Tertiary Lymphoid Structure and CD8 T Cell Exclusion in Minimally Invasive Adenocarcinoma

Jin Wang[1], Dongbo Jiang[2], 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], Fan Zhang[4], Kun Yang[2], Peng Zhang[8]

Affiliation

1. Clinical Translational Research Center, Shanghai Pulmonary Hospital, School of Life Sciences and Technology, Tongji University, Shanghai, China.
2. Department of Immunology, School of Basic Medicine, Air-Force Medical University. Xi'an, China.
3. Department of Mathematics, Shanghai Normal University, Shanghai, China.
4. Department of Thoracic Surgery, Shanghai Pulmonary Hospital, Tongji University School of Life Sciences and Technology, Shanghai, China.
5. Tissue Bank, Department of Pathology, Experimental Animal Center, Shanghai Pulmonary Hospital, Tongji University School of Medicine, Shanghai, China.
6. Academy of Medical Engineering and Translational Medicine, Tianjin University, China
7. Department of Gynecologic Oncology, Chinese PLA General Hospital, Beijing, China
8. Department of Thoracic Surgery, Shanghai Pulmonary Hospital, Tongji University School of Medicine, Shanghai, China.

Abstract

Lung adenocarcinoma is the leading cause of cancer death. To characterize the tumor microenvironment (TME) of early-stage lung adenocarcinoma, we performed RNA-seq profiling on 59 pairs of minimally invasive adenocarcinoma (MIA) tumors and matched adjacent normal lung tissues from Chinese patients. We observed mucin over-expression and glycosylation, and altered cytokine-cytokine interactions in MIA tumors, which also had distinct adaptive immune TME of higher CD4+ T cell infiltration, higher plasma B cell activation, and lower CD8+ T cell infiltration. The high expression of markers for B cells, activated CD4 T cells, and follicular helper T (Tfh) cells in MIA implicated the formation of tertiary lymphoid structures (TLS), which were supported by two independent single-cell RNA-seq data. Multiplex immunohistochemistry (mIHC) staining of 22 MIA tumors 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.

Significance: We discovered a potential immunosuppressive phenotype in MIA mediated by follicular regulatory T cells (Tfr) in TLS formation, which could be a potential mechanism of CD8+ T cell exclusion. Our study demonstrated how integrating tumor transcriptome and pathology can characterize the 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 computed tomography (LDCT) scan has 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 (1). 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, National Cancer Institute of the United States has 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 (2). 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 (2-5). For example, one study examined the mutation patterns of in situ and invasive adenocarcinoma by whole-exome sequencing and compared the results with TCGA (2). Another study used targeted sequencing and droplet PCR to examine the mutation patterns in both early tumors and the matched blood samples (3). Most of these studies focused on identifying driver mutations potentially responsible for tumor initiation, which have been summarized in a recent review (6).

Besides tumor mutations, there has been growing appreciations of the important roles tumor microenvironment (TME) plays in cancer development, progression, and outcome. The TME consists of cancer cells, tumor-infiltrating lymphocytes, stromal and vascular cells, along with extra-cellular molecules such as cytokines and chemokines that modulate their interactions. Changes in cellular and extra-cellular components of the TME can facilitate tumor behaviors such as unchecked tumor growth and immune evasion (7). To examine such TME changes, gene expression profiles from RNA-seq are more informative than DNA-based tumor profiling.

Besides characterizing differentially expressed genes and pathways, RNA-seq has reasonable power to detect somatic mutations (8), gene fusions (9) and microsatellite instability (10) , especially if matched normal samples are also available. There have also been multiple studies that use tumor RNA-seq data to infer immune cell populations and subtypes, and to associate immune infiltration with clinical outcomes or driver mutations (11,12). However, few such tumor profiling studies include matched adjacent tissues, and often the normal samples are sequenced at much lower coverage. One NSCLC study, which combined Whole Exome Sequencing (WES) and RNA-seq to discover the association between EGFR mutation and tumor-infiltrating B lymphocytes, did not profile RNA-seq from the matched normal, thus limiting the ability to characterize tumor-normal differences (13). Therefore, knowledge of the tumor immune microenvironment in early lung cancer patients, 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 (14,15). How cytotoxic T lymphocytes (CTL) influence immune response to tumor progression has been under intensive investigation in recent years (16). However, the role of B-T cell interactions in anti-tumor immune response has not been completely understood. Inflammation is known to be a potential cause and hallmark of early cancer development (17), and clusters of B and T lymphocytes can form tertiary lymphoid structure (TLS) at the site of inflammation(18). There has been increasing research interest in understanding TLS formation and function in human pathology, especially related to clinical implications and potential therapeutic targeting (19). A recent review summarized studies on the presence, composition, and functions of TLS in cancers and how TLS is correlated with therapeutic response (20). Three recent studies also reported TLS formation to be associated with response to immune checkpoint blockade (21-24). 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 59 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 hematoxylin and eosin (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.

125 # Results

126 **Genes in cytokine-cytokine receptor interactions and mucin glycosylation are over**
127 **expressed in MIA**

128 We collected 59 pairs of pathology-confirmed MIA tumors and matched adjacent normal
129 tissues **(Fig. S1, Table S1)**. All the MIA tumors were confirmed by pathology using the 2015
130 World Health Organization classification criteria of lung tumors (25). Hematoxylin and eosin
131 (H&E) staining confirmed the tumor lepidic growth pattern and the less than 5 mm in diameter
132 of invasive components **(Figure 1A)**. We conduced RNA-seq of all the samples at a minimum
133 depth of 100 million paired-end 150bp fragments, roughly three times the sequencing depth
134 of TCGA samples. We mapped the RNA-seq reads with STAR (26), quantified gene
135 expression with HTSeq (27), and called differentially expressed genes between paired tumor
136 and normal samples using edgeR (28). In total, we identified 1,471 up-regulated and 431
137 down-regulated genes with over 2-fold change and a false discovery rate (FDR) of less than
138 0.05 **(Table S2)**. We then applied gene set enrichment analysis (GSEA) (29) based on Kyoto
139 Encyclopedia of Genes and Genomes (KEGG) (30) to examine altered pathways in MIA
140 tumors **(Figure 1B)**. The most enriched up-regulated pathway in MIA is ABC-transporters
141 which has an established tumor-promoting role in regulating cell survival, proliferation,
142 migration, invasion, angiogenesis, and inflammation (31). In addition, widely studied cancer-
143 associated pathways, such as Ras signaling and PI3K-AKT signaling (32), are also
144 upregulated. Furthermore, MIA tumors exhibit stronger cytokine::cytokine-receptor
145 interactions than adjacent tissues. One particularly overexpressed cytokine, *CXCL13*,
146 encodes a chemokine that has been reported as a major B lymphocyte chemoattractant (33)
147 **(Fig. S2)**. Another pathway showing significant enrichment in MIA tumors is mucin-type O-
148 glycan biosynthesis. Many genes in the mucin family are over-expressed in MIA, especially
149 *MUC21*, *MUC5B*, and *MUC4* **(Figure 1C)**, a phenomenon also observed in TCGA lung
150 adenocarcinoma (LUAD) **(Fig. S3)**. Over-expression of mucin family proteins have been
151 reported to contribute to the carbohydrate-enriched coating of cancer cells, which could
152 promote integrin-mediated tumor growth and survival (34). In addition, overexpression and
153 glycosylation of mucins, especially of MUC1, have been reported in multiple cancers and
154 implicated to promote tumor metastasis and drug resistance (35,36). The overexpression of
155 mucin family members likely accounts for the enrichment of mucin-type O-glycan pathway in
156 MIA tumors.

157

158 **MIA tumors harbor EGFR and Mucin gene mutations**

159 We next sought to detect mutations in MIA tumors by comparing RNA-seq data between tumor
160 and adjacent normal tissue using Varscan2 (37). Although this is less sensitive than mutation

161 calling from DNA-based approaches such as whole genome or whole exome sequencing, we
162 still managed to find approximately 70 (median) variants per tumor, most of which were
163 missense mutations **(Fig. S4A)**. The most enriched type of single nucleotide variants (SNV)
164 is C>T transition **(Fig. S4B)**, which is known to be a signature of mismatch repair deficiency
165 resulting in elevated microsatellite instability (MSI) (38). To computationally assess the MSI
166 level of these MIA samples, we applied MSIsensor (10) to compare the RNA-seq data between
167 tumor and normal. Although this is also less sensitive than DNA-based approaches, we indeed
168 observed higher MSI score in MIA tumors with higher CT transition rate (**Fig. S4C**).
169
170 *EGFR* was found to be mutated in 31% of the MIA samples **(Figure 1D)**, consistent with the
171 previous report of the prevalent *EGFR* mutations in progressive stages of lung
172 adenocarcinoma (13). *EGFR* mutations occur mainly at the mutational hotspot "L858R/M" (39)
173 **(Figure 1E)** in the kinase domain, in agreement observations in TCGA LUAD **(Fig. S4D)**.
174 *EGFR* is known to participate in the signaling of the RTK-RAS pathway (32). The gain-of-
175 function EGFR mutations primarily contributed to the oncogenic activation of the RTK-RAS
176 pathway. Another important gene in the RTK-RAS pathway, *KRAS*, is mutated in only two
177 tumors (**Fig. S4E**), and as expected, occurs mutually exclusively with *EGFR* mutations (40).
178
179 In addition to RTK-RAS pathway mutations, *MUC6* and *MUC4* genes were also highly mutated
180 in MIA tumors. These two genes, as well as other MUC family members *MUC16*, *MUC5B* and
181 *MUC17*, are frequently mutated in TCGA LUAD **(Fig. S3)**. Most of the MUC family SNVs are
182 missense mutations, and together with their frequent over expression, suggest that tumor-
183 associated mucins might create neoantigens in MIA tumors. We therefore applied OptiType
184 (41) to infer the patient MHC Class I type and pVAC-Seq (42) to predict the likely neo-antigens
185 in these MIA tumors. Indeed, our analysis supports mutated MUC6, MUC4, MUC20 as being
186 potential neoantigens in MIA tumors (Fig. S5).
187
188 **Higher B cell and CD4+ T cell infiltration and activation but lower CD8+ T cell infiltration**
189 **in MIA tumors**
190 To decipher immune cell infiltration in the MIA TME, we estimated the immune cell composition
191 from bulk tumor RNA-seq data using TIMER (43), CIBERSORT (11). Relative to adjacent
192 normal tissues, MIA samples have significantly higher B cells and CD4+ T cells, and
193 significantly lower CD8+ T cells and macrophages **(Figure 2A and Fig. S6)**. Consistent with
194 the observation in MIA, TCGA LUAD tumors also showed higher B-cell and lower CD8+ T cell
195 infiltration in tumors than the adjacent normal tissues. Interestingly, B cell infiltration gradually
196 decreases towards normal during tumor progression from Stage I to Stage IV **(Figure 2B)** in
197 TCGA LUAD, suggesting that increased B cell increase is associated with early LUAD onset.

198 Next we characterized the immune repertoires in these samples to investigate the activation
199 and clonal expansion of tumor-infiltrating B and T cells. We ran TRUST, a computational
200 algorithm we developed (44,45), to assemble the T-cell and B-cell receptor (TCR and BCR)
201 repertoires from RNA-seq data. Significantly higher fractions of raw reads were mapped to
202 TCR and BCR in MIA tumors than normal tissues (**Fig S7A-B**). Consistent with this
203 observation, TRUST assembled significantly higher number TCR and BCR unique clonotypes
204 of Complementary Determining Region 3 (CDR3) sequences in MIA tumors (**Figure 2C-D**).
205 Since TRUST has better power to call the abundant and clonal-expanded TCRs and BCRs,
206 the fact that TRUST assembled more TCR/BCR clonotypes already suggest that T cells and
207 B cells have clonal expansion in the MIA tumors. Based on the CD4+ and CD8+ T cell
208 infiltration trend in MIA vs adjacent normal, we expect most of the T cell activation and clonal
209 expansion to occur in CD4+ T cells. Indeed, MIA tumors have significantly higher expression
210 of CD4+ T cell activation markers CD69, CD40LG, CD25 (IL2RA), and CD62 (SELL) (46-
211 49)(**Figure 2E**). However, since TCR diversity and clonality calculation could be confounded
212 since we couldn't distinguish TCRs from CD4+ vs CD8+ T cells, we turned to BCR to further
213 evaluate immune cell clonal expansion. For B cells, we not only observed higher total number
214 of BCR CDR3 clonotypes (**Figure 2D**) and BCR clusters (**Figure 2F**), but also lower BCR
215 diversity (**Figure 2G**) and higher BCR clonality (**Figure 2H**) in MIA than adjacent normal, which
216 strongly support B cell clonal expansion. In addition, B cells, upon antigen recognition and
217 clonal expansion, undergo somatic hype-mutation (SHM) to evolve BCRs with better antigen
218 binding affinity, and class-switch recombination (CSR) to transition to antibody-producing
219 plasma cells. Indeed, we also observed higher BCR SHM rate (**Figure 2I**), higher CSR (**Fig**
220 **S7C**) in MIA than adjacent normal, further supporting higher activated and clonal expanded B
221 cells.
222
223 To check whether the activated tumor-infiltrating B and T cells could generate more diverse
224 clonotypes for binding to different antigens in the TME, we compared the clonotypes detected
225 between tumors and adjacent normal tissues. Majorities of the BCR (58.6%, **Figure 3A**) or
226 TCR (57.8%) clonotypes were uniquely detected in tumors (**Figure 3B**), suggesting more B
227 cell and T cell infiltration and clonal expansion in MIA tumors. We measured the repertoire
228 similarity between each pair of MIA and adjacent tissue using Jaccard index (50,51), and found
229 within-individual repertoire similarity to be much higher than the across-patient repertoire
230 similarity (**Figure 3C**). Interestingly, the TCR and BCR similarities between tumor and normal
231 (**Figure 3C**) are correlated, suggesting that T cells and B cells have coordinated clonal
232 expansion in the MIA tumors. In addition, repertoire similarities among normal samples (N-N)
233 are higher than that among MIA samples (T-T) from differential patients (**Figure 3D**),
234 suggesting that tumor-specific antigens stimulate the expansion of different TCR and BCR

235 clones in different tumors. Furthermore, tumors with higher mutation burden, as inferred from
236 Varscan2, have significantly lower BCR (p-value < 0.01) and TCR (p-value < 0.001) repertoire
237 similarity between tumor and adjacent tissues (**Fig. 3E**). Collectively, these observations
238 suggest that TRUST-inferred immune receptor repertoires represent activated and clonal
239 expanded B cells and potentially CD4+ T cells for recognizing and binding to tumor antigens
240 in MIA.
241
242 **Mucin mutation and overexpression is associated with CD4+ T cell-mediated B cell**
243 **activation in MIA tumors**
244 Since B cells could be activated by direct antigen stimulation and interaction with CD4+ T cells,
245 we further checked how B cell activation was induced in MIA. It has been reported that MUC1
246 glycosylated peptides can induce MHCII-restricted B-T cell interactions (52). In addition, a
247 synthetic peptide from the MUC1 tandem repeat region was shown to be presented by MHCII
248 molecules which resulted in the initiation of T helper cell responses (53). In our MIA samples,
249 we observed a positive association between mucin level and MHCII level (**Figure 4A**), raising
250 the possibility that higher mucin expression in cancer cells might be associated with higher
251 MHCII presentation. In addition, CD40 and its ligand CD40L are well-known to mediate the
252 interaction between B cells and CD4+ T cells (54). We observed significant positive correlation
253 between CD40 and CD40L in MIA tumors **(Figure 4B)**, which suggests CD4+ T cell-mediated
254 B cell activation. The average level of CD40 and CD40L is associated with mucin expression
255 (**Figure 4C**). Furthermore, we observed MIA tumors with over-expressed mucins have higher
256 levels of IgG and IgA isotypes, suggesting higher level of B cell activation and transition to
257 plasma cells (**Figure 4D**). Taken together, these results implicate the potential interactions
258 between the mutation or overexpression on mucins, MHCII presentation to CD4+ T cells, and
259 B cell activation and plasma cell transition in MIA tumors.
260
261 To further investigate B cell activation among heterogeneous tumors, we characterized the
262 overall immune status among the MIA samples using five signatures derived from the TCGA
263 pan-cancer study (12). These signatures include “IFN-g” (representing IFN-g response),
264 “Lymphocyte” (representing overall lymphocyte infiltration by B and T cells), “Macrophage”
265 (representing the activation of macrophages/monocytes), “TGF-β” (representing TGF-β
266 response), and “Wound healing” (representing core serum response). These five signatures
267 seem to have correlated signals in the MIA tumors, so we used the average of the five to rank
268 the MIA tumors from “immune-low” to “immune-high” **(Figure 4E, Methods)**. The “immune-
269 high” tumors are associated with increased expression of MHCII (**Fig. S8A**) and mucins (**Fig.
270 S8B**). In addition, the “immune-high” tumors are also associated with increased expression of
271 CD40 (**Fig. S8C**) and CD40L (**Fig. S8D**), suggesting stronger interactions between B cells

272 and CD4+ T cells. Furthermore, “immune-high” tumors are associated with higher B cell
273 signals, such as IgG and IgA expression (**Fig. S8E-F**) and unique BCR clonotypes (**Fig. S8H**).
274 Collectively, our observations suggest that mucin expression is associated with CD4+ T cell-
275 mediated B cell activation in MIA adaptive immunity.
276
277 **Formation of tertiary lymphoid structure (TLS) in MIA tumors**
278 Previous literature (18) reported that B-cell and CD4+ T-cell aggregation could lead to the
279 formation of tertiary lymphoid structures (TLS). TLS starts from scattered B and T cell
280 aggregation, progresses to follicles with separate T-Zones and B-Zones, and finally matures
281 into ectopic germinal centers (EGC) with follicular dendritic cells (FDC) in the light-Zone for
282 better B cell activation (20) (**Figure 5A**). In the EGC of mature TLS, B cells activates somatic
283 hypermutations, isotype class-switches, and antibody production(19). So far our analysis
284 revealed higher B-cell and CD4+ T-cell infiltration, lower CD8+ T-cell infiltration **(Figure 2A)**,
285 and higher CD4+ T cell-mediated B cell activation **(Figure 3, Figure 4**) in MIA tumors. This
286 motivated us to examine the expression of typical TLS markers for evidence of TLS formation
287 in MIA. CD19 and MS4A1 (CD20), which are typical B-cell markers (55,56), were highly
288 expressed in MIA tumors. The ligand-receptor pair CXCL13 and CXCR5, which are
289 conventional indicators of follicles (57), were also significantly higher expressed in MIA tumors.
290 Furthermore, the ligand-receptor pair CCL19 and CCR7, which are important for promoting T
291 cell and B cell homing to lymphoid structures (58), were also highly expressed in tumors
292 **(Figure 5B)**. These six TLS markers were all positively correlated with TIMER-estimated B
293 cell and CD4+ T cell abundance but negatively correlated with CD8+ T cell abundance **(Figure**
294 **5B)**.
295
296 Besides the six TLS markers we evaluated above, another 12-chemokine gene signature has
297 also been proposed to predict the presence of TLS (59,60). We observed highly correlated
298 expression signals between our six TLS markers and the 12-chemokine signatures **(Fig. S9A)**.
299 As a result, the two approaches give concordant TLS predictions in MIA tumors, and found
300 TLS signals to be associated with immune-high tumors with stronger CD4+ T cell mediated B
301 cell activation **(Fig. S9B-C)**. To further confirm the presence of aggregated B cells and CD4+
302 T cells in non-small-cell lung cancer (NSCLC), we also examined two recently published
303 NSCLC single-cell RNA-seq datasets with tumors and adjacent normal tissues (61,62). Indeed,
304 NSCLC tumors have higher level of MS4A1+ B cells, CD4+CXCR5+ follicular helper T cells
305 (Tfh) than adjacent normal tissues (**Fig. S10A-B, Fig. S11A-B**). In addition, we used a
306 computational algorithm CellPhoneDB (63) infer cell–cell communication from combined
307 expression of multi-subunit ligand–receptor complexes to confirm the interaction between

