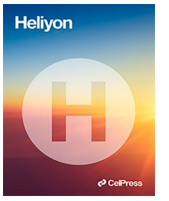
Heliyon 10 (2024) e37726

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon



Research article

Germinal center B-cell subgroups in the tumor microenvironment cannot be overlooked: Their involvement in prognosis, immunotherapy response, and treatment resistance in head and neck squamous carcinoma

Li Lin [a,b,1], Jiani Zou [d,1], Shengbin Pei [e,1], Wenyi Huang [b], Yichi Zhang [b], Zhijie Zhao [b,*], Yantao Ding [c,f,**], Can Xiao [a,***]

[a] Department of Stomatology, the First Affiliated Hospital of Soochow University, 188 Shi Zi Rd, Suzhou, 215006, China
[b] Department of Plastic and Reconstructive Surgery, Shanghai 9th People's Hospital, School of Medicine, Shanghai Jiao Tong University, 639 Zhi Zao Ju Rd, Shanghai, 200011, China
[c] Institute of Dermatology and Department of Dermatology, The First Affiliated Hospital, Anhui Medical University, No. 81 Meishan Road, Hefei, Anhui, 230032, China
[d] China Eastern Airlines, Comprehensive Management Department, Aviation Health Department, China
[e] Department of Breast Surgical Oncology, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China
[f] China bKey Laboratory of Dermatology, Anhui Medical University, Ministry of Education, Hefei, Anhui, 230032, China

ARTICLE INFO

Keywords:
Head and neck squamous carcinoma
Germinal center B cells
Single-cell RNA-Seq
Immune cell infiltration
Drug sensitivity

ABSTRACT

Background: More than 60 % of patients with head and neck squamous carcinoma (HNSCC) are diagnosed at advanced stages and miss radical treatment. This has prompted the need to find new biomarkers to achieve early diagnosis and predict early recurrence and metastasis of tumors.
Methods: Single-cell RNA sequencing (scRNA-seq) data from HNSCC tissues and peripheral blood samples were obtained through the Gene Expression Omnibus (GEO) database (GSE164690) to characterize the B-cell subgroups, differentiation trajectories, and intercellular communication networks in HNSCC and to construct a prognostic model of the associated risks. In addition, this study analyzed the differences in clinical features, immune cell infiltration, functional enrichment, tumor mutational burden (TMB), and drug sensitivity between the high- and low-risk groups.
Results: Using scRNA-seq of HNSCC, we classified B and plasma cells into a total of four subgroups: naive B cells (NBs), germinal center B cells (GCBs), memory B cells (MBs), and plasma cells (PCs). Pseudotemporal trajectory analysis revealed that NBs and GCBs were at the early stage of B cell differentiation, while MBs and PCs were at the end. Cellular communication revealed that GCBs acted on tumor cells through the CD99 and SEMA4 signaling pathways. The independent prognostic value, immune cell infiltration, TMB and drug sensitivity assays were validated for the MEF2B^{+} GCB score groups.

* Corresponding author.
** Corresponding author.
*** Corresponding author.
E-mail addresses: 15836015662@163.com (Z. Zhao), ding2008yantao@163.com (Y. Ding), canxiao0511@163.com (C. Xiao).
[1] Co-First Author.

https://doi.org/10.1016/j.heliyon.2024.e37726
Received 5 March 2024; Received in revised form 9 September 2024; Accepted 9 September 2024
Available online 11 September 2024

Conclusions: We identified GCBs as B cell-specific prognostic biomarkers for the first time. The $MEF2B^{+}$ GCB score fills the research gap in the genetic prognostic prediction model of HNSCC and is expected to provide a theoretical basis for finding new therapeutic targets for HNSCC.

1. Introduction

Head and neck squamous carcinoma (HNSCC), an immunosuppressive disease characterized by molecular heterogeneity and tumor-host interactions, has been increasing in incidence and mortality year by year and has now become the sixth most common cancer and the eighth leading cause of cancer deaths worldwide [1]. About 90 % of head and neck malignancies are squamous cell carcinomas, and early diagnosis and treatment can lead to a better prognosis [2,3]. The diagnosis and treatment of HNSCC have made significant progress in the past few decades. Surgical treatment, radiotherapy, and chemotherapy can significantly improve the disease control rate and survival of patients with HNSCC. However, the overall survival (OS) of patients with HNSCC is still relatively low, with the 5-year survival rate at all stages around 40 %–50 %, mainly due to the recurrence of tumors and metastasis after treatment [4–6]. However, in the treatment of HNSCC, due to the specificity of the tumor site, surgical treatment may change the corresponding anatomical structure, and radiotherapy will form fibrous scars in the corresponding area, thus making it particularly difficult to achieve early detection of local recurrence directly through physical examination or re-biopsy [7]. The rest of the diagnostic tools, such as Computed Tomography (CT), ultrasound, and Magnetic Resonance Imaging (MRI), are not sensitive to early lesions, so we must rely on specific biomarkers to predict early recurrence and metastasis of tumors. Improved survival in patients with HNSCC depends partly on disease stage, lymph node involvement, patient response to treatment, and early diagnosis of local tumor recurrence. Nevertheless, it has been suggested that the clinical staging of tumors is not a good reflection of the behavior of the tumor and the outcome of the disease [8]. This is why a growing number of studies have concluded, by evaluating clinical and pathological indicators, that non-advanced tumors tend to be accompanied by early recurrence and that advanced tumors tend to have longer disease-free intervals [9]. These factors have prompted the need to search for essential tumor markers for preoperative assessment of the likelihood of regional metastasis, the significance of postoperative adjuvant therapy, assessment of the patient's response to treatment, and prognostic analysis of the patient.

Tumor immune microenvironment (TIME), which consists of immune-related components such as immune cells and their secreted immunomodulatory molecules, has become a hot topic in oncology research in recent years due to its critical regulatory role in the growth of tumors as well as metastasis [10,11]. As an essential component of TIME, there is growing evidence that B cells can participate in the immune response to tumors as well as immune evasion through the secretion of specific antibodies, the processing of presenting antigens, and the maintenance of the structure and function of tertiary lymphatic structures, which are biologically important [12,13]. However, research on TIME has focused mainly on T cells, with relatively few B cell-related studies [14]. The role of B cells is dual. On the one hand, most B cells can inhibit tumor growth through antibody production, secretion of anti-tumor cytokines, and direct killing. On the other hand, some B cell subgroups may also play a tumor-promoting role. For example, a group of regulatory B cell subgroups (Bregs) may promote tumor growth and invasion through the secretion of specific tumor-promoting cytokines (e.g., IL-10, TGF-β, IL-35) and other mechanisms [15]. In addition, B cells are closely associated with the prognosis of various tumors, including skin melanoma, gastric cancer, and lung cancer [16,17]. An early study of 33 HNSCC patients demonstrated improved outcomes with increased peritumoral B-cells in lymph node metastases. Recent analyses of B-cell phenotypes and responses in HNSCC have delineated notable heterogeneity [18]. Nevertheless, the impact of B cells on survival or their influence on responses to immunotherapy or radiation remained unexplored [19,20], possibly due to the lack of uniformity in the choice of B cell markers.

In recent decades, the emergence of bulk RNA-sequencing (bulk RNA-seq) has revolutionized oncology research and has been widely used in diagnosing and prognostic assessing cancer patients. However, there are individual differences between different cell types in tumor tissues. Still, bulk sequencing can only reflect the overall average gene expression level in the whole sample and cannot accurately characterize the gene expression at the individual cell level. Single-cell RNA-sequencing (scRNA-seq) allows access to the genetic signature of a particular cell type by extracting, reverse transcribing, amplifying, and sequencing RNA at the level of a single cell, better-revealing heterogeneity in the intra-tumor microenvironment and providing assistance in the development of individualized therapeutic regimens [21]. Although scRNA-seq technology can characterize genes in different cell types more accurately than traditional bulk RNA-seq, bulk RNA-seq datasets have larger sample sizes and more complete follow-up information. Therefore, there is a need to integrate scRNA-seq and bulk RNA-seq datasets to identify new tumor biomarkers and develop more accurate tumor prognosis-related models.

In summary, the biological functions of B cells are complex, and their prognostic value in HNSCC remains unknown. Screening for appropriate B-cell markers in HNSCC and in-depth exploration of their biological functions and prognostic value could help to assess the prognosis. It may provide new targets for the treatment of patients with HNSCC.

2. Materials and method

2.1. Data collection and processing

The GPL18573 platform-based Head and neck squamous cell carcinomas (HNSCCs) -related single-cell RNA sequencing (scRNA-seq) data GSE164690 [PMID: 34921143] was downloaded from Gene Expression Omnibus (GEO, https://www.ncbi.nlm.nih.gov/geo/

) and analyzed (Fig. S1). In addition, bulk RNA sequencing data of HNSCC tissue samples were downloaded from the Cancer Genome Atlas (TCGA, https://portal.gdc.cancer.gov/) database, along with clinical information including age, gender, TNM stage, race, and survival time.

2.2. Single-cell sequencing data analysis

Using the Seurat R package (version 4.3.0), raw gene expression substrates were imported into R software (version 4.2.2). Using the DoubletFinder R package (version 2.0.3), the scRNA-seq data underwent rigorous quality control.

The percentage of mitochondrial and erythrocyte-related genes in each cell was calculated using the "PercentageFeatureSet" function in the Seurat package, followed by cell quality control, with the following exclusion criteria: (1) the number of genes in the cell was <300; (2) the number of genes in the cell was >7500; (3) the percentage of mitochondrial genes in the cell was >25.

Then use the "NormalizeData" function for normalization (normalization.method = "LogNormalize", scale.factor = 10000). Use the "CellCycleScoring" function to determine cell cycle conditions. Find the top 2000 top variable genes for subsequent analysis using the "FindVariableFeatures" function. The data are standardized using the "ScaleData" function. The harmony R package (version 0.1.1) was used to remove batch effects between samples. The dim value was set to 30 and the resolution to 1.2. The "RunPCA" function was used to cluster the dataset and to downscale the UMAP visualization [22]. Marker genes for each B-cell subgroup were compiled by searching previous literature, and the clusters were manually annotated after the downscaling of the subgroups. Next, the "FindAllMarkers" function was used to find differentially expressed genes in each B-cell subpopulation, where min.pct was set to 0.25 and logfc.threshold was set to 0.25.

2.3. Single cells copy number variation (CNV) evaluation

The inferCNV (version 1.6.0) available from the GitHub repository of the Broad Institute (https://github.com/broadinstitute/inferCNV) was used to analyze the CNV of cells (Fig. S2).

2.4. Pseudotime analysis

We performed pseudotime trajectory analysis of B and plasma cell subgroups using the Monocle R package (version 2.24.0) [23]. The UMAP method was employed to reduce dimensionality, and the functions "plot_cell_trajectory" was utilized for visualization. Different cell subpopulations were arranged in a proposed chronological order, and pseudotemporal heatmaps were used to identify and display genes that changed simultaneously with pseudotime.

2.5. Cell-cell communication analysis

The CellChat R package (version 1.6.1) was applied to infer cell-to-cell interactions among cell subgroups and establish a regulatory network based on ligand-receptor crosstalk [24]. The "netVisual_circle", "netVisual_aggregate", "netVisual_individual", "netVisual_heatmap", "netVisual_bubble", "netVisual_chord_gene", "netAnalysis_river", "netAnalysis_dot", and "netAnalysis_signalingRole_heatmap" functions were used to visualize cell-to-cell interactions.

