data and the pre-trained scoring model were downloaded from the CRI-iAtlas (<https://github.com/CRI-iAtlas/shiny-iatlas>). Based on the model, we computed the five representative immune signatures score from the expression data, which was normalized by paired normal tissues and median-centered by gene level. Then, we ranked MIA tumors from immune “low” to immune “high” based on the averaged immune signature score.

Definitions of BCR repertoire metrics

We applied TRUST3 to infer the immune repertoire from bulk RNA-seq data (Li et al., 2017; Hu et al., 2019). We measured basic metrics to quantify immune repertoire and further performed paired comparison of TCR and BCR repertoire between tumor and tumor-adjacent tissues. The fraction of TCR or BCR reads was defined as the library size located in BCR loci, which is normalized by total mapped reads (E1). The number of unique TCR or BCR CDR3 sequences summarized the total unique CDR3 size. The number of TCR or BCR clonotypes per kilo reads measures TCR or BCR diversity. In addition, B cells undergo clonal expansions with a high rate of somatic hypermutation by single base substitutions in CDR3 sequences. Thus, we built BCR lineages referred to the method of a previous study (Hu et al., 2019), which generates 8-mer CDR3 motifs and an upper triangular matrix of counting the substitutions between DNA sequences from any two clonotypes in one BCR lineage to cluster partial and complete CDR3 sequences. For each sample, BCR or TCR clonality is defined by Shanno entropy (E2). Somatic hypermutation rate was defined as the total count of harboring one base substitution normalized by the total length of aligned DNA sequence (E3). Ig class switch recombination (CSR) were built for the BCR clonotypes with different Ig isotypes from the same BCR lineage. Any two Ig isotypes within the same cluster motif could be switched based on the specific order, which is IgM, IgD, IgG3, IgG1, IgA1, IgG2, IgG4, IgE, IgA2 (Looney et al., 2016). Here we only remain the Ig isotypes with abundance more than 1% and check CSR on IgG and IgA for measuring B-cell activation signal. The ratio of CSR was computed referring to the method in the previous study (Hu et al., 2019). The are shown below.

$$\text{Fraction of TCR or BCR reads} = \frac{I}{M} \quad (\text{Equation 1})$$

I denotes library size located in BCR loci; M denotes total mapped reads.

$$\text{Clonality} = 1 - \frac{-\sum_{i=1}^n p_i \log_2(p_i)}{\log_2(n)} \quad (\text{Equation 2})$$

$$p_i = \frac{f_i}{\sum_{i=1}^n f_i}$$

n denotes total number of TCR clonotypes or total number of BCR clusters and non-clustered clonotypes, f_i denotes clonal frequency for non-clustered clonotypes or clustered clonotypes.

$$\text{SHM rate} = \frac{\sum_{i=1}^c S_i}{\sum_{i=1}^c \max(L_i)} \quad (\text{Equation 3})$$

c denotes the number of BCR clusters, S_i denotes the sum of clonotype pairs harboring one base substitution, L_i denotes the max number of aligned non-gapped DNA sequence in one cluster.

Tumor purity evaluation

Tumor purity was estimated by the algorithm of EPIC (Racle et al., 2017) (<http://epic.gfellerlab.org>), which is designed for estimating the absolute fraction of immune cells and cancer cells simultaneously from RNA-seq expression data. We selected “tumor infiltrating cells” as reference profile to predict cell fractions.

Single-cell RNA-seq data analysis

Three independent cohorts with single-cell RNAseq of tumors and adjacent normal tissues were analyzed to validate our conclusions, including two independent single-cell RNA-seq datasets collected from public NSCLC cohorts (GSE139555, E-MTAB-6149) and another cohort from the same hospital as MIA samples (He et al., 2020).

For the two public cohorts, Single-cell RNA-seq dataset pre-processing was performed by the comprehensive pipeline MAESTRO (Wang et al., 2020). Single cell clustering and cell type annotation are available to be visualized in TISCH (Sun et al., 2020). Cluster averaged or cell-level expression data of tumor and normal tissues were available to perform analysis. Cell-cell interaction analysis was inferred utilizing a public method CellPhoneDB (Efremova et al., 2020). Only receptor-ligand pairs significantly enriched in less than 5% of all cell-cell interactions were exhibited for the cell interactions on B cells or T cells in the dot plot.

