Prognostic and therapeutic potential of gene profiles related to tertiary lymphoid structures in colorectal cancer

Jinglu Yu [1] , **Yabin Gong** [2] , **Xiaowei Huang** [3] , **Yufang Bao** [Corresp. 4]

[1] PuDong Traditional Chinese Medicine Hospital, Shanghai University of Traditional Chinese Medicine, Shanghai, China, Shanghai, Pudong New Area, China

[2] Department of Oncology,, Yueyang Hospital of Integrated Traditional Chinese and Western Medicine, Shanghai University of Traditional Chinese Medicine, Shanghai, China., Shanghai, China

[3] Department of Oncology, Longhua Hospital, Shanghai University of Traditional Chinese Medicine, Shanghai, China, Shanghai, China

[4] PuDong Traditional Chinese Medicine Hospital, Shanghai University of Traditional Chinese Medicine, Shanghai, China., Shanghai, Pudong New Area, China

Corresponding Author: Yufang Bao
Email address: YufangBaosarly@126.com

The role of Tertiary Lymphoid Structures (TLS) in oncology is gaining interest, particularly in colorectal carcinoma, yet a thorough analysis remains elusive. This study pioneered a novel TLS quantification system for prognostic and therapeutic response prediction in colorectal carcinoma, alongside a comprehensive depiction of the TLS landscape. Utilizing single-cell sequencing, we established a TLS score within the Tumor Immune Microenvironment (TIME). Analysis of Tertiary Lymphoid Structure-related Genes (TLSRGs) in 1184 patients with Colon Adenocarcinoma/Rectum Adenocarcinoma (COADREAD) from TCGA and GEO databases led to the identification of two distinct molecular subtypes. Differentially Expressed Genes (DEGs) further segregated these patients into gene subtypes. A TLS score was formulated using Gene Set Variation Analysis (GSVA) and its efficacy in predicting immunotherapy outcomes was validated in two independent cohorts. High-scoring patients exhibited a 'hot' immune phenotype, correlating with enhanced immunotherapy efficacy. Key genes in our model, including *C5AR1*, *APOE*, *CYR1P1*, and *SPP1*, were implicated in COADREAD cell proliferation, invasion, and *PD-L1* expression. These insights offer a novel approach to colorectal carcinoma treatment, emphasizing TLS targeting as a potential anti-tumor strategy.

1

Prognostic and therapeutic potential of gene profiles related tertiary lymphoid structures in colorectal cancer

5

6 Jinglu Yu [1], Yabin Gong [2], Xiaowei Huang [2], Yufang Bao [1] *

7 [1] PuDong Traditional Chinese Medicine Hospital, Shanghai University of Traditional Chinese
8 Medicine, Shanghai, China.
9 [2] Department of Oncology, Yueyang Hospital of Integrated Traditional Chinese and Western
10 Medicine, Shanghai University of Traditional Chinese Medicine, Shanghai, China.
11 [3] Department of Oncology, Longhua Hospital, Shanghai University of Traditional Chinese
12 Medicine, Shanghai, China

13

14 Corresponding Author:
15 Yufang Bao [1]
16 No. 399 Pingchuan Road, Pudong New Area, Shanghai, 201299, China
17 Email address: YufangBaosarly@126.com

18

Abstract

20 The role of Tertiary Lymphoid Structures (TLS) in oncology is gaining interest, particularly in
21 colorectal carcinoma, yet a thorough analysis remains elusive. This study pioneered a novel TLS
22 quantification system for prognostic and therapeutic response prediction in colorectal carcinoma,
23 alongside a comprehensive depiction of the TLS landscape. Utilizing single-cell sequencing, we
24 established a TLS score within the Tumor Immune Microenvironment (TIME). Analysis of
25 Tertiary Lymphoid Structure-related Genes (TLSRGs) in 1184 patients with Colon
26 Adenocarcinoma/Rectum Adenocarcinoma (COADREAD) from TCGA and GEO databases led
27 to the identification of two distinct molecular subtypes. Differentially Expressed Genes (DEGs)
28 further segregated these patients into gene subtypes. A TLS score was formulated using Gene Set
29 Variation Analysis (GSVA) and its efficacy in predicting immunotherapy outcomes was
30 validated in two independent cohorts. High-scoring patients exhibited a 'hot' immune phenotype,
31 correlating with enhanced immunotherapy efficacy. Key genes in our model, including *C5AR1*,
32 *APOE*, *CYR1P1*, and *SPP1*, were implicated in COADREAD cell proliferation, invasion, and
33 *PD-L1* expression. These insights offer a novel approach to colorectal carcinoma treatment,
34 emphasizing TLS targeting as a potential anti-tumor strategy.

35

Introduction

37 Colorectal cancer, a predominant gastrointestinal malignancy, ranks third in cancer-related
38 mortality globally. The survival rates at five years hover around 65%[1,2]. Surgical removal is the

39 cornerstone for treating early-stage colorectal cancer, while advanced stages often benefit from

40 an integrated approach of chemotherapy and targeted therapies[3]. The advent of immune

41 checkpoint inhibitors (ICIs) has marked a revolutionary shift in oncological therapeutics,

42 offering significant clinical advantages for patients[4,5]. Yet, the response to ICIs is limited to a

43 small patient subset[6,7], underscoring the urgent need for biomarker research. This research is

44 essential for refining patient selection and developing strategies to overcome immune resistance.

45 Historically, biomarker investigations have predominantly relied on RNA sequencing (RNA-

46 Seq) from whole tumor tissues[8], providing a collective genetic snapshot of diverse cell

47 populations. Unfortunately, biomarkers identified through this method have demonstrated

48 restricted predictive capability. The emergence of single-cell RNA sequencing (scRNA-Seq) has

49 revolutionized this landscape, facilitating gene expression analysis at an individual cell level and

50 paving the way for the discovery of more effective biomarkers[9].

51 TLS are spontaneously formed ectopic lymphoid formations that arise in chronic inflammation

52 sites, including tumor environments. These structures, which bear a structural similarity to

53 secondary lymphoid organs, are chiefly comprised of an array of immune cells - B cells, T cells,

54 dendritic cells, neutrophils, and macrophages[10]. Additionally, TLSs encompass high endothelial

55 venules and lymphatic vessels, crucial for directing immune cell movement into TIME. Notably,

56 the presence of TLSs has been associated with improved prognosis in a range of solid tumors,

57 such as melanoma[11], renal cell carcinoma[12], and colorectal cancer[13]. Furthermore, emerging

58 studies highlight a significant correlation between TLSs and the efficacy of immunotherapy[14].

59 The existence of TLSs has been identified as a predictive biomarker for the response to ICI

60 therapy in advanced-stage cancers, including bladder cancer[15] and head and neck squamous cell

61 carcinoma[16]. Consequently, the induction of TLSs is increasingly being explored as a novel

62 therapeutic approach in oncology [17].

63 Additionally, numerous investigations have demonstrated that prevalent anti-cancer therapies

64 are capable of prompting intratumoral TLS formation in murine models. Chelvanambi utilized

65 the STING agonist ADU S-100 in treating B16.F10 melanomas, led to STING activation within

66 the TIME, correlating with the reduction of melanoma progression and concurrent development

67 of TLSs[18]. Lee JW employed Stromal Vascular Fraction Spheroid-Based Immunotherapy, which

68 resulted in the formation TLS-like structures, thereby enhancing antigen-specific immune

69 responses and anti-tumor immunity in mice[19]. While immunohistochemistry remains the

70 conventional method for TLS detection[20], there's an escalating need for more definitive

71 approaches to assess TLS levels through transcriptomic analysis of tumor biopsies or excised

72 samples. A gene signature encompassing *CCL2*, *CCL4*, *CCL5*, *CCL8*, *CXCL9*, *CXCL10*,

73 *CXCL11*, *CXCL13*, *CCL18*, *CCL19*, and *CCL21* has been formulated and employed to classify

74 tumors as either TLS+ or TLS-[21]. Tumors categorized as TLS+ have shown a reduced risk of

75 early recurrence in hepatocellular carcinoma[21]. This has led to our initiative to establish a refined

76 TLS signature for COADREAD, aiming to analyze TLS neogenesis through a multi-omics lens

77 and thereby enhance the scope of future clinical research.

78   Nevertheless, the intricacies of how TLSs interact with COADREAD, as well as the interplay
79 among immune cells within TLSs, especially on the single cell level, remain enigmatic.
80 Moreover, the contribution of TLSs in modulating responses to immune therapy is yet to be fully
81 elucidated.
82   In this research endeavor, our objective was to amalgamate single-cell and bulk RNA
83 sequencing data to elucidate the role of TLSs, particularly their potential as biomarkers for
84 predicting ICI outcomes in patients with COADREAD. Additionally, our study sought to unravel
85 the molecular and immune characteristics of TLSs, assessing their impact on the prognosis for
86 COADREAD patients. To this end, we have devised TLS Score, which could serve as an
87 invaluable tool in the realm of personalized precision medicine, aiding clinicians in tailoring
88 treatment strategies.
89

Materials & Methods

90

91 **Data Sources and Analysis Workflow**
92 The study's analytical process was depicted in Figure 1. We obtained the CRC_EMTAB8107
93 scRNA-seq dataset for COADREAD from the GEO Database
94 (https://www.ncbi.nlm.nih.gov/geo/)[22], which includes 7 instances of colorectal carcinoma.
95 Furthermore, the TCGA database (https://portal.gdc.cancer.gov) was used to obtain the bulk
96 RNA-seq data and clinical profiles of 380 COADREAD patients[23]. We also sourced datasets
97 GSE17538 (n =238) and GSE39582 (n =566) from the GEO repository.
98 **Single-Cell Dataset Processing and Cluster Identification**
99 For the scRNA-seq dataset, we employed Seurat (v4.1.1) in the R programming environment[24].
100 Initial steps included quality assessment using the "Seurat" package, leading to the exclusion of
101 cells not meeting specific criteria: (1) nFeature_RNA range of 200 to 7500; (2) nCount_RNA
102 range of 200 to 35,000; (3) a mitochondrial gene content of $\geq$10%. This filtration yielded a
103 dataset comprising 66,050 cells and 21,753 genes. Following actions involved batch
104 normalization using the 'harmony' package and data scaling with the 'ScaleData' function. The
105 top 28 principal components were extracted by performing Principal Component Analysis (PCA)
106 on the 2,000 genes with the highest variability. UMAP was used for unsupervised clustering to
107 visualize clusters on a two-dimensional plane[25]. The function 'FindAllMarkers' was used to
108 evaluate gene expression variations among clusters, considering the criteria of |log2 (fold
109 change)| > 1, an adjusted p-value < 0.05, and a resolution of 0.5. Cell subpopulations were
110 annotated employing the "SingleR" package, CellMarker database[26], and PanglaoDB database[27].
111   Following this, the "FindAllMarkers" function in the R environment was employed to identify
112 genes differentially expressed across the various cell types. We used the GSVA package (version
113 1.40.1) to calculate GSVA scores for 50 hallmark gene sets obtained from the Molecular
114 Signatures Database (MSigDB) (https://www.gsea-msigdb.org/gsea/msigdb) [28]. Concurrently,
115 we obtained a set of 39 TLSRGs as outlined in PMID: 31092904[29], and applied the GSVA
116 package to calculate their scores. Based on the median score, cell clusters were segregated into
117 groups with high and low TLS scores.

118   Furthermore, to discern the relationship between TLS and Hallmark scores, and to analyze the
119   clustering of TLS scores with Hallmark pathways, we used a heat map representation. The
120   creation of heat maps and correlation heat maps was facilitated by the "pheatmap" and "corrplot"
121   packages. All threshold values for these analyses were set as per the default specifications
122   detailed in the package vignettes.

123   **Elaborate Analysis of Bulk RNA-Sequencing Data**
124   **Formulation of the Prognostic TLSRGs Signature**
125   We initiated our analysis by amalgamating the bulk RNA-seq datasets, rectifying batch effects
126   using the "limma" and "sva" packages. Subsequently, a log2 transformation was applied to
127   standardize the data. The relationship between TLSRGs expression levels and patient overall
128   survival (OS) was scrutinized using univariate Cox regression analysis, utilizing the "survival"
129   package in R (version 3.2–13) (https://CRAN.Rproject.org/package=survival).