308 CD40LG in CD4+ T cells with CD40 in B cells (**Fig. S11C**). Collectively, our MIA RNA-seq
309 data and NSCLC scRNA-seq data support TLS formation in lung adenocarcinoma (**Figure 4**).
310
311 With the transcriptome evidences of TLS formation in MIA tumors, we next sought to confirm
312 TLS formation from pathology imaging. Firstly, we examined the hematoxylin and Eosin
313 staining (H&E) pathology slides and observed lymphocytes aggregation **(Fig. S12),** which
314 support TLS formation in MIA tumors. Then, we used a six-marker multiplex
315 immunohistochemistry (mIHC) to stain for TLS in 22 MIA tumors with high Tfh phenotype,
316 which included the markers CD20 for B cells, CD4 and CXCR5 for follicular helper T cells (Tfh),
317 CD8 for CD8+ T cells, and CD35 for follicular dendritic cells (FDC) **(Figure 5C)**. mIHC
318 confirmed the formation of TLS in MIA tumors, although different samples showed varying
319 level of TLS maturation **(Fig. S13)**. In the samples with mature EGC, we observed TLS B-cell
320 zone (TLS-BZ) as marked by CD20, TLS T-cell zone (TLS-TZ) as marked by CD4, and EGC
321 light zone (EGC-LZ) as marked by CD20, CD4, CXCR5, and CD35 **(Figure 5C, Fig. S14, Fig.**
322 **S15)**. Furthermore, we segmented each observation field into different areas
323 (Adenocarcinoma tissue, Non-Follicle, TLS T zone, TLS B zone, EGC light zone) to quantify
324 and compare the immune cell density between TLS follicles and non-follicle areas. Here TLS
325 follicle area includes TLS-TZ, TLS-BZ, and EGC-LZ, and non-follicle area includes scattered
326 T cells and B cells outside of TLS. In MIA tumors, CD20+ B cells (**Figure 5D**, p-value < 0.0001),
327 CD4+ T cells (**Figure 5E**, p-value < 0.0001) and CD35+ FDC (**Figure 5F,** p-value < 0.0001)
328 were significantly higher in TLS than in non-follicular area. In contrast, CD8+ T cell showed
329 significantly lower density (p-value < 0.0001) in TLS follicles than in non-follicle areas **(Figure**
330 **5G)**.
331
332 **Treg recruitment as potential mechanism of CD8+ T cell exclusion in tertiary lymphoid**
333 **structures**
334 Our transcriptome and mIHC data indicated that MIA tumors had lower CD8+ T cell infiltration
335 (**Figure 2A**) and lower CD8+ T cell density in TLS (**Figure 5G**), respectively. In addition, MIA
336 tumors have lower PD-L1 (CD274) expression than adjacent normal, suggesting PD-L1
337 independent mechanisms for CD8+ T cell suppression **(Fig. S16)**. This prompted us to explore
338 the potential mechanism of CD8+ T cell exclusion in TLS and MIA. We examined immune
339 suppressive cells, such as Tregs, M2 macrophages, neutrophils, monocytes, and fibroblasts
340 (64) for possible cause of CD8+ T cell exclusion. Most suppressive cells were lower in MIA
341 tumors than adjacent normal, with the exception of Treg cells (**Figure 6A**), suggesting a Treg-
342 mediated CD8 T cell exclusion. From two independent scRNA-seq datasets, we also observed
343 CD4+CXCR5+FOXP3+ follicular T regulatory (Tfr) cells to be enriched in NSCLC tumors (**Fig.**
344 **S10B, Fig. S11B**). These data suggest that Treg cells might be responsible to suppress 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) (65), 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 MIA.

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 mediated by 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 **(Fig. S17)**. These results suggest TLS formation to be associated with CD8+ T cell exclusion through Tfr mediated immune suppression.

Discussion

In this study, 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 (hematoxylin and eosin staining and multiplex immune-histochemistry). We observed varying levels of TLS formation, from the initial T cells 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 as a potential CD8+ T cell exclusion mechanism in MIA tumors.

Our analysis revealed mucins in MIA samples to be highly mutated, over-expressed, and highly glycosylated. Mucins synthesized by epithelial cells are well-known to undergo glycosylation alterations during neoplasia (66-68). Aberrant glycosylation metabolism is common in cancers and known to facilitate cell survival and proliferation (66). Mucin overexpression may promote tumorigenicity and metastasis (66), which was proposed as an independent predictor for poor prognosis in lung adenocarcinoma (34). Glycosylated tumor-associated mucin peptides have been reported 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 (52,69). 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 up-regulation of six key TLS indicators (19,70) of B cell infiltration (CD19, CD20), follicle phenotype (CXCR5, CXCL13), and lymphocyte homing (CCR7 and CCL19). The results were corroborated using another 12-chemokine TLS signature (59,60). 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. EGC are known to activate B cells for somatic hypermutation, antibody affinity maturation, class switching, and antibody production (71). 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 (64), inhibit CD8+ T cell infiltration in MIA. Further examination of bulk RNA-seq on our MIA tumors implicated Treg as the culprit of CD8 T cell exclusion in MIA tumors. This hypothesis was supported by previous auto-immune studies in human and mouse (72-76). 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 exclusion in MIA.

Controversy still exists regarding the impact of TLS on clinical prognosis. For example, TLS was reported to be associated with poor prognosis in human hepatocellular carcinoma (77) but good prognosis in colorectal cancer and breast cancer (78,79). In NSCLC cohorts, TLS in the tumors was observed to be associated with favorable prognosis (80-82), 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 (21,22,24). 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 exclusion, 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. Our current study is still limited in sample size and statistical power, although another study on medRxiv reported similar findings in early lung cancers (83). Currently, our analysis only revealed associations without having functional validation of the causal effects. Therefore, future research and clinical studies are needed to fully elucidate the mechanism of CD8 T cell exclusion and explore better immunotherapy strategies in cancer.

Methods

Sample collection and RNA sequencing

All 118 samples were collected from 59 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 S**1. 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 MIA RNA-Seq data

The transcriptome data were processed by the standard RNA-Seq pipeline. We performed sequence mapping by STAR alignment (26), reads quantifying by HT-Seq Count (27), and gene expression measurement by TPM converting from read count. Differential expression analysis on paired samples was performed using edgeR (28). Gene set enrichment analysis was performed using GSEA (29). 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 (37) were annotated by Variant Effect Predictor(VEP) (84) and vcf2maf (https://github.com/mskcc/vcf2maf). Variants with low quality were removed based on the criteria which GDC pipeline recommends (https://docs.gdc.cancer.gov/Data/File_Formats/MAF_Format/). Mutation visualization was implemented by Maftools (85) and Complexheatmap (86).

469 **Normalization for altered mucin expression**
470 Z-score normalization of mRNA expression for patients was performed. One over-expressed
471 or down-expressed gene was expected to have a difference of at least two standard deviations
472 from the mean of expression in the reference samples.
473
474 **Immune infiltration estimation from RNA-seq data**
475 We applied Tumor IMmune Estimation Resource (TIMER) (43) to deconvolve the immune cell
476 components in the TME using TPM expression data. To ensure robust abundance estimation,
477 CIBERSORT (11) according to the predefined LM22 signature and XCELL (87) were both
478 used to estimate immune cell infiltration.
479
480 **The immune signature score for classifying "immune-low" and "immune-high" MIA**
481 **tumors**
482 From the pan-cancer paper (12), the immune signature data and the pre-trained scoring model
483 were downloaded from the CRI-iAtlas (https://github.com/CRI-iAtlas/shiny-iatlas/). Based on
484 the model, we computed the five representative immune signatures score from the expression
485 data, which was normalized by paired normal tissues and median-centered by gene level.
486 Then, we ranked MIA tumors from immune "low" to immune "high" based on the averaged
487 immune signature score.
488
489 **Definitions of BCR repertoire metrics**
490 We applied TRUST3 to infer the immune repertoire from bulk RNA-seq data (44,45).We
491 measured basic metrics to quantify immune repertoire and further performed paired
492 comparison of TCR and BCR repertoire between tumor and tumor-adjacent tissues. The
493 fraction of TCR or BCR reads was defined as the library size located in BCR loci, which is
494 normalized by total mapped reads (E1). The number of unique TCR or BCR CDR3 sequences
495 summarized the total unique CDR3 size. The number of TCR or BCR clonotypes per kilo reads
496 measures TCR or BCR diversity. In addition, B cells undergo clonal expansions with a high
497 rate of somatic hypermutation by single base substitutions in CDR3 sequences. Thus, we built
498 BCR lineages referred to the method of a previous study (45), which generates 8-mer CDR3
499 motifs and an upper triangular matrix of counting the substitutions between DNA sequences
500 from any two clonotypes in one BCR lineage to cluster partial and complete CDR3 sequences.
501 For each sample, BCR or TCR clonality is defined by Shanno entropy (E2) (88). Somatic
502 hypermutation rate was defined as the total count of harboring one base substitution
503 normalized by the total length of aligned DNA sequence (E3). Ig class switch recombination
504 (CSR) were built for the BCR clonotypes with different Ig isotypes from the same BCR lineage.
505 Any two Ig isotypes within the same cluster motif could be switched based on the specific

506 order, which is IgM, IgD, IgG3, IgG1, IgA1, IgG2, IgG4, IgE, IgA2 (89,90). Here we only remain
507 the Ig isotypes with abundance more than 1% and check CSR on IgG and IgA for measuring
508 B cell activation signal. The ratio of CSR was computed referring to the method in the previous
509 study (45). The equations are shown below.

510 $$Fraction\ of\ TCR\ or\ BCR\ reads = \frac{l}{M} \qquad (E1)$$

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

512 $$Clonality = 1 - \frac{-\sum_{i=1}^{n} p_i log2(p_i)}{log2(n)} \qquad (E2)$$

513 $$p_i = \frac{f_i}{\sum_{i=1}^{n} f_i}$$

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

516 $$SHM\ rate = \frac{\sum_{i=1}^{c} S_i}{\sum_{i=1}^{c} \max(L_i)} \qquad (E3)$$

517 c denotes the number of BCR clusters, S_i denotes the sum of clonotype pairs harboring one
518 base substitution, L_i denotes the max number of aligned non-gapped DNA sequence in one
519 cluster.
520

521 **Tumor purity evaluation**

522 Tumor purity was estimated by the algorithm of EPIC (91) (http://epic.gfellerlab.org), which is
523 designed for estimating the absolute fraction of immune cells and cancer cells simultaneously
524 from RNA-seq expression data. We selected "tumor infiltrating cells" as reference profile to
525 predict cell fractions.
526

527 **Single-cell RNA-seq data analysis**

528 Two independent single-cell RNA-seq datasets were collected from NSCLC cohorts with
529 tumors and adjacent normal tissues (GSE139555, EMTAB6149). Single-cell RNA-seq dataset
530 pre-processing was performed by the comprehensive pipeline MAESTRO (92) . Single cell
531 clustering and cell type annotation are available to be visualized in TISCH (http://tisch.comp-
532 genomics.org/). In the dataset of NSCLC_ GSE139555, only cells from tumor and normal
533 tissues were extracted from raw count data for comparing gene expression of cells between
534 tumor and normal. In the dataset of NSCLC_ EMTAB6149, cluster averaged expression data
535 was downloaded from the paper supplementary data due to no available tissue information on
536 cell barcode level. Cell-cell interaction analysis was inferred utilizing a public method
537 CellPhoneDB (63). Only receptor-ligand pairs significantly enriched in less than 5% of all cell-
538 cell interactions were exhibited for the cell interactions on B cells or T cells in the dot plot.

539

540 **Immunohistochemistry**

541 Slides at 4-μm thickness were cut from paraffin blocks from MIA biopsies (n=59) and paired
542 normal lung tissue. Slides were then stained for the interested markers in certain condition of
543 antigen retrieval and incubation (**Table S3**) with Dako Real kit according to IHC-P protocol.
544 Tonsil tissue were performed as control for staining optimization.

545

546 **Multiplex immune-histochemistry staining for lymphocyte infiltrating in MIA**

547 Twenty-two matched paraffin blocks of the 59 samples were processed through the 4-μm-
548 thick section, deparaffinization in xylene, rehydration in an ethanol gradient. Slides were
549 stained according to Opal 7-plex technology (PerkinElmer), yielding to the simultaneous
550 visualization of seven markers (containing DAPI for nucleus) on the same slide. During each
551 of the six cycles of staining, antigen retrieval (AR) was conducted via microwave treatment in
552 AR solution pH 6 or pH 9 (AR6 or AR9) suggested by IHC validation; blocking was followed
553 by instructions for 15 mins at room temperature (RT); and primary antibodies were then
554 incubated for 1 hr at RT or overnight at 4 ℃. Next, HRP labeled polymer goat anti-mouse and
555 rabbit antibodies were incubated at RT for 10 min followed by TSA opal fluorophores (Opal
556 520, Opal 540, Opal 570, Opal 620, Opal 650 or Opal 690) 10 min incubation. Microwave
557 treatment was performed to remove the antibody TSA complex at each cycle of staining with
558 AR solution (pH 9 or pH 6). Finally, all slides were counterstained with DAPI for 5 min and
559 enclosed in ProLong Antifade Mountant (Solarbio). The slides were scanned using the
560 PerkinElmer Mantra System, and the multispectral images obtained were unmixed using
561 spectral libraries that were previously built from images stained for each fluorophore (mono-
562 plex), using the inForm Advanced Image Analysis software (inForm 2.4.1 PerkinElmer).

563

564 **Multispectral image analysis for tumor microenvironment of MIA**

565 The entire area of the sections was stained, scanned, scored, and quantified. The scoring
566 protocol consisted of several automated steps: tissue categorization, cell segmentation, and
567 cell phenotyping. Using the integrated InForm image analysis software, multispectral images
568 that were representative of different samples were selected and used to train the InForm
569 software for categorization of the tissue into epithelium and stroma. Microinvasive epithelium
570 without infiltration was tagged as “ACC” for cancer cell of adenocarcinoma, while lymphocyte
571 infiltrating stroma area was further recognized as “Non-FoI”, “TLS-BZ” and “TLS-TZ”,
572 representing scattered lymphocytes infiltration, B-cell zone of TLS formation, and T-cell zone
573 of TLS formation. The settings learned from the training on the representative images from
574 different samples were saved within an algorithm, which enabled a batch analysis of all the
575 tissue slides. We designed six-plex TLS panel for TLS component cells and seven-plex Tfr/h

panel for suppressive/stimulatory immune cells. They were used on two sequential slides to characterize the immune microenvironment of MIA lesions. The TLS panel included CD4, CD8, CD20, CD35, DAPI, and CXCR5 while the Tfr/h panel included CD3, CD4, CD8, FOXP3, BCL-6, DAPI, and CXCR5. The software classified the tumor (epithelial) and infiltrated stromal regions. The classification was verified and approved by a certified pathologist (W.D.). 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.

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. The thresholds used in the differential expression analysis were 2.0 for fold-change and 0.05 for the adjusted P value. 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.

Authors' Contributions

Conception and design: Jin Wang, Dongbo Jiang
Development of methodology: Jin Wang, Dongbo Jiang
Acquisition of data (acquired and managed patients, provided facilities): Peng Zhang, Gening Jiang, Fan Zhang, Wang Li, Tian Zhao
Analysis and interpretation of data (e.g., statistical analysis, computational analysis, experiment analysis): Jin Wang, Dongbo Jiang, Xiaoqi Zheng, Jian Zhang, Dongqing Sun, Zhe Zhang, Kun Yang
Writing and review the manuscript: Jin Wang, Dongbo Jiang, Ziyi Li, Kun Yang
Administrative, technical, or material support: Huansha Yu, Di Wang
Study supervision: Peng Zhang, Kun Yang

Data Availability

The RNA-seq data is deposited at Genome Sequence Archive (GSA), which is hosted by National Genomics Data Center (NGDC) under the submission code: subCRA003749. All the other data could be found at manuscript or available from the authors upon request.

Competing Interests

No competing interests.

Acknowledgments

The authors acknowledge Dr. X. Shirley Liu in Dana-Farber Cancer Institute and Dr. Arlene Sharpe in Harvard Medical School 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. 81872290, 81972172, and 81772763), the Shanghai Science and Technology Committee (No. 19XD1423200, 18140903900).

632 ## References

633 1. Pinsky PF. Assessing the benefits and harms of low-dose computed tomography
634 screening for lung cancer. Lung Cancer Manag **2014**;3(6):491-8 doi
635 10.2217/LMT.14.41.
636 2. Vinayanuwattikun C, Le Calvez-Kelm F, Abedi-Ardekani B, Zaridze D, Mukeria A,
637 Voegele C*, et al.* Elucidating Genomic Characteristics of Lung Cancer Progression
638 from In Situ to Invasive Adenocarcinoma. Sci Rep **2016**;6:31628 doi
639 10.1038/srep31628.
640 3. Izumchenko E, Chang X, Brait M, Fertig E, Kagohara LT, Bedi A*, et al.* Targeted
641 sequencing reveals clonal genetic changes in the progression of early lung neoplasms
642 and paired circulating DNA. Nat Commun **2015**;6:8258 doi 10.1038/ncomms9258.
643 4. Lavin Y, Kobayashi S, Leader A, Amir ED, Elefant N, Bigenwald C*, et al.* Innate
644 Immune Landscape in Early Lung Adenocarcinoma by Paired Single-Cell Analyses.
645 Cell **2017**;169(4):750-65 e17 doi 10.1016/j.cell.2017.04.014.
646 5. Mascaux C, Angelova M, Vasaturo A, Beane J, Hijazi K, Anthoine G*, et al.* Immune
647 evasion before tumour invasion in early lung squamous carcinogenesis. Nature
648 **2019**;571(7766):570-5 doi 10.1038/s41586-019-1330-0.
649 6. Inamura K. Clinicopathological Characteristics and Mutations Driving Development of
650 Early Lung Adenocarcinoma: Tumor Initiation and Progression. International Journal
651 of Molecular Sciences **2018**;19(4) doi 10.3390/ijms19041259.
652 7. Binnewies M, Roberts EW, Kersten K, Chan V, Fearon DF, Merad M*, et al.*
653 Understanding the tumor immune microenvironment (TIME) for effective therapy. Nat
654 Med **2018**;24(5):541-50 doi 10.1038/s41591-018-0014-x.
655 8. Graw S, Meier R, Minn K, Bloomer C, Godwin AK, Fridley B*, et al.* Robust gene
656 expression and mutation analyses of RNA-sequencing of formalin-fixed diagnostic
657 tumor samples. Sci Rep **2015**;5:12335 doi 10.1038/srep12335.
658 9. Heyer EE, Deveson IW, Wooi D, Selinger CI, Lyons RJ, Hayes VM*, et al.* Diagnosis of
659 fusion genes using targeted RNA sequencing. Nat Commun **2019**;10(1):1388 doi
660 10.1038/s41467-019-09374-9.
661 10. Niu B, Ye K, Zhang Q, Lu C, Xie M, McLellan MD*, et al.* MSIsensor: microsatellite
662 instability detection using paired tumor-normal sequence data. Bioinformatics
663 **2014**;30(7):1015-6 doi 10.1093/bioinformatics/btt755.
664 11. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y*, et al.* Robust enumeration
665 of cell subsets from tissue expression profiles. Nat Methods **2015**;12(5):453-7 doi
666 10.1038/nmeth.3337.
667 12. Thorsson V, Gibbs DL, Brown SD, Wolf D, Bortone DS, Ou Yang T-H*, et al.* The
668 Immune Landscape of Cancer. Immunity **2018**;48(4):812-30.e14 doi
669 10.1016/j.immuni.2018.03.023.
670 13. Wang C, Yin R, Dai J, Gu Y, Cui S, Ma H*, et al.* Whole-genome sequencing reveals
671 genomic signatures associated with the inflammatory microenvironments in Chinese
672 NSCLC patients. Nat Commun **2018**;9(1):2054 doi 10.1038/s41467-018-04492-2.
673 14. Hennecke J, Wiley DC. T cell receptor-MHC interactions up close. Cell **2001**;104(1):1-
674 4 doi 10.1016/s0092-8674(01)00185-4.
675 15. Heesters BA, van der Poel CE, Das A, Carroll MC. Antigen Presentation to B Cells.
676 Trends in Immunology **2016**;37(12):844-54 doi 10.1016/j.it.2016.10.003.
677 16. Tscharke DC, Croft NP, Doherty PC, La Gruta NL. Sizing up the key determinants of
678 the CD8+ T cell response. Nature Reviews Immunology **2015**;15(11):705-16 doi
679 10.1038/nri3905.
680 17. Coussens LM, Werb Z. Inflammation and cancer. Nature **2002**;420(6917):860-7 doi
681 10.1038/nature01322.
682 18. Corsiero E, Nerviani A, Bombardieri M, Pitzalis C. Ectopic Lymphoid Structures:
683 Powerhouse of Autoimmunity. Front Immunol **2016**;7:430 doi
684 10.3389/fimmu.2016.00430.