2.6. Construction of a risk score in HNSCC patients

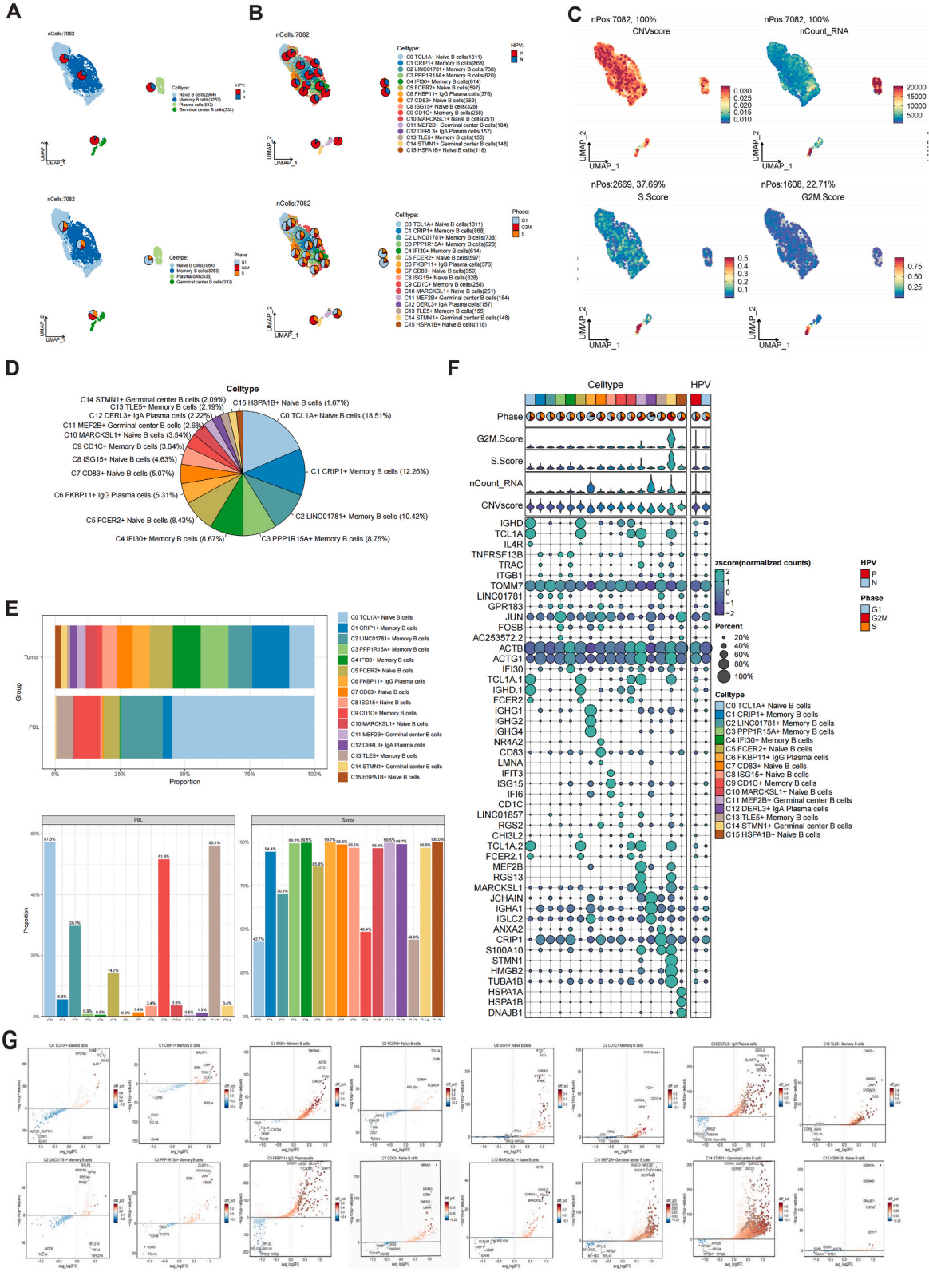
The expression values of signature genes of patients in the GDC TCGA Head and Neck Cancer (HNSC) cohort were extracted after obtaining the list of MBs or GCBs signature genes (top 100 genes). Signature genes associated with OS of HNSCC patients were screened using univariate Cox regression analysis. To avoid the multicollinearity of these genes, we further filtered them using LASSO regression (glmnet, version 4.1–6) to fileted these genes. The risk coefficients for each gene were then calculated using multivariate Cox regression. Risk scores of the prognostic risk score model (X: coefficient, Y: gene expression level) = $\sum_{i}^{n} Xi \times Yi$.

Patients were categorized into high and low-risk groups based on median of risk scores. Kaplan-Meier survival analysis (version 3.3-1) was performed in the TCGA-HNSC cohort to compare the prognosis of patients in different risk groups. The predictive performance of risk score was further evaluated by plotting the area under the curve of ROC (AUC) over time using the timeROC R package (version 0.4).

Risk score, age, TNM stage, and molecular subtypes were subjected to were subjected to multivariate Cox regression analysis to screen out independent prognostic factors in HNSCC patients. The nomogram column-line graph prediction model for predicting 1-, 3-, and 5-year overall survival of HNSCC patients was constructed using the "rms" package (version 6.7-0). The constructed nomogram prediction model was validated using the ROC curve and calibration curve.

2.7. Immune infiltration analysis

Previous literature was queried to obtain marker genes for different immune cells, and ssGSEA enrichment analysis was performed to assess the level of immune cell infiltration for each sample in the TCGA-HNSC cohort [25]. The relative proportion of immune cells was quantified using the CIBERSORT R package (version 0.1.0) [26]. The Wilcox.test was used to compare the levels of immune cell



(caption on next page)

Fig. 1. | Integration and analysis of scRNAseq data of B and plasma cells in HNSCC. (A–B) The Uniform Manifold Approximation and Projection (UMAP) of the expression profiles of the 7082 B and plasma cells derived from tumor tissues and peripheral blood (PBL) of HNSCC. B cells are classified into 16 distinct transcriptional clusters. HPV infection status (HPV^+ vs HPV^-), Phase (G1, S, G2M). **(C)** The distribution of CNVscore, nCount_RNA, S.Score, G2M.Scorce for each subgroup was visualized using UMAP. **(D)** Pie chart showing the proportions of each subgroup in the HNSCC samples. **(E)** The proportion of subgroups derived from PBL and tumor tissues. **(F)** Visualization of sample characteristics associated with different B and plasma cell subgroups and different HPV infection status (HPV^+ vs HPV^-), Phase (G1, S, G2M), G2M.Score, S.Score, nCount_RNA, CNVscore, differential gene expression. **(G)** Volcano plots showing DEGs in subgroups.

infiltration and the expression levels of joint exempted checkpoints between different risk groups. Correlation analyses between immune cell infiltration levels of each type and between immune cell infiltration levels and risk genes were performed using the "corrplot" R package (version 0.92). Stromal and immune cells in HNSCC tissues were estimated using the ESTIMATE R package (version 1.0.13), and stromal score, immune score, ESTIMATE score, and tumor purity were calculated [27].

2.8. Functional enrichment analysis

The gene expression matrices of HNSCC patients in the high and low-risk subgroups were analyzed differently by the limma R package (version 3.54.0). The GO [28] and KEGG [29] were used to analyze the DEGs between different risk subgroups by using clusterProfiler R package (version 4.6.0). We downloaded the list of pathways gene sets from the MSIGDB database (https://www.gsea-msigdb.org/) and performed GSEA [30] enrichment analyses on these pathways.

2.9. Tumor mutation burden analysis

The maftools package was utilized to analyze the overall mutation status of the tumor, TMB, as well as the transformation status of SNPs. The Wilcox.test was used to compare the TMB levels between risk subgroups of patients, and the Spearman correlation test was performed to determine the relationship between risk score and TMB.

2.10. Prediction of drug sensitivity

The Genomics of Drug Sensitivity in Cancer (GDSC) (https://www.cancerrxgene.org/) is a public dataset containing information about drug sensitivity and molecular markers of drug response in cancer [31]. The dataset information was downloaded as reference files, and the pRRophetic R package (version 0.5) [32] was used to predict the sensitivity to chemotherapy and targeted therapies for each sample. The Wilcox.test was performed to determine whether there was a statistically significant difference in medicine sensitivity between the high and low risk groups. Pearson's correlation was utilized to examine the relationship between risk scores and drug sensitivity.

2.11. Multiplexed immunofluorescence assay

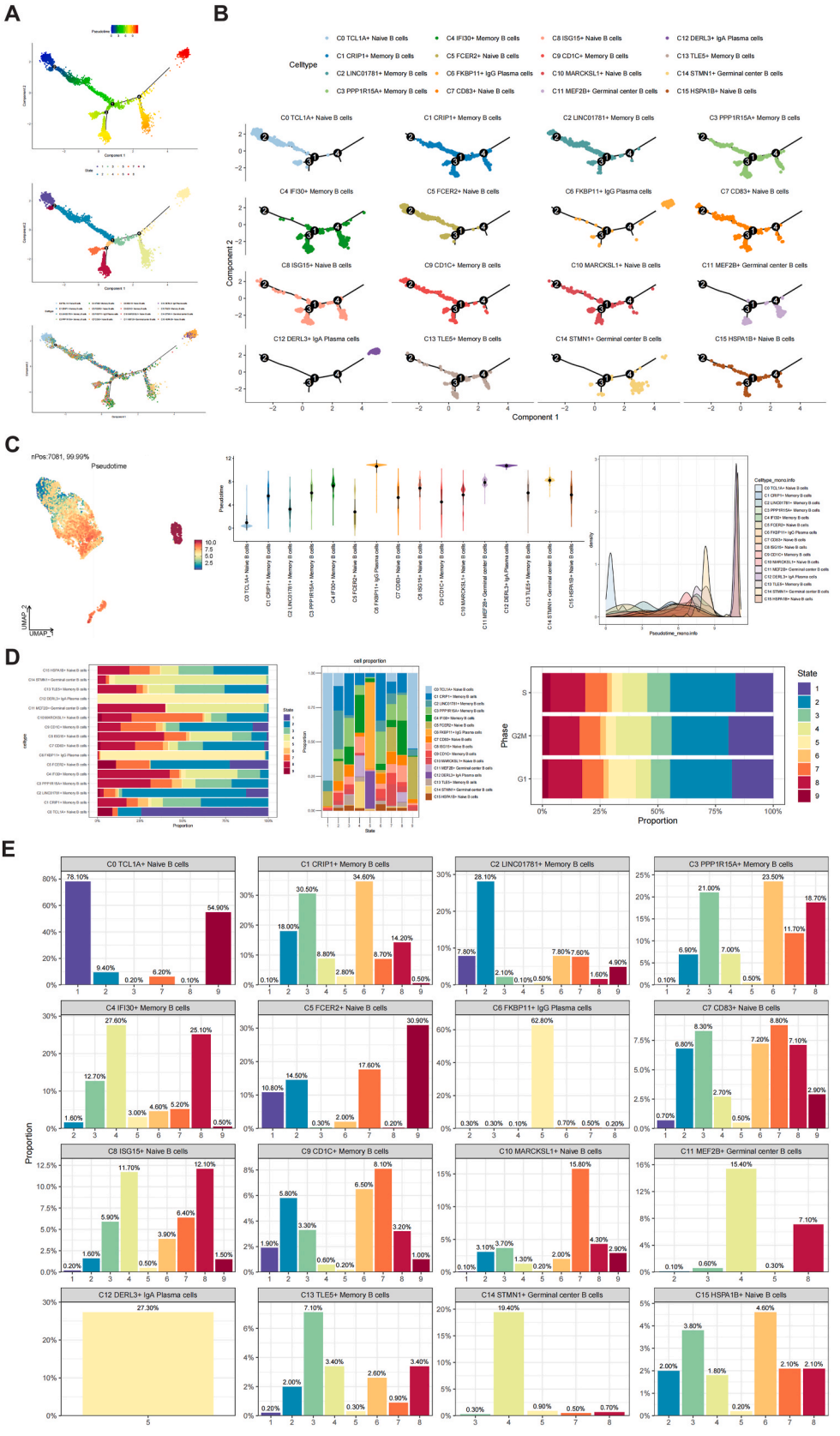
Multicolor IHC staining was performed using a Multiple Immunofluorescence Kit (AiFang biological, AFIHC035) according to the manufacturer's instructions. Briefly, the paraffin-embedded tissues were dewaxed, then the antigen retrieval was performed by sodium citrate at 98 °C for 20 min. Ater infiltrating in 3 % H_2O_2 for 15 min, the tissues were blocked by 5 % BSA. Then the tissues were incubated by primary antibody at 4 °C overnight. Then the tissues were incubated by secondary antibody for 30 min and stained by TSA for 5 min. For each targe, the steps of antigen retrieval, blocking, incubation of primary antibody, incubation of secondary antibody, and TSA staining are required. Finally, the nuclei were stained with DAPI. PBS was used to wash tissues between each step. These images were obtained by the LEICA DMi8 system and analyzed by the ImageJ software.