130   **Consensus Clustering for Prognostic TLSRGs**
131   We used the R package 'ConsensusClusterPlus' to perform unsupervised clustering analysis on
132   COADREAD patients, stratifying them based on the expression of prognostic-TLSRGs. By
133   employing the K-means algorithm, we identified the optimal number of clusters ranging from 2
134   to 10. This process was reiterated 1,000 times to affirm the stability and reliability of our results.

135   **Clinicopathological Traits**
136   Investigating the clinical implications, we analyzed the correlation between the molecular
137   subtypes and clinicopathological features, subsequently visualized in a heatmap. The
138   clinicopathological features that were examined included the patient's age, sex, tumor grade, and
139   stage of tumor node metastasis (TNM). The "survminer" packages in R facilitated our survival
140   analysis.

141   **Pathway Enrichment Analysis**
142   To elucidate the TIME characteristics in various molecular subtypes, GSVA was employed
143   utilizing gene sets from hallmark, KEGG, and Reactome obtained from the MSigDB database.

144   **Immune Infiltration**
145   The stromal score, immune score, and ESTIMATE score in tumor tissues were quantified using
146   the "estimate" R package[30] by employing the ESTIMATE algorithm from the ESTIMATE
147   project (https://sourceforge.net/projects/estimateproject/). Furthermore, we utilized ssGSEA to
148   delineate the immune infiltration pattern in COADREAD by employing marker genes for 23
149   immune cell types [31].

150   **Analysis of Differentially Expressed Genes**
151   The "limma" package was utilized to identify differentially expressed genes (DEGs) among
152   clusters, specifically targeting those with an adjusted p-value <0.05 and |log2 (Fold Change)>1.
153   The "clusterProfiler" R package was used to perform Gene Ontology (GO) and Kyoto
154   Encyclopedia of Genes and Genomes (KEGG) analyses. To discern DEGs with prognostic
155   significance, univariate Cox regression analysis was executed, with results presented in a forest
156   plot. Additionally, the GSCA datasets (https: //bioinfo.life.hust.edu.cn/GSCA/#/based) [32] were
157   used to perform mutation, SNV, CNV, and Methylation analyses on these prognostic genes.

158 **Refinement of the TLS Score System**

159 We employed the "ConsensusClusterPlus" package in R for unsupervised consensus clustering
160 to categorize COADREAD patients into distinct molecular subtypes. The classification was
161 determined by analyzing the expression patterns of previously identified prognostic TLS-DEGs.
162 The clustering process involved the K-means algorithm, assessing optimal cluster numbers from
163 $k = 2-10$. This procedure was iteratively performed 1,000 times for robust stability of results.
164 Kaplan-Meier plots elucidated significant prognostic variations between the clusters.
165 Additionally, a heatmap delineated the interplay between clinical attributes, gene expression
166 profiles, and clustering outcomes. The TLS score, computed via the GSVA package using 16
167 pivotal prognostic TLS-DEGs, facilitated subsequent survival analysis. This analysis discerned
168 the prognostic predictive capability of the TLS score in COADREAD cases. A Sankey diagram
169 illustrated the intricate associations among signature genes, scoring, and prognostic
170 categorization.

171 **Comprehensive Immune Landscape Analysis Utilizing the TLS Score**

172 To further understand the TLS score's implications, we created a correlation heatmap
173 showcasing its association with immune cell infiltration levels. Moreover, the interrelations
174 between the TLS score and various cytokines, chemokines, and their receptors were visualized
175 through additional heatmaps. The GSVA method was pivotal in correlating the TLS score with
176 50 hallmark pathways, with the findings depicted using R's "ggplot2" package. We also gathered
177 and analyzed mRNA expression data of immune checkpoints and gene mutation information
178 from the TCGA-COADREAD cohorts. This led to the calculation of mean normalized
179 expression levels of immune checkpoints per sample, subsequently normalized to convey their
180 relative expression magnitudes.

181 **Analysis of MSI, TMB, CSCs, and Somatic Mutations**

182 We conducted an in-depth analysis of MSI, CSC markers, and TMB across varying TLS score
183 groups. For this, somatic mutation data specific to COADREAD patients were extracted from the
184 TCGA database. To discern somatic mutations between different TLS score groups, we utilized
185 the "maftools" package to construct detailed waterfall plots. A comparative analysis of the
186 somatic mutation variances across the TLS score groups was then conducted using forest plots.

187 **Evaluation of Immunotherapy Response and Chemotherapy Sensitivity**

188 The predictive relevance of the TLS score for immunotherapy response was investigated using
189 transcriptomic data from two distinct cohorts. The dataset GSE135222 includes information from
190 patients with advanced non-small cell lung carcinoma who received treatments targeting anti-
191 PD-1/PD-L1. Meanwhile, GSE176307 includes transcriptomic information from patients with
192 Metastatic Urothelial Cancer. The prognostic significance of the TLS score in COADREAD was
193 evaluated using Kaplan-Meier analysis in the R package. Moreover, by utilizing the
194 CancerRxGene database (https://www.cancerrxgene.org/) [33], we calculated the IC50 values for
195 different target-therapeutic medications in relation to the TLS score. This estimation was
196 performed using the "pRRophetic" package in R, offering insights into the chemotherapy
197 sensitivity linked to the TLS score.

RNA Extraction and Quantitative Real-Time PCR

Total RNA from colorectal cells was isolated employing the Trizol method (Invitrogen) as per the manufacturer's instructions. Subsequent reverse transcription to synthesize cDNA was conducted using Prime Script RT reagent Kit (RR047A, Takara). Gene expression was measured using TB Green Premix Ex Taq (RR420A, Takara), and GAPDH was used as the reference for normalization. Details of the primers used are provided in Supplementary Table 1.

Cultivation of Cell Lines and Reagent Preparation

The Chinese Academy of Sciences Cell Bank (Shanghai, China) provided several colorectal cancer cell lines, namely NCM460, HCT116, HT29, and LOVO. Each cell line underwent STR profiling for authentication and was confirmed to be free of mycoplasma contamination (https://www.cellb ank.org.cn/). NCM460 and HCT116 were cultured in RPMI 1640 medium (Gibco) with the addition of 10% FBS, whereas HT29 and LOVO were maintained in DMEM (Gibco) with a 10% FBS enrichment. Cultures were incubated at 37°C in a humidified atmosphere with 5% CO_2.

Targeted Knockdown of TLS-Related Genes

To conduct gene silencing experiments, HCT116 and HT29 cells were plated at a density of 3×10^5 cells per 60 mm dish. Following 24 hours of incubation, the medium was replaced with fresh growth medium. Next, the cells were transfected with siRNAs that targeted C5AR1, APOE, CYP1B1, and SPP1, or a control siRNA obtained from Genepharma in Shanghai, China. This transfection was performed using Lipofectamine RNAiMax (13778075, Life Technologies). After transfection, the cells were cultured in RPMI 1640 or DMEM, with the addition of 10% FBS, for at least 24 hours. The siRNA sequences are listed in Supplementary Table 1.

Cell Proliferation Assessment

Cells were seeded in 96-well plates at a concentration of 3,000 cells per well. The growth of cells was observed for a period of 24 hours using a Cell Counting Kit-8 (CCK-8) from Dojindo, following the manufacturer's provided instructions. A microplate reader (Thermo Fisher, USA) was used to acquire absorbance measurements at 450 nm. GraphPad Prism version 9.5.0 was utilized for conducting data analysis.

Transwell Migration Assay

To perform migration experiments, a total of 100,000 cells were placed in the top compartment of a Transwell device on a Matrigel layer (Corning), utilizing either RPMI 1640 or DMEM excluding FBS. The bottom section contained 600 μL of the corresponding solution enriched with 20% FBS. After a 24-hour incubation period, cells were fixed and stained with crystal violet. Cells remaining in the upper chamber were carefully removed, and the migrated cells were imaged for further analysis.

Statistical Methodology

The statistical analysis was conducted using R software (version 4.1.2). Correlation assessments were conducted using Pearson or Spearman correlation coefficients. The Wilcoxon test was applied for comparisons between two groups. The overall survival (OS) across various subgroups was analyzed using Kaplan-Meier survival curves and log-rank tests.

238   Univariate Cox regression analyses were utilized to estimate prognostic values. The data from

239   the qRT-PCR analysis were examined utilizing the t-test of Student. A *p*-value less than 0.05 was

240   deemed to be statistically significant, with an asterisk (*) representing *p*-values less than 0.05,

241   two asterisks (**) representing *p*-values less than 0.01, and three asterisks (***) representing *p*-

242   values less than 0.001.

243   **Ethical Consideration**

244   The authors conducted this study without the participation of any humans or animals.

245

246   ## Results

247   Flowchart of this study was shown in figure 1.

248   **Analysis of Single Cell Sequencing Data**

249   **Dimensionality Reduction**

250   In our initial step, the single-cell sequencing data from the CRC_EMTAB8107 dataset, were

251   subjected to an integration analysis to consolidate information across different samples,

252   Supplementary Figure 1 exhibited the minimal batch effects and 19 distinct clusters categorized

253   by the UMAP algorithm. Further, we examined the expression of specific surface marker genes

254   across these clusters. This analysis enabled the identification of 10 unique cell types,

255   encompassing malignant cells, CD4+ T cells, plasma cells, fibroblasts, CD8+ T cells,

256   monocytes-macrophages, B cells, endothelial cells, mast cells, and dendritic cells, as depicted in

257   Figure 2A.

258   **Landscape of Transcriptome**

259   In the realm of gene expression, Figure 2B highlighted the top five genes making the most

260   significant contributions. Specifically, in malignant cells, the genes exhibiting the highest

261   expression levels included *C19orf33*, *KRT19*, *TSPAN8*, *AGR2*, and *LGALS4*. In stark contrast,

262   genes such as *SRGN*, *RGS1*, *TSC22D3*, *CXCR4*, and *CD52* were among the least expressed.

263   Within CD8+ T cells, genes like *CCL5*, *NKG7*, *GZMA*, *CCL4*, and *GNLY* were highly expressed,

264   while others including *CALD1*, *CST3*, *SPARCL1*, *GSN*, and *IFITM3* showed minimal expression.

265   For monocytes-macrophages, the top expressed genes were *AIF1*, *TYROBP*, *C1QA*, *FCER1G*,

266   and *C1QB*, whereas genes like *TRAC*, *CALD1*, *IL32*, *CD69*, and *FKBP11* were the least

267   expressed.

268   **Pathway Enrichment Analysis**

269   The GSVA brought to light Hallmark pathways that were predominantly enriched in each cell

270   type (see Figure 2C). In malignant cells, pathways such as mTORC1 signaling, Glycolysis,

271   Xenobiotic metabolism, fatty acid metabolism, were notably prevalent. Fibroblasts showed a

272   significant enrichment of the Epithelial mesenchymal transition pathway. The monocytes-

273   macrophages were enriched in pathways such as Complement, IL2-STAT5 signaling,

274   Inflammatory response, and Interferon response pathways.

275   **Cell-Type Specific Expression of TLS Insights**

276   Employing the "FindAllMarkers" function, a total of 39 TLSRGs were analyzed. This detailed

277   examination culminated in deciphering the expression patterns of TLSRGs across various cell

278　types (Figure 2D). Focusing on cell-specific expression, *MS4A1* emerged as a gene

279　predominantly expressed in B cells, while its expression remained minimal in other cell types. In

280　the realm of CD8+ T cells, genes like *CCL4* and *CCL5* showcased high expression levels,

281　contrasting with their subdued expression in other cells. Additionally, *TNFRSF17* was notably

282　expressed in plasma cells, yet exhibited low expression in other cell types.