19. Pitzalis C, Jones GW, Bombardieri M, Jones SA. Ectopic lymphoid-like structures in infection, cancer and autoimmunity. Nat Rev Immunol **2014**;14(7):447-62 doi 10.1038/nri3700.
20. Sautès-Fridman C, Petitprez F, Calderaro J, Fridman WH. Tertiary lymphoid structures in the era of cancer immunotherapy. Nature Reviews Cancer **2019**;19(6):307-25 doi 10.1038/s41568-019-0144-6.
21. Cabrita R, Lauss M, Sanna A, Donia M, Skaarup Larsen M, Mitra S*, et al.* Tertiary lymphoid structures improve immunotherapy and survival in melanoma. Nature **2020**;577(7791):561-5 doi 10.1038/s41586-019-1914-8.
22. Helmink BA, Reddy SM, Gao J, Zhang S, Basar R, Thakur R*, et al.* B cells and tertiary lymphoid structures promote immunotherapy response. Nature **2020**;577(7791):549-55 doi 10.1038/s41586-019-1922-8.
23. Bruno TC. New predictors for immunotherapy responses sharpen our view of the tumour microenvironment. Nature **2020**;577(7791):474-6 doi 10.1038/d41586-019-03943-0.
24. Petitprez F, de Reynies A, Keung EZ, Chen TW, Sun CM, Calderaro J*, et al.* B cells are associated with survival and immunotherapy response in sarcoma. Nature **2020**;577(7791):556-60 doi 10.1038/s41586-019-1906-8.
25. Travis WD, Brambilla E, Nicholson AG, Yatabe Y, Austin JHM, Beasley MB*, et al.* The 2015 World Health Organization Classification of Lung Tumors: Impact of Genetic, Clinical and Radiologic Advances Since the 2004 Classification. J Thorac Oncol **2015**;10(9):1243-60 doi 10.1097/JTO.0000000000000630.
26. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S*, et al.* STAR: ultrafast universal RNA-seq aligner. Bioinformatics **2013**;29(1):15-21 doi 10.1093/bioinformatics/bts635.
27. Anders S, Pyl PT, Huber W. HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics **2015**;31(2):166-9 doi 10.1093/bioinformatics/btu638.
28. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics **2010**;26(1):139-40 doi 10.1093/bioinformatics/btp616.
29. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA*, et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A **2005**;102(43):15545-50 doi 10.1073/pnas.0506580102.
30. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res **2000**;28(1):27-30 doi 10.1093/nar/28.1.27.
31. Fletcher JI, Haber M, Henderson MJ, Norris MD. ABC transporters in cancer: more than just drug efflux pumps. Nature Reviews Cancer **2010**;10(2):147-56 doi 10.1038/nrc2789.
32. Sanchez-Vega F, Mina M, Armenia J, Chatila WK, Luna A, La KC*, et al.* Oncogenic Signaling Pathways in The Cancer Genome Atlas. Cell **2018**;173(2):321-37 e10 doi 10.1016/j.cell.2018.03.035.
33. Legler DF, Loetscher M, Roos RS, Clark-Lewis I, Baggiolini M, Moser B. B cell-attracting chemokine 1, a human CXC chemokine expressed in lymphoid tissues, selectively attracts B lymphocytes via BLR1/CXCR5. J Exp Med **1998**;187(4):655-60 doi 10.1084/jem.187.4.655.
34. Lakshmanan I, Ponnusamy MP, Macha MA, Haridas D, Majhi PD, Kaur S*, et al.* Mucins in lung cancer: diagnostic, prognostic, and therapeutic implications. J Thorac Oncol **2015**;10(1):19-27 doi 10.1097/JTO.0000000000000404.
35. Bouillez A, Adeegbe D, Jin C, Hu X, Tagde A, Alam M*, et al.* MUC1-C promotes the suppressive immune microenvironment in non-small cell lung cancer. Oncoimmunology **2017**;6(9):e1338998 doi 10.1080/2162402X.2017.1338998.
36. Maeda T, Hiraki M, Jin C, Rajabi H, Tagde A, Alam M*, et al.* MUC1-C Induces PD-L1 and Immune Evasion in Triple-Negative Breast Cancer. Cancer Res **2018**;78(1):205-15 doi 10.1158/0008-5472.CAN-17-1636.

740 37. Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L*, et al.* VarScan 2:
741 somatic mutation and copy number alteration discovery in cancer by exome
742 sequencing. Genome Res **2012**;22(3):568-76 doi 10.1101/gr.129684.111.
743 38. Harris RS. Cancer mutation signatures, DNA damage mechanisms, and potential
744 clinical implications. Genome Med **2013**;5(9):87 doi 10.1186/gm490.
745 39. Paez JG, Janne PA, Lee JC, Tracy S, Greulich H, Gabriel S*, et al.* EGFR mutations in
746 lung cancer: correlation with clinical response to gefitinib therapy. Science
747 **2004**;304(5676):1497-500 doi 10.1126/science.1099314.
748 40. Li S, Li L, Zhu Y, Huang C, Qin Y, Liu H*, et al.* Coexistence of EGFR with KRAS, or
749 BRAF, or PIK3CA somatic mutations in lung cancer: a comprehensive mutation
750 profiling from 5125 Chinese cohorts. Br J Cancer **2014**;110(11):2812-20 doi
751 10.1038/bjc.2014.210.
752 41. Szolek A, Schubert B, Mohr C, Sturm M, Feldhahn M, Kohlbacher O. OptiType:
753 precision HLA typing from next-generation sequencing data. Bioinformatics
754 **2014**;30(23):3310-6 doi 10.1093/bioinformatics/btu548.
755 42. Hundal J, Carreno BM, Petti AA, Linette GP, Griffith OL, Mardis ER*, et al.* pVAC-Seq:
756 A genome-guided in silico approach to identifying tumor neoantigens. Genome Med
757 **2016**;8(1):11 doi 10.1186/s13073-016-0264-5.
758 43. Li B, Severson E, Pignon JC, Zhao H, Li T, Novak J*, et al.* Comprehensive analyses
759 of tumor immunity: implications for cancer immunotherapy. Genome Biol
760 **2016**;17(1):174 doi 10.1186/s13059-016-1028-7.
761 44. Li B, Li T, Wang B, Dou R, Zhang J, Liu JS*, et al.* Ultrasensitive detection of TCR
762 hypervariable-region sequences in solid-tissue RNA-seq data. Nat Genet
763 **2017**;49(4):482-3 doi 10.1038/ng.3820.
764 45. Hu X, Zhang J, Wang J, Fu J, Li T, Zheng X*, et al.* Landscape of B cell immunity and
765 related immune evasion in human cancers. Nat Genet **2019**;51(3):560-7 doi
766 10.1038/s41588-018-0339-x.
767 46. Ivetic A, Hoskins Green HL, Hart SJ. L-selectin: A Major Regulator of Leukocyte
768 Adhesion, Migration and Signaling. Front Immunol **2019**;10:1068 doi
769 10.3389/fimmu.2019.01068.
770 47. Gorabi AM, Hajighasemi S, Kiaie N, Gheibi Hayat SM, Jamialahmadi T, Johnston TP*,*
771 *et al.* The pivotal role of CD69 in autoimmunity. J Autoimmun **2020**;111:102453 doi
772 10.1016/j.jaut.2020.102453.
773 48. Greenwald RJ, Freeman GJ, Sharpe AH. The B7 family revisited. Annu Rev Immunol
774 **2005**;23:515-48 doi 10.1146/annurev.immunol.23.021704.115611.
775 49. Grewal IS, Flavell RA. The CD40 ligand. At the center of the immune universe?
776 Immunol Res **1997**;16(1):59-70 doi 10.1007/BF02786323.
777 50. Thomas N, Best K, Cinelli M, Reich-Zeliger S, Gal H, Shifrut E*, et al.* Tracking global
778 changes induced in the CD4 T-cell receptor repertoire by immunization with a complex
779 antigen using short stretches of CDR3 protein sequence. Bioinformatics
780 **2014**;30(22):3181-8 doi 10.1093/bioinformatics/btu523.
781 51. Shugay M, Bagaev DV, Turchaninova MA, Bolotin DA, Britanova OV, Putintseva EV*,*
782 *et al.* VDJtools: Unifying Post-analysis of T Cell Receptor Repertoires. PLoS Comput
783 Biol **2015**;11(11):e1004503 doi 10.1371/journal.pcbi.1004503.
784 52. Cai H, Degliangeli F, Palitzsch B, Gerlitzki B, Kunz H, Schmitt E*, et al.* Glycopeptide-
785 functionalized gold nanoparticles for antibody induction against the tumor associated
786 mucin-1 glycoprotein. Bioorg Med Chem **2016**;24(5):1132-5 doi
787 10.1016/j.bmc.2016.01.044.
788 53. Cai H, Chen MS, Sun ZY, Zhao YF, Kunz H, Li YM. Self-adjuvanting synthetic
789 antitumor vaccines from MUC1 glycopeptides conjugated to T-cell epitopes from
790 tetanus toxoid. Angew Chem Int Ed Engl **2013**;52(23):6106-10 doi
791 10.1002/anie.201300390.
792 54. Schoenberger SP, Toes RE, van der Voort EI, Offringa R, Melief CJ. T-cell help for
793 cytotoxic T lymphocytes is mediated by CD40-CD40L interactions. Nature
794 **1998**;393(6684):480-3 doi 10.1038/31002.

795 55. Ginaldi L, De Martinis M, Matutes E, Farahat N, Morilla R, Catovsky D. Levels of
796 expression of CD19 and CD20 in chronic B cell leukaemias. J Clin Pathol
797 **1998**;51(5):364-9 doi 10.1136/jcp.51.5.364.
798 56. Wang K, Wei G, Liu D. CD19: a biomarker for B cell development, lymphoma diagnosis
799 and therapy. Exp Hematol Oncol **2012**;1(1):36 doi 10.1186/2162-3619-1-36.
800 57. Havenar-Daughton C, Lindqvist M, Heit A, Wu JE, Reiss SM, Kendric K*, et al.* CXCL13
801 is a plasma biomarker of germinal center activity. Proc Natl Acad Sci U S A
802 **2016**;113(10):2702-7 doi 10.1073/pnas.1520112113.
803 58. Kobayashi D, Endo M, Ochi H, Hojo H, Miyasaka M, Hayasaka H. Regulation of CCR7-
804 dependent cell migration through CCR7 homodimer formation. Sci Rep
805 **2017**;7(1):8536 doi 10.1038/s41598-017-09113-4.
806 59. Coppola D, Nebozhyn M, Khalil F, Dai H, Yeatman T, Loboda A*, et al.* Unique ectopic
807 lymph node-like structures present in human primary colorectal carcinoma are
808 identified by immune gene array profiling. Am J Pathol **2011**;179(1):37-45 doi
809 10.1016/j.ajpath.2011.03.007.
810 60. Messina JL, Fenstermacher DA, Eschrich S, Qu X, Berglund AE, Lloyd MC*, et al.* 12-
811 Chemokine gene signature identifies lymph node-like structures in melanoma:
812 potential for patient selection for immunotherapy? Sci Rep **2012**;2:765 doi
813 10.1038/srep00765.
814 61. Lambrechts D, Wauters E, Boeckx B, Aibar S, Nittner D, Burton O*, et al.* Phenotype
815 molding of stromal cells in the lung tumor microenvironment. Nat Med
816 **2018**;24(8):1277-89 doi 10.1038/s41591-018-0096-5.
817 62. Wu TD, Madireddi S, de Almeida PE, Banchereau R, Chen YJ, Chitre AS*, et al.*
818 Peripheral T cell expansion predicts tumour infiltration and clinical response. Nature
819 **2020**;579(7798):274-8 doi 10.1038/s41586-020-2056-8.
820 63. Efremova M, Vento-Tormo M, Teichmann SA, Vento-Tormo R. CellPhoneDB: inferring
821 cell-cell communication from combined expression of multi-subunit ligand-receptor
822 complexes. Nat Protoc **2020**;15(4):1484-506 doi 10.1038/s41596-020-0292-x.
823 64. Cassetta L, Pollard JW. Targeting macrophages: therapeutic approaches in cancer.
824 Nat Rev Drug Discov **2018**;17(12):887-904 doi 10.1038/nrd.2018.169.
825 65. Xu B, Wang S, Zhou M, Huang Y, Fu R, Guo C*, et al.* The ratio of circulating follicular
826 T helper cell to follicular T regulatory cell is correlated with disease activity in systemic
827 lupus erythematosus. Clin Immunol **2017**;183:46-53 doi 10.1016/j.clim.2017.07.004.
828 66. Tuccillo FM, de Laurentiis A, Palmieri C, Fiume G, Bonelli P, Borrelli A*, et al.* Aberrant
829 glycosylation as biomarker for cancer: focus on CD43. Biomed Res Int
830 **2014**;2014:742831 doi 10.1155/2014/742831.
831 67. Kufe DW. Mucins in cancer: function, prognosis and therapy. Nat Rev Cancer
832 **2009**;9(12):874-85 doi 10.1038/nrc2761.
833 68. Hollingsworth MA, Swanson BJ. Mucins in cancer: protection and control of the cell
834 surface. Nat Rev Cancer **2004**;4(1):45-60 doi 10.1038/nrc1251.
835 69. van Kooten C, Banchereau J. CD40-CD40 ligand. J Leukoc Biol **2000**;67(1):2-17 doi
836 10.1002/jlb.67.1.2.
837 70. Nerviani A, Pitzalis C. Role of chemokines in ectopic lymphoid structures formation in
838 autoimmunity and cancer. J Leukoc Biol **2018**;104(2):333-41 doi
839 10.1002/JLB.3MR0218-062R.
840 71. Domeier PP, Schell SL, Rahman ZS. Spontaneous germinal centers and autoimmunity.
841 Autoimmunity **2017**;50(1):4-18 doi 10.1080/08916934.2017.1280671.
842 72. Sage PT, Sharpe AH. T follicular regulatory cells in the regulation of B cell responses.
843 Trends Immunol **2015**;36(7):410-8 doi 10.1016/j.it.2015.05.005.
844 73. Sage PT, Ron-Harel N, Juneja VR, Sen DR, Maleri S, Sungnak W*, et al.* Suppression
845 by TFR cells leads to durable and selective inhibition of B cell effector function. Nat
846 Immunol **2016**;17(12):1436-46 doi 10.1038/ni.3578.
847 74. Wing JB, Tekguc M, Sakaguchi S. Control of Germinal Center Responses by T-
848 Follicular Regulatory Cells. Front Immunol **2018**;9:1910 doi
849 10.3389/fimmu.2018.01910.