2.12. Cell culture

HN-5 cells and UMSCC-47 cells originated from the American Type Culture Collection (ATCC). HN-5 cells underwent cultivation utilizing RPMI 1640 medium, whereas UMSCC-47 cells were subjected to DMEM medium (Gibco BRL, United States). Both cell lines were nurtured within an incubation environment at 37 °C, featuring a 5 % CO2 concentration.

2.13. Cell transfection and real-time fluorescent quantitative PCR (RT-qPCR)

Two small interfering RNAs (siRNAs) targeting the PLXNB2 gene, along with their corresponding negative controls (si-NC), were synthesized by Ribobio Guangzhou (China). The transfection procedure adhered to the methodologies delineated in the Lipofectamine 3000 instructions provided by Invitrogen, USA. Thermo's TRIzol reagent (15596018) was employed to extract total RNA from the cellular lines, and RNA concentrations were standardized. Subsequently, cDNA synthesis was carried out using Vazyme's PrimeScript™ RT kit (R232-01). Quantitative reverse transcription-polymerase chain reaction (qRT-PCR) was executed utilizing the SYBR Green Kit from TaKaRa Biotechnology (Dalian, China), with GAPDH as the internal reference control. Further details regarding primer and siRNA sequences are available in Supplemental Table 1.



(caption on next page)

Fig. 2. | Trajectory of B and plasma cells. (A) The differentiation trajectory of B and plasma cells shows the derivation of HNSCC cells, based on the pseudotime of cells (from left to right). **(B)** Pseudotime differentiation trajectory plots of B and plasma cell subgroups. **(C)** The UMAP visualization delineates the pseudotime scores of the entire spectrum of cells. The violin plot illustrates the pseudotime scores attributed to each distinct subgroup. The ridge map provides an insight into the density distribution of pseudotime values. **(D–E)** Bar charts of the changes in the proportion of B and plasma cell subgroups in the pseudotime state, respectively.

2.14. Cell counting

A total of 6000 transfected cells were seeded into individual wells of 96-well plates (Corning, USA, 3599). Following cellular adhesion to the well surfaces, the cells underwent treatment with the CCK-8 labeling reagent (A311-01, Vazyme) on days 1, 2, 3, 4, and 5, respectively. Incubation of the cells occurred in conditions devoid of light for a duration of 2 h, during which absorbance (OD) values were meticulously recorded.

2.15. Wound healing

Transfected cells were nurtured in a six-well culture plate (Corning, USA, 3516). Upon achieving a cellular density of approximately 90–95 %, a 200 μl sterile pipette was employed to delicately abrade the cell layer in a linear fashion. Subsequently, a gradual rinsing with PBS was conducted to eliminate unattached cells and debris. Following this, the serum-free cell medium was substituted to sustain optimal cell proliferation. Consistent photographic documentation was undertaken at identical locations, both at the initiation of the culture and 48 h subsequent to the initiation.

2.16. Transwell

In the Transwell experimental setup, the Transwell chamber (Corning, USA, 341201) was positioned within a 24-well plate. Subsequently, 8000 cells were uniformly distributed in each chamber and cultured utilizing 200 μl of serum-free DMEM or RPMI 1640 medium. To the lower chamber, 600 μl of complete DMEM or RPMI 1640 medium, enriched with 10 % serum, was added. The experimental plate was then placed in an incubation apparatus and cultured for a duration of 36 h, allowing for the migration of cells from the upper chamber to the lower chamber. Following this incubation period, the Transwell chamber underwent meticulous cleaning, fixation, and staining procedures.

3. Results

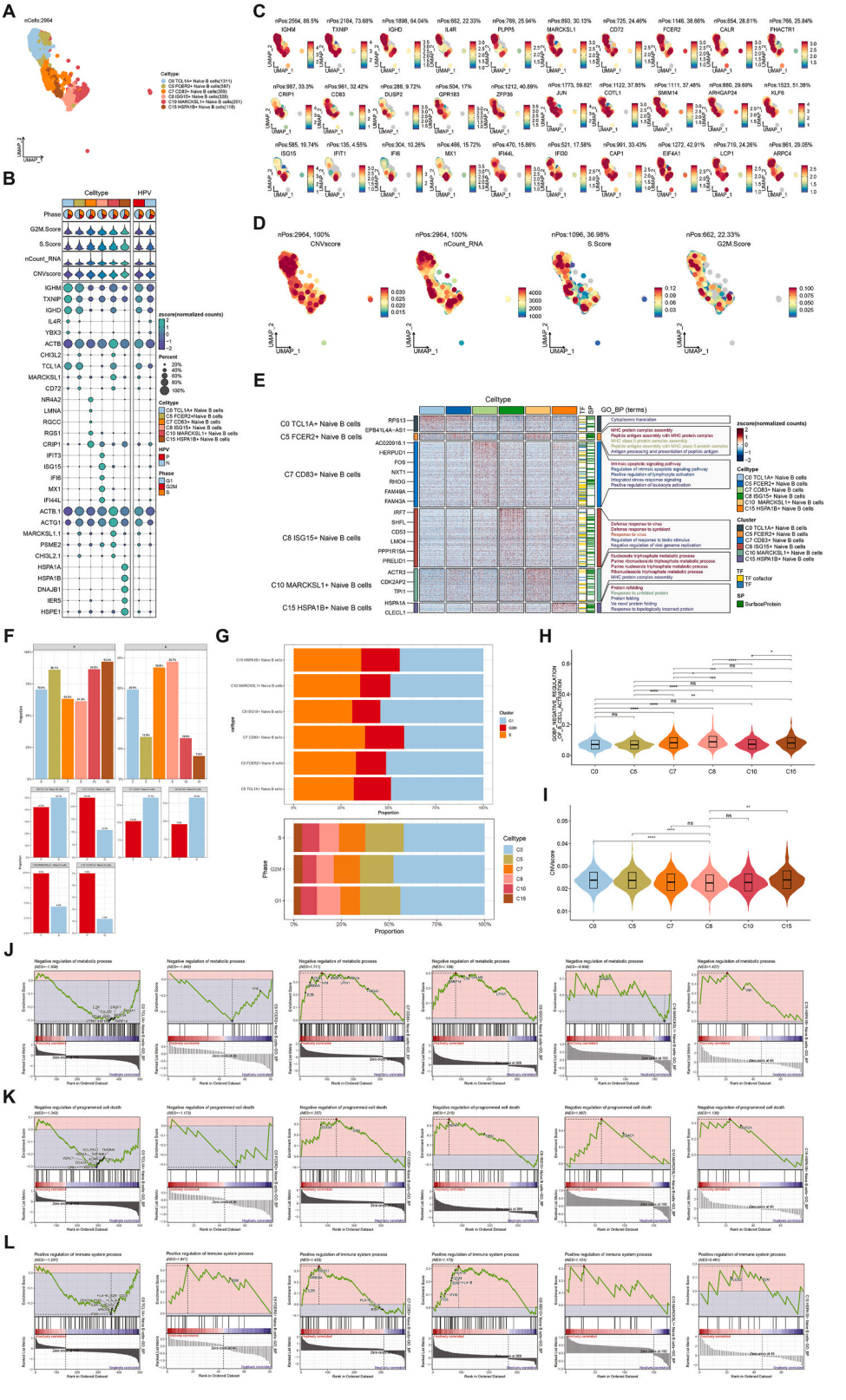
3.1. Phenotypic clusters of B cells according to scRNA-seq

To investigate the variations in the immune microenvironment across all cell subgroups in HNSCC, we collected scRNA-seq datasets of HNSCC. By utilizing these datasets, we aim to gain a more profound comprehension of the molecular mechanisms at the individual cell level.

Through the standard workflow of single-cell analysis, we identified four cell types in the integrated data, including NBs, MBs, GCBs, and PCs. We plotted the distribution of HPV infection status and phase (Fig. 1A). To further elucidate the heterogeneity and intricacy of HNSCC, we have subdivided them into 16 distinct subgroups. Among these subgroups, the population of NBs encompasses C0 TCL1A^{+} Naive B cells, C5 FER2^{+} Naive B cells, C7 CD83^{+} Naive B cells, C8 ISG15^{+} Naive B cells, C10 MARCKSL1^{+} Naive B cells, and C15 HSPA1B^{+} Naive B cells. Additionally, the category of MBs includes C1 CRIP1^{+} Memory B cells, C2 LINC01781^{+} Memory B cells, C3 PPP1R15A^{+} Memory B cells, C4 IFI30^{+} Memory B cells, C9 CD1C^{+} Memory B cells, and C13 TLE5^{+} Memory B cells. Furthermore, the subgroup of GCBs comprises C11 MEF2B^{+} Germinal center B cells and C14 STMN1^{+} Germinal center B cells. Lastly, we have identified PCs represented by C6 FKBP11^{+} IgG Plasma cells and C12 DERL3^{+} IgA Plasma cells (Fig. 1B). The distribution of CNVscore, nCount_RNA, S.Score, and G2M. The score for each subpopulation is shown in Fig. 1C, in which GCBs on the right side of the uniform manifold approximation and projection (UMAP) plot have a high CNVscore, a high nCount_RNA, and a more active cell cycle state. In addition, we analyzed the proportion of subgroups derived from peripheral blood (PBL) and tumor tissues. The pie chart displayed the proportion of B and plasma cell subtypes (Fig. 1D). The results showed that most subgroups belonged to malignant cells, with a small proportion of NBs and MBs originating from the PBL (Fig. 1E). Bubble maps and volcano plots demonstrated the expression of differentially expressed genes (DEGs) in each subgroup and the associated features of different HPV infection statuses (HPV^{+} vs HPV^{-}) (Fig. 1F and G).

3.2. Pseudotime trajectories unveiled the differentiation status of B and plasma cell subgroups

To understand the dynamic transitional processes of B and plasma cells in HNSCC, we further performed a pseudotime analysis to trace the transcriptional trajectory. As shown in Fig. 2A and B, with the progression of pseudotime, the number of NBs gradually decreased, while GCBs and MBs gradually differentiated toward PCs. The decrease in GCBs and MBs may be due to their gradual transformation into PCs, most of PCs were at the end of the cell developmental trajectory. Cellular composition was identified by colouring the pseudotemporal map with tissue of origin. By analysing Fig. 2A and B & Fig. 2C, it was found that the pseudotime of NBs



(caption on next page)

Fig. 3. | Landscape of NBs subgroups in the HNSCC. (A) The distribution of UMAP for the 6 subgroups in NBs. **(B)** Visualization of sample characteristics associated with different NBs subgroups and different HPV infection status (HPV$^+$ vs HPV$^-$), Phase (G1, S, G2M), G2M.Scorce, S.Score, nCount_RNA, CNVscore, differential gene expression. **(C)** Expression of top5 marker genes in different subgroups. **(D)** The distribution of CNVscore, nCount_RNA, S.Score, G2M.Scorce for each subgroup was visualized using UMAP. **(E)** Results of GO-BP enrichment analysis of differential genes and their corresponding transcription factors and surface proteins in 6 NBs subgroups. **(F)** Histogram of HPV infection status of NBs subgroups. **(G)** The cell cycle of NBs subgroups. **(H–I)** Violin plots of GO-BP Negative regulation of B cell activation and CNV score. **(J–L)** GSEA enrichment analysis among NBs subgroups.

was significantly lower than that of other B cell subgroups, and the pseudotime of PCs represented by C6 FKBP11$^+$ IgG Plasma cells and C12 DERL3$^+$ IgA Plasma cells was the highest. An enhanced graphical representation, a bar plot, elucidated that the NBs subgroups predominantly comprised neoplastic cells originating from State1. Likewise, the MBs subgroups exhibited a preponderance of cells stemming from State2. Furthermore, the GCBs subgroups was characterized by a prevailing presence of neoplastic cells deriving from State4. Lastly, the PCs subgroups exhibited a notable predominance of cells originating from State5 (Fig. 2D and E). These observations collectively unveil that the NBs subgroups is positioned within the incipient stage of cellular differentiation, contrasting with the PCs, which signifies the end of differentiation. Additionally, cells within the GCBs and MBs subgroups display an augmented propensity for proliferation and differentiation capacity, thereby highlighting their distinctive biological characteristics.