283　**GSVA-Based TLS Score Calculation**

284　Utilizing the UMAP cell annotation map, we calculated the TLS score for each cell through the

285　GSVA method, as illustrated in Figure 2E. Subsequently, cells were categorized into clusters

286　with low and high TLS scores based on their median expression levels, depicted in Figure 2F.

287　Predominantly, malignant cells and fibroblasts were found in the low TLS score cluster, whereas

288　T cells and monocyte-macrophages were more prevalent in the high TLS score cluster.

289　**Cell Cluster Distribution and TLS Score Visualization**

290　This distribution was further detailed through box plots representing TLS scores across different

291　cell clusters (Figure 2G), where CD8+ T cells received the highest scores, and Mast cells the

292　lowest. The distribution patterns of cell types within these clusters were effectively represented

293　in a Composition chart based on cell count and proportion (Figure 2H-I), highlighting the

294　tendency of malignant cells and fibroblasts to fall into the low scoring group, in contrast to

295　monocyte-macrophages and CD8+ T cells which were more commonly found in the high scoring

296　group.

297　**TLS Score Correlation with Hallmark Pathways**

298　Further analysis revealed a significant correlation between the TLS score and specific Hallmark

299　pathway scores (Figure 2J). In malignant cells, the TLS score showed a positive correlation with

300　pathways such as apoptosis and KRAS signaling up. In monocyte-macrophages, a positive

301　correlation was observed with the Inflammatory response, interferon gamma response, IL2-

302　STAT5 signaling pathways. For dendritic cells, the TLS score positively correlated with

303　pathways including Inflammatory response, Interferon gamma response, Interferon alpha

304　response. We identified hallmark pathways that were differentially expressed between high- and

305　low-TLS score clusters, as shown in Figure 2K. These enriched pathways in high-score group

306　included Inflammatory response, Interferon gamma response, Interferon alpha response,

307　markedly distinct compared to the low-TLS score group.

308　**Comprehensive Analysis of TLSRGs in COADREAD**

309　**TLS Molecular Subtypes: Identification and Analysis**

310　Our exploration focused on discerning the TLS patterns in COADREAD tumorigenesis. To this

311　end, we examined 1,184 COADREAD patient samples sourced from the TCGA and GEO

312　databases (including GSE39582 and GSE17538). Survival analysis pinpointed 12 TLSRGs

313　(*CCL2, CCL20, CXCR3, IL1R1, TIGIT, SGPP2, ICOS, CCL8, CXCL9, CXCL11, CXCL10,*

314　*CXCL13*) significantly correlated with the overall survival (OS) in COADREAD patients (Figure

315　3B, $p < 0.05$). We then constructed a TLS network to delineate the interrelationships among

316　TLSRGs and their prognostic values (Figure 3A). Univariate Cox regression analysis on these

317　genes revealed positive associations of *CCL20, CXCR3, TIGIT, SGPP2, ICOS, CXCL9, CXCL11,*

318 *CXCL10,* and *CXCL13* with patient survival, while *CCL2, CCL8,* and *IL1R1* showed negative
319 associations.
320   Based on the 12 prognostic TLSRGs, we employed a consensus clustering algorithm to divide
321 COADREAD patients into two distinct molecular subtypes, namely Cluster A and Cluster B,
322 with the analysis suggesting k=2 as the optimal cluster number (Figure 3D). Patients in subtype
323 B exhibited significantly higher survival rates compared to those in subtype A, as demonstrated
324 by the log-rank test ($p = 0.001$; Figure 3C). A notable divergence in the expression patterns of
325 TLSRGs was observed between clusters A and B (Figure 3E). Particularly, a subset of TLSRGs,
326 including *CXCL9, CXCL10, CXCL11, CXCL13, ICOS, TIGIT,* among others, were found to be
327 predominantly upregulated in cluster B ($p < 0.05$). Further, we profiled the associations of
328 TLSRG expression with various clinical characteristics such as age, disease stage, gender,
329 recurrence, metastasis, fustat, and futime across the different subtypes (Figure 3F). This
330 profiling provided insights into the clinical relevance of TLSRG expression variations in distinct
331 patient subgroups. To enhance the distinct biological traits of the subtypes, our study
332 incorporated a comprehensive GSVA enrichment analysis using Hallmark, KEGG, and
333 Reactome pathway databases. Collectively, these comprehensive pathway analyses consistently
334 underscored the prominence of interferon gamma response, interferon alpha response,
335 chemokine signaling pathway, and cytokine-cytokine receptor interaction pathways in subtype B.
336 These pathways were critically linked to immune activation [34], further delineating the distinct
337 immunological landscape of subtype B. (Figure 3G-I).
338   The PCA analysis reinforced the significant transcriptional distinctions between subtype A
339 and B (Figure 3 J). To delve deeper into the role of TLSRGs in shaping the TIME, we quantified
340 human immune cell subsets in each COADREAD sample. As depicted in Figure 3 K, cluster B
341 was characterized by elevated stromal, immune, and ESTIMATE scores. A more granular
342 examination through ssGSEA revealed marked disparities in immune cell infiltration between
343 subtypes A and B (Figure 3L). Specifically, subtype B demonstrated notably higher infiltration
344 levels of activated B cells, activated CD4+ T cells, activated CD8+ T cells, activated dendritic
345 cells, macrophages, natural killer T cells, natural killer cells, T follicular helper cells, Type 1 T-
346 helper cells, and Type 2 T-helper cells compared to subtype A.
347   From these observations, it can be inferred that subtype B was significantly enriched in
348 immune activation pathways, which appears to be intricately linked with the dynamics of TIME.
349 In the quest to decode the biological intricacies of TLS subtypes, we scrutinized 189 DEGs
350 between subtypes A and B. These were visually represented through a volcano plot (Figures 4A,
351 |logFC|>1, $p < 0.05$). To identify pertinent biological pathways, we engaged in comprehensive
352 GO and KEGG enrichment analyses. The GO and KEGG enrichment analysis pinpointed the
353 DEGs' significant involvement in pathways such as cytokine-cytokine receptor interaction,
354 upregulation of cytokine production, lymphocyte-mediated immunity, and adaptive immune
355 responses founded on somatic leukocyte proliferation (Figures 4 B, C). A cnetplot was
356 constructed to visualize the interplay within these pathways, highlighting the top 5 pathways
357 (Figure 4 D).

Identification of TLS gene subtypes

In the continuum of our research, a univariate Cox regression analysis was conducted to evaluate the prognostic significance of the 189 identified DEGs. This rigorous analysis led to the isolation of 16 key signature genes closely associated with overall survival (OS) in COADREAD patients ($p < 0.001$). These genes, namely *HOXC6*, *SFRP2*, *CYP1B1*, *GAS1*, *SPOCK1*, *SPP1*, *POSTN*, *MARCO*, *VSIG4*, *THBS2*, *OLR1*, *COL10A1*, *FAP*, *GZMB*, *C5AR1*, and *APOE*, were depicted in a forest plot for clarity (Sup-Figure 2 A).

Subsequently, COADREAD patients were stratified into two distinct groups based on the expression profiles of these 16 prognostic genes, utilizing a consensus clustering algorithm. The partitioning suggested that k=2 was an optimal division, forming two gene clusters, namely A and B (Figure 4 E). Survival analysis revealed a noteworthy finding: patients in subtype B exhibited a significantly enhanced survival probability compared to those in subtype A (log-rank test, $p < 0.001$; Figure 4 F). The box-map analysis highlighted that most of the signature genes, including *APOE*, *VSIG4*, *C5AR1*, *CYP1B1*, and *SPP1*, were predominantly 11 upregulated in cluster B (Figure 4 G, $p < 0.05$). Moreover, we extended our analysis to explore the correlation of these gene expressions with various clinical features such as age, stage, gender, recurrence, metastasis, fustat, and futime, stratifying them across the molecular subtypes (Figure 4 H).

Construction of TLS score

In advancing our study, a TLS score was formulated based on GSVA of the 16 identified signature genes. This scoring system facilitated the stratification of patients into two distinct genomic subtypes: those with low and high TLS scores. Survival analysis revealed a critical insight; patients categorized in the low-score group exhibited a markedly higher survival probability compared to their high-score counterparts (Figure 4 I, $p < 0.001$). A Sankey diagram was employed to eloquently depict the patient distribution across the two TLS molecular subtypes, gene subtypes, and score groups, alongside their respective prognostic implications (Figure 4 J).

Association of TLS score with immune infiltration

Spearman correlation analysis established a positive correlation between the TLS score and the infiltration of 23 different immune cells. This correlation suggested an augmented immune component presence in the TIME of patients within the high-score cluster, potentially indicating a more favorable immune prognosis (Figure 4 K, $p < 0.05$). Building upon these findings, we delved deeper into the association of TLS with cytokine-chemokine interaction networks. Notably, the high-score TLS cluster demonstrated significant enrichment in chemokines and their receptors, interleukins, and interferons along with their respective receptors (Fig. 4 L). The hallmark pathway analysis revealed a significant positive correlation12 of the TLS score with pathways indicative of immune activation. These pathways included interferon gamma response, interferon alpha response, inflammatory response, and IL-2 STAT5 signaling. Conversely, an inverse correlation was observed between the TLS score and pathways typically associated with malignancy, such as peroxisome proliferation, MYC target activation, oxidative phosphorylation,

397 fatty acid metabolism, and DNA repair mechanisms (Fig. 4 M). These findings underscored the
398 strong association of the TLS score with predominantly immune-activated pathways.
399 **Association of TLS score with MSI, TMB, CSC, and somatic mutations**
400 Cancer Stem Cells (CSCs), which were acknowledged for their significant roles due to their
401 selfrenewal capacity and differentiation potential [35]. We delved into the relationship between the
402 TLS score and CSC indices, including mDNAsi and mRNAsi. Our analysis, as depicted in
403 Figure 5 A, revealed a notable negative linear correlation between the TLS score and the
404 mDNAsi index (R = -0.18, p = 6e-04). This trend was further demonstrated by the lower
405 mDNAsi observed in the high TLS score cluster compared to the low TLS score cluster ($p < 0.01$;
406 Figure 5 B). Similarly, Figure 5 D presented a negative linear correlation between the TLS score
407 and the mRNAsi index (R = -0.63, p = 0), with the high TLS score cluster exhibiting
408 significantly lower mRNAsi than its low-score counterpart ($p < 0.0001$; Figure 5 E). Heat maps
409 further corroborate these findings, illustrating the inverse relationship between both mDNAsi and
410 mRNAsi index values and the TLS score (Figure 5 C, Figure 5 F). These insights suggested that
411 COADREAD patients with a higher TLS score may exhibit reduced stem cell characteristics and
412 diminished cell differentiation capabilities.
413 In the context of Microsatellite Instability (MSI), a promising genomic biomarker for
414 gauging13 patient responsiveness to immunotherapy[36], our findings revealed a significantly
415 higher MSI in the high TLS score cluster compared to the low ($p < 0.05$; Figure 5 H). Spearman
416 correlation analysis further substantiates this, indicating a positive association between MSI and
417 TLS score (R = 0.18, p = 7e-04; Figure 5 G). The positive correlation between MSI expression
418 and TLS score was also visually represented in a heat map (Figure 5 I). This pattern suggested a
419 potential link with enhanced immunotherapy efficacy.
420 Regarding Tumor Mutational Burden (TMB), which served as an indicator of cancer mutation
421 volume and has clinical relevance with ICIs outcomes[37], our data analysis from
422 TCGACOADREAD cohorts indicated a higher TMB in the high TLS score cluster than in the
423 low score cluster ($p < 0.001$; Figure 5 K). Spearman's analysis corroborated a positive
424 relationship between TMB and TLS score (R = 0.2, p = 1e-04; Figure 5 J), with a heat map
425 further illustrating this correlation (Figure 5 L). These findings were particularly noteworthy for
426 patients with higher TMB who typically show more benefit from ICIs.
427 Our analysis extended to the somatic mutation landscape within the two distinct TLS score
428 clusters. In the high TLS score cluster, the predominant mutated genes were *TP53*, *TTN*, *APC*,
429 *KRAS*, *SYNE1*, *MUC16*, *PIK3CA*, *OBSCN*, *FAT4*, and *RYR2*. Conversely, the low TLS score
430 cluster was characterized by mutations primarily in *APC*, *TP53*, *KRAS*, *TTN*, *SYNE1*, *PIK3CA*,
431 *FAT4*, *MUC16*, *OBSCN*, and *MUC4*. Notably, the mutation frequency across these genes was
432 more pronounced in the high TLS score cluster, as illustrated in Figures 5 M-N. Additionally, a
433 forest plot detailed the top 12 genes exhibiting the most significant differences in mutation
434 frequency between the high and low score clusters (Figures 5 O). This led to the inference that
435 gene mutation frequency in the high TLS score cluster was not only higher but also more14
436 diverse. Detailed SNV, CNV, and methylation profiles of prognostic TLS-DEGs in

437 COADREAD patients were depicted in supplementary Figures 2-3. Analysis revealed significant
438 disparities in recurrence/ metastasis, stage, and fustat between the low- and high-TLS score
439 clusters (Figure 5 P-S).