850 75. Viguier M, Lemaitre F, Verola O, Cho MS, Gorochov G, Dubertret L*, et al.* Foxp3
851 expressing CD4+CD25(high) regulatory T cells are overrepresented in human
852 metastatic melanoma lymph nodes and inhibit the function of infiltrating T cells. J
853 Immunol **2004**;173(2):1444-53 doi 10.4049/jimmunol.173.2.1444.
854 76. Clement RL, Daccache J, Mohammed MT, Diallo A, Blazar BR, Kuchroo VK*, et al.*
855 Follicular regulatory T cells control humoral and allergic immunity by restraining early
856 B cell responses. Nat Immunol **2019**;20(10):1360-71 doi 10.1038/s41590-019-0472-4.
857 77. Finkin S, Yuan D, Stein I, Taniguchi K, Weber A, Unger K*, et al.* Ectopic lymphoid
858 structures function as microniches for tumor progenitor cells in hepatocellular
859 carcinoma. Nat Immunol **2015**;16(12):1235-44 doi 10.1038/ni.3290.
860 78. Di Caro G, Bergomas F, Grizzi F, Doni A, Bianchi P, Malesci A*, et al.* Occurrence of
861 tertiary lymphoid tissue is associated with T-cell infiltration and predicts better
862 prognosis in early-stage colorectal cancers. Clin Cancer Res **2014**;20(8):2147-58 doi
863 10.1158/1078-0432.CCR-13-2590.
864 79. Gu-Trantien C, Loi S, Garaud S, Equeter C, Libin M, de Wind A*, et al.* CD4(+) follicular
865 helper T cell infiltration predicts breast cancer survival. J Clin Invest **2013**;123(7):2873-
866 92 doi 10.1172/JCI67428.
867 80. Dieu-Nosjean MC, Antoine M, Danel C, Heudes D, Wislez M, Poulot V*, et al.* Long-
868 term survival for patients with non-small-cell lung cancer with intratumoral lymphoid
869 structures. J Clin Oncol **2008**;26(27):4410-7 doi 10.1200/JCO.2007.15.0284.
870 81. Remark R, Lupo A, Alifano M, Biton J, Ouakrim H, Stefani A*, et al.* Immune contexture
871 and histological response after neoadjuvant chemotherapy predict clinical outcome of
872 lung cancer patients. Oncoimmunology **2016**;5(12):e1255394 doi
873 10.1080/2162402X.2016.1255394.
874 82. Germain C, Gnjatic S, Tamzalit F, Knockaert S, Remark R, Goc J*, et al.* Presence of
875 B cells in tertiary lymphoid structures is associated with a protective immunity in
876 patients with lung cancer. Am J Respir Crit Care Med **2014**;189(7):832-44 doi
877 10.1164/rccm.201309-1611OC.
878 83. Dejima H, Hu X, Chen R, Zhang J, Junya F, Parra EC*, et al.* Immune evolution from
879 preneoplasia to invasive lung adenocarcinomas and underlying molecular features.
880 **2020**:2020.07.11.20142992 doi 10.1101/2020.07.11.20142992 %J medRxiv.
881 84. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GR, Thormann A*, et al.* The Ensembl
882 Variant Effect Predictor. Genome Biol **2016**;17(1):122 doi 10.1186/s13059-016-0974-
883 4.
884 85. Mayakonda A, Lin DC, Assenov Y, Plass C, Koeffler HP. Maftools: efficient and
885 comprehensive analysis of somatic variants in cancer. Genome Res
886 **2018**;28(11):1747-56 doi 10.1101/gr.239244.118.
887 86. Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in
888 multidimensional genomic data. Bioinformatics **2016**;32(18):2847-9 doi
889 10.1093/bioinformatics/btw313.
890 87. Aran D, Hu Z, Butte AJ. xCell: digitally portraying the tissue cellular heterogeneity
891 landscape. Genome Biol **2017**;18(1):220 doi 10.1186/s13059-017-1349-1.
892 88. Sims JS, Grinshpun B, Feng Y, Ung TH, Neira JA, Samanamud JL*, et al.* Diversity and
893 divergence of the glioma-infiltrating T-cell receptor repertoire. Proc Natl Acad Sci U S
894 A **2016**;113(25):E3529-37 doi 10.1073/pnas.1601012113.
895 89. Kitaura K, Yamashita H, Ayabe H, Shini T, Matsutani T, Suzuki R. Different Somatic
896 Hypermutation Levels among Antibody Subclasses Disclosed by a New Next-
897 Generation Sequencing-Based Antibody Repertoire Analysis. Front Immunol
898 **2017**;8:389 doi 10.3389/fimmu.2017.00389.
899 90. Looney TJ, Lee JY, Roskin KM, Hoh RA, King J, Glanville J*, et al.* Human B-cell isotype
900 switching origins of IgE. J Allergy Clin Immunol **2016**;137(2):579-86 e7 doi
901 10.1016/j.jaci.2015.07.014.
902 91. Racle J, de Jonge K, Baumgaertner P, Speiser DE, Gfeller D. Simultaneous
903 enumeration of cancer and immune cell types from bulk tumor gene expression data.
904 Elife **2017**;6 doi 10.7554/eLife.26476.

92. Wang CS, Dongqing. MAESTRO: Model-based AnalysEs of Single-cell Transcriptome and Regulome. **2020** doi 10.5281/zenodo.3862813.

909 # FIGURE LEGEND

910

911 **Figure 1. MIA tumors are characterized by highly mutated and over-expressed mucins,**
912 **cytokine::cytokine-receptor interactions, and EGFR mutations**

913 **(A)** Representative HE images of MIA. Top at 4X magnification, scale bar 300 μm. Boxed area
914 of top image shown below (20x magnification, scale bar 50 μm). Lymphocyte infiltration
915 (dashed box, bottom) and MIA tumor cells (solid box, bottom) are shown. **(B)** Normalized
916 enrichment scores (x-axis) of tumor-upregulated KEGG pathways. Colors indicate different
917 KEGG pathway categories. Asterisks denote significance level (*** FDR < 0.001; ** FDR <
918 0.01; * FDR < 0.05; NS FDR > 0.05). **(C)** Alterations of mucin genes in MIA patients. Each bar
919 represents a MIA patient. Shown are mutation classifications, and expression status. Values
920 on the right indicate the proportion of patients in which the gene was either mutated or had
921 altered expression. **(D)** Most commonly mutated genes in MIA patients. Colors indicate variant
922 type. Shown are the proportion of tumors containing a mutation in the indicated gene (left);
923 total number of mutations in the indicated gene (right). **(E)** Schematic of the EGFR gene,
924 showing mutations (stick and ball markers) and protein domains (color blocks).

925

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

928 **(A)** Comparison of immune components between tumor (red) and adjacent normal tissue (blue)
929 in MIA. Immune Cell Abundance values (y-axis) were generated from TIMER. Significance
930 levels are indicated (*** P < 0.001; ** P < 0.01; * P < 0.05; ns P> 0.05). **(B)** Immune infiltration
931 among TCGA lung adenocarcinoma (LUAD) samples separated by clinical stage (colors).
932 Comparison of fraction of TCR reads **(C)** and fraction of BCR reads (**D**) between MIA tumor
933 and adjacent normal tissues. (**E**) Comparison of gene expressions on CD4+ T cell activation
934 markers (CD69, CD40LG, IL2RA, and CD62) between MIA tumor and adjacent normal tissues.
935 Comparison of the number of BCR clusters (**F**), BCR clonotypes per kilo reads **(G)**, BCR
936 clonality **(H), and** somatic hyper-mutation rate of BCR **(I)** between MIA tumor and adjacent
937 normal tissues. Asterisks denote significance on difference between MIA tumors and adjacent
938 normal tissues (*** P < 0.001; ** P < 0.01; * P < 0.05).

939

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

941 **(A)** Normalized BCR clonotype frequency on patient level. Each dot represents a BCR
942 clonotype and the normalized clonotype abundance in tumor and adjacent normal tissue. The
943 three percentage values correspond to the overall percentage of tumor-specific clonotypes
944 (58.6% in BCR), normal-specific clonotypes (38.6% in BCR), and shared (2.86% in BCR)

945 clonotypes. **(B)** Normalized TCR clonotype frequency on patient level. **(C)** Jaccard similarity
946 of the BCR/TCR clonotypes between tumor and adjacent normal tissues. Shown are shared
947 clonotypes within each patient (dark brown for BCR, dark green for TCR) and average shared
948 clonotypes between tumors and unmatched normal samples (light brown for BCR, light green
949 for TCR). **(D)** Jaccard similarity of BCR and TCR repertoire among tumors (solid line) and
950 normal tissues (dashed line) shown in top (BCR) and bottom (TCR) panel, respectively. **(E)**
951 Jaccard similarity in groups with high (blue) and low (red) mutation load (*** $P < 0.001$; ** $P <$
952 0.01; * $P < 0.05$).
953
954 **Figure 4. Mucin mutation and overexpression is associated with CD4+ T cell-mediated**
955 **B cell activation in MIA tumors**
956 **(A)** The correlation between mucins and MHCII expression. **(B)** The correlation between
957 CD40 and CD40LG expression. **(C)** The correlation between mucins and averaged level of
958 CD40 and CD40LG. **(D)** The correlation between mucin expression with IgG (green) and IgA
959 (brown) expression in MIA patients. **(E)** MIA tumors (x-axis) ranked by immune-related
960 signature score (heatmap colors). The top bar shows the average immune signature score for
961 each patient. The bottom bars show the tumor purity, expression level of mucins, MHCII
962 molecules, CD40, CD40LG, IgG, and IgA.
963
964 **Figure 5. Formation of tertiary lymphoid structure (TLS) in MIA tumors**
965 **(A)** The schema of TLS formation. The first stage is B & T cell aggregation. The second stage
966 is follicle formation with T-Zone and B-Zone. The third stage is ectopic germinal center
967 formation with formation with follicular dendritic cell in light-Zone. **(B)** Hierarchical clustering
968 of Z-score normalized gene expression of B-cell infiltration markers (CD19, CD20), follicle
969 phenotype markers (CXCR5, CXCL13), and lymphocyte homing signal markers (CCR7 and
970 CCL19). Color of the top bar represents tumor (red) or adjacent normal (blue) tissue.
971 Correlations between expression levels from these six TLS markers and immune cell
972 abundance corrected by tumor purity are shown in the right bar. **(C)** Typical observation field
973 of MIA six-plex TLS panel staining. The merged-channel multispectral image was captured
974 after 5-cycle TSA staining and DAPI counterstain, revealing a classic lymphoid structure
975 residing in MIA stromal tissue. Single channels are shown for the area bounded by the white
976 rectangle. Fake colors are assigned as DAPI: blue, CD8: yellow, CXCR5: pink, CD4: green,
977 CD20: red, CD35: cyan. Tfh cells are highlighted with white arrows (lower right corner images).
978 **(D)** Paired comparisons of CD20+ B cell density, CD4+ T cell density **(E),** CD35+ follicle
979 dendritic cell density **(F)**, and CD8+ T cell density **(G)** between Non-Follicle (Non-Fol) and TLS
980 areas.

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

(A) Comparison of immune suppressive cell infiltration estimated from XCELL between MIA tumor and adjacent normal tissues (*** $P < 0.001$; ** $P < 0.01$; * $P < 0.05$). **(B)** Plot of Tfr/Tfh marker ratio vs CD8+ T cell infiltration. FoxP3, PDCD1, and IL2RA are markers of Tfr cells, while BCL6, ICOS, IL7R are markers of Tfh cells. See text. **(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.

997 **Figure S1. The workflow of data processing.**
998 In this workflow, five sections were performed for analyzing transcriptome data and imaging
999 data, including expression profiling EdgR and GSEA, mutation calling from Varscan2,
1000 estimating immune cell infiltration from TIMER and CIBERSORT, quantifying immune
1001 repertoire from TRUST3, and Morphological validation from H&E staining, multiplex IHC,
1002 multispectral imaging and spatial statistics.
1003
1004 **Figure S2. Differential gene expression analysis.**
1005 Colors of dots represent the tumor up-regulated genes (red) and tumor down-regulated genes
1006 (blue). Top differentially expressed mucin family members, cytokines, and cytokine receptors
1007 are highlighted in Volcano plot.
1008
1009 **Figure S3. Genetic alterations of mucin family members in TCGA LUAD.**
1010 Mutations and gene alterations in mucin family members are highlighted with different colors,
1011 including inframe mutation (pink inside), missense mutation (green inside), truncating mutation
1012 (blue inside), mRNA high expression (pink outside), mRNA low expression (blue outside), and
1013 no alterations (grey).
1014
1015 **Figure S4. Mutation overview of MIA.**
1016 The summary of variants from MIA, including the number of different variant classifications
1017 and SNV classes for each patient (**A**) and the frequency of nucleotide changes (**B**). (**C**)
1018 Comparison of msi score between high and low CT transition. (**D**) The mutation loci in EGFR.
1019 (**E**) The mutation pattern of EGFR, KRAS and TP53 in MIA.
1020
1021 **Figure S5. Neo-epitope prediction for top mutated genes.**
1022 (**A**) Mutation rate (dark blue bar) and neo-epitope rate (MHC-I bound, light blue bar) among
1023 59 patients for top 15 mutate genes. (**B**) Neo-epitope binding affinity and gene expression for
1024 top 15 mutated genes in patients. Each dot represents a predicted neo-epitope in a patient.
1025 Epitopes predicted from MUC20 (green), MUC4 (yellow), and MUC6 (purple) are highlighted
1026 with text. The color of dot represents the types of gene expression changes, including up-
1027 regulation (red), down-regulation (blue) and no change (grey).
1028
1029 **Figure S6. Immune infiltration estimation from CIBERSORT.**
1030 Comparison of Immune cell infiltration estimated from CIBERSORT ("Absolute" mode)
1031 between MIA tumor and adjacent normal tissues.

1032 **Figure S7. Immune repertoire metrics in MIA.** Comparison of faction of TCR reads (**A**),
1033 faction of BCR reads (**B**) between MIA tumor and normal tissues. (**C**) The ratio of Ig class
1034 switch recombination between IgG and IgA in MIA tumor and normal tissues. (*** P < 0.001;
1035 ** P < 0.01; * P < 0.05).
1036
1037 **Figure S8. Association between molecules mediated B-CD4+ T activation and immune-**
1038 **related signature score.**
1039 Association between MHCII level (**A**), mucin level (**B**), CD40 level (**C**), CD40LG level (**D**), IgG
1040 level (**E**), IgA level (**F**), and percentage of tumor unique BCR clonotypes (**G**) with immune-
1041 related signature score. MIA tumors are divided into “immune-low” and “immune-high” groups
1042 based on the median immune-related signature score. (*** P < 0.001; ** P < 0.01; * P < 0.05).
1043
1044 **Figure S9. 6 TLS markers and a 12-chemokine signature for predicting TLS existence.**
1045 (**A**) Association between 6 TLS markers expression and a 12-chemokine signature score. (**B**)
1046 Association between 6 TLS markers expression and immune-related signature score. (**C**)
1047 Association between a 12-chemokine signature score and immune-related signature score.
1048 MIA tumors are divided into “immune-low” and “immune-high” groups based on the median
1049 immune-related signature score. (*** P < 0.001; ** P < 0.01; * P < 0.05).
1050
1051 **Figure S10. B, follicular T helper, and follicular T regulatory cells are enriched in NSCLC**
1052 **tumors in an independent single-cell RNA-seq dataset NSCLC_GSE139555.** (**A**) (left) cell
1053 distribution in different tissues and each color represents a type of tissue (left), including blood,
1054 NAT (normal adjacent tissue), and tumor;(right) cell type annotation and each color represents
1055 a cell type. (**B**) Gene expressions for CD4, CXCR5, FOXP3, and MS4A1 are compared
1056 between tumor and NAT across different cell types. (*** P < 0.001; ** P < 0.01; * P < 0.05).
1057
1058 **Figure S11. B, follicular T helper, and follicular T regulatory cells are enriched in NSCLC**
1059 **tumors in an independent single-cell RNA-seq dataset NSCLC_EMTAB6149.** (**A**) cell type
1060 annotations. Each color represents a cell type. (**B**) Cluster averaged signature expression
1061 comparison between tumor (red) and normal tissues (blue) of different cell types. (**C**)
1062 Receptor-ligand pairs enrichment of different cell-cell interactions. The dot color represents
1063 the expression level of the receptor-ligand pairs and the dot size represents the enrichment
1064 significance of interaction pairs.
1065
1066 **Figure S12. Evidence of TLS presence from H&E staining.** The red circles denote the
1067 inferred TLS B and T cell aggregation.
1068

Figure S13. Different maturation level of TLS formation and corresponding CD8+ T cell infiltration. TLS formation stained by CD4(green), CD20(red), and CD35(cyan) co-localization in the left panel and CD8+ T cell infiltration in the right panel from the same observation field are shown.

Figure S14. Normal germinal center in tonsil and ectopic germinal center in MIA. Normal germinal center in tonsil (**A**) and ectopic germinal center in MIA (**B**). T-cell zone, B-cell zone, dark zone, and light zone consist of GC in tonsil or EGC in MIA stained by immunofluorescence (left) and immunohistochemistry (right).

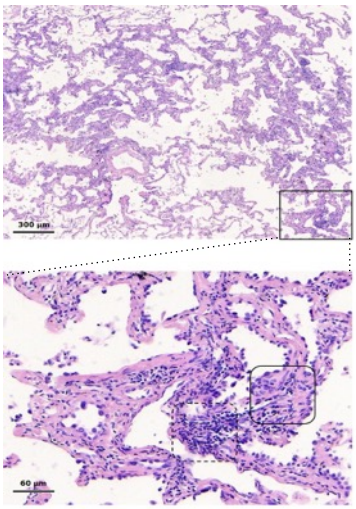
Figure S15. Multispectral analysis for segmenting tissues, cells and phenotypes. Segmentation for six-plex panel for TLS (**A**) and seven-plex panel for Tfr and Tfh (**B**). Each observation field could be segmented into 5 tissue categories as adenocarcinoma tissue (ACT), Non-Follicle area (Non-Fol), TLS T-cell area (TLS-T), TLS B-cell area (TLS-B) and light zone of germinal center (EGC-LZ) in both six-plex panel (top) and seven-plex panel (bottom). Five phenotypes were contained in six-plex panel, including CC (blue), CD8+ T (yellow), CD4+ T (green), CD20+ B (red), and CD35+ FDC (cyan). Seven phenotypes were contained in seven-plex, including CC (blue), CTL (yellow), Th (green), Tfh (cyan), Treg (pink), Tfr (red), and oLeu (black). ACC: Adenocarcinoma cancer cell; CC: cancer cell; Non-Fol: Non-follicular area; CTL: cytotoxic T lymphocyte; Th: T helper; Tfh: follicular T helper; Treg: regulatory T-cell; Tfr: follicular Treg; FDC: follicular dendritic cell; oLeu: other leukocyte.

Figure S16. Down-regulated PD-L1 in MIA tumors. Comparison of the expression of PD-L1 (CD274) in tumor (red) and adjacent-normal (blue) samples (*** $P < 0.001$; ** $P < 0.01$; * $P < 0.05$).

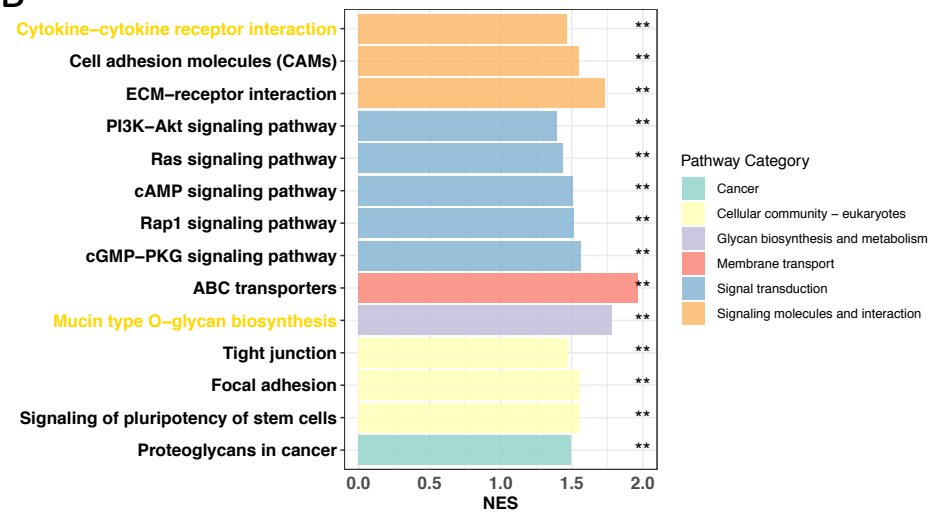
Figure S17. The ratio of regulatory T to helper T and CD8+ T cell infiltration in Non-follicle areas. CD8+ T cell comparison between groups with high (red) and low (blue) ratio of regulatory T to helper T.