3.3. NBs are in the initial differentiation stage of pseudotime trajectories

According to our research, B and plasma cells could be classified into NBs, GCBs, MBs, and PCs. Each of these subgroups manifests distinct phenotypic characteristics and undertakes specific roles within the TME. NBs can be further classified into 6 subgroups (Fig. 3A), and the expression of top5 marker genes in different subgroups is also shown (Fig. 3B and C). By analyzing the differences in cell cycle, gene expression and CNV among samples with different NBs subgroups and HPV infection states, it was found that the CNV score of C0 TCL1A$^+$ Naive B cells and C15 HSPA1B$^+$ Naive B cells was higher, the proportion of G2M and S phase in the cell cycle was higher, and the expression of differential genes was higher, indicating that their cell proliferation was more active and genetic instability (Fig. 3D–I). The DEGs of 6 NBs subgroups and their corresponding transcription factors and surface proteins were analyzed by GO$^-$BP enrichment analysis, and the active biological process of each tumor cell was obtained. It was found that C0 TCL1A$^+$ Naive B cells was active in the biological process of Cytoplasmic translation, and C8 ISG15$^+$ Naive B cells was active in the biological process of Negative regulation of b cell activation (Fig. 3E–H). Immune landscape studies showed that HPV$^+$ HNSCC were enriched with more NBs, HPV$^-$ HNSCC were infiltrated by fewer NBs, and C0 TCL1A$^+$ Naive B cells had the highest percentage of HPV$^+$ cells (Fig. 3F). In the analysis of NBs subgroups and cell cycle, it was found that the proportion of NBs within the G1 phase of cellular growth was the highest, which may be related to the fact that NBs is in the initial stage of B cell differentiation (Fig. 3G). Through GSEA enrichment analysis, it was found that C0 TCL1A$^+$ Naive B cells subgroup was negatively correlated with Negative regulation of metabolic process and Negative regulation of programmed cell death pathway, indicating that cells of C0 TCL1A$^+$ Naive B cells subgroup had higher ability of cell metabolism and proliferation (Fig. 3J–L).

3.4. GCBs subgroup may have high immune response ability

In the germinal center, follicular helper T cells (T_{fh}) stimulate the B cell receptor (BCR) and induce the differentiation of NBs into a variety of fates, mainly including GCBs, MBs and PCs [33]. Complete transcriptional characteristics at the single-cell level would be particularly helpful in determining transcriptional trajectories and heterogeneity in the process of differentiation from NBs to PCs. C11 MEF2B$^+$ Germinal center B cells (C11 GCBs) and C14 STMN1$^+$ Germinal center B cells (C14 GCBs) were obtained by further subgroup analysis of GCBs (Fig. 4A). The bubble chart and UMAP showed the expression of GCBs top marker genes (Fig. 4B and C). Among the two GCBs subgroups, the CNVscore of C11 was higher, and the nCount_RNA, G2M.Score and S.Score of C14 were higher ($P < 0.001$) (Fig. 4D–F). The GCBs in patients with HNSCC mainly comes from tumor tissues, and some from PBL (Fig. 4G). The percentage of C11 GCBs in HPV$^+$ samples was higher than in C14 GCBs (Fig. 4H). In C11 GCBs, there were more cells with G1 cell cycle, and most of C14 GCBs were in G2M and S phase (Fig. 4I). The proportion of cells in G1 phase was more in HPV$^+$ group, and the proportion of GCBs cells in G2M and S phase was more in HPV$^-$ group (Fig. 4J). The DEGs of C11 GCBs and C14 GCBs were screened (Fig. 4L), and the enrichment results of GO-BP among C11 GCBs and C14 GCBs was shown by heatmap. It was found that GO-BP terms of C11 GCBs were up-regulated in biological processes such as cell activation, leukocyte activation, immune response, lymphocyte activation, regulation of immune system process, regulation of immune response, indicating that the DEGs of C11 GCBs may play a promoting role in the regulation of immune system (Fig. 4K). Similarly, in the GO-BP enrichment analysis results of C11 GCBs, C11 GCBs was more active in cytoplasmic translation, b cell activation, b cell proliferation, production of molecular mediator of immune response and immunoglobulin production pathways. The high immune response ability of C11 GCBs subgroup was further verified, which was consistent with the previous research results (Fig. 4M and N).

3.5. MBs plays a key role in the proliferation and differentiation of B cells

MBs, the subgroup with the largest number of B cells (Fig. 5A), could be further divided into six subgroups and the expression of top5 marker genes in different subgroups are also shown in Fig. 5B and C. UMAP showed that the samples in the C4 IFI30$^+$ Memory B cells (C4 MBs) had higher CNVscore, higher nCount_RNA and more active cell cycle state. Among them, the chromosome stability,

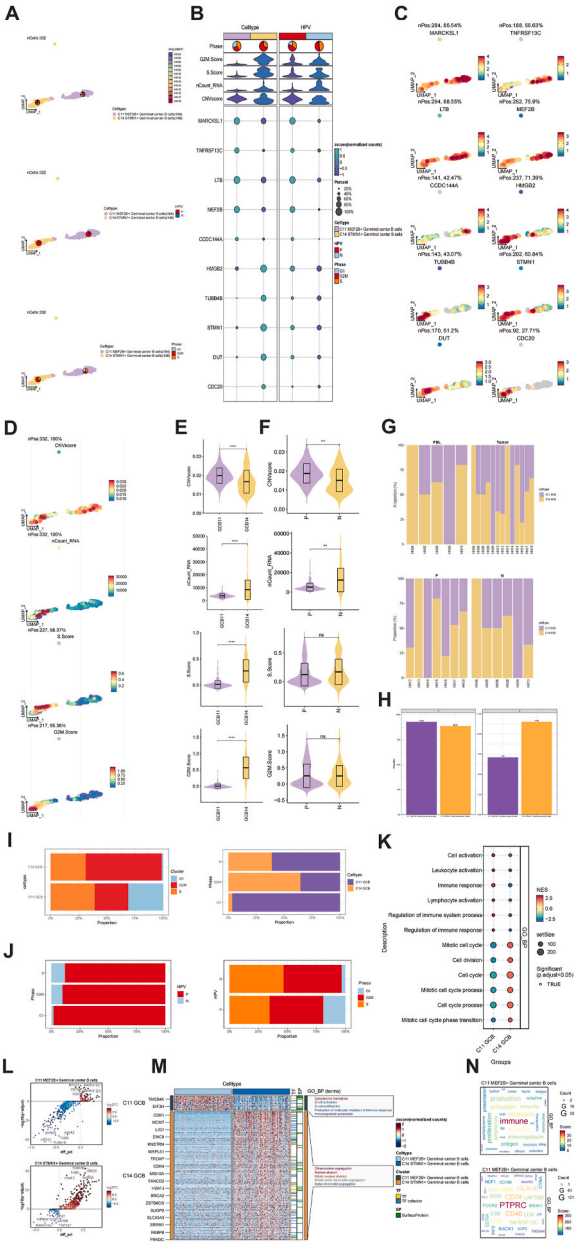
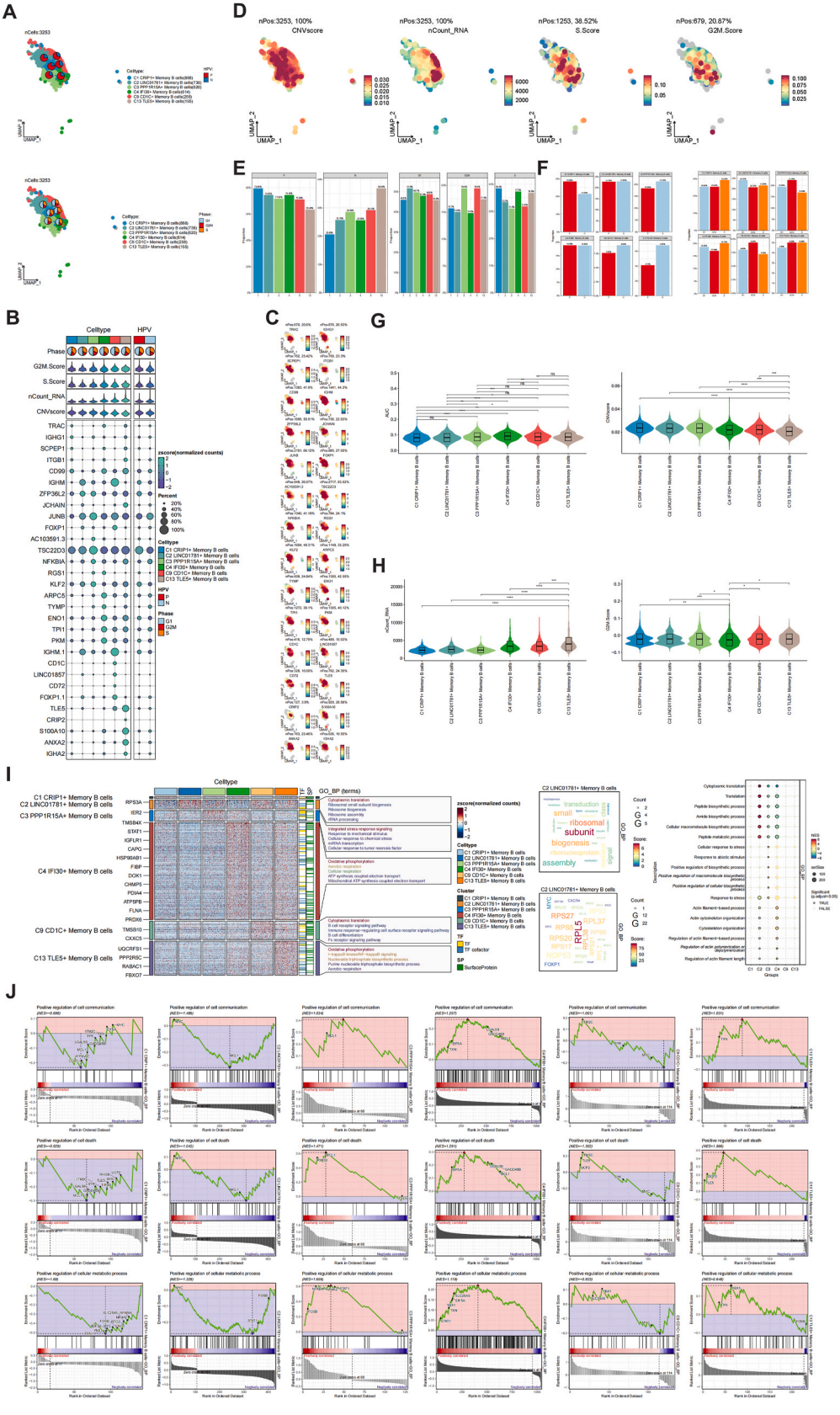


Fig. 4. | Landscape of GCBs subgroups in the tumor microenvironment of HNSCC. (A) 332 GCBs were subjected to UMAP dimension reduction, divided into two cell clusters including C11 MEF2B^{+} Germinal center B cells (C11 GCBs, 184) and C14 STMN1^{+} Germinal center B cells (C14 GCBs, 148). The contribution of pie chart to patients (13 patients), HPV infection status (HPV^{+} vs HPV-) and Phase (G1, S, G2M) were visualized. **(B–C)** The bubble chart and UMAP showed the expression of GCBs top marker genes. **(D)** Visualization of relevant sample features for 332 GCBs using UMAP dimensionality reduction clustering: nCount_RNA, CNVscore, G2M.Scorce, S.Score. **(E)** The violin plot shows the nCount_RNA, CNVscore, G2M.Scorce, S.Score of C11 GCBs and C14 GCBs and their differences. *P < 0.05, **P < 0.01, ***P < 0.001, means significant difference, ns means no significant difference. The boxplots showed the median (middle line), upper and lower quantiles (boxplot), and range of the data (violin plot). **(F)** nCount_RNA, CNVscore, G2M.Scorce, S.Score of GCBs with different HPV infection status (HPV^{+} vs HPV-) and their differences. **P < 0.05, **P < 0.01, ***P < 0.001, means significant difference, ns means no significant difference. The boxplots showed the median (middle line), upper and lower quantiles (boxplot), and range of the data (violin plot). **(G)** The proportion of C11 GCBs and C14 GCBs in PBL or tumor group among patients (above). The ratio of C11 GCBs and C14 GCBs in different groups of HPV infection status (HPV^{+} vs HPV-) among patients (below). **(H)** The C11 GCBs and C14 GCBs ratios between the HPV^{+} and HPV-groups. **(I–J)** Bar graphs depicted the difference in cell proportion between subgroups and phases, as well as the difference in cell proportion between different HPV infection statuses and phases. **(K)** GSEA enrichment analysis of GCBs subgroups. The DEGs of C11 GCBs and C14 GCBs were screened, and the enrichment items of GOBP were scored by GSEA. NES >0 was up-regulated and <0 was down-regulated. NES, N stands for standardization and ES for enrichment scores. **(L)** Volcano plots showing expression of DEGs in GCBs subgroups. **(M)** The heat map showed the active biological processes of each subgroup of GCBs. **(N)** Word cloud for GO-BP enrichment analysis of GCBs DEG.