Interplay of TLS Score and Immune Checkpoints Expression

441 Subsequent research focused on the expression levels of immune checkpoints in relation to TLS
442 scores. Elevated expression of key immune checkpoints, including *PDCD1* (*PD-1*), *CD274* (*PD-*
443 *L1*), *CTLA4*, *TIGIT*, and *LAG3*, was predominantly observed in the high-score cluster, as
444 depicted in Figure 5 T. This trend suggested that TLS score was potentially served as a
445 predictive marker for immunotherapy efficacy.

Evaluating TLS Score as a Predictor for PD-L1 Blockade Immunotherapy

447 In the realm of T-cell immunotherapy, a promising approach for cancer treatment, we explored
448 the significance of the TLS score in COADREAD, building on prior findings and our current
449 results[38]. Within the GSE135222 cohort, a notable observation was the higher progression rate
450 (100%) in patients with lower TLS scores compared to those with higher scores (75%) (Figure 5
451 V). Furthermore, a markedly prolonged progression-free survival (PFS) was evident in patients
452 with higher TLS scores (Figure 5 U, $p = 0.0061$). Analyzing the GSE176307 cohort, which
453 comprised 90 patients treated with anti-PD-L1 receptor blockers, revealed a spectrum of
454 responses ranging from complete response (CR) and partial response (PR) to stable disease (SD)
455 and progressive disease (PD). Significantly, patients exhibiting CR/PR had elevated TLS scores
456 compared to those with SD/PD (Figure 5 X). In this cohort, patients in the high-TLS cluster
457 demonstrated substantial clinical benefits and notably longer overall survival (OS) than
458 15 their low-TLS counterparts (Figure 5 W, $p = 0.018$).

Analysis of Drug Susceptibility in Relation to TLS Score

460 Subsequent to establishing the TLS score as an effective predictor for immunotherapy
461 response, our study delved into the variances in IC50 values for various targeted anticancer drugs
462 across low- and high-TLS clusters. Intriguingly, patients within the high-TLS cluster exhibited
463 notably lower IC50 values for a range of anticancer agents, namely A.770041, AP.24534,
464 AG.014699, AICAR, ABT.263, and AMG.706 (Figure 5 Y). These findings underscored the
465 potential utility of the TLS score as a guiding metric in selecting appropriate anticancer drugs.

Impact of TLSRGs on Colorectal Cancer Cell Behavior In Vitro

467 In our investigation, RT-q PCR was utilized to ascertain the expression levels of *C5AR1*, *APOE*,
468 *CYP1B1*, and *SPP1* in a standard colon cell line and three colorectal cancer cell lines (Figure 6
469 A-D). These four genes were markedly upregulated in cancer cells compared to normal cells. We
470 employed siRNA techniques targeting *C5AR1*, *APOE*, *CYP1B1*, and *SPP1*. This was conducted
471 in HCT116 and HT29 cell lines, selecting siRNA-1 and siRNA-2 for further analysis based on
472 their transfection efficiency exceeding 70% (Sup-Figure 4 A-H). A CCK8 assay was
473 subsequently performed to evaluate cell proliferation. Results demonstrated that the knockdown
474 of *APOE*, *C5AR1*, *CYP1B1*, and *SPP1* significantly impeded the proliferation of both
475 HCT116 and HT29 cells (Figure 6 E-L). The Transwell assay was conducted to assess cellular
476 invasion capabilities. Notably, the suppression of *APOE*, *C5AR1*, *CYP1B1*, and *SPP1* led to a

477 marked reduction in the invasion capacity of HCT116 and HT29 cells (Figure 6 M-N).
478 Furthermore, an unexpected observation was the considerable decrease in *PD-L1* expression in
479 both HCT116 and HT29 cells following the knockdown of these genes. This reduction in *PD-L1*
480 expression could hint at the potential of combining TLS inducers with PD-L1 inhibitors as a
481 therapeutic strategy (Figure 6 P-Q). Additionally, the immunohistochemical data from the HPA
482 database revealed an elevated protein expression of *APOE*, *C5AR1*, *CYP1B1*, and *SPP1* in the
483 stroma of COADREAD tissue (Figure 6 O).
484

Discussion

486 Globally, the prevalence of COADREAD, a leading form of cancer, is escalating in numerous
487 nations. Recent endeavors in COADREAD management have been substantial; however, the
488 disease's heterogeneous nature and its aggressive tendencies continue to pose challenges in
489 prognostic evaluations [1,3,5]. Identifying new biomarkers is thus imperative and urgent, offering a
490 pathway to tailor patient-specific treatments and enhance prognostic accuracy.
491 While immune checkpoint inhibitors and chimeric antigen receptor T cell therapies are
492 established as safe for various cancer types, their efficacy limitations underscore the need to
493 explore additional cellular pathways for more effective treatment strategies[39]. Recent studies
494 have linked the development of tertiary lymphoid structures (TLS) with improved clinical
495 outcomes across several cancer types, and these structures are believed to enhance the efficacy of
496 immunotherapies[15-17]. In colorectal cancers, which exhibit significant tumor heterogeneity, only
497 certain molecular subtypes respond to immunotherapy[4-6]. Therefore, TLS could potentially
498 provide insights into predicting and selecting COADREAD patients for immunotherapy, and in
499 addressing immunotherapy resistance, particularly in groups with traditionally low response rates.
500 In contrast to bulk RNA-sequencing, which gauges average gene expression levels in cell
501 populations, scRNA-seq has risen as a pivotal technique. It facilitates the delineation of cellular
502 subpopulations, pinpointing distinctive biomarkers, and understanding the heterogeneity across
503 different cell types in an array of cancers[9].
504 Thus, this study undertook an extensive analysis combining both bulk RNA-seq and scRNA-
505 seq. Our approach encompassed clustering analysis, examination of TLS-associated DEGs,
506 assessment of immune infiltration and mutation landscapes, and screening for prognostic genes.
507 This comprehensive methodology enabled us to formulate a TLS score, demonstrating
508 substantial predictive accuracy for immunotherapy effectiveness in COADREAD. Initially, our
509 scRNA-seq analysis on COADREAD samples delineated two distinct TLS score clusters. In the
510 cluster with elevated scores, we noted a significant enrichment of critical immune activation
511 markers, such as inflammatory response, interferon-gamma and alpha responses, and TNFA
512 signaling through NFKB pathways, aligning with earlier findings[34]. Further, a higher prevalence
513 of monocytes, macrophages, and CD8+ T cells was evident in the high score group compared to
514 the low-score group, which predominantly consisted of malignant cells and fibroblasts. This
515 distribution implied that the high-score group harbors a more robust anti-tumor immune cell
516 congregation, thereby intensifying anti-tumor responses. This observation was in concordance

517 with recent literature suggesting TLS's role in invigorating antitumor immunity and boosting
518 immunotherapy efficacy[11,15,18], reinforcing the concept of TLSs as indicative of "hot tumors".
519 Subsequently, our investigation led to the identification of 12 TLSRGs with prognostic
520 significance. Utilizing these genes, we stratified COADREAD patients into two distinct clusters.
521 Our study distinguished 189 DEGs between the two identified clusters, leading to the isolation of
522 16 genes intimately linked with overall survival (OS). we segregated COADREAD patients into
523 two gene clusters based on these 16 prognostic genes. Leveraging this groundwork, we crafted a
524 TLS score model using GSVA. This model incorporated 16 pivotal TLS-DEGs: *HOXC6*, *SFRP2*,
525 *CYP1B1*, *GAS1*, *SPOCK1*, *SPP1*, *POSTN*, *MARCO*, *VSIG4*, *THBS2*, *OLR1*, *COL10A1*, *FAP*,
526 *GZMB*, *C5AR1*, and *APOE*. Their prognostic relevance in various malignancies was further
527 corroborated by previous research.
528 The role of *HOXC6* in nonmetastatic CRC is noteworthy, its heightened expression correlates
529 significantly with increased immunogenicity[40]. Within an aging microenvironment, *sFRP2* has
530 been identified as a driver of melanoma metastasis and resistance to therapies[41]. The enzyme
531 *CYP1B1* played a crucial role in the advancement and progression of castration-resistant prostate
532 cancer[42]. Interestingly, *GAS1* impeded colorectal tumorigenesis through WNT signaling,
533 influencing CD143+ cancer-associated fibroblasts[42]. As a potential cancer prognostic marker,
534 *SPOCK1* enhanced proliferation and metastasis in gallbladder cancer cells via the PI3K/AKT
535 pathway[43]. The protein *SPP1*, known for its upregulation of *PD-L1*, contributed to lung cancer's
536 ability to evade immune detection[44]. *POSTN*, by augmenting M2 macrophages and cancer-
537 associated fibroblasts, facilitated ovarian cancer metastasis[45]. In lung cancer, targeting *MARCO*
538 on immunosuppressive macrophages has shown promise, hindered regulatory T cells while
539 bolstering cytotoxic lymphocyte functions[46]. Finally, *VSIG4*, which regulated immune
540 homeostasis by modulating complement pathways and T-cell differentiation, presented a double-
541 edged sword. While it curbed immune-mediated inflammatory diseases, it simultaneously aided
542 in cancer progression, positioning it as a novel target in cancer immunotherapy[47]. In the context
543 of early-stage lung adenocarcinoma, *THBS2*+ cancer-associated fibroblasts have been pinpointed
544 as crucial facilitators of the disease's aggressiveness through multi-scale integrative analyses[48].
545 Research revealed that silencing *OLR1* attenuates glycolytic metabolism, thereby inhibiting
546 proliferation and chemoresistance in colon cancer cells[49]. *COL10A1* has been identified as a
547 promoter of gastric cancer invasion and metastasis, primarily through the process of epithelial-
548 to-mesenchymal transition[50]. *FAP* appeared to exacerbate malignancy outcomes through various
549 mechanisms, including extracellular matrix remodeling, intracellular signaling, angiogenesis,
550 epithelial-to-mesenchymal transition, and immunosuppression [51]. *GZMB* stood out as a primary
551 indicator of cytotoxic T-cell activity, with *GZMB*+ T-cells being pivotal in antitumor immunity
552 [52]. The intracellular *C5a/C5aR1* complex was found to stabilize β-catenin, thereby facilitating
553 colorectal tumorigenesis [53]. Furthermore, *APOE*, secreted by prostate tumor cells, has been
554 linked to cellular senescence and is associated with adverse prognoses [54]. This study presented,
555 for the first time, an extensive exploration of the roles of these TLS signature genes in
556 COADREAD.