Figure1

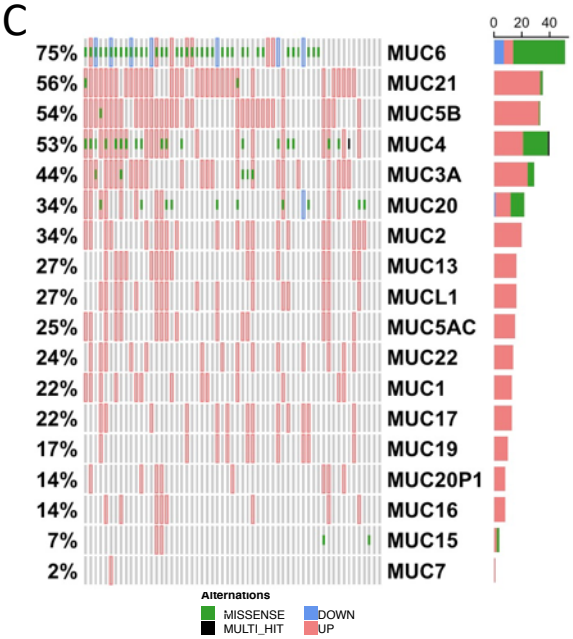
A



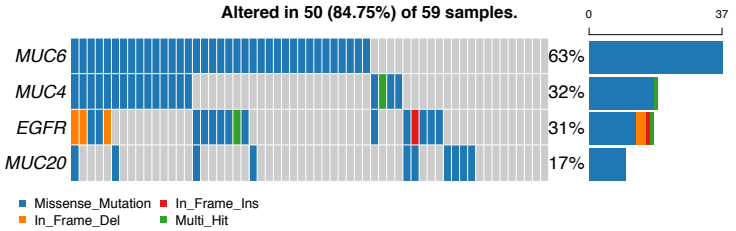
B



C



D



E

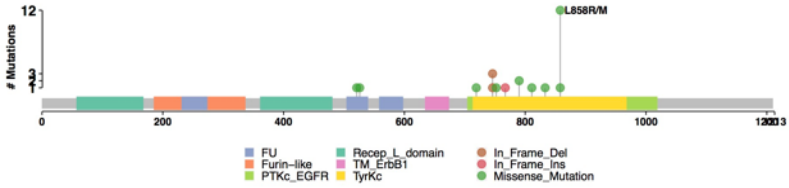


Figure 2

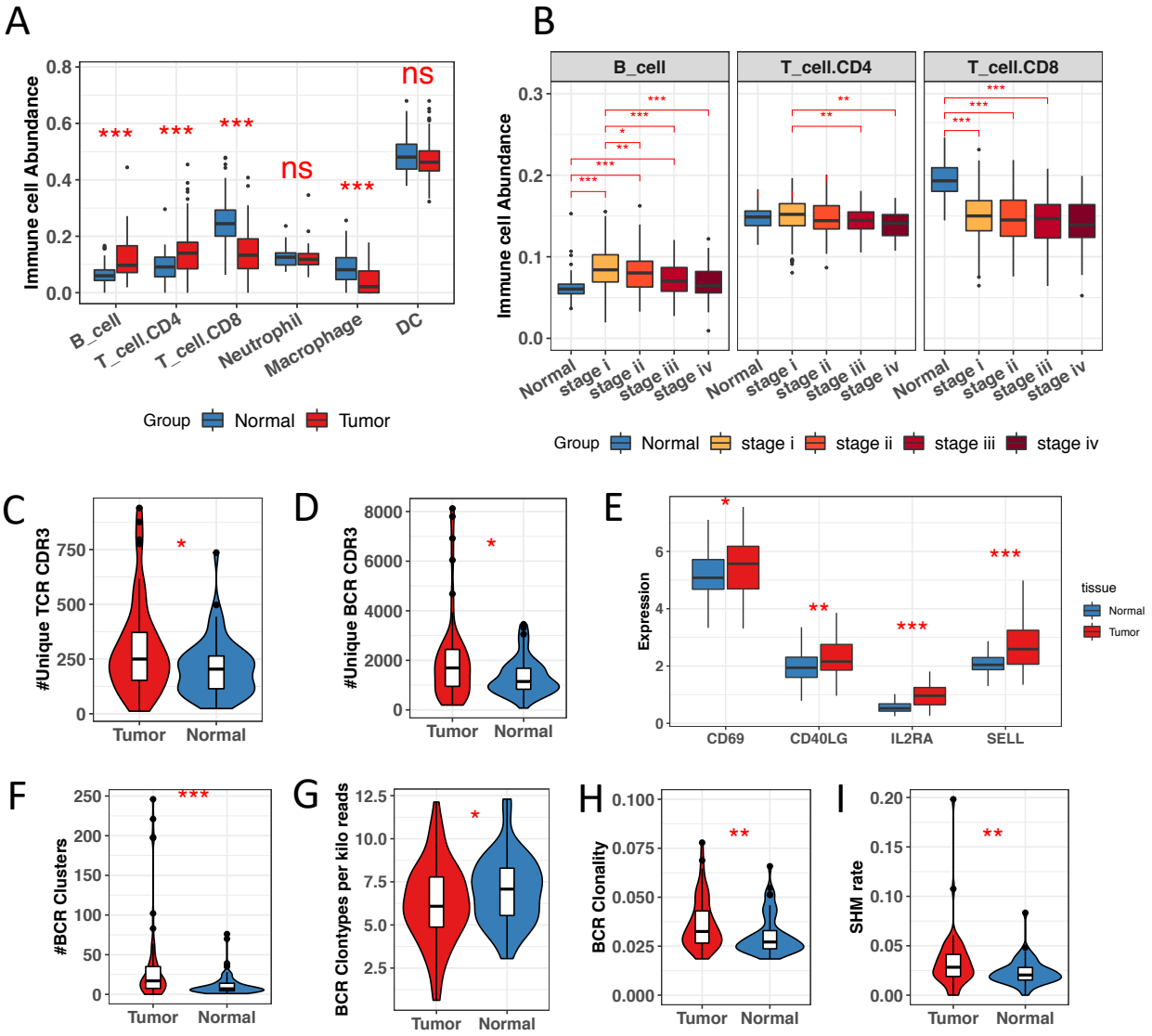

A
Immune cell Abundance
0.8
0.6
0.4
0.2
0.0
***
***
***
ns
***
ns
B_cell
T_cell.CD4
T_cell.CD8
Neutrophil
Macrophage
DC
Group
Normal
Tumor
B
B_cell
T_cell.CD4
T_cell.CD8
Immune cell Abundance
0.3
0.2
0.1
0.0
Normal
stage i
stage ii
stage iii
stage iv
Group
Normal
stage i
stage ii
stage iii
stage iv
C
#Unique TCR CDR3
750
500
250
0
*
Tumor
Normal
D
#Unique BCR CDR3
8000
6000
4000
2000
0
*
Tumor
Normal
E
Expression
6
4
2
0
*
**
***
***
CD69
CD40LG
IL2RA
SELL
tissue
Normal
Tumor
F
#BCR Clusters
250
200
150
100
50
0
***
Tumor
Normal
G
BCR Clontypes per kilo reads
12.5
10.0
7.5
5.0
2.5
*
Tumor
Normal
H
BCR Clonality
0.100
0.075
0.050
0.025
0.000
**
Tumor
Normal
I
SHM rate
0.20
0.15
0.10
0.05
0.00
**
Tumor
Normal

Figure 3

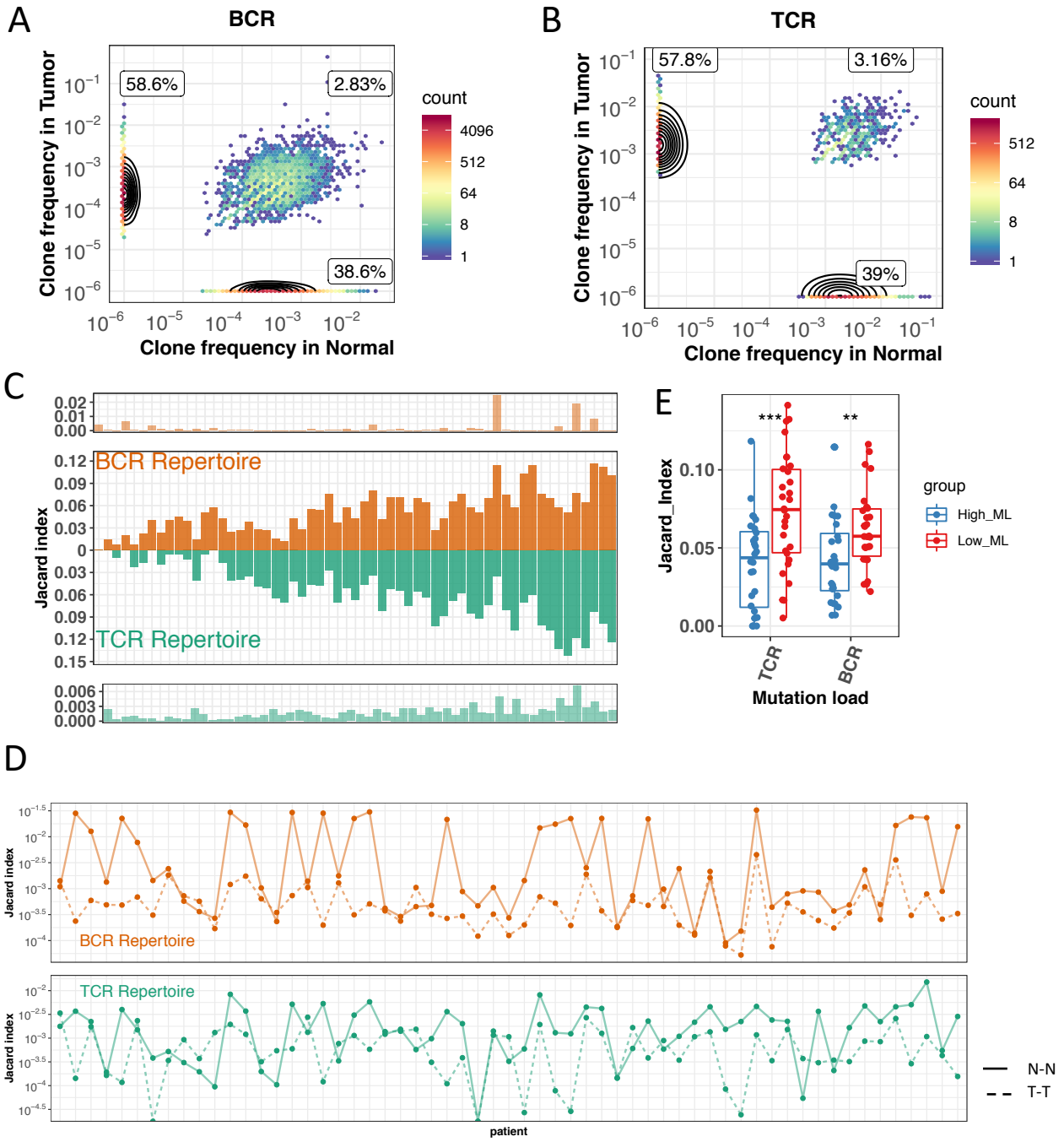

A
BCR
58.6%
2.83%
38.6%
Clone frequency in Tumor
Clone frequency in Normal
count
4096
512
64
8
1
B
TCR
57.8%
3.16%
39%
Clone frequency in Tumor
Clone frequency in Normal
count
512
64
8
1
C
BCR Repertoire
TCR Repertoire
Jacard index
D
BCR Repertoire
TCR Repertoire
Jacard index
patient
N-N
T-T
E
Jacard_Index
Mutation load
TCR
BCR
***
**
group
High_ML
Low_ML

Figure 4

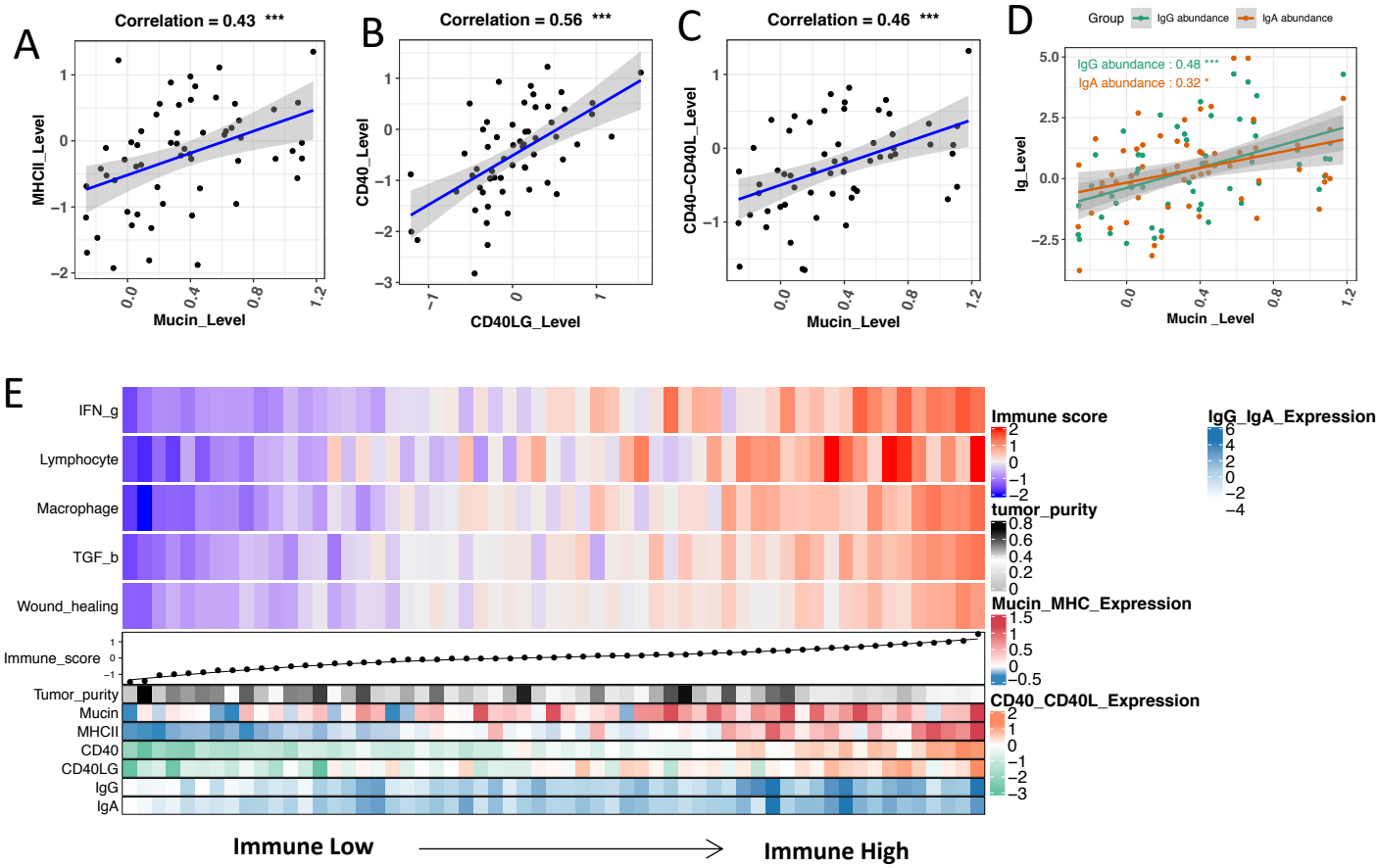

A
Correlation = 0.43 ***
MHCII_Level
Mucin_Level
B
Correlation = 0.56 ***
CD40_Level
CD40LG_Level
C
Correlation = 0.46 ***
CD40−CD40L_Level
Mucin_Level
D
Group
IgG abundance
IgA abundance
IgG abundance : 0.48 ***
IgA abundance : 0.32 *
Ig_Level
Mucin _Level
E
IFN_g
Lymphocyte
Macrophage
TGF_b
Wound_healing
Immune_score
Tumor_purity
Mucin
MHCII
CD40
CD40LG
IgG
IgA
Immune score
tumor_purity
Mucin_MHC_Expression
CD40_CD40L_Expression
IgG_IgA_Expression
Immune Low
Immune High

Figure 5

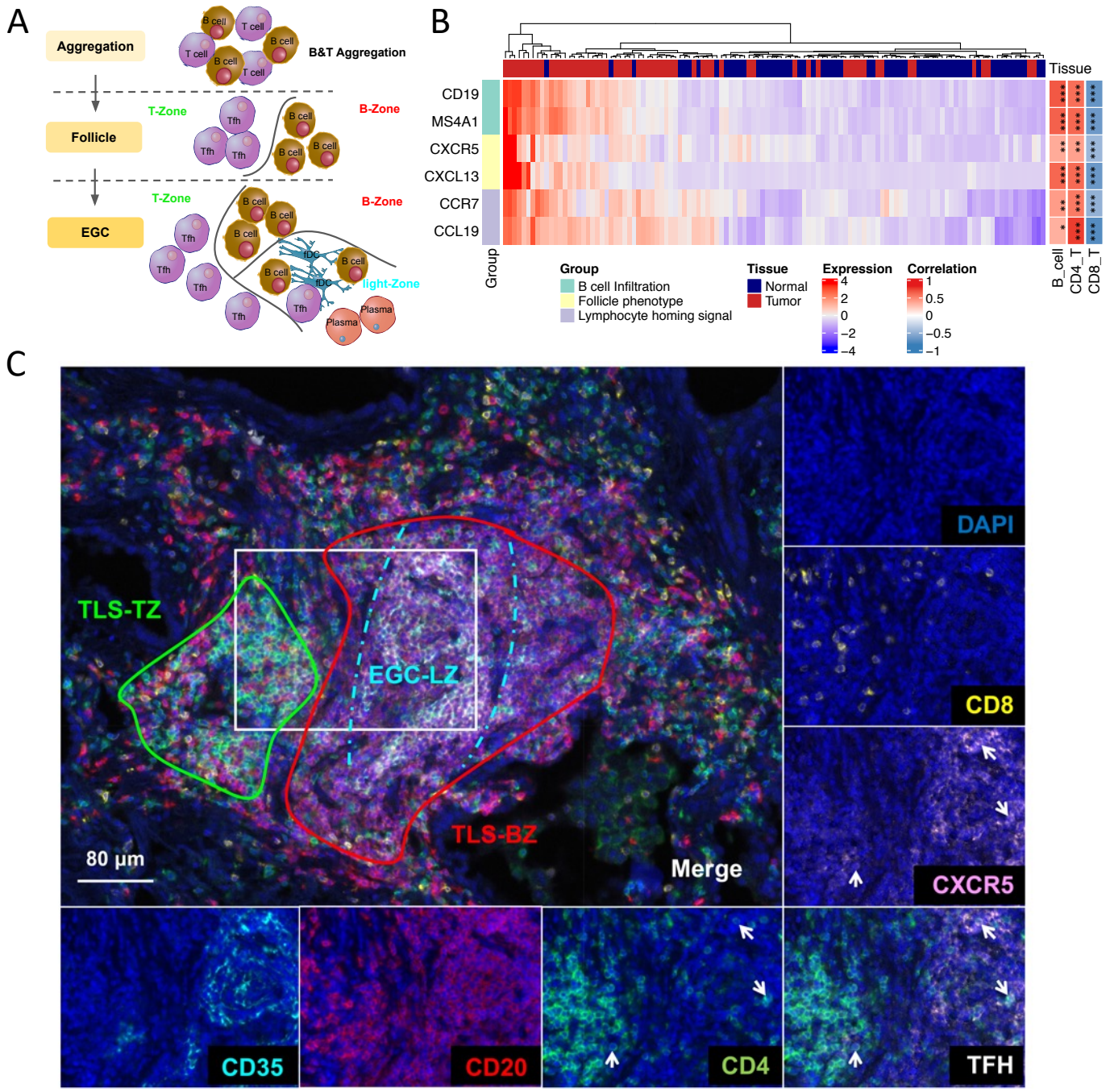

A
Aggregation
B&T Aggregation
T-Zone
B-Zone
Follicle
EGC
light-Zone
B cell
T cell
Tfh
fDC
Plasma
B
Tissue
CD19
MS4A1
CXCR5
CXCL13
CCR7
CCL19
Group
B_cell
CD4_T
CD8_T
Group
B cell Infiltration
Follicle phenotype
Lymphocyte homing signal
Tissue
Normal
Tumor
Expression
Correlation
C
TLS-TZ
EGC-LZ
TLS-BZ
80 μm
Merge
DAPI
CD8
CXCR5
CD35
CD20
CD4
TFH
D
p < 0.0001
CD20+ B density
Non-Fol
TLS
E
p < 0.0001
CD4+ T density
F
p < 0.0001
CD35+ FDC density
G
p < 0.0001
CD8+ T density