(caption on next page)

Fig. 5. | Landscape of MBs subgroups in the HNSCC microenvironment. (A) 3253 MBs were subjected to UMAP dimension reduction, divided into 6 cell clusters including C1 CRIP1$^+$ Memory B cells(C1 MBs, 868), C2 LINC01781$^+$ Memory B cells(C2 MBs, 738), C3 PPP1R15A$^+$ Memory B cells(C3 MBs, 620), C4 IFI30$^+$ Memory B cells(C4 MBs, 614), C9 CD1C$^+$ Memory B cells(C9 MBs, 258), C13 TLE5$^+$ Memory B cells(C13 MBs, 155). The contribution of pie chart to HPV infection status (HPV$^+$ vs HPV$^-$) and Phase (G1, S, G2M) were visualized. (B–C) The bubble chart and UMAP showed the expression of MBs top5 marker genes. (D) Visualization of relevant sample features for 3253 MBs using UMAP dimensionality reduction clustering: nCount_RNA, CNVscore, G2M.Scorce, S.Score. (E–F) The histogram showed the distribution of HPV infection status (HPV$^+$ vs HPV$^-$) and Phase (G1, G2M, S) in each subgroup of MBs intuitively. (G–H) The violin plot shows the nCount_RNA, CNVscore, G2M.Scorce, S.Score of MBs subgroups and their differences. **P < 0.05, **P < 0.01, ***P < 0.001, means significant difference, ns means no significant difference. The boxplots showed the median (middle line), upper and lower quantiles (boxplot), and range of the data (violin plot). (I) The activity of different pathways in MBs subgroups was demonstrated by GO-BP enrichment analysis and word cloud. (J) The correlation of DEGs among MBs subgroups in positive regulation of cell communication, positive regulation of cell death, positive regulation of cellular metabolic process was analyzed by GSEA enrichment analysis.

transcription level and cell cycle state of different subgroups of MBs are also different, which is of great significance for the study of B cell heterogeneity (Fig. 5D–H). The bar charts revealed that the majority of MBs subgroups were associated with HPV$^+$ infection, whereas the cell cycle mostly occurred in the G1 and S phases (Fig. 5E and F). The AUCell analysis of the GOBP_POSITIVE_REGULATION_OF_B_CELL_PROLIFERATION in several subgroups of MBs revealed that the C4 MBs had a higher AUC for this GOBP term compared to other subtypes (Fig. 5G, left). The CNV score of C13 TLE5$^+$ Memory B (C13 MBs) cells exhibited a lower value compared to the other subtypes (Fig. 5G, right), while the nCount_RNA of C13 MBs was higher than other subtypes (Fig. 5H, left). The G2M.Score of C4 MBs was lower than other subtypes of MBs (Fig. 5H, right). The results of GO-BP enrichment analysis among Memory B cells was shown in Fig. 5I. In GSEA enrichment analysis, the DEGs of C4 MBs were up-regulated in positive regulation of cellular biosynthetic process, which was positively correlated with positive regulation of cell communication, positive regulation of cell death, positive regulation of cellular metabolic process (Fig. 5J). These results suggest that the DEGs of C4 MBs plays a key role in the differentiation of B cells and promotes the regulation of immune system process, which may reflect the high immune response ability of tumor cells.

3.6. B cells eventually differentiate into mature PCs

GCBs that have a strong attraction to tumor-associated antigens ultimately develop into fully developed plasma cells (PCs) that generate immunoglobulins (IgG, IgA, IgM, and IgE). These immunoglobulins have distinct purposes within the given situation [34]. According to the difference in the expression of tumor-associated antibodies in PB, we divided the subgroups into C6 FKBP11$^+$ IgG Plasma cells (C6 IgG, 376), C12 DERL3$^+$ IgA Plasma cells (C12 IgA, 157) (Fig. 6A). The UMAP visualization showed the difference between C6 IgG and C12 IgA in terms of CNVscore, nCount_RNA, S.Score, and G2M (Fig. 6B). The bubble chart and UMAP visualization displayed the marker genes for C6 IgG and C12 IgA (Fig. 6C and D). The bar chart depicted the distribution of PCs among various categories (Fig. 6E–G). To investigate the molecular diversity among PCs subgroups and its functionality. The GO-BP enrichment analysis of the DEGs in PCs revealed that C6 IgG was involved in the biosynthesis of amyloid precursor protein and glycoprotein. Additionally, C12 IgA was found to be active in the biological processes of responding to endoplasmic reticulum stress, the ERAD pathway, and topologically incorrect protein (Fig. 6H and I).

3.7. Exploring the intercellular communication between B cells, plasma cells, and tumor cells

Numerous investigations have demonstrated the existence of a diverse spectrum of mutual influence between B cells, plasma cells, and tumor cells. This interplay exerts a profound influence on the progression of cancer [34]. To elucidate the intricate mechanism of cell-cell interaction between these cells, we established a separate intercellular communication network between B cells, plasma cells, and tumor cells in HNSCC (Fig. 7A). Three patterns of interactions and 20 significant signaling pathways were identified (Fig. 7B). We generally describe the inferred strengths of incoming and outcoming interaction, in which C11 GCBs and C14 GCBs are the main cell type expressing ligands and receptors that actively participate in cell-cell interaction. These results suggest that GCBs is associated with tumors and plays a central role in TME (Fig. 7C). As the observed C11 GCBs and C14 GCBs played a central role in the intercellular communications at HNSCC, we further explore the interactions between incoming and outgoing communication patterns of target cells and secreting cells. Notably, in outgoing communication patterns of secreting cells of GCBs, CD99 and SEMA4 are shown as the most predominant signaling in C11 GCBs compared to other signaling pathways (Fig. 7D and E). Consistent with these observations, GCBs and MBs are the core of the interaction between B cells and tumor cells (Fig. 7G and H). In summary, the intercellular interaction based on CD99 and SEMA4 signaling pathways is the molecular basis of GCBs-centered TME observed in HNSCC.

3.8. GCBs based regulatory network detect tumor-specifc signaling pathways in HNSCC

To unveil the intricate signaling pathways that contribute to the multifaceted intercellular communications, we computed the CD99 and SEMA4 signaling pathways (Fig. 8A–F). Tumor-specific C11 GCBs could express CD99 and SEMA4 to interact with some receptors of tumor cells (Fig. 8B–G). Regarding the CD99 signaling pathway, we observed that C11 MEF2B$^+$ Germinal center B cells and C14 STMN1$^+$ Germinal center B cells highly expressed CD99 to interact with the tumor cell receptors, C4 and C13 of memory B cells (Fig. 8C–E). In SEMA4 signaling pathway network, the specific interaction network of C11 and C14 GCBs with tumor cells was also observed (Fig. 8H–J). These phenomena show that CD99 and SEMA4 could potentially serve as pivotal pathways mediating the

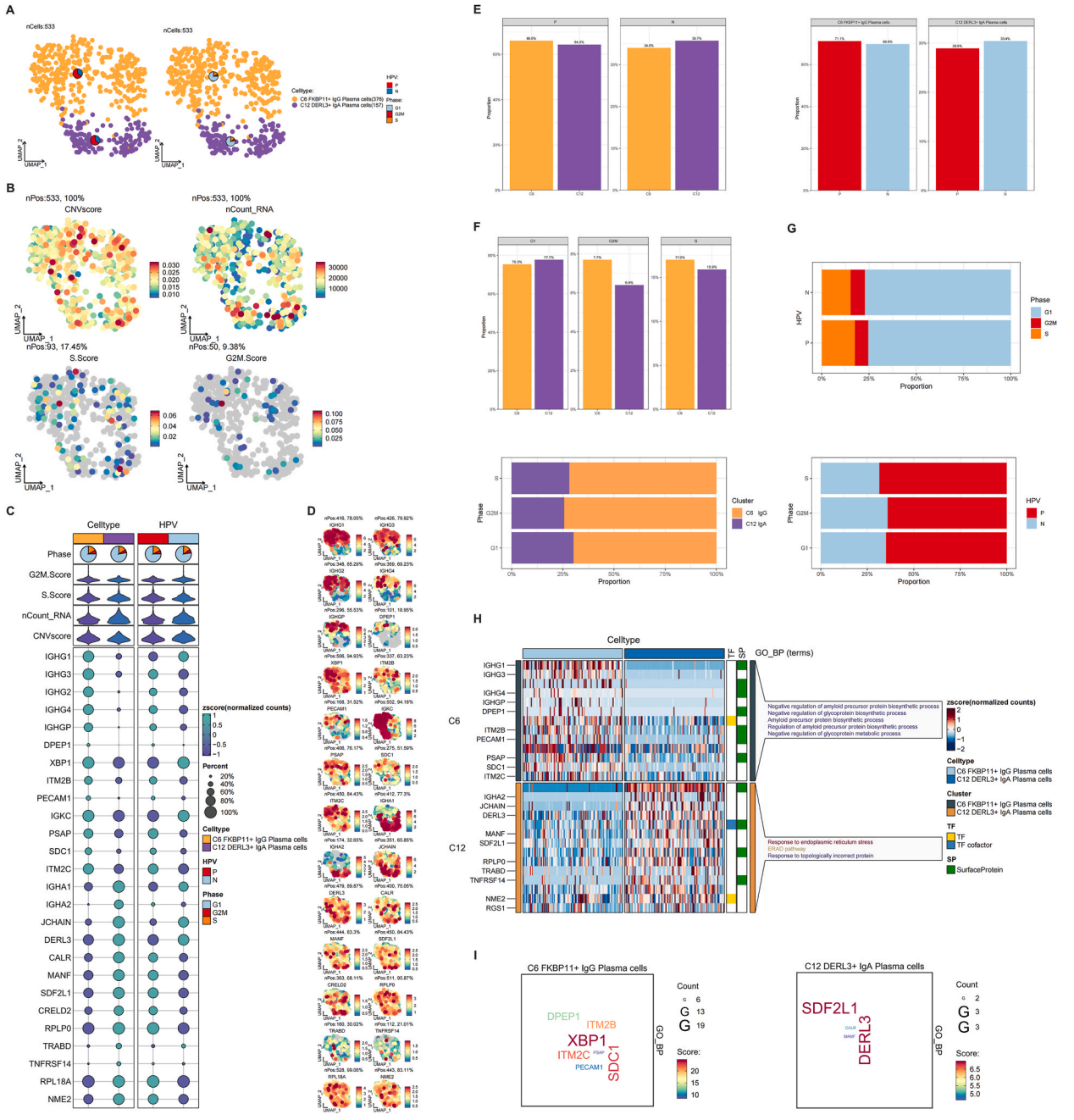


Fig. 6. | Landscape of PCs subgroups in the HNSCC microenvironment. (A) The distribution of UMAP for the 2 subgroups in PCs. (B) UMAP was used to illustrate the distribution of CNVscore, nCount_RNA, S.Score, and G2M.Score for each subgroup. (C) Visualization of sample characteristics associated with different PCs and different HPV infection status (HPV^+ vs HPV^-): Phase (G1, S, G2M), G2M.Scorce, S.Score, nCount_RNA, CNVscore, differential gene expression. (D) Expression of marker genes in different PCs. (E–G) The histogram displayed the differences in cell percentage across different HPV groups and phase groups. (H–I) Results of GO-BP enrichment analysis of differential genes and their corresponding transcription factors and surface proteins in 2 PCs.

interplay between GCBs and tumor cells. Furthermore, these signaling pathways possess the capacity to regulate B cell differentiation, consequently exercising control over tumor growth. Overall, our findings underscore that the C11 GCBs manifest intricate intercellular interplays that exacerbate the virulence of the tumor. Targeting tumor-specific cell interaction networks could be a promising strategy for treating HNSCC.