557 Subsequent findings indicated that individuals with higher TLS scores potentially derive
558 greater benefit from anti-PD-1 therapies and targeted treatment approaches. Intriguingly, we
559 observed a contrasting prognosis associated with TLS scores in patients undergoing
560 immunotherapy versus those who were not. This aligns with the observed higher rates of
561 immune cell infiltration, immune checkpoint expression, MSI index, TMB index, and somatic
562 mutations in the high-TLS score cohort. Furthermore, a significant positive correlation was noted
563 between most immune activation-related pathways and TLS scores. Existing literature supported
564 the notion that both immunotherapy and chemotherapy can promote the development of tertiary
565 lymphoid structures, including the proliferation of CD8+ T cells, thereby bolstering anti-tumor
566 immunity[11,14,17,19]. This aligned with our RNA-seq data analysis. Similarly, our single-cell data
567 model corroborated this, indicating enhanced anti-tumor immune cell infiltration and pathway
568 activation in the high-score group. These insights underscored the pivotal role of TLS status in
569 determining the efficacy of immunotherapy and other immune-based treatments.

570 This study provides important insights into the prognostic and therapeutic potential of TLS-
571 related gene profiles in colorectal cancer. However, several limitations must be acknowledged to
572 fully contextualize the findings and guide future research efforts. A primary limitation is the
573 reliance on RNA sequencing data without direct histopathological validation of TLS presence.
574 While TLS-related gene signatures were identified, their correlation with actual TLS structures
575 in the tumor microenvironment was not confirmed. Future studies should incorporate
576 histopathological validation, such as immunohistochemistry or immunofluorescence, to directly
577 assess TLS formation and establish a clearer link between gene signatures and TLS presence.

578 Additionally, while this study provides a comprehensive analysis of TLS-related gene
579 signatures, it does not explore the underlying mechanisms that drive TLS formation or their
580 functional impact on immune responses. Mechanistic insights into how TLSs form and interact
581 with immune cells within the tumor microenvironment are crucial for understanding their role in
582 modulating immune activity and influencing patient outcomes. Future research should focus on
583 functional assays and mechanistic studies using in vitro and in vivo models to investigate these
584 pathways.

585 Another limitation stems from the reliance on publicly available datasets from GEO and
586 TCGA, which introduces potential biases. These datasets may not fully represent the diversity of
587 colorectal cancer populations, particularly in terms of ethnicity, disease stage, and treatment
588 history, which could limit the generalizability of the findings. Validation in more diverse patient
589 cohorts, ideally through prospective multicenter studies, will be necessary to ensure the
590 applicability of the TLS score across broader clinical settings.

591 Moreover, the study did not explicitly control for confounding variables such as tumor stage,
592 treatment history, or patient demographics, which could influence the TLS score's correlation
593 with immunotherapy efficacy. These factors might have affected the outcomes and should be
594 considered in future research. Stratifying patients or adjusting for confounders in statistical
595 analyses will help ensure more accurate conclusions about the clinical utility of the TLS score.

596   A further limitation is the absence of functional validation, which makes some conclusions
597 speculative. Although the bioinformatics analysis suggests that TLS presence correlates with
598 immune responses, direct experimental validation is needed to confirm these findings. Future
599 studies should employ functional assays, such as co-culture systems or animal models, to test the
600 biological role of TLSs in modulating immune responses and treatment outcomes.
601   The clinical relevance of the TLS score remains somewhat unclear, as we did not fully clarify
602 its association with key biomarkers such as immune responses, tumor mutational burden (TMB),
603 microsatellite instability (MSI), and specific gene signatures. While immune response markers
604 were shown to correlate with higher TLS scores, the relationships with TMB and MSI need
605 further exploration. Future research should refine the TLS score, better delineating its
606 associations with these biomarkers and validating its use in clinical decision-making.
607   The complexity of the bioinformatics approach used in this study may also pose challenges for
608 replication. However, we have provided all necessary methodological details, along with
609 publicly available data and code, to ensure that other researchers can replicate and build upon
610 these findings.
611   Additionally, the findings of this study have not been validated in clinical trials, limiting their
612 immediate applicability in clinical practice. Prospective clinical trials will be essential to
613 evaluate the TLS score's ability to predict treatment responses and guide personalized therapy in
614 colorectal cancer patients. Such trials would also help assess the long-term utility of TLS-related
615 biomarkers in predicting recurrence and survival outcomes, which were not covered by the
616 current study.
617   Finally, while this study offers a broad assessment of TLS-related gene profiles, the global
618 nature of the analysis may dilute the focus. A more targeted approach, investigating specific
619 hypotheses related to TLS function or its relevance in immunotherapy, would yield more
620 actionable conclusions. Future research should aim for a deeper exploration of particular aspects
621 of TLS biology that have the most clinical significance.
622   In conclusion, while this study provides valuable insights into TLS-related gene profiles in
623 colorectal cancer, addressing these limitations in future research will significantly enhance the
624 robustness and clinical relevance of TLS as a prognostic and therapeutic marker.

625 Conclusions

626 In this study, we elucidated the distinct features of TLS within the COADREAD framework.
627 Utilizing a quadratic clustering-based TLS score model, our approach enabled the assessment of
628 TLS dynamics at an individual level. This method also facilitated predictions regarding the
629 effectiveness of immunotherapy and other immune-based treatments in COADREAD cases.
630 Consequently, a deeper comprehension of TLS patterns, combined with the application of the
631 TLS scoring system, held promise for guiding clinical decisions and enhancing the prognosis for
632 patients with colorectal cancer.

633

634 Acknowledgements

636

References

638   1   Morgan E, Arnold M, Gini A, Lorenzoni V, Cabasag CJ, Laversanne M, Vignat J, Ferlay J,
639   Murphy N, and Bray F. 2023. Global burden of colorectal cancer in 2020 and 2040:
640   incidence and mortality estimates from GLOBOCAN. *Gut* 72:338-344. 10.1136/gutjnl-
641   2022-327736
642   2   Siegel RL, Wagle NS, Cercek A, Smith RA, and Jemal A. 2023. Colorectal cancer
643   statistics, 2023. *CA Cancer J Clin* 73:233-254. 10.3322/caac.21772
644   3   Morris VK, Kennedy EB, Baxter NN, Benson AB, 3rd, Cercek A, Cho M, Ciombor KK,
645   Cremolini C, Davis A, Deming DA, Fakih MG, Gholami S, Hong TS, Jaiyesimi I, Klute K,
646   Lieu C, Sanoff H, Strickler JH, White S, Willis JA, and Eng C. 2023. Treatment of
647   Metastatic Colorectal Cancer: ASCO Guideline. *J Clin Oncol* 41:678-700.
648   10.1200/jco.22.01690
649   4   André T, Shiu KK, Kim TW, Jensen BV, Jensen LH, Punt C, Smith D, Garcia-Carbonero
650   R, Benavides M, Gibbs P, de la Fouchardiere C, Rivera F, Elez E, Bendell J, Le DT,
651   Yoshino T, Van Cutsem E, Yang P, Farooqui MZH, Marinello P, and Diaz LA, Jr. 2020.
652   Pembrolizumab in Microsatellite-Instability-High Advanced Colorectal Cancer. *N Engl J*
653   *Med* 383:2207-2218. 10.1056/NEJMoa2017699
654   5   Ludford K, Ho WJ, Thomas JV, Raghav KPS, Murphy MB, Fleming ND, Lee MS, Smaglo
655   BG, You YN, Tillman MM, Kamiya-Matsuoka C, Thirumurthi S, Messick C, Johnson B,
656   Vilar E, Dasari A, Shin S, Hernandez A, Yuan X, Yang H, Foo WC, Qiao W, Maru D,
657   Kopetz S, and Overman MJ. 2023. Neoadjuvant Pembrolizumab in Localized
658   Microsatellite Instability High/Deficient Mismatch Repair Solid Tumors. *J Clin Oncol*
659   41:2181-2190. 10.1200/jco.22.01351
660   6   Gurjao C, Liu D, Hofree M, AlDubayan SH, Wakiro I, Su MJ, Felt K, Gjini E, Brais LK,
661   Rotem A, Rosenthal MH, Rozenblatt-Rosen O, Rodig S, Ng K, Van Allen EM, Corsello
662   SM, Ogino S, Regev A, Nowak JA, and Giannakis M. 2019. Intrinsic Resistance to
663   Immune Checkpoint Blockade in a Mismatch Repair-Deficient Colorectal Cancer. *Cancer*
664   *Immunol Res* 7:1230-1236. 10.1158/2326-6066.Cir-18-0683
665   7   Marabelle A, Fakih M, Lopez J, Shah M, Shapira-Frommer R, Nakagawa K, Chung HC,
666   Kindler HL, Lopez-Martin JA, Miller WH, Jr., Italiano A, Kao S, Piha-Paul SA, Delord JP,
667   McWilliams RR, Fabrizio DA, Aurora-Garg D, Xu L, Jin F, Norwood K, and Bang YJ.
668   2020. Association of tumour mutational burden with outcomes in patients with advanced
669   solid tumours treated with pembrolizumab: prospective biomarker analysis of the
670   multicohort, open-label, phase 2 KEYNOTE-158 study. *Lancet Oncol* 21:1353-1365.
671   10.1016/s1470-2045(20)30445-9
672   8   Thind AS, Monga I, Thakur PK, Kumari P, Dindhoria K, Krzak M, Ranson M, and Ashford
673   B. 2021. Demystifying emerging bulk RNA-Seq applications: the application and utility of
674   bioinformatic methodology. *Brief Bioinform* 22. 10.1093/bib/bbab259
675   9   Kotsiliti E. 2022. Single-cell RNA-seq keeps cells alive. *Nat Biotechnol* 40:1432.
676   10.1038/s41587-022-01515-8
677   10   Schumacher TN, and Thommen DS. 2022. Tertiary lymphoid structures in cancer.
678   *Science* 375:eabf9419. 10.1126/science.abf9419
679   11   Cabrita R, Lauss M, Sanna A, Donia M, Skaarup Larsen M, Mitra S, Johansson I, Phung
680   B, Harbst K, Vallon-Christersson J, van Schoiack A, Lövgren K, Warren S, Jirström K,
681   Olsson H, Pietras K, Ingvar C, Isaksson K, Schadendorf D, Schmidt H, Bastholt L,
682   Carneiro A, Wargo JA, Svane IM, and Jönsson G. 2020. Tertiary lymphoid structures
683   improve immunotherapy and survival in melanoma. *Nature* 577:561-565.
684   10.1038/s41586-019-1914-8