Figure 6

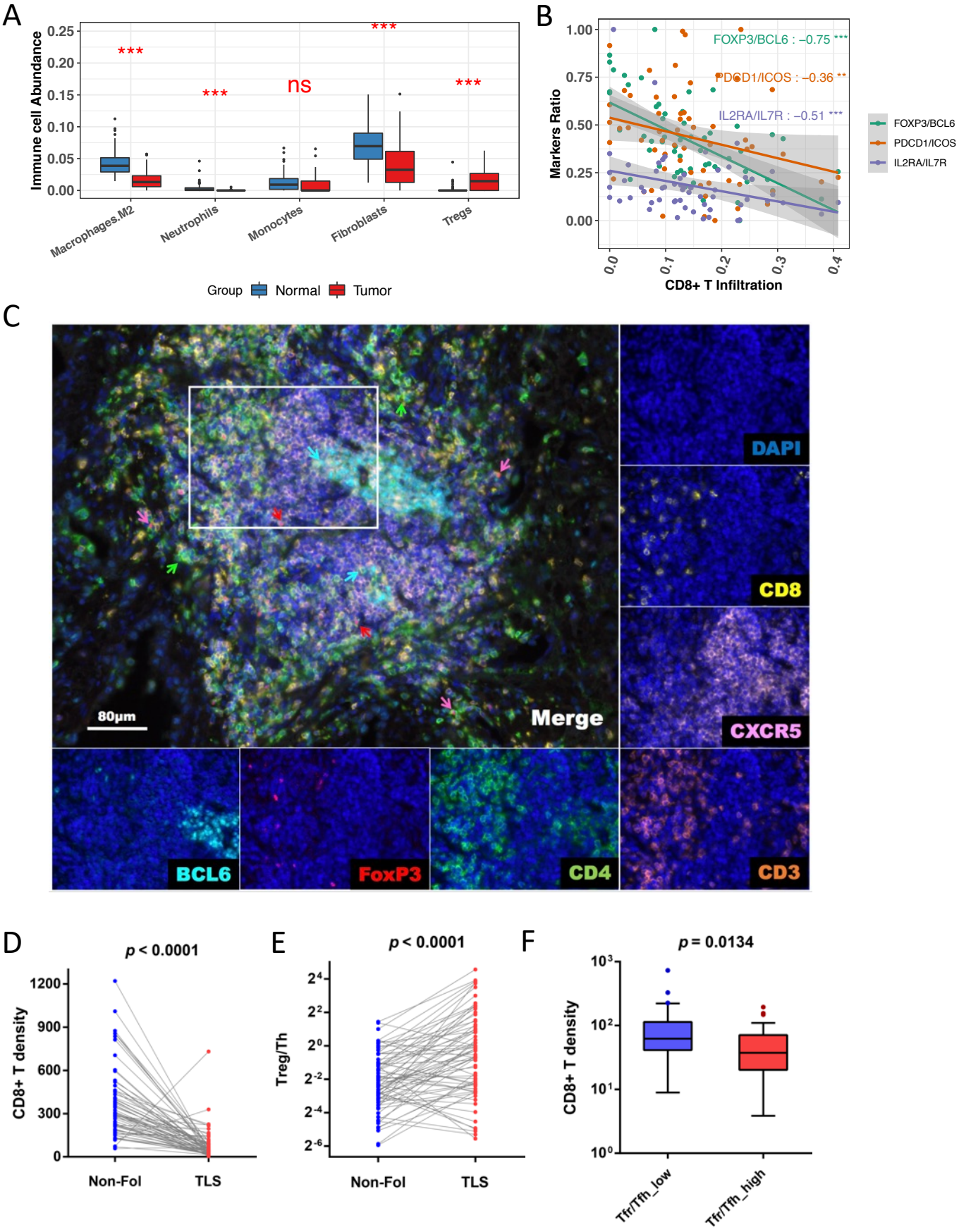

A
Immune cell Abundance
0.25
0.20
0.15
0.10
0.05
0.00
***
***
ns
***
***
Macrophages.M2
Neutrophils
Monocytes
Fibroblasts
Tregs
Group
Normal
Tumor
B
Markers Ratio
1.00
0.75
0.50
0.25
0.00
FOXP3/BCL6 : −0.75 ***
PDCD1/ICOS : −0.36 **
IL2RA/IL7R : −0.51 ***
FOXP3/BCL6
PDCD1/ICOS
IL2RA/IL7R
0.0
0.1
0.2
0.3
0.4
CD8+ T Infiltration
C
DAPI
CD8
CXCR5
CD3
Merge
80μm
BCL6
FoxP3
CD4
D
p < 0.0001
CD8+ T density
1200
900
600
300
0
Non-Fol
TLS
E
p < 0.0001
Treg/Th
2^4
2^2
2^0
2^-2
2^-4
2^-6
Non-Fol
TLS
F
p = 0.0134
CD8+ T density
10^3
10^2
10^1
10^0
Tfr/Tfh_low
Tfr/Tfh_high

Figure S1

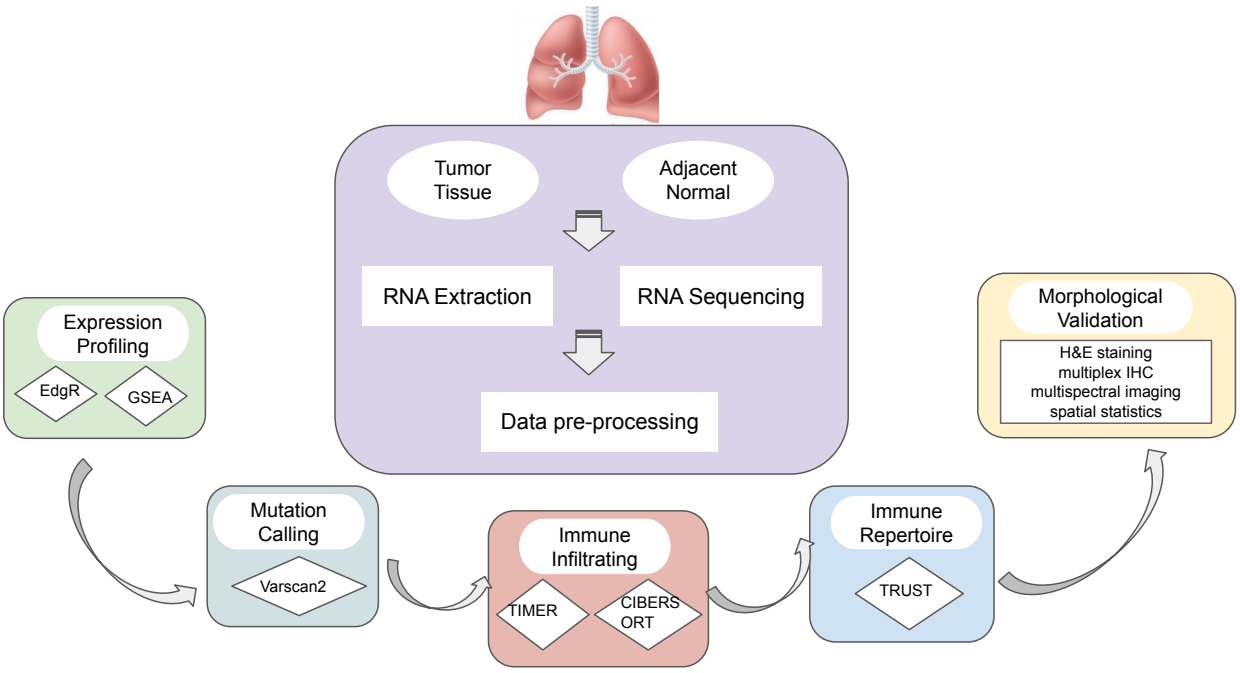

Tumor Tissue
Adjacent Normal
RNA Extraction
RNA Sequencing
Data pre-processing
Expression Profiling
EdgR
GSEA
Morphological Validation
H&E staining
multiplex IHC
multispectral imaging
spatial statistics
Mutation Calling
Varscan2
Immune Infiltrating
TIMER
CIBERSORT
Immune Repertoire
TRUST

Figure S2

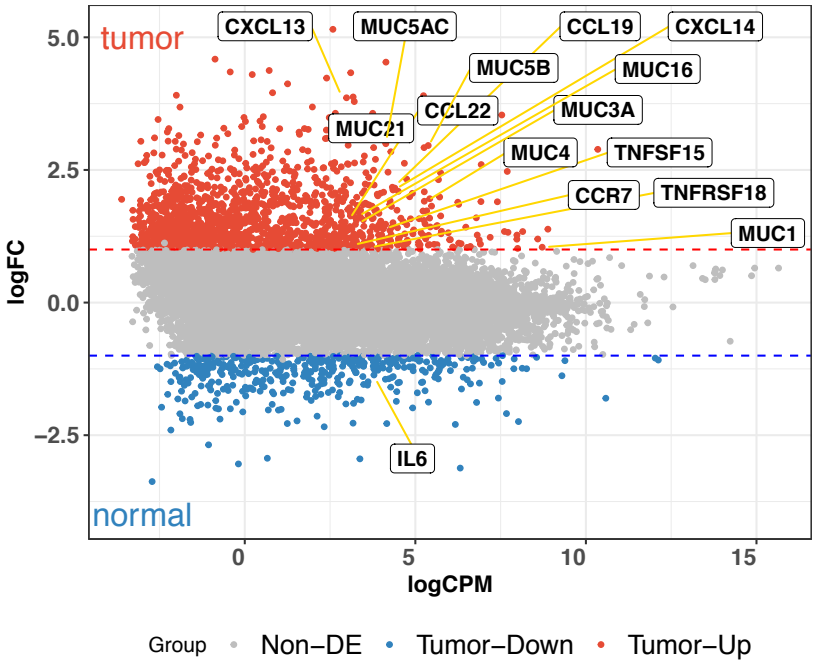

tumor
normal
CXCL13
MUC5AC
CCL19
CXCL14
MUC5B
MUC16
CCL22
MUC3A
MUC21
MUC4
TNFSF15
CCR7
TNFRSF18
MUC1
IL6
5.0
2.5
0.0
−2.5
logFC
0
5
10
15
logCPM
Group
Non−DE
Tumor−Down
Tumor−Up

Figure S3

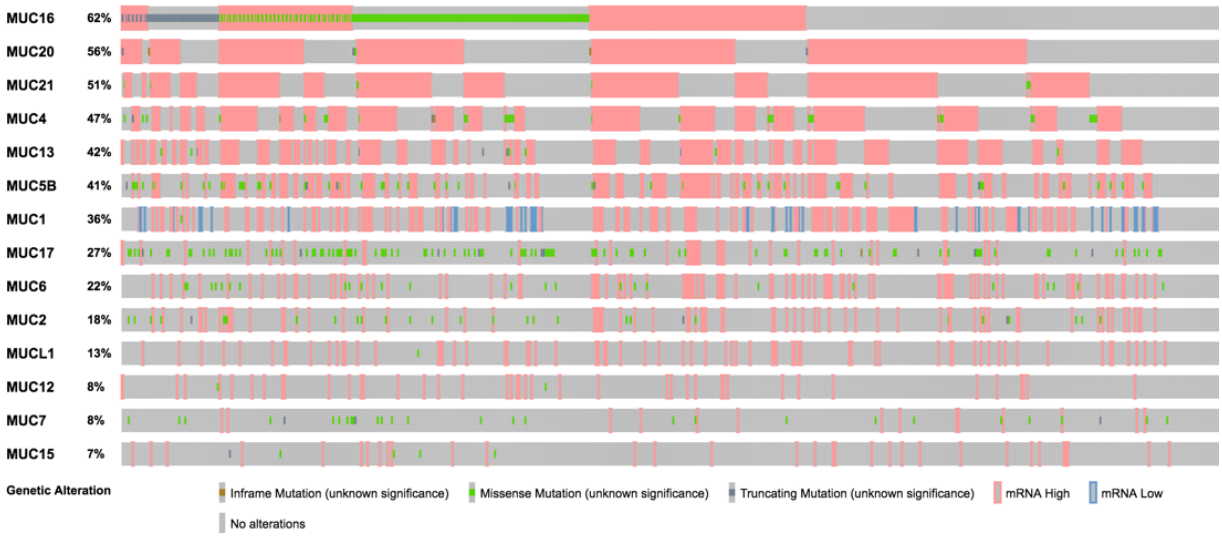

MUC16 62%
MUC20 56%
MUC21 51%
MUC4 47%
MUC13 42%
MUC5B 41%
MUC1 36%
MUC17 27%
MUC6 22%
MUC2 18%
MUCL1 13%
MUC12 8%
MUC7 8%
MUC15 7%
Genetic Alteration
Inframe Mutation (unknown significance)
Missense Mutation (unknown significance)
Truncating Mutation (unknown significance)
mRNA High
mRNA Low
No alterations

Figure S4

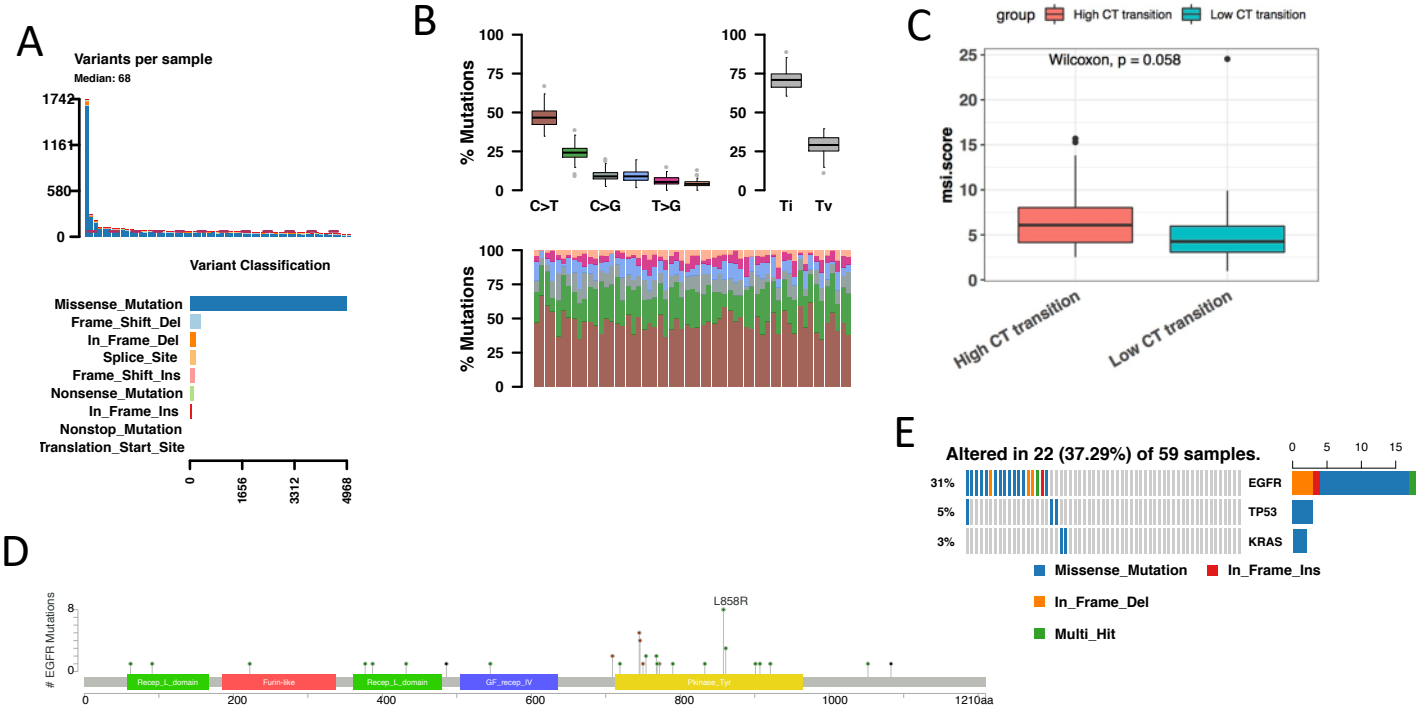
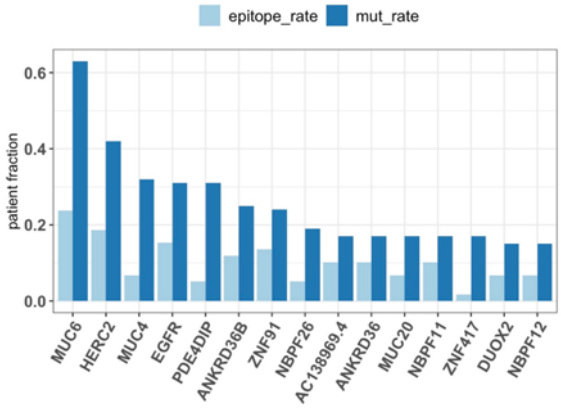

A
Variants per sample
Median: 68
1742
1161
580
0
Variant Classification
Missense_Mutation
Frame_Shift_Del
In_Frame_Del
Splice_Site
Frame_Shift_Ins
Nonsense_Mutation
In_Frame_Ins
Nonstop_Mutation
Translation_Start_Site
0
1656
3312
4968
B
% Mutations
C>T
C>G
T>G
Ti
Tv
C
group
High CT transition
Low CT transition
Wilcoxon, p = 0.058
msi.score
High CT transition
Low CT transition
E
Altered in 22 (37.29%) of 59 samples.
31%
5%
3%
EGFR
TP53
KRAS
Missense_Mutation
In_Frame_Ins
In_Frame_Del
Multi_Hit
D
EGFR Mutations
L858R
Recep_L_domain
Furin-like
Recep_L_domain
GF_recep_IV
Pkinase_Tyr
1210aa

Figure S5

A


epitope_rate
mut_rate
patient fraction
0.6
0.4
0.2
0.0
MUC6
HERC2
MUC4
EGFR
PDE4DIP
ANKRD36B
ZNF91
NBPF26
AC138969.4
ANKRD36
MUC20
NBPF11
ZNF417
DUOX2
NBPF12

B

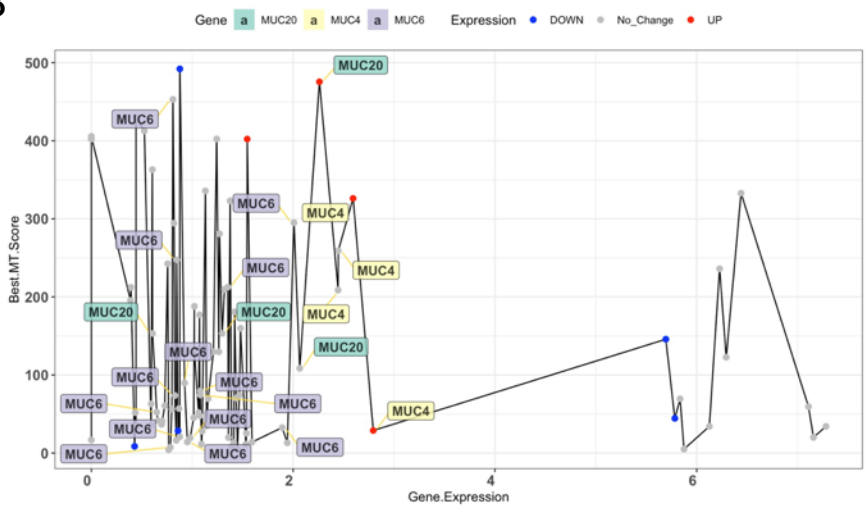

Gene
MUC20
MUC4
MUC6
Expression
DOWN
No_Change
UP
Best.MT.Score
Gene.Expression
500
400
300
200
100
0
0
2
4
6