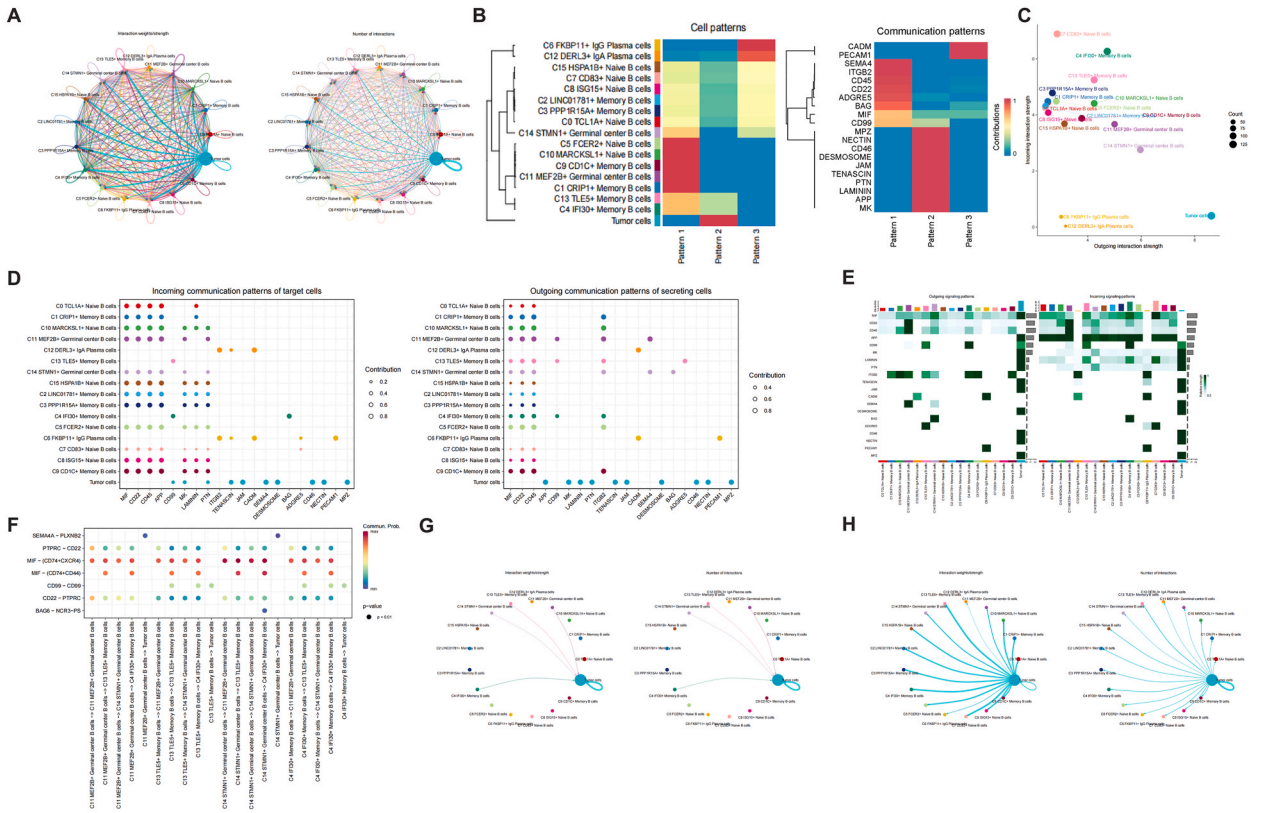


Fig. 7. | Cell communication network. (A) Circle visualization of interaction strength (left) and number (right) among the B and plasma cell subgroups and tumor cells. (B) Cell and communication patterns. (C) Illustration of the incoming and outgoing interaction strengths for each of the B and plasma cell subgroups and tumor cells. (D) Incoming communication patterns of target cells (left). Outgoing communication patterns of secreting cells (right). (E) Heatmap shows an overview of cell-cell interactions in both outgoing and incoming signaling patterns. (F) Interaction bubble diagram among GCBs, MBs and tumor cells. (G) The circle plots show the interaction strength (left) and number (right) of B and plasma cell subgroups on tumor cells. (H) Circle visualization of interaction strength (left) and number (right) about tumor cells on B cell subgroups.

3.9. Screening and validation of prognostic genes in GCBs and MBs

Considering the presence of tumor-specific signaling pathways in MBs and GCBs mentioned earlier, we will conduct an in-depth examination of these subgroups. A total of 499 HNSCC patients from the TCGA database were finally included in this study. Univariate Cox regression analysis was performed to identify the prognosis-related genes in MBs and GCBs (Fig. 9A, F, 10A, 10F). The candidate genes were recruited into the Least Absolute Shrinkage and Selection Operator (LASSO) Cox regression model to screen out the robust prognosticators (Fig. 9B, G, 10B, 10G). Ultimately, using multivariate Cox regression analysis, we were able to construct four prognostic risk scores consisting of prognostic genes specific to the MBs and GCBs subgroups (Fig. 9D, I, 10D, 10I). The individual genes that constituted risk scores, which showed an important association with overall survival in MBs and GCBs, were thoroughly shown (Fig. 9E, J, 10E, 10J, 10K). The expression of CD38, SMIM14, and TCL1A in C11 germinal center B-cells (GCBs) varied considerably over distinct N stages ($P < 0.05$).

3.10. MEF2B^{+} GCB score based on the prognosis-related genes in C11 GCBs

We then investigated the possible mechanisms of prognostic effects of 18 prognosis-related genes that comprised the MEF2B^{+} GCB score. Bubble maps demonstrated the expression of 18 prognosis-related genes in each GCBs subgroup and the associated features of different HPV infection statuses (HPV^{+} vs HPV^{-}), Phase (G1, S, G2M), G2M.Score, S.Score, nCount_RNA, CNVscore (Fig. 11A). The expression of 18 prognostic genes in C11 and C14 GCBs clusters were further visualized in Fig. 11B and C. Risk scores were computed for individuals diagnosed with HNSCC within the TCGA datasets. The collected data were subsequently partitioned into cohorts of low and high risk groups, demarcated by the median risk score (Fig. 11D above). The heatmap specifically showed the expression of genes in high and low risk groups (Fig. 11D below). More HNSCC patients had higher survival rates ($p < 0.05$) observed in the low-risk group, while the distribution of stage and TNM phase was not statistically different (Fig. 11E). The bar chart shows the risk coefficients that make up the risk score genes (Fig. 11F). Moreover, the time-dependent ROC curve demonstrated the proficient prognostic capacity of this model concerning the OS of HNSCC patients. The calculated Area Under the ROC Curve (AUC) values stood at 0.690 (1 year),

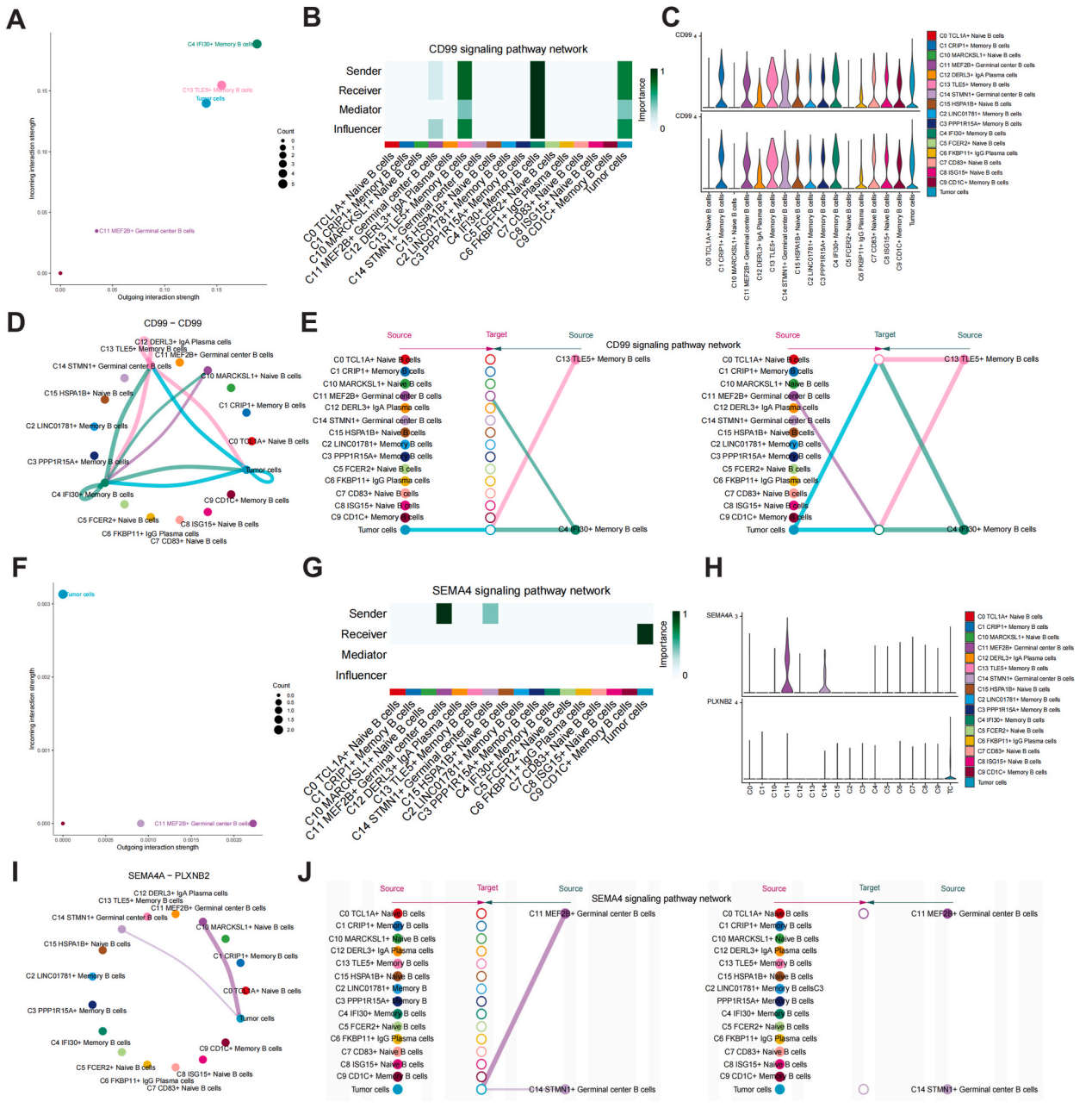


Fig. 8. | GCBs based regulatory network detect tumor-specifc signaling pathways in HNSCC. (A, F) Illustration of the incoming and outgoing interaction strengths for CD99 and SEMA4 signaling pathways. (B, G) Heatmaps show dominant senders, receivers, mediators and infuencers in CD99 and SEMA4 signaling pathways of C11 GCBs inferred by network centrality score. (C, H) The violin diagram depicts the gene expression levels in the CD99 and SEMA4 signaling pathway networks including B and plasma cells as well as malignant cells. (D) The circle diagram shows the interaction of CD99 ligand on MBs and tumor cells. (I) The circle plot shows the interaction of SEMA4 ligand on GCBs and tumor cells. (E, J) The hierarchy plots of signaling pathway networks indicate the regulation of CD99 and SEMA4 on B cells and tumor cells in HNSCC, as predicted by CellChat anlysis.