685 12   Meylan M, Petitprez F, Becht E, Bougoüin A, Pupier G, Calvez A, Giglioli I, Verkarre V,
686   Lacroix G, Verneau J, Sun CM, Laurent-Puig P, Vano YA, Elaïdi R, Méjean A, Sanchez-
687   Salas R, Barret E, Cathelineau X, Oudard S, Reynaud CA, de Reyniès A, Sautès-
688   Fridman C, and Fridman WH. 2022. Tertiary lymphoid structures generate and
689   propagate anti-tumor antibody-producing plasma cells in renal cell cancer. *Immunity*
690   55:527-541.e525. 10.1016/j.immuni.2022.02.001
691 13   Overacre-Delgoffe AE, Bumgarner HJ, Cillo AR, Burr AHP, Tometich JT, Bhattacharjee
692   A, Bruno TC, Vignali DAA, and Hand TW. 2021. Microbiota-specific T follicular helper
693   cells drive tertiary lymphoid structures and anti-tumor immunity against colorectal
694   cancer. *Immunity* 54:2812-2824.e2814. 10.1016/j.immuni.2021.11.003
695 14   Helmink BA, Reddy SM, Gao J, Zhang S, Basar R, Thakur R, Yizhak K, Sade-Feldman
696   M, Blando J, Han G, Gopalakrishnan V, Xi Y, Zhao H, Amaria RN, Tawbi HA, Cogdill
697   AP, Liu W, LeBleu VS, Kugeratski FG, Patel S, Davies MA, Hwu P, Lee JE,
698   Gershenwald JE, Lucci A, Arora R, Woodman S, Keung EZ, Gaudreau PO, Reuben A,
699   Spencer CN, Burton EM, Haydu LE, Lazar AJ, Zapassodi R, Hudgens CW, Ledesma
700   DA, Ong S, Bailey M, Warren S, Rao D, Krijgsman O, Rozeman EA, Peeper D, Blank
701   CU, Schumacher TN, Butterfield LH, Zelazowska MA, McBride KM, Kalluri R, Allison J,
702   Petitprez F, Fridman WH, Sautès-Fridman C, Hacohen N, Rezvani K, Sharma P, Tetzlaff
703   MT, Wang L, and Wargo JA. 2020. B cells and tertiary lymphoid structures promote
704   immunotherapy response. *Nature* 577:549-555. 10.1038/s41586-019-1922-8
705 15   Groeneveld CS, Fontugne J, Cabel L, Bernard-Pierrot I, Radvanyi F, Allory Y, and de
706   Reyniès A. 2021. Tertiary lymphoid structures marker CXCL13 is associated with better
707   survival for patients with advanced-stage bladder cancer treated with immunotherapy.
708   *Eur J Cancer* 148:181-189. 10.1016/j.ejca.2021.01.036
709 16   Ruffin AT, Cillo AR, Tabib T, Liu A, Onkar S, Kunning SR, Lampenfeld C, Atiya HI,
710   Abecassis I, Kürten CHL, Qi Z, Soose R, Duvvuri U, Kim S, Oesterrich S, Lafyatis R,
711   Coffman LG, Ferris RL, Vignali DAA, and Bruno TC. 2021. B cell signatures and tertiary
712   lymphoid structures contribute to outcome in head and neck squamous cell carcinoma.
713   *Nat Commun* 12:3349. 10.1038/s41467-021-23355-x
714 17   Vanhersecke L, Brunet M, Guégan JP, Rey C, Bougouin A, Cousin S, Moulec SL, Besse
715   B, Loriot Y, Larroquette M, Soubeyran I, Toulmonde M, Roubaud G, Pernot S, Cabart M,
716   Chomy F, Lefevre C, Bourcier K, Kind M, Giglioli I, Sautès-Fridman C, Velasco V,
717   Courgeon F, Oflazoglu E, Savina A, Marabelle A, Soria JC, Bellera C, Sofeu C, Bessede
718   A, Fridman WH, Loarer FL, and Italiano A. 2021. Mature tertiary lymphoid structures
719   predict immune checkpoint inhibitor efficacy in solid tumors independently of PD-L1
720   expression. *Nat Cancer* 2:794-802. 10.1038/s43018-021-00232-6
721 18   Chelvanambi M, Fecek RJ, Taylor JL, and Storkus WJ. 2021. STING agonist-based
722   treatment promotes vascular normalization and tertiary lymphoid structure formation in
723   the therapeutic melanoma microenvironment. *J Immunother Cancer* 9. 10.1136/jitc-
724   2020-001906
725 19   Lee JW, Park BC, Jang NY, Lee S, Cho YK, Sharma P, Byun SW, Jeon K, Jeon YH,
726   Park U, Ro HJ, Park HR, Kim Y, Lee DS, Chung S, Kim YK, and Cho NH. 2022.
727   Inducing Ectopic T Cell Clusters Using Stromal Vascular Fraction Spheroid-Based
728   Immunotherapy to Enhance Anti-Tumor Immunity. *Adv Sci (Weinh)* 9:e2203842.
729   10.1002/advs.202203842
730 20   Buisseret L, Desmedt C, Garaud S, Fornili M, Wang X, Van den Eyden G, de Wind A,
731   Duquenne S, Boisson A, Naveaux C, Rothé F, Rorive S, Decaestecker C, Larsimont D,
732   Piccart-Gebhart M, Biganzoli E, Sotiriou C, and Willard-Gallo K. 2017. Reliability of
733   tumor-infiltrating lymphocyte and tertiary lymphoid structure assessment in human
734   breast cancer. *Mod Pathol* 30:1204-1212. 10.1038/modpathol.2017.43

735 21 Calderaro J, Petitprez F, Becht E, Laurent A, Hirsch TZ, Rousseau B, Luciani A,
736   Amaddeo G, Derman J, Charpy C, Zucman-Rossi J, Fridman WH, and Sautès-Fridman
737   C. 2019. Intra-tumoral tertiary lymphoid structures are associated with a low risk of early
738   recurrence of hepatocellular carcinoma. *J Hepatol* 70:58-65. 10.1016/j.jhep.2018.09.003
739 22 Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, Marshall KA,
740   Phillippy KH, Sherman PM, Holko M, Yefanov A, Lee H, Zhang N, Robertson CL, Serova
741   N, Davis S, and Soboleva A. 2013. NCBI GEO: archive for functional genomics data
742   sets--update. *Nucleic Acids Res* 41:D991-995. 10.1093/nar/gks1193
743 23 Blum A, Wang P, and Zenklusen JC. 2018. SnapShot: TCGA-Analyzed Tumors. *Cell*
744   173:530. 10.1016/j.cell.2018.03.059
745 24 Mangiola S, Doyle MA, and Papenfuss AT. 2021. Interfacing Seurat with the R tidy
746   universe. *Bioinformatics* 37:4100-4107. 10.1093/bioinformatics/btab404
747 25 Becht E, McInnes L, Healy J, Dutertre CA, Kwok IWH, Ng LG, Ginhoux F, and Newell
748   EW. 2018. Dimensionality reduction for visualizing single-cell data using UMAP. *Nat*
749   *Biotechnol*. 10.1038/nbt.4314
750 26 Hu C, Li T, Xu Y, Zhang X, Li F, Bai J, Chen J, Jiang W, Yang K, Ou Q, Li X, Wang P,
751   and Zhang Y. 2023. CellMarker 2.0: an updated database of manually curated cell
752   markers in human/mouse and web tools based on scRNA-seq data. *Nucleic Acids Res*
753   51:D870-d876. 10.1093/nar/gkac947
754 27 Franzén O, Gan LM, and Björkegren JLM. 2019. PanglaoDB: a web server for
755   exploration of mouse and human single-cell RNA sequencing data. *Database (Oxford)*
756   2019. 10.1093/database/baz046
757 28 Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, and Tamayo P. 2015.
758   The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst*
759   1:417-425. 10.1016/j.cels.2015.12.004
760 29 Sautès-Fridman C, Petitprez F, Calderaro J, and Fridman WH. 2019. Tertiary lymphoid
761   structures in the era of cancer immunotherapy. *Nat Rev Cancer* 19:307-325.
762   10.1038/s41568-019-0144-6
763 30 Scire J, Huisman JS, Grosu A, Angst DC, Lison A, Li J, Maathuis MH, Bonhoeffer S, and
764   Stadler T. 2023. estimateR: an R package to estimate and monitor the effective
765   reproductive number. *BMC Bioinformatics* 24:310. 10.1186/s12859-023-05428-4
766 31 Hänzelmann S, Castelo R, and Guinney J. 2013. GSVA: gene set variation analysis for
767   microarray and RNA-seq data. *BMC Bioinformatics* 14:7. 10.1186/1471-2105-14-7
768 32 Liu CJ, Hu FF, Xie GY, Miao YR, Li XW, Zeng Y, and Guo AY. 2023. GSCA: an
769   integrated platform for gene set cancer analysis at genomic, pharmacogenomic and
770   immunogenomic levels. *Brief Bioinform* 24. 10.1093/bib/bbac558
771 33 Yang W, Soares J, Greninger P, Edelman EJ, Lightfoot H, Forbes S, Bindal N, Beare D,
772   Smith JA, Thompson IR, Ramaswamy S, Futreal PA, Haber DA, Stratton MR, Benes C,
773   McDermott U, and Garnett MJ. 2013. Genomics of Drug Sensitivity in Cancer (GDSC): a
774   resource for therapeutic biomarker discovery in cancer cells. *Nucleic Acids Res*
775   41:D955-961. 10.1093/nar/gks1111
776 34 Yenyuwadee S, Aliazis K, Wang Q, Christofides A, Shah R, Patsoukis N, and Boussiotis
777   VA. 2022. Immune cellular components and signaling pathways in the tumor
778   microenvironment. *Semin Cancer Biol* 86:187-201. 10.1016/j.semcancer.2022.08.004
779 35 Nassar D, and Blanpain C. 2016. Cancer Stem Cells: Basic Concepts and Therapeutic
780   Implications. *Annu Rev Pathol* 11:47-76. 10.1146/annurev-pathol-012615-044438
781 36 Vilar E, and Gruber SB. 2010. Microsatellite instability in colorectal cancer-the stable
782   evidence. *Nat Rev Clin Oncol* 7:153-162. 10.1038/nrclinonc.2009.237
783 37 Chan TA, Yarchoan M, Jaffee E, Swanton C, Quezada SA, Stenzinger A, and Peters S.
784   2019. Development of tumor mutation burden as an immunotherapy biomarker: utility for
785   the oncology clinic. *Ann Oncol* 30:44-56. 10.1093/annonc/mdy495

786 38 Waldman AD, Fritz JM, and Lenardo MJ. 2020. A guide to cancer immunotherapy: from
787   T cell basic science to clinical practice. *Nat Rev Immunol* 20:651-668. 10.1038/s41577-
788   020-0306-5
789 39 Marisa L, Svrcek M, Collura A, Becht E, Cervera P, Wanherdrick K, Buhard O,
790   Goloudina A, Jonchère V, Selves J, Milano G, Guenot D, Cohen R, Colas C, Laurent-
791   Puig P, Olschwang S, Lefèvre JH, Parc Y, Boige V, Lepage C, André T, Fléjou JF,
792   Dérangère V, Ghiringhelli F, de Reynies A, and Duval A. 2018. The Balance Between
793   Cytotoxic T-cell Lymphocytes and Immune Checkpoint Expression in the Prognosis of
794   Colon Tumors. *J Natl Cancer Inst* 110. 10.1093/jnci/djx136
795 40 Qi L, Ye C, Zhang D, Bai R, Zheng S, Hu W, and Yuan Y. 2021. The Effects of
796   Differentially-Expressed Homeobox Family Genes on the Prognosis and HOXC6 on
797   Immune Microenvironment Orchestration in Colorectal Cancer. *Front Immunol*
798   12:781221. 10.3389/fimmu.2021.781221
799 41 Kaur A, Webster MR, Marchbank K, Behera R, Ndoye A, Kugel CH, 3rd, Dang VM,
800   Appleton J, O'Connell MP, Cheng P, Valiga AA, Morissette R, McDonnell NB, Ferrucci L,
801   Kossenkov AV, Meeth K, Tang HY, Yin X, Wood WH, 3rd, Lehrmann E, Becker KG,
802   Flaherty KT, Frederick DT, Wargo JA, Cooper ZA, Tetzlaff MT, Hudgens C, Aird KM,
803   Zhang R, Xu X, Liu Q, Bartlett E, Karakousis G, Eroglu Z, Lo RS, Chan M, Menzies AM,
804   Long GV, Johnson DB, Sosman J, Schilling B, Schadendorf D, Speicher DW, Bosenberg
805   M, Ribas A, and Weeraratna AT. 2016. sFRP2 in the aged microenvironment drives
806   melanoma metastasis and therapy resistance. *Nature* 532:250-254.
807   10.1038/nature17392
808 42 Lin Q, Cao J, Du X, Yang K, Yang X, Liang Z, Shi J, and Zhang J. 2022. CYP1B1-
809   catalyzed 4-OHE2 promotes the castration resistance of prostate cancer stem cells by
810   estrogen receptor α-mediated IL6 activation. *Cell Commun Signal* 20:31.
811   10.1186/s12964-021-00807-x
812 43 Shu YJ, Weng H, Ye YY, Hu YP, Bao RF, Cao Y, Wang XA, Zhang F, Xiang SS, Li HF,
813   Wu XS, Li ML, Jiang L, Lu W, Han BS, Jie ZG, and Liu YB. 2015. SPOCK1 as a
814   potential cancer prognostic marker promotes the proliferation and metastasis of
815   gallbladder cancer cells by activating the PI3K/AKT pathway. *Mol Cancer* 14:12.
816   10.1186/s12943-014-0276-y
817 44 Zhang Y, Du W, Chen Z, and Xiang C. 2017. Upregulation of PD-L1 by SPP1 mediates
818   macrophage polarization and facilitates immune escape in lung adenocarcinoma. *Exp
819   Cell Res* 359:449-457. 10.1016/j.yexcr.2017.08.028
820 45 Lin SC, Liao YC, Chen PM, Yang YY, Wang YH, Tung SL, Chuang CM, Sung YW, Jang
821   TH, Chuang SE, and Wang LH. 2022. Periostin promotes ovarian cancer metastasis by
822   enhancing M2 macrophages and cancer-associated fibroblasts via integrin-mediated
823   NF-κB and TGF-β2 signaling. *J Biomed Sci* 29:109. 10.1186/s12929-022-00888-x
824 46 La Fleur L, Botling J, He F, Pelicano C, Zhou C, He C, Palano G, Mezheyeuski A, Micke
825   P, Ravetch JV, Karlsson MCI, and Sarhan D. 2021. Targeting MARCO and IL37R on
826   Immunosuppressive Macrophages in Lung Cancer Blocks Regulatory T Cells and
827   Supports Cytotoxic Lymphocyte Function. *Cancer Res* 81:956-967. 10.1158/0008-
828   5472.Can-20-1885
829 47 Liu B, Cheng L, Gao H, Zhang J, Dong Y, Gao W, Yuan S, Gong T, and Huang W. 2023.
830   The biology of VSIG4: Implications for the treatment of immune-mediated inflammatory
831   diseases and cancer. *Cancer Lett* 553:215996. 10.1016/j.canlet.2022.215996
832 48 Yang H, Sun B, Fan L, Ma W, Xu K, Hall SRR, Wang Z, Schmid RA, Peng RW, Marti
833   TM, Gao W, Xu J, Yang W, and Yao F. 2022. Multi-scale integrative analyses identify
834   THBS2(+) cancer-associated fibroblasts as a key orchestrator promoting aggressiveness
835   in early-stage lung adenocarcinoma. *Theranostics* 12:3104-3130. 10.7150/thno.69590