Figure S6

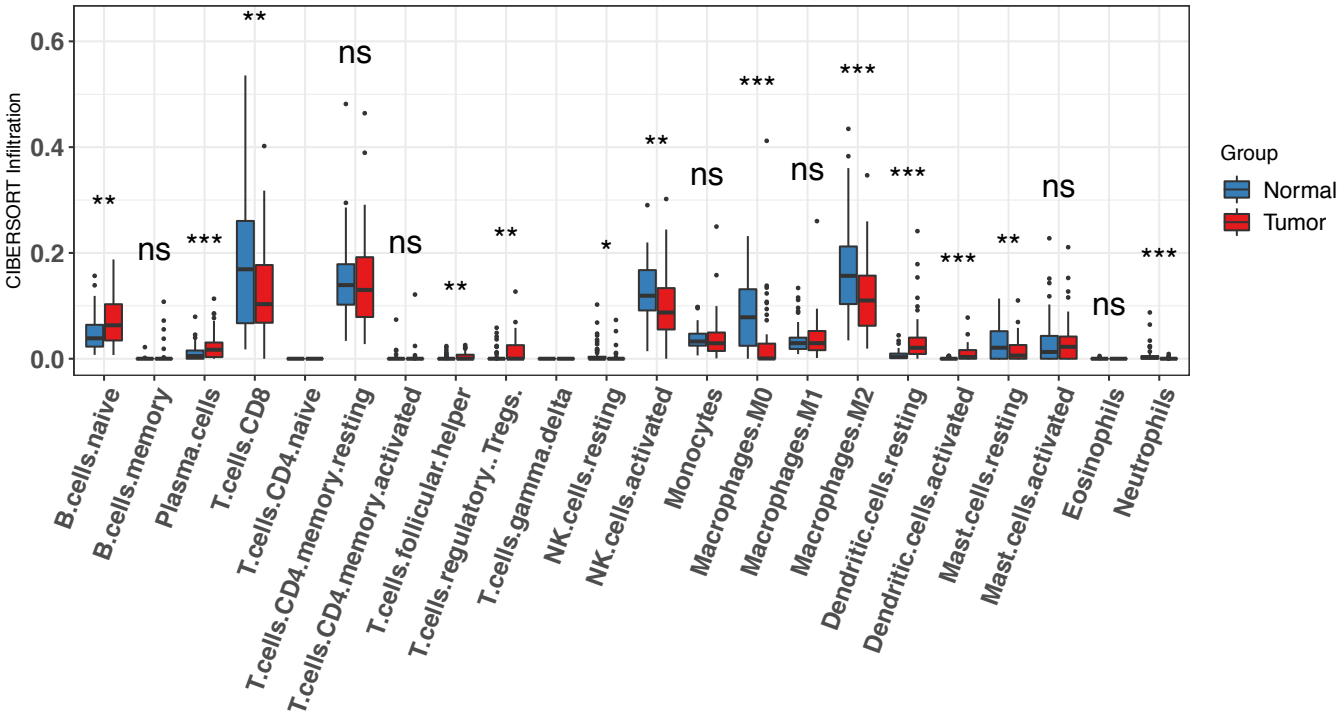

CIBERSORT Infiltration
0.6
0.4
0.2
0.0
**
ns
***
**
ns
ns
**
**
*
**
ns
***
ns
***
***
**
ns
ns
***
B.cells.naive
B.cells.memory
Plasma.cells
T.cells.CD8
T.cells.CD4.naive
T.cells.CD4.memory.resting
T.cells.CD4.memory.activated
T.cells.follicular.helper
T.cells.regulatory..Tregs.
T.cells.gamma.delta
NK.cells.resting
NK.cells.activated
Monocytes
Macrophages.M0
Macrophages.M1
Macrophages.M2
Dendritic.cells.resting
Dendritic.cells.activated
Mast.cells.resting
Mast.cells.activated
Eosinophils
Neutrophils
Group
Normal
Tumor

Figure S7

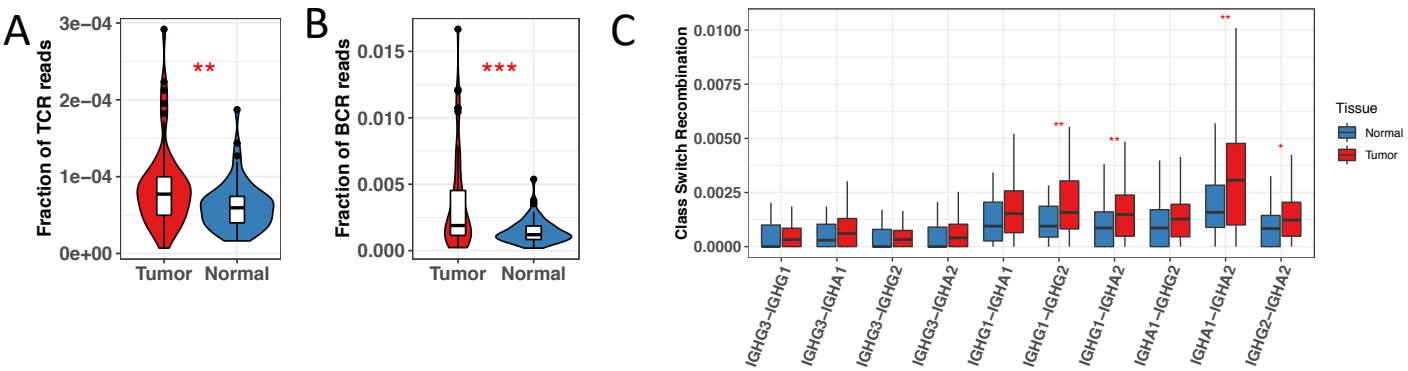

A
Fraction of TCR reads
3e−04
2e−04
1e−04
0e+00
**
Tumor
Normal
B
Fraction of BCR reads
0.015
0.010
0.005
0.000
***
Tumor
Normal
C
Class Switch Recombination
0.0100
0.0075
0.0050
0.0025
0.0000
IGHG3–IGHG1
IGHG3–IGHA1
IGHG3–IGHG2
IGHG3–IGHA2
IGHG1–IGHA1
IGHG1–IGHG2
IGHG1–IGHA2
IGHA1–IGHG2
IGHA1–IGHA2
IGHG2–IGHA2
Tissue
Normal
Tumor

Figure S8

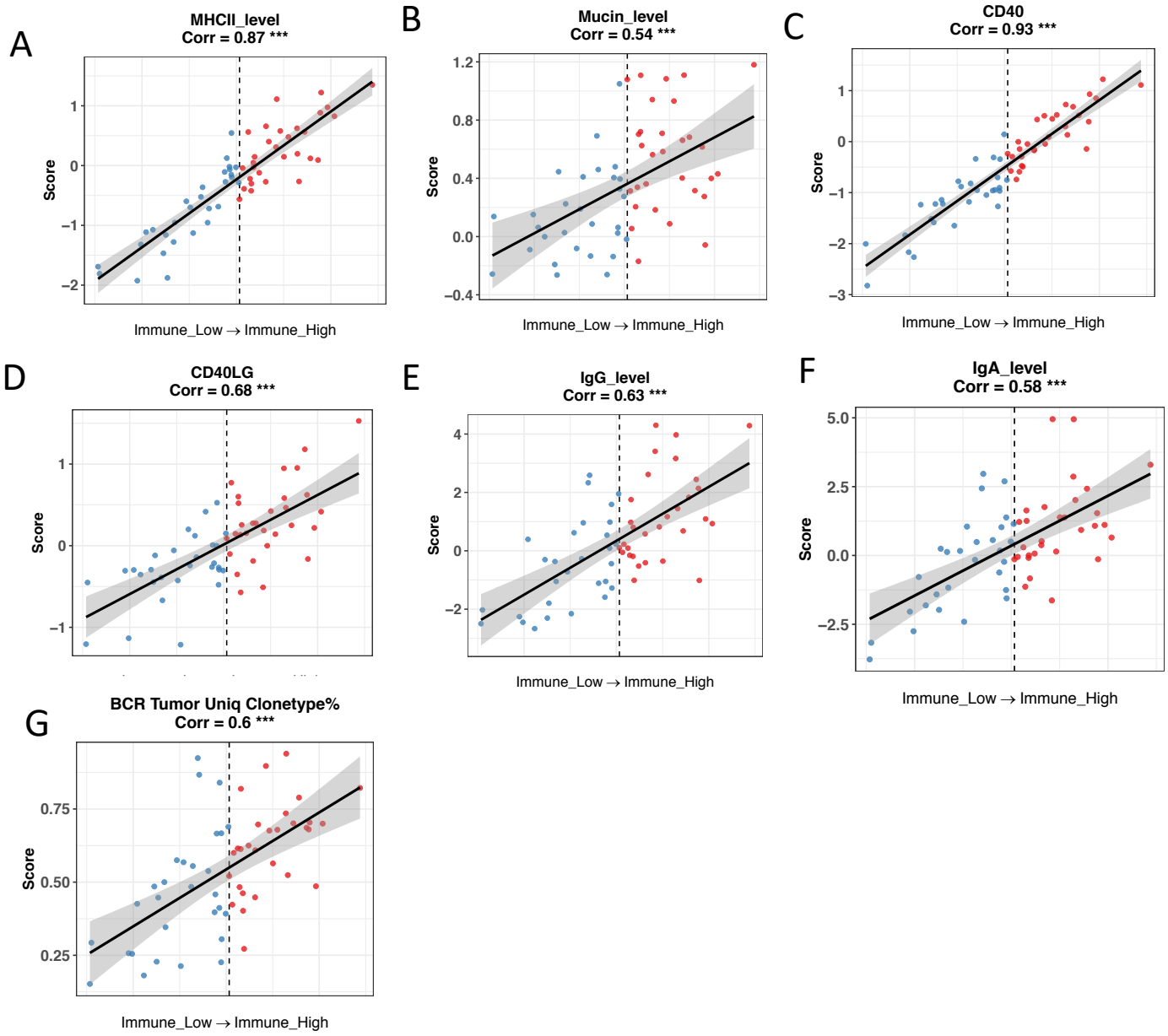

A
MHCII_level
Corr = 0.87 ***
Score
Immune_Low → Immune_High
B
Mucin_level
Corr = 0.54 ***
Score
Immune_Low → Immune_High
C
CD40
Corr = 0.93 ***
Score
Immune_Low → Immune_High
D
CD40LG
Corr = 0.68 ***
Score
E
IgG_level
Corr = 0.63 ***
Score
Immune_Low → Immune_High
F
IgA_level
Corr = 0.58 ***
Score
Immune_Low → Immune_High
G
BCR Tumor Uniq Clonetype%
Corr = 0.6 ***
Score
Immune_Low → Immune_High

Figure S9

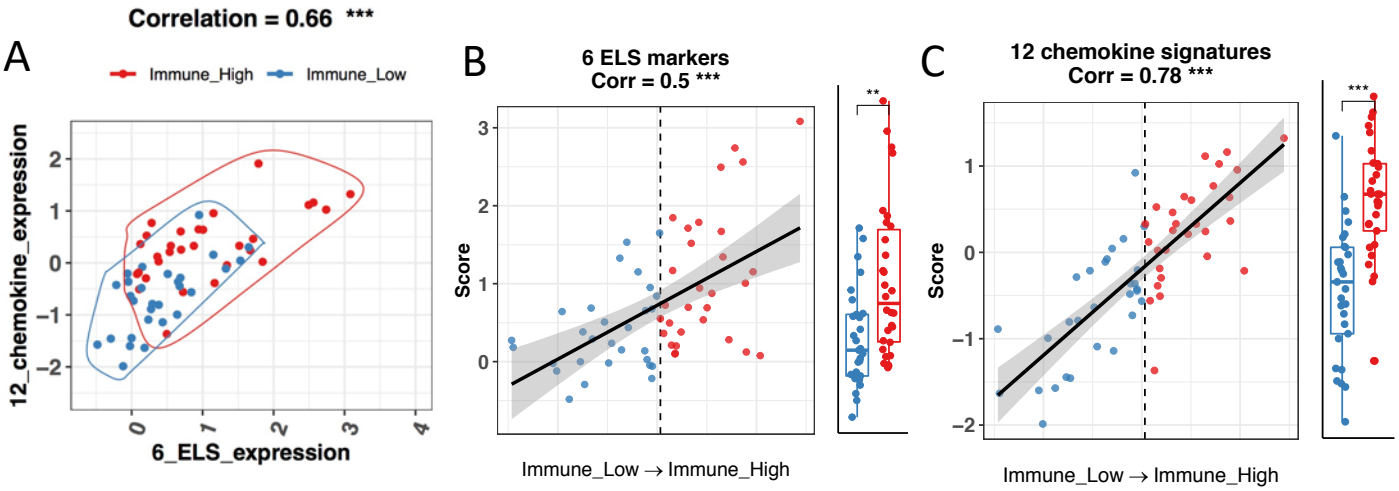
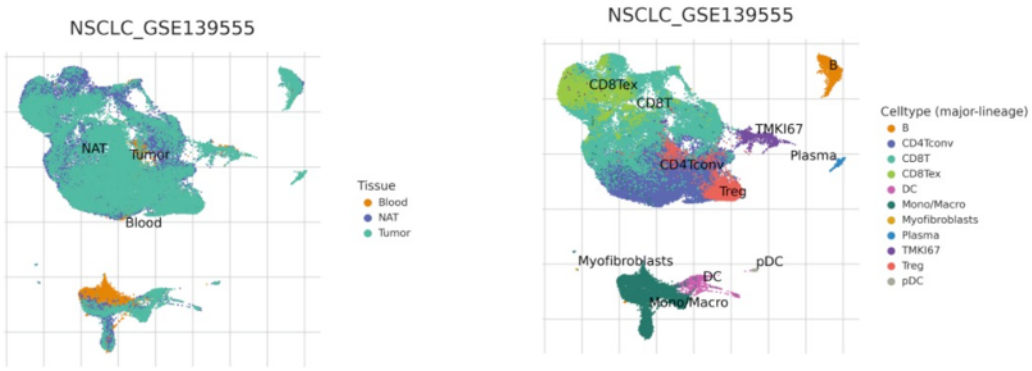

A
Correlation = 0.66 ***
Immune_High
Immune_Low
12_chemokine_expression
6_ELS_expression
B
6 ELS markers
Corr = 0.5 ***
Score
Immune_Low → Immune_High
**
C
12 chemokine signatures
Corr = 0.78 ***
Score
Immune_Low → Immune_High
***

Figure S10

A



B

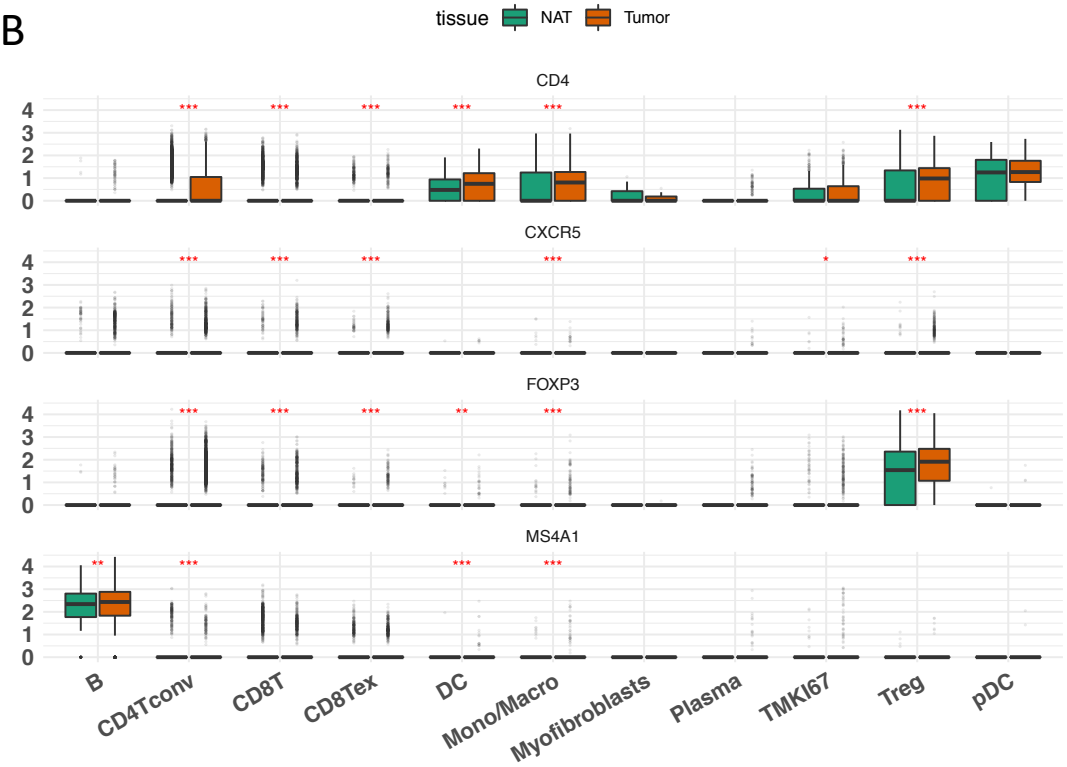
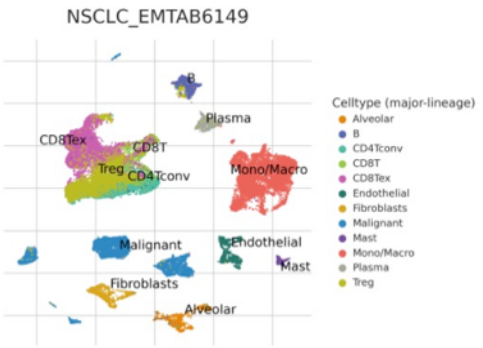
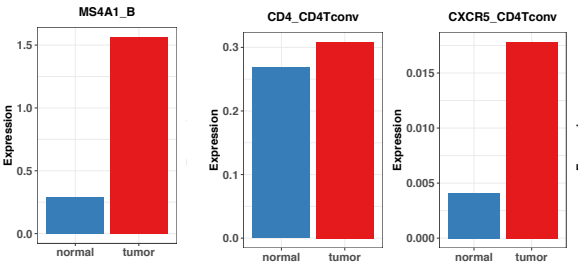