0.711 (3 years) and 0.677 (5 years), as evidenced in Fig. 11G. Then, we further assessed the correlation between gene expression and risk score in the prognostic model (Fig. 11H). The expression of five genes, ACTB, BASP1, EZR, PRDX6, and RRAS2, was positively correlated with the risk scores (Fig. 11I), which was in agreement with the results of the previous prognostic studies (Fig. 10J), indicating that these five genes may be tumor-related risk genes in HNSCC. Meanwhile, the five genes of ACTB, BASP1, EZR, PRDX6, RRAS2 were highly expressed in the high risk group, and the difference between the high and low risk groups was statistically significant ($P < 0.001$) (Fig. 11J). They were negatively correlated with OS (Fig. 11K). The correlation analysis between these genes was

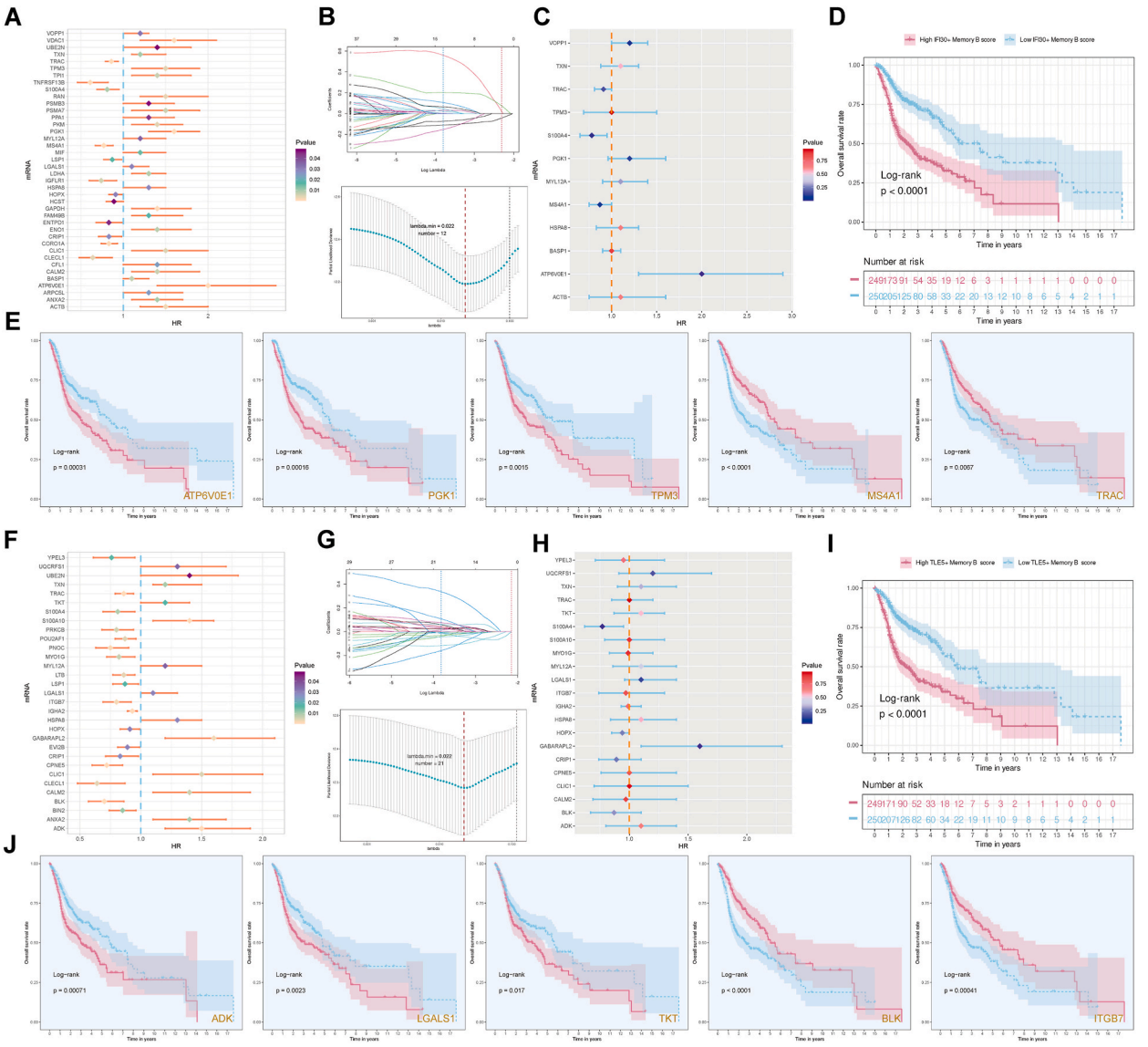


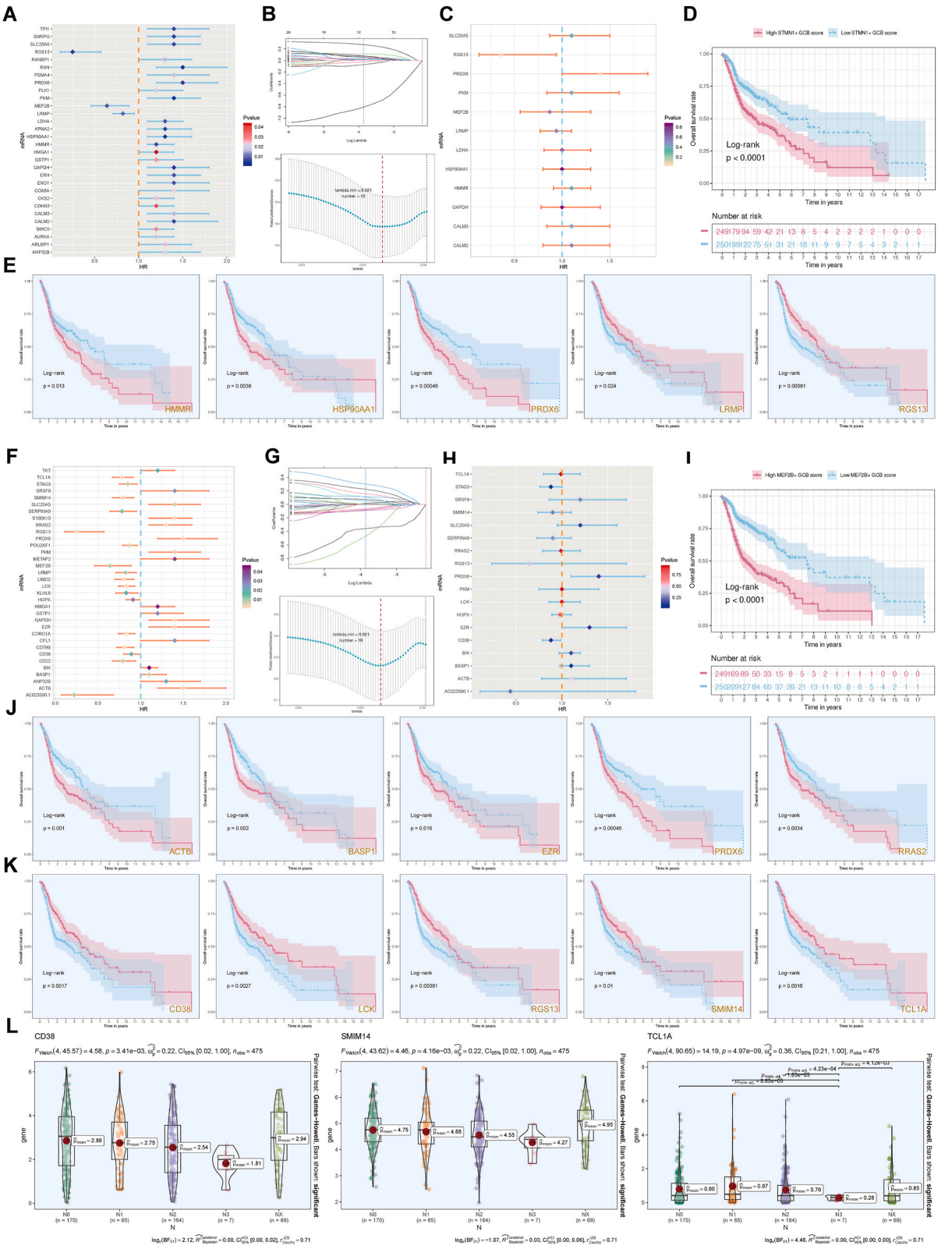
Fig. 9. | Screening for prognosis-related memory B cells in HNSCC patients. (A, F) Univariate Cox regression analysis was performed to identify the prognosis-related genes. (B, G) Lasso regression analysis of prognosis-related genes. (C, H) Multivariate COX regression analysis on OS in the TCGA cohort. (HR > 1 protective factor, <1 risk factor.) (D, I) Kaplan–Meier curves were created to estimate OS for high- and low-score groups from TCGA cohort ($P < 0.0001$). (E, J) Kaplan-Meier curve showed that the prognosis of HNSCC with high expression of ATP6V0E1, PGK1, TPM3, ADK, LGALS1, TKT was poor in OS, while that of HNSCC with high expression of MS4A1, TRAC, BLK, ITGB7 was better in OS.

shown in Fig. 11L and M. We further investigated the differences of EZR and RRAS2 in high- and low-risk groups, different ages, different races, and different TNM stages, and found that the gene expression was higher in both high-risk groups (Fig. 11N).

We observed that the risk group based on the MEF2B^{+} GCB score had independent prognostic value ($P < 0.001$) (Fig. 12A). A nomogram plot was established to get a better clinical application ability by integrating the risk group based on the MEF2B^{+} GCB score and clinical features (Fig. 12B). The time-dependent ROC curve demonstrated the proficient prognostic capacity of this model. The AUC values were 0.738 (1 year), 0.769 (3 years) and 0.634 (5 years) (Fig. 12C and D). For the 1-, 3-, and 5-year survival, a satisfactory fit was observed between the survival predicted by the nomogram and the actual survival (Fig. 12E).

3.11. Immune cells play a dual role in the dynamic ecosystem of the HNSCC microenvironment

The initiation, progression, and invasion of HNSCC are controlled by the complex network of signaling pathways in the tumor microenvironment (TME) as well as cell-cell interactions between the tumor stroma, immune cells, and tumor cells [35]. HNSCC microenvironment contained multiple types of immune cells, most of which are T cells and macrophages (Fig. 12F). The high risk



(caption on next page)

Fig. 10. | Screening for prognosis-related germinal center B cells in HNSCC patients. (A, F) Univariate Cox regression analysis was performed to identify the prognosis-related genes. (B, G) Lasso regression analysis of prognosis-related genes. (C, H) Multivariate COX regression analysis on OS in the TCGA cohort. (HR > 1 protective factor, <1 risk factor.). (D, I) Kaplan–Meier curves were created to estimate OS for high- and low-score groups from TCGA cohort ($P < 0.0001$). (E, J, K) Kaplan-Meier curve showed that the prognosis of HNSCC with high expression of HMMR, HSP90AA1, PRDX6, ACTB, BASP1, EZR, RRAS2 was poor in OS, while that of HNSCC with high expression of LRMP, RGS13, CD38, LCK, SMIM14, TCL1A was better in OS. (L) The bar diagrams show the distribution of the expression of CD38, SMIM14 and TCL1A in N0, N1, N2, N3, NX.