836 49 Zhao T, Li Y, Shen K, Wang Q, and Zhang J. 2021. Knockdown of OLR1 weakens
837   glycolytic metabolism to repress colon cancer cell proliferation and chemoresistance by
838   downregulating SULT2B1 via c-MYC. *Cell Death Dis* 13:4. 10.1038/s41419-021-04174-
839   w
840 50 Li T, Huang H, Shi G, Zhao L, Li T, Zhang Z, Liu R, Hu Y, Liu H, Yu J, and Li G. 2018.
841   TGF-β1-SOX9 axis-inducible COL10A1 promotes invasion and metastasis in gastric
842   cancer via epithelial-to-mesenchymal transition. *Cell Death Dis* 9:849. 10.1038/s41419-
843   018-0877-2
844 51 Fitzgerald AA, and Weiner LM. 2020. The role of fibroblast activation protein in health
845   and malignancy. *Cancer Metastasis Rev* 39:783-803. 10.1007/s10555-020-09909-3
846 52 Bassez A, Vos H, Van Dyck L, Floris G, Arijs I, Desmedt C, Boeckx B, Vanden Bempt M,
847   Nevelsteen I, Lambein K, Punie K, Neven P, Garg AD, Wildiers H, Qian J, Smeets A,
848   and Lambrechts D. 2021. A single-cell map of intratumoral changes during anti-PD1
849   treatment of patients with breast cancer. *Nat Med* 27:820-832. 10.1038/s41591-021-
850   01323-8
851 53 Ding P, Xu Y, Li L, Lv X, Li L, Chen J, Zhou D, Wang X, Wang Q, Zhang W, Liao T, Ji
852   QH, Lei QY, and Hu W. 2022. Intracellular complement C5a/C5aR1 stabilizes β-catenin
853   to promote colorectal tumorigenesis. *Cell Rep* 39:110851. 10.1016/j.celrep.2022.110851
854 54 Bancaro N, Calì B, Troiani M, Elia AR, Arzola RA, Attanasio G, Lai P, Crespo M, Gurel
855   B, Pereira R, Guo C, Mosole S, Brina D, D'Ambrosio M, Pasquini E, Spataro C, Zagato
856   E, Rinaldi A, Pedotti M, Di Lascio S, Meani F, Montopoli M, Ferrari M, Gallina A, Varani
857   L, Pereira Mestre R, Bolis M, Gillessen Sommer S, de Bono J, Calcinotto A, and Alimonti
858   A. 2023. Apolipoprotein E induces pathogenic senescent-like myeloid cells in prostate
859   cancer. *Cancer Cell* 41:602-619.e611. 10.1016/j.ccell.2023.02.004
860

Figure 1

Flowchart of this study

(A) The CRC_EMTAB8107 scRNA-seq dataset was downloaded from the GEO Database (https://www.ncbi.nlm.nih.gov/geo/), and the Hallmark pathways were retrieved from the MsigDB database (https://www.gsea-msigdb.org/gsea/msigdb/index.jsp). (B-E) The GSE17538 and GSE39582 datasets were downloaded from the GEO database, and the COADREAD dataset was retrieved from the TCGA database (https://portal.gdc.cancer.gov). Hallmark, KEGG, and Reactome pathways were all sourced from the MsigDB database. (F) The datasets for analyzing MSI, TMB, CSCs, and somatic mutations were sourced from the TCGA_COADREAD database. (G) The immunotherapy datasets GSE135222 and GSE176307 were obtained from the GEO database. Drug sensitivity analyses were performed using the GSE17538, GSE39582 datasets from the GEO database, the COADREAD dataset from TCGA, and data from the CancerRxGene database (https://www.cancerrxgene.org/). (H) The immunohistochemistry results were retrieved from the HPA database (https://www.proteinatlas.org).

A

Construction of the TLS score in Single-cell transcriptome

B

Construction of the TLS molecular clusters

C

Characteristics of the TLS clusters

D

Construction of the TLS gene clusters

E

Construction of the TLS score

F

Landscape of the TLS score

G

Immunotherapy and drug susceptibility

H

Experiment verification

Figure 2

Figure 2

Figure 2 Construction of TLS score revealing high cellular heterogeneity in COADREAD based on single cell RNA sequencing (scRNA-seq) data. (A) Reduced dimensionality and cluster analysis. The COADREAD dataset cells might be classified into 19 clusters by UMAP, which included B cells, dendritic cells, endothelial cells, fibroblasts, malignant cells, mono-macrophages, mast cells, plasma cells, CD8+T cells and CD4+T cells. (B) The volcano plot displaying differential expression genes across various cell types, with the top five shown. (C) The heatmap displayed the score of significantly enriched Hallmark pathways across each tumor infiltrating cell type. The bar represented the row-scaled pathway enrichment level. (D) The dot plot displayed the expression levels of 39 TLSRGs across various cell types. (E) Identification and landscape of TLS score. The TLS score of each cell type was calculated by GSVA. The expression of TLS gene cluster in each cell type was visualized by UMAP plot. (F) Grouping by TLS score. The cell clusters were divided into high- and low score groups and displayed by UMAP plot. (G) Comparison of TLS score among cell clusters. Box plots showed the TLS score of every cell cluster and were sorted by ascending order. (H-I) Quantization of cell number and proportion in TLS score clusters. Cell number and proportion of each cell cluster was compared between high- and low- TLS groups by composition chart. (J) Correlation of TLS score with Hallmark pathways. The heatmap showed the correlation of TLS score with Hallmark pathway scores across all cell clusters. (K) Enrichment of Hallmark pathways. The comparison of Hallmark pathway score between high- and low- score group was showed by heatmap.
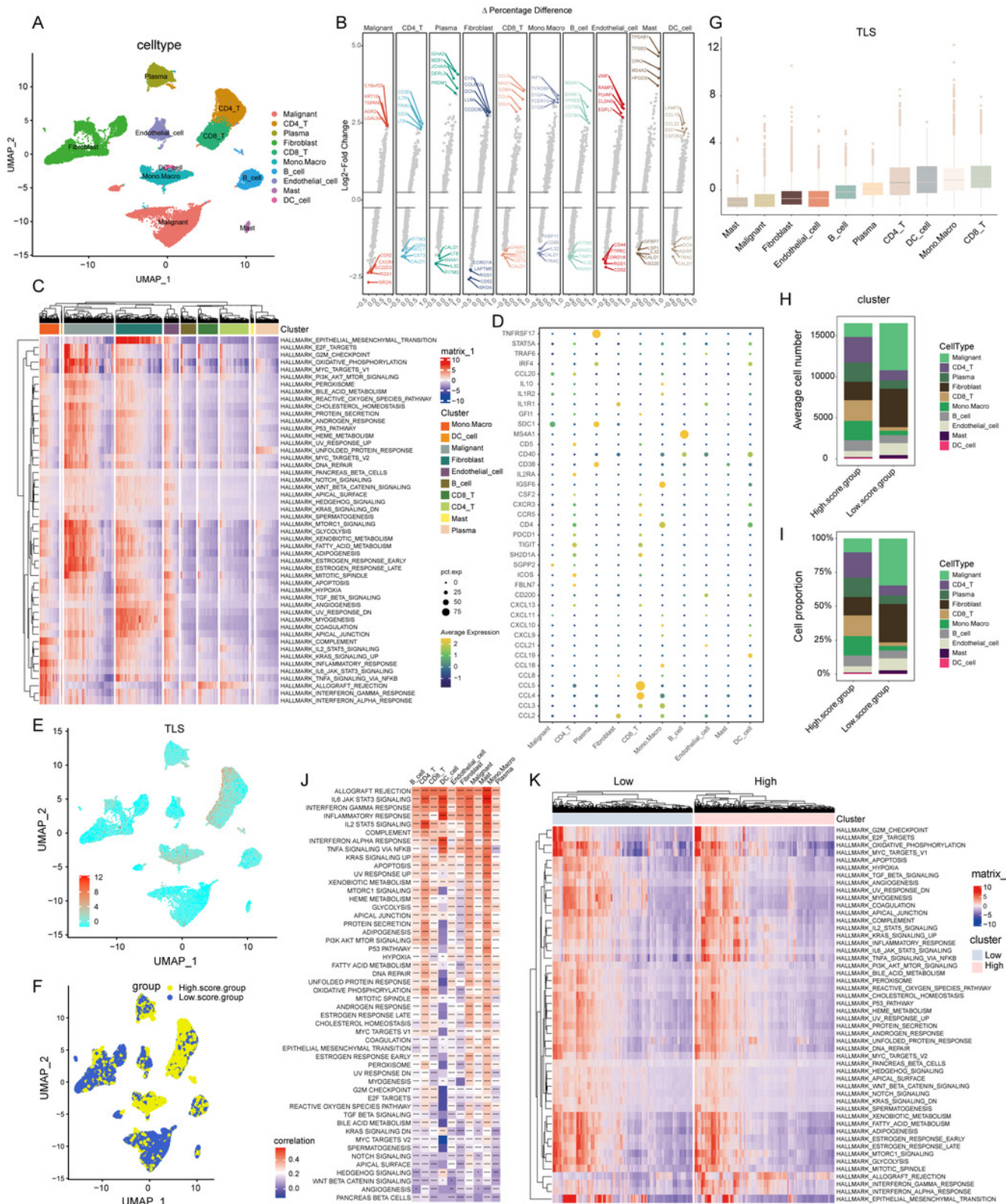
Figure 3

Figure 3

Figure 3 Construction and biological analysis of TLS molecular subtypes in colorectal carcinoma. (A) UniCox regression analysis and mutual correlations among TLSRGs in 1184 colorectal cancer samples. Spearman correlation analyses were used. The line between two TLSRGs indicated their interaction, and the stronger the correlation, the thicker the line. green line represented positive correlation and yellow line represented negative correlation. (B) Kaplan–Meier plot and log-rank tests were conducted for survival analyses of those TLSRGs (CCL2, CCL20, CXCR3, IL1R1, TIGIT, SGPP2, ICOS, CCL8, CXCL9, CXCL11, CXCL10, CXCL13). (C) Survival analyses for two different clusters. (D) Consensus clustering matrix for k = 2. (E) The abundance of TLSRGs in two clusters. (F) The heat-map showed the associations of clinicopathologic characteristics with molecular subtypes. Purple color indicated up-regulation and green color indicated down-regulation. GSVA enrichment analyses of Hallmark (G), KEGG (H), and Reactome (I) pathways in subtype A and B. Purple color indicated more enriched pathways and green color indicated less enriched pathways. (J) PCA presented a great difference in transcriptomes between different TLS molecular subtypes. (K) Differences in the stromal, immune and ESTIMATE score between different TLS clusters. (L) The abundance of tumor infiltrating immune cells in cluster A and B.
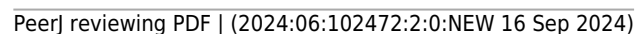
Figure 4