For the single-cell RNAseq cohort with five pairs of early-stage LUAD tumors with matched normal from the same hospital, raw data was processed by CellRanger (Zheng et al., 2017) then cells with more than 1000 UMI counts, 500 genes covered, and less than 5% mitochondrial genes were retained for analysis (102,221 high-quality cells and totally 22643 genes covered). The Dimension reduction step was done with top principle components for all merged samples with the largest variance (at resolution=1.0) based on the parameter criteria in TISCH pre-processing module. Next, curated cell-type annotation based on the classical cell type markers in this study (He et al., 2020) was performed. To confirm the lymphocyte infiltration difference in tumor and normal samples, only 18691 cells annotated as T and B lymphocytes were extracted for downstream analysis. To decipher CD4⁺ T cell sub-types, we annotated Tfh cells using ICOS, BCL6, IL7R, CXCR5, and CXCL13. Similarly, Tfr cells were annotated using PDCD1, FOXP3, IL2RA.

Immunohistochemistry

Paraffin sections at 4- μ m thickness were prepared from MIA biopsies (n = 58) and paired normal lung tissue. Antigen retrieval were performed and the interested markers were stained in certain condition (Table S5) with Dako Real kit according to IHC-P protocol. Tonsil tissue were set as control for staining optimization.

Multiplex immunohistochemistry staining for lymphocyte infiltrating in MIA

Twenty-two of the 58 samples were stained according to Opal 7-plex technology (PerkinElmer), yielding to the simultaneous visualization of seven markers (containing DAPI for nucleus) on the same slide. During each cycles of staining, microwave antigen retrieval (AR) was conducted in AR solution pH 6 or pH 9 (AR6 or AR9) suggested by IHC validation; blocking for 15 mins at room temperature (RT) was followed by primary antibodies incubated for 1 hr at RT or overnight at 4°C. Next, HRP labeled polymer goat anti-mouse and rabbit antibodies were incubated at RT for 10 min followed by TSA opal fluorophores (Opal 520, Opal 540, Opal 570, Opal 620, Opal 650 or Opal 690) 10 min incubation. Microwave AR was performed to remove the antibody TSA complex at each cycle. Finally, all slides were counterstained with DAPI for 5 min and enclosed in ProLong Antifade Mountant (Solarbio). The slides were scanned, and the multi-spectral images obtained were unmixed using spectral libraries that were previously built from images stained for each fluorophore (mono-plex).

Multispectral image analysis for tumor microenvironment of MIA

The entire area of the sections was stained, scanned, scored, and quantified using the PerkinElmer Mantra System and inForm Advanced Image Analysis software (inForm 2.4.1 PerkinElmer). The multispectral images were unmixed followed with several machine learning automated steps: tissue categorization, cell segmentation, and cell phenotyping. Microinvasive epithelium without infiltration was tagged as "ACC" for cancer cell of adenocarcinoma, while lymphocyte infiltrating stroma area was further recognized as "Non-Fol", "TLS-BZ" and "TLS-TZ", representing scattered lymphocytes infiltration, B-cell zone of TLS formation, and T-cell zone of TLS formation. The classification was verified and approved by a certified pathologist (W.D.). The algorithm enabled a batch analysis of all the slides. We designed a six-plex TLS panel for TLS component cells including CD4, CD8, CD20, CD35, DAPI, and CXCR5, and a seven-plex Tfr/h panel for suppressive/stimulatory immune cells including CD3, CD4, CD8, FOXP3, BCL-6, DAPI, and CXCR5. For six-plex mIHC panel, CD20⁺ B density is defined as the percentage of CD20⁺ cells in total immune cells, including CD8⁺ cells, CD20⁺ cells, CD35⁺ cells, and CD4⁺ cells. Similarly, CD4⁺ T density is defined as the percentage of CD4⁺ cells in total immune cells. CD35⁺ density is

defined as the percentage of CD35⁺ cells in total immune cells and CD8⁺ T density is defined as the percentage of CD8⁺ cells in total immune cells. For seven-plex mIHC panel, we defined the regulatory T-cell phenotype with CD3⁺CD4⁺CXCR5⁺FOXP3⁺Tfr and CD3⁺CD4⁺CXCR5⁺FOXP3⁺Treg and stimulatory T-cell phenotype with CD3⁺CD4⁺CXCR5⁺FOXP3⁺Tfh and CD3⁺CD4⁺CXCR5⁺FOXP3⁺Th. Either cell densities or absolute counts were enrolled for calculation.

QUANTIFICATION AND STATISTICAL ANALYSIS

For all the statistical test for comparing the difference between tumor and match adjacent tissues, Mann-Whitney U test was used to evaluate the P values and $P < 0.05$ was considered statistically significant. Multiple hypothesis correction was computed in the form of adjusted p-value through the Benjamini-Hochberg procedure. Spearman's rank correlation coefficient was presented as the association of two independent features. A partial correlation test was used to measure the correlation between genes and immune cell infiltration when corrected by tumor purity. All the statistical tests were implemented using the R programming language. All of the statistical details of experiments can be found in the [results](#) and figures.