Figure S11

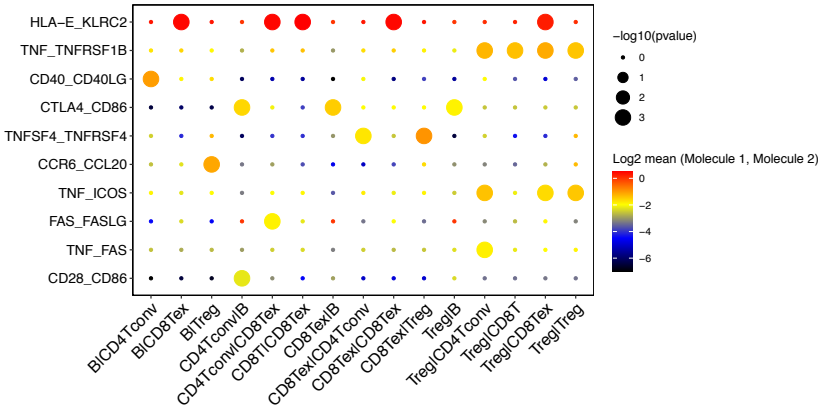
A



B



C



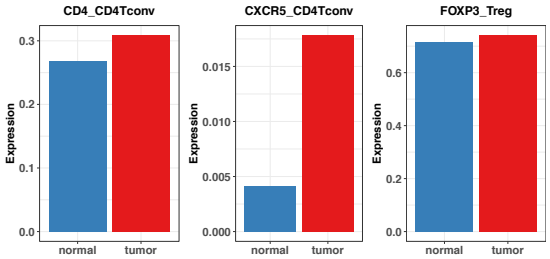


Figure S12

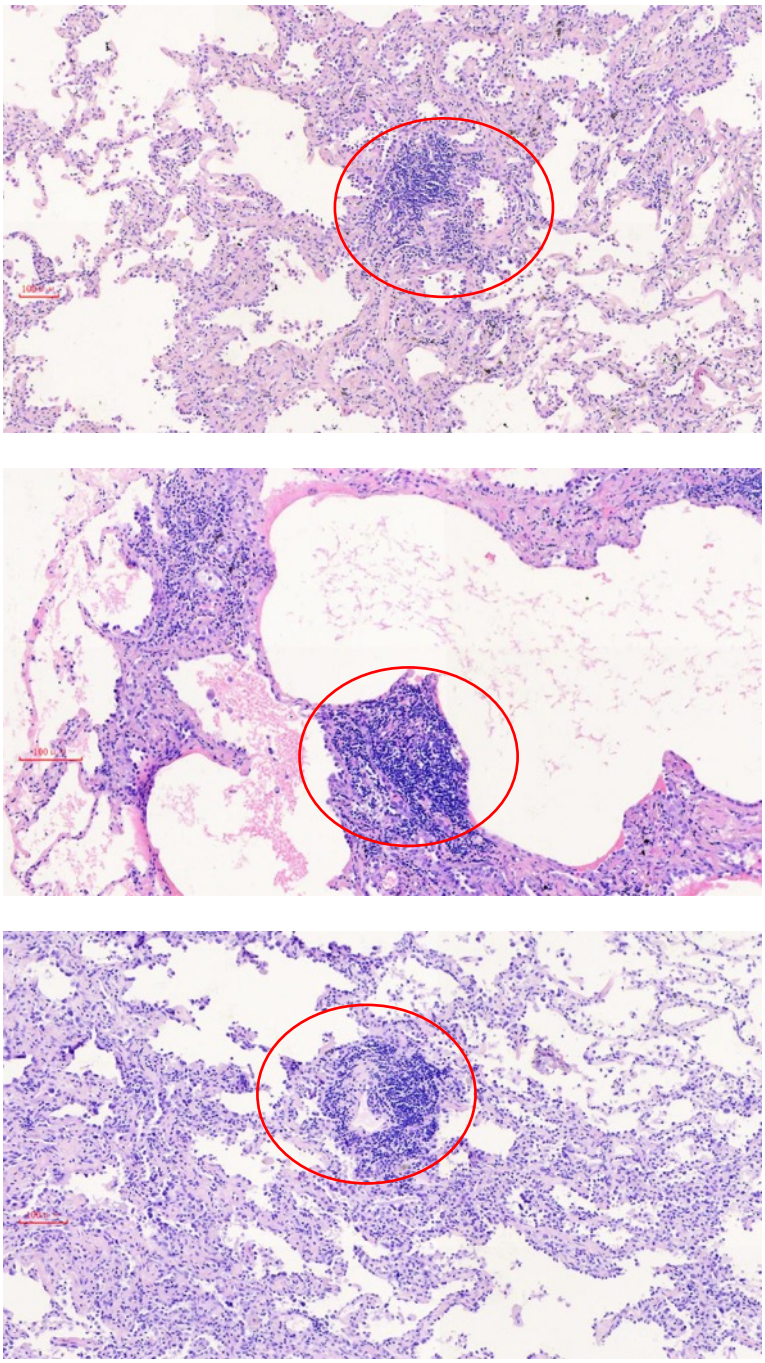


Figure S13

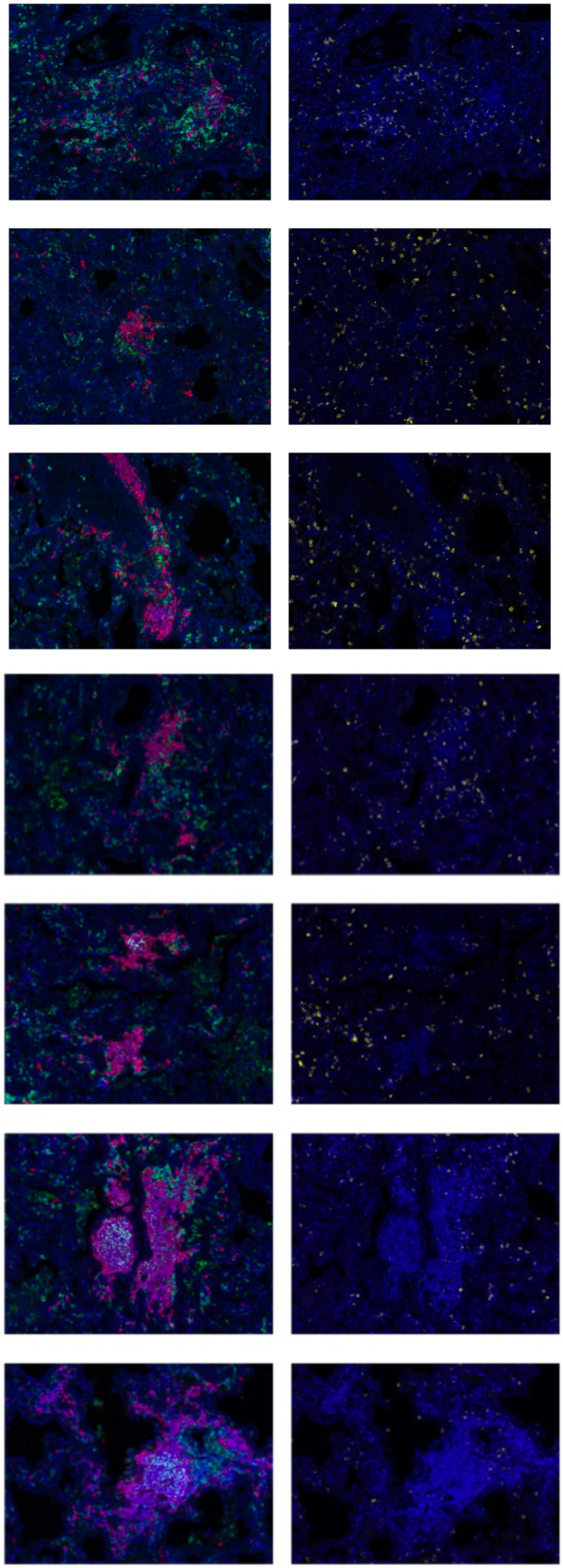
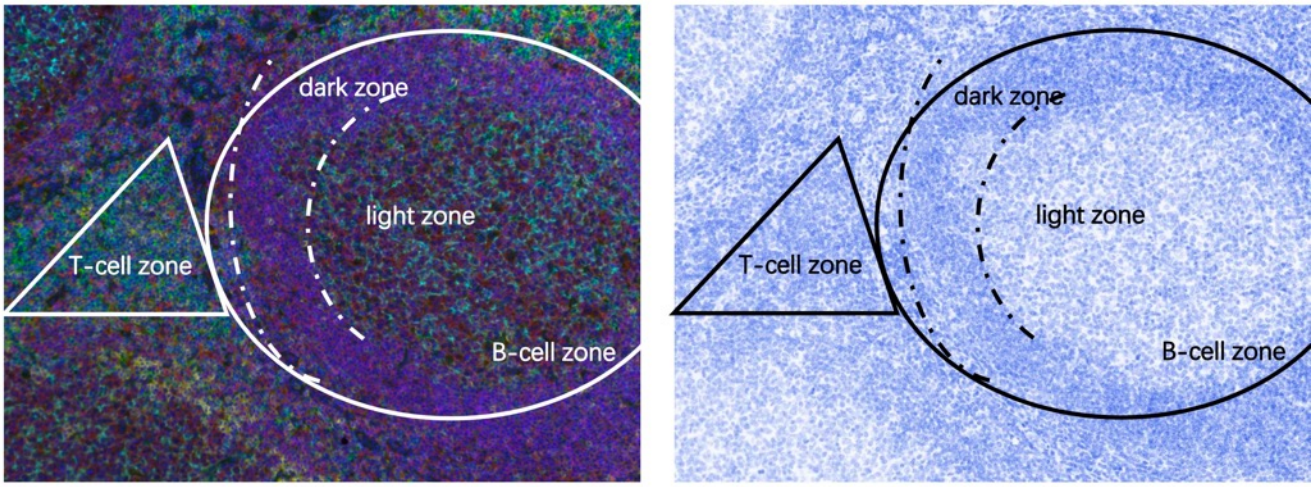


Figure S14

A



B

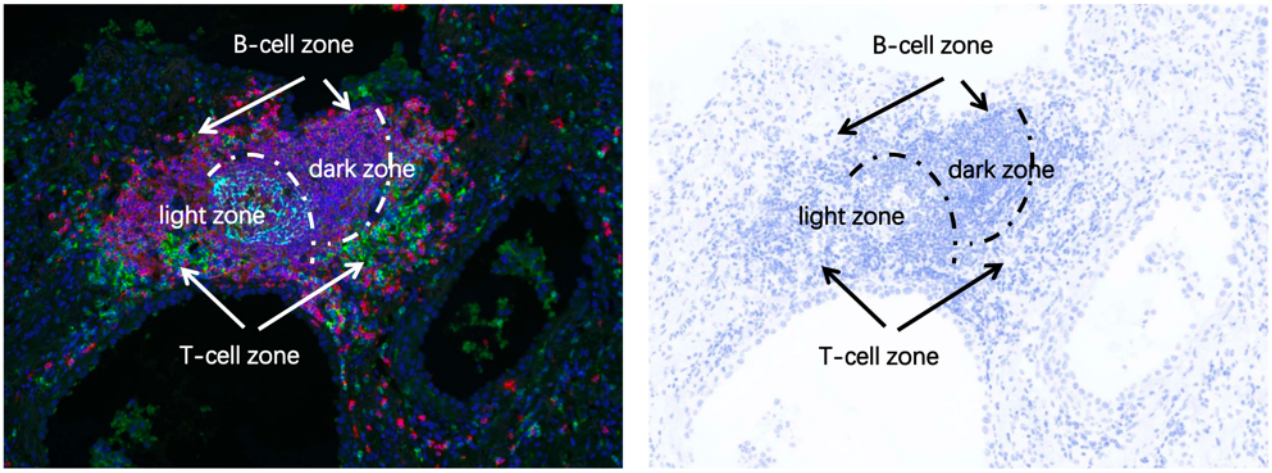
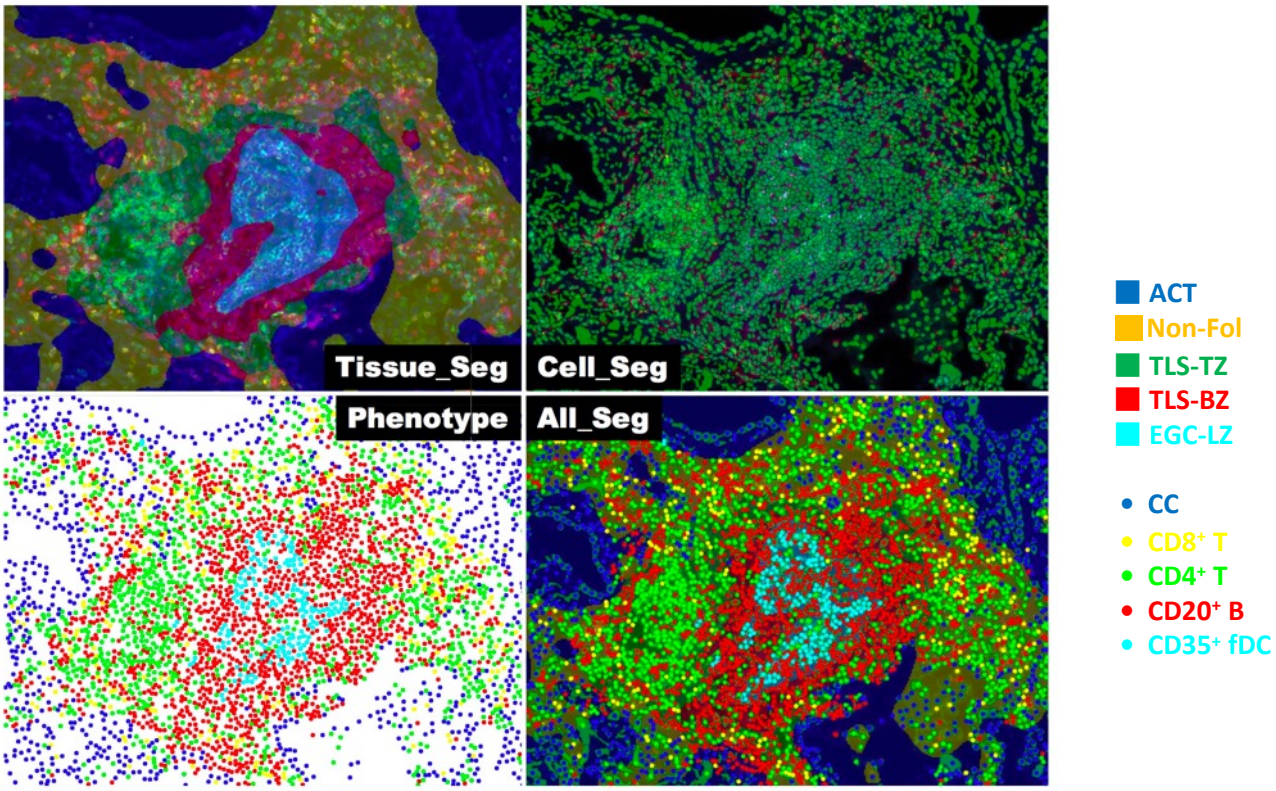


Figure S15

A



B

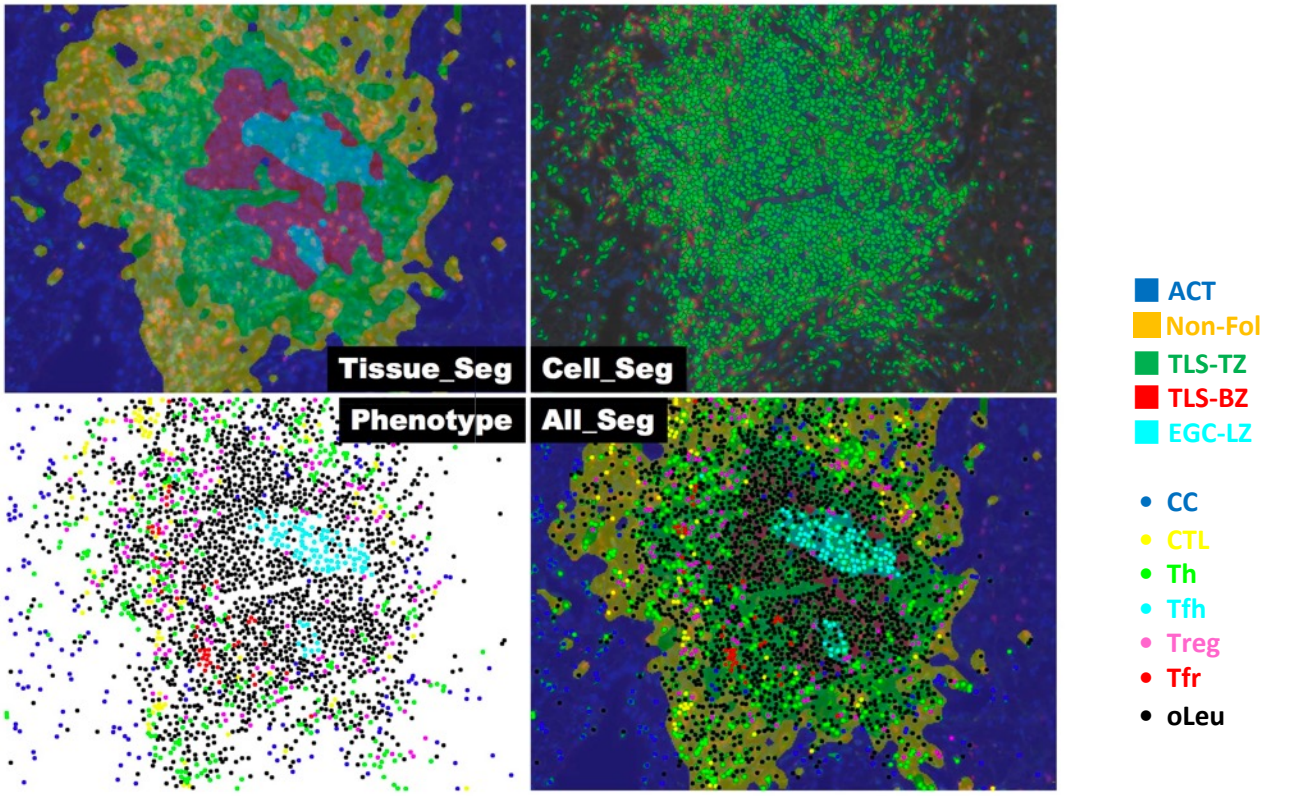


Figure S16

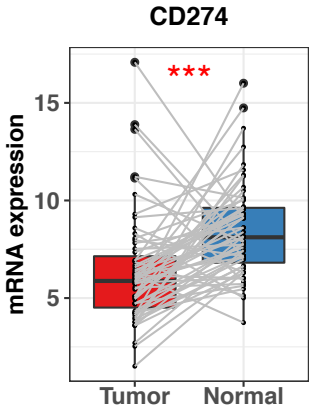
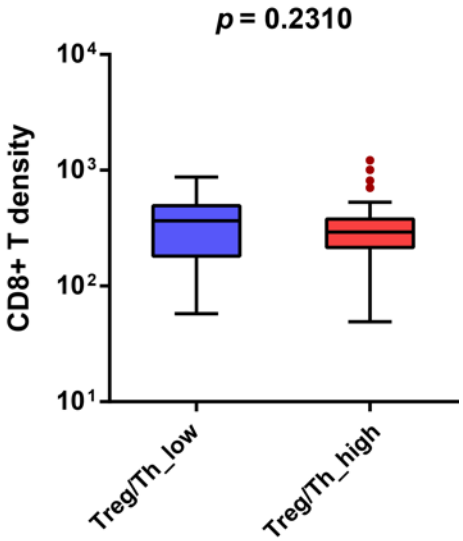



Figure S17