Figure 4

Figure 4 Construction of TLS gene subtypes and TLS score, following integrative analysis. (A)189 TLS-DEGs shown in the volcano plot. (B) Functional annotation for TLS-DEGs using GO enrichment analysis. The size of the plots represented the number of genes enriched. (B) The pathways were grouped by biological process (BP), cellular component (CC), and molecular function (MF). (C) Functional annotation for TLS-DEGs using KEGG enrichment analysis. The size of the plots represented the number of genes enriched. (D) The chord graph showed the 5 vital pathways and corresponding genes of KEGG analysis. (E) Identification of two gene subtypes (k=2) and their correlation area through consensus clustering analysis of the 16 genes. (F) Survival analysis of the two gene subtypes. (G) Expression of the 16 genes in different subtypes. (H) The heat-map showed the associations of clinicopathologic characteristics with gene subtypes. (I) TLS scores were calculated based on the 16 TLS-DEGs by GSVA, and survival analyses were performed in high- and low-score group. (J) Sankey diagram displayed the relationship of molecular subtypes, gene subtypes, TLS scores and survival outcomes. (K) The TLS score was positively associated with all over tumor infiltrating immune cells. (L) GSVA analysis for cytokines, chemokines, and their receptors in high- and low-TLS score clusters. (M) Correlation of TLS score with 50 hallmark pathway scores by GSVA.
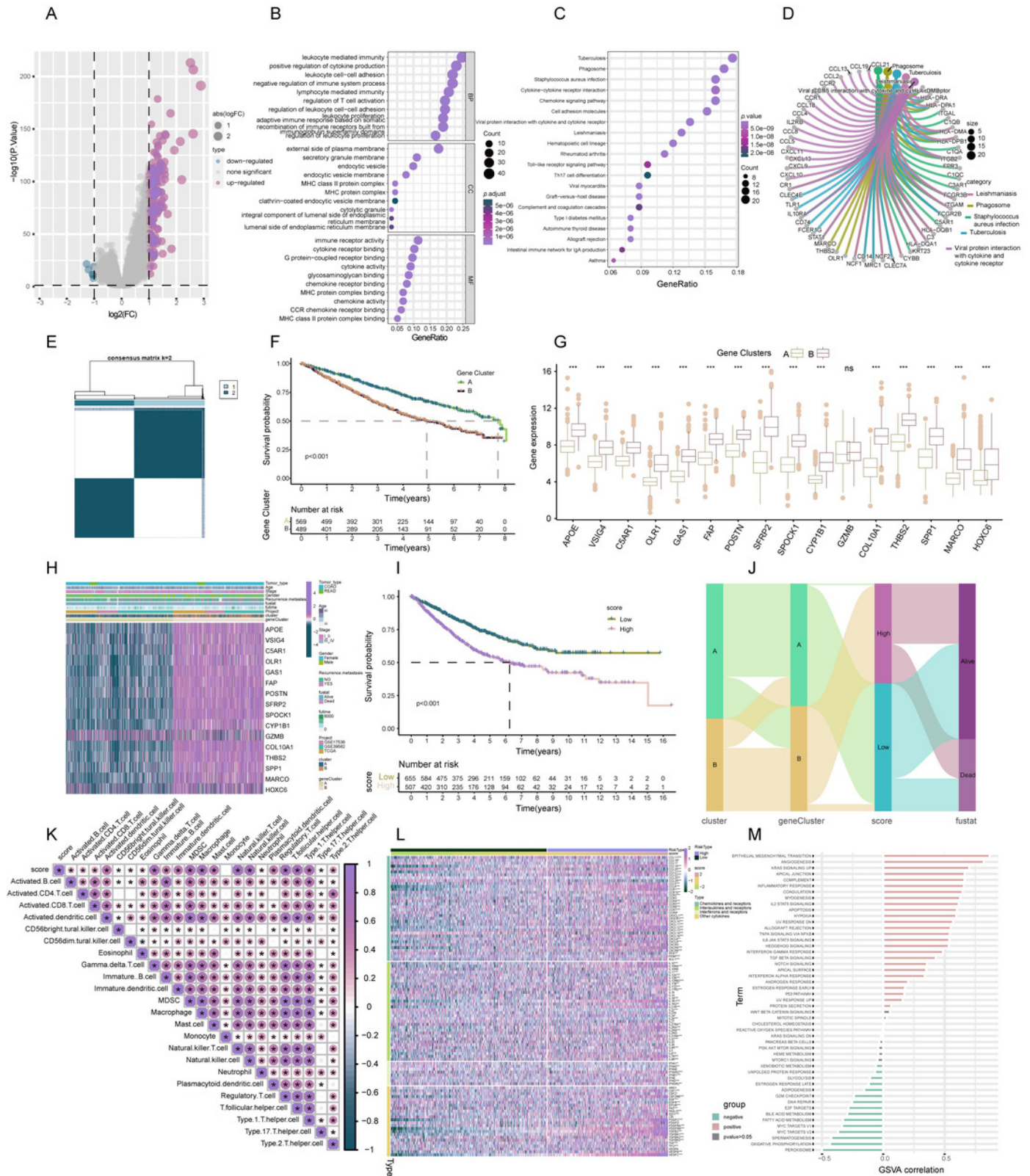
Figure 5

Figure 5

Figure 5 Mutations landscape, Clinical significance, Immunotherapy response and drug susceptibility of TLS score. Correlation (A), expression (B) and the distribution (C) of TLS score with mDNAsi index. Correlation (D), expression (E) and the distribution (F) of TLS score with mRNAsi index. Correlation (G), expression (H) and the distribution (I) of TLS score with MSI index. Correlation (J), expression (K) and the distribution (L) of TLS score with TMB index. The waterfall plot of somatic mutation characteristics in high-and low-TLS score groups. High score group contained 172 COADREAD samples (M) and low score group contained 198 COADREAD samples (N). (O) Mutational variation in high-and low-score clusters, Forest plot shows the HRs of 2 score clusters. Differences between TLS score clusters in fustat (P), gender (Q), recurrence/ metastasis (R), stage (S) and the expression of immune checkpoints (T). Difference of survival analyses (U), response to anti-PD-1/PD-L1 therapy (V) between low- and high-TLS clusters in non-small cell lung carcinoma cohort (GSE135222). Differences of survival analyses (W), response to anti-PD-1 therapy (X) between low- and high-TLS clusters in Metastatic Urothelial carcinoma cohort (GSE176307). (Y) The Box diagram showed the differences of drug sensitivity (IC50) to targeted therapy between high-and low-TLS score clusters.
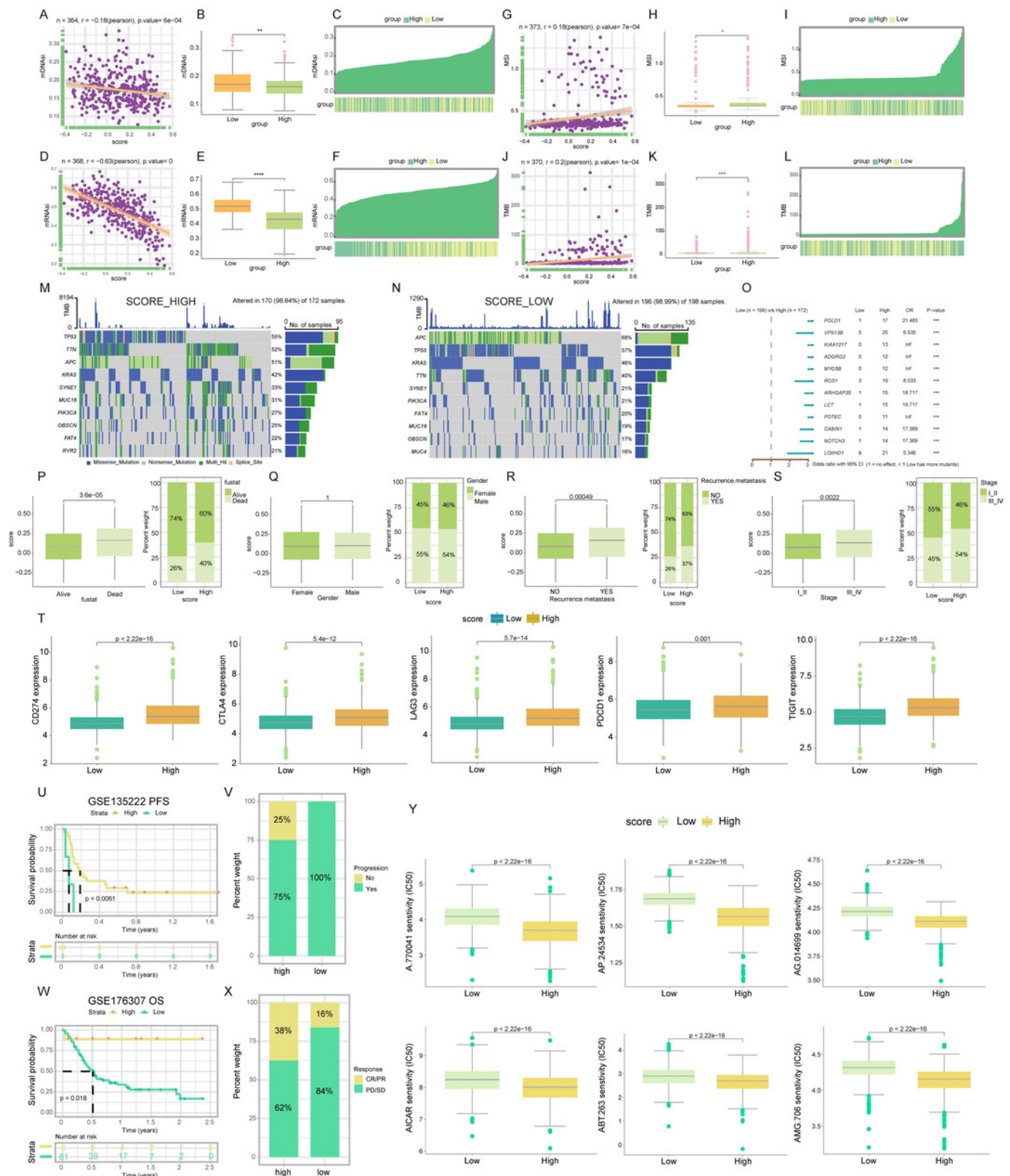
Figure 6

Cell experiment and tissue expression verification.

Validation of the expression of C5AR1 (A), APOE (B), CYP1B1 (C), SPP1 (D) in a normal colon cell line (NCM460) and three COADREAD cell lines (HCT116, HCT29, and VOLO) by qRT-PCR. (EL) CCK8 assay. After knockdown of C5AR1 (E, I), APOE (F, J), CYP1B1 (G, K), and SPP1 (H, L), the HCT 116 cells and HT 29 cells showed significant reduction in viability separately. The invasion capacity of HCT116 cells and HCT29 cells decreased significantly after APOE, C5AR1, CYP1B1, and SPP1 knockdown (M-N). Immunohistochemistry showing the protein expressions of APOE, C5AR, CYP1B1, and SPP1 based on the Human Protein Atlas (HPA) database (O). After knockdown of APOE, C5AR, CYP1B1, and SPP1, the HCT 116 cells and HT 29 cells showed significant reduction of PD-L1 expression separately (P-Q).