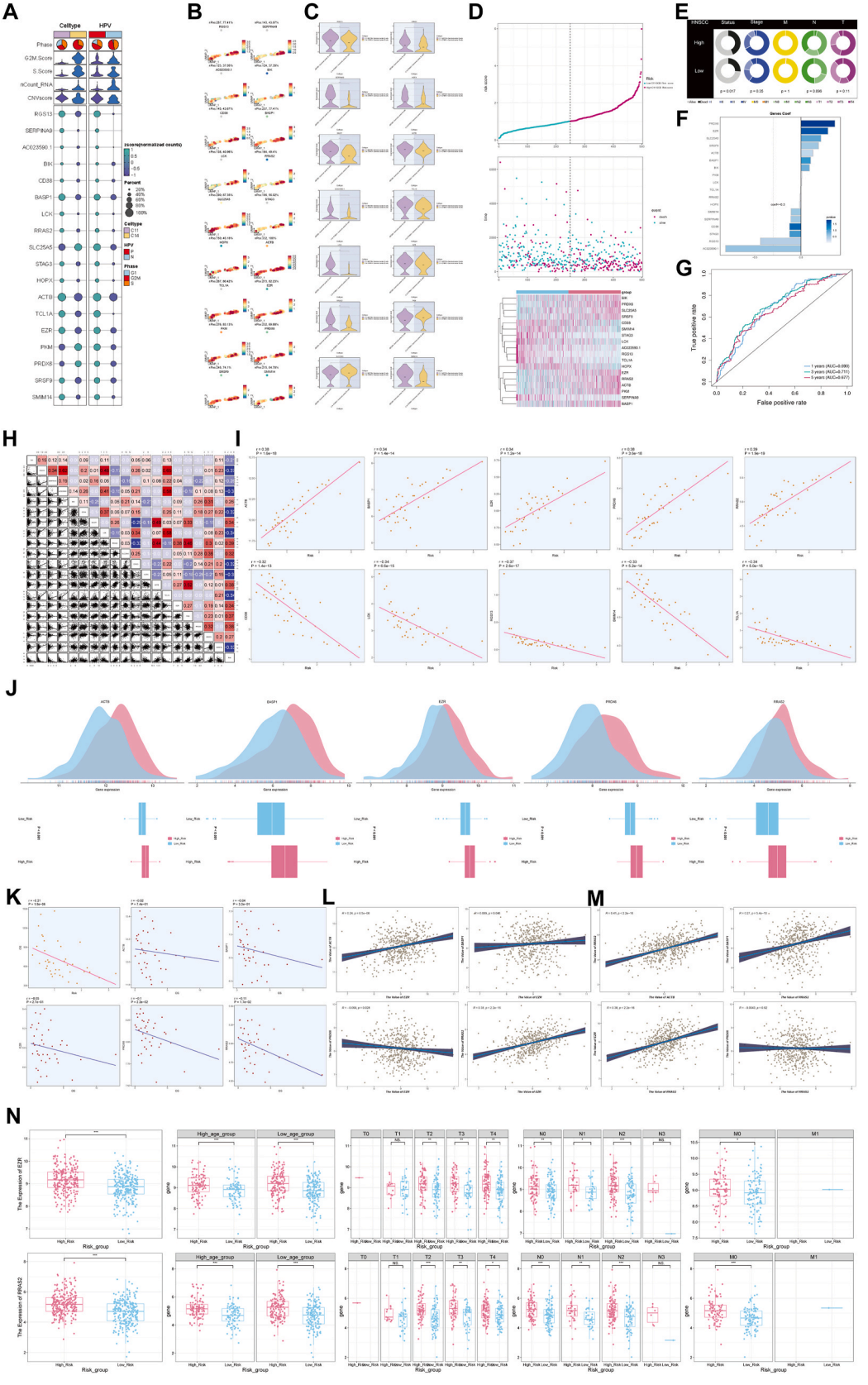
group had lower immune scores and higher tumor purity than the low risk group (both $P < 0.001$), which may be related to the fact that immune cells infiltrate into the tumor stroma within the HNSCC and interact with tumor cells (Fig. 12G and H). By analyzing the principal components of immune cells in HNSCC, it was found that there were significant differences in the infiltration of many kinds of immune cells between the two risk subgroups ($P < 0.05$). There was more CD8 T cell infiltration in the low risk group and more macrophage M2 infiltration in the high risk group (Fig. 12I and J). In the correlation analysis between risk score and immune cells, mast cells, NK cells, and macrophages were significantly positively correlated with the risk score, whereas various T cells such as T cells regulatory (Tregs), T cells follicular helper, and CD8 T cells were significantly negatively correlated with the risk score (Fig. 12K). In addition, we also analyzed the correlation between the expression of 18 prognostic genes and the proportion of immune cells in HNSCC. The results showed that the risk score, the gene expression of ACTB, BASP1, PRDX6 was positively correlated with macrophage M2 (all $P < 0.05$), and the risk score, the gene expression of ACTB, BASP1, RRAS2 was negatively correlated with CD8 T cells (all $P < 0.01$) (Fig. 12L–N). The gene expression of RRAS2 was negatively correlated with naive B cells, plasma cells, stromal score, and immunity score (all $P < 0.05$), and was positively correlated with tumor purity ($P < 0.001$) (Fig. 12O and *P*). Furthermore, RRAS2 was positively correlated with dendritic cells activated, macrophages M0, CD4 memory T cells resting, and NK cells resting (all $P < 0.05$). RRAS2 was negatively correlated with CD4 memory T cells activated, Mast cells resting, CD8 T cells, and regulatory T cells (Tregs) (Fig. 12Q). A study has defined gene expression characteristics associated with CD4 T follicular helper cells, which are correlated with longer progression-free survival in HNSCC patients [36].

3.12. Functional enrichment and GSEA analysis of differentially expressed genes in HNSCC patients

To scrutinize the molecular heterogeneity and explore the biological mechanism distinguishing the high-risk from the low-risk subgroups, we discerned DEGs within the TCGA cohort [$|\log_2(\textit{fold change})| > 1$, fdr<0.05] (Fig. 13A and B). The results of Gene Ontology (GO) enrichment analysis revealed that the DEGs primarily participated in biological processes such as immunoglobulin complex, circulating immunoglobulin complex, and external side of plasma membrane. Their products were predominantly engaged in immune components, encompassing immunoglobulin production, production of molecular mediator of immune response, B cell activation, etc. Furthermore, these genes played a pivotal role in biological molecular functions, which covered antigen binding and immunoglobulin receptor binding, etc. (Fig. 13C). The Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways mainly included the B cell receptor signaling pathway, Ras signaling pathway, PI3K−Akt signaling pathway, Cytokine−cytokine receptor interaction, Primary immunodeficiency, Hematopoietic cell lineage, etc. (Fig. 13D). Our GSEA results, presented in Fig. 13E, revealed that Ribosomal large subunit biogenesis, Mitochondrial translation, Ncrna 3 End processing, Maturation of ssu Rrna from tricistronic Rrna transcript ssu Rrna 5 8s Rrna lsu Rrna, amino acid activation were activated, while antigen receptor mediated signaling, regulation of B cell activation, B cell mediated immunity, positive regulation of B cell activation, B cell receptor signaling pathway were suppressed in the GCBs subgroup. According to the results of GSEA, many B-cell-associated pathways were suppressed in GCBs, highlighting the presence of a B-cell response in patients with HNSCC.

3.13. Relationship of the OS in HNSCC patients among the TMB subgroups and drug sensitivity

In the gene mutation spectrum of the risk and age groups, 466 (91.37 %) of 510 HNSCC patients had the gene mutation (Fig. 13F left). Among the 18 prognostic genes, 31 (6.08 %) had gene mutations in 510 HNSCC patients (Fig. 13F right). We further investigated the frequency of CNV in a total of 526 samples. As shown in Fig. 13G, ACTB, RRAS2, SERPINA9, TCL1A, PKM, EZR, LCK, CD38, and BIK had significant CNV amplification frequencies. In addition, the probability of simultaneous mutations in two groups of genes, PKM and SERPINA9, TCL1A and STAG3, was higher ($P < 0.05$) (Fig. 13H). We selected the top 6 genes of somatic mutation rate for single nucleotide polymorphism (SNP) analysis and found that SERPINA9, STAG3, ACTB, CD38, LCK, and PKM were dominated by missense mutation (Fig. 13I). In the two risk subgroups, the TMB of the high risk group was higher than that of the low risk group ($P < 0.001$) (Fig. 13J), and TMB was positively correlated with the risk score ($P < 0.001$) (Fig. 13K). Combined with the survival analysis in Fig. 13L, the survival time of patients in high risk-high TMB group was the shortest, while that in low risk-low TMB group was the longest, and the OS of patients in each group was significantly different ($P < 0.0001$). The AUC values were 0.541 (1 year), 0.532 (3 years), and 0.5I2 (5 years) (Fig. 13M). This indicates that the TMB model has a discernible predictive ability for the prognosis of HNSCC patients. This indicates that the TMB model has a discernible predictive ability for the prognosis of HNSCC patients. Furthermore, in drug sensitivity analysis, HNSCC patients in the low-risk subgroup demonstrated enhanced responsiveness to cisplatin and docetaxel in chemotherapy (Fig. 13N). In contrast, the high-risk subgroup was more sensitive to Nilotinib and Rapamycin, both targeted drugs (Fig. 13O). The IC50 values of ABT.263, AKT.inhibitor.VIII, CCT007093 were higher in the high-risk subgroup, and Bosutinib, Epothilone. B and Pazopanib were higher in the low-risk subgroup (Fig. 13P). These discoveries not only furnish invaluable insights into selecting suitable chemotherapy and targeted drugs following the risk score of HNSCC patients afflicted with HNSCC but



(caption on next page)

Fig. 11. MEF2B+GCB score. (A) Visualization of genes constituted the MEF2B$^+$ GCB score in different GCB subgroups and different HPV infection statuses (HPV$^+$ vs. HPV-), Phase (G1, S, G2M), and G2M.Scorce, S.Score, nCount_RNA, CNVscore, differential gene expression. (B–C) Expression of 18 prognosis-related genes in C11 and C14 subgroups. (D) Distribution and median value of risk scores of HNSCC patients. Distribution of OS status and risk scores of HNSCC patients. The heatmap showed the expression of 18 prognosis-related genes in high and low-risk groups. (E) Pie charts depict the differences in clinical characteristics (status, stage, TNM, phase) between high and low-risk groups in HNSCC. (F) The bar chart shows the risk coefficients that make up the risk score genes. (G) Time-dependent ROC curves and AUC values evaluate the prognostic performance of the risk score model at 1, 3, and 5 years. (H) An overview of the correlation between 18 prognosis-related genes. Red indicates a positive correlation, and blue indicates a negative correlation. (I) ACTB, BASP1, EZR, PRDX6, and RRAS2 were positively correlated with risk scores. CD38, LCK, RGS13, SMIM14, and TCL1A were negatively correlated with risk score ($P < 0.001$). (J) The differential expression of ACTB, BASP1, EZR, PRDX6, and RRAS2 in high and low-risk groups. (K) ACTB, BASP1, EZR, PRDX6, RRAS2, and risk score were negatively correlated with OS. (L–M) The correlation analysis among ACTB, BASP1, EZR, PRDX6, and RRAS2. (N) The differences of EZR and RRAS2 in high and low risk groups, different age groups, different races and different TNM stages.

also facilitate informed clinical therapeutic deliberations.

3.14. In vitro experiment

Based on the cell communication analysis findings using cellchat between MEF2B$^+$ B cells (C11 GCBs) and tumor cells, we conducted in vitro knockdown experiments targeting the PLXNB2 gene in the head and neck cancer cell line. The results, illustrated in Fig. 14A and B, manifest a conspicuous reduction in the proliferation capacity of both cell lines, discernibly contrasting with the control group. Furthermore, plate cloning experiments unveiled a pronounced suppression in both the quantity and size of cell colonies after the knockdown of the PLXNB2 gene in both cell lines (Fig. 14C and D).

Subsequent wound healing experiments showcased a statistically significant decrease in the cell migration rate within the PLXNB2 knockdown group (Fig. 14E and F). Moreover, Transwell experiments demonstrated a marked reduction in the number of cells penetrating the lower chamber following PLXNB2 gene knockout, indicative of a substantial inhibition of the migration and invasion capabilities of head and neck cancer cells (Fig. 14G and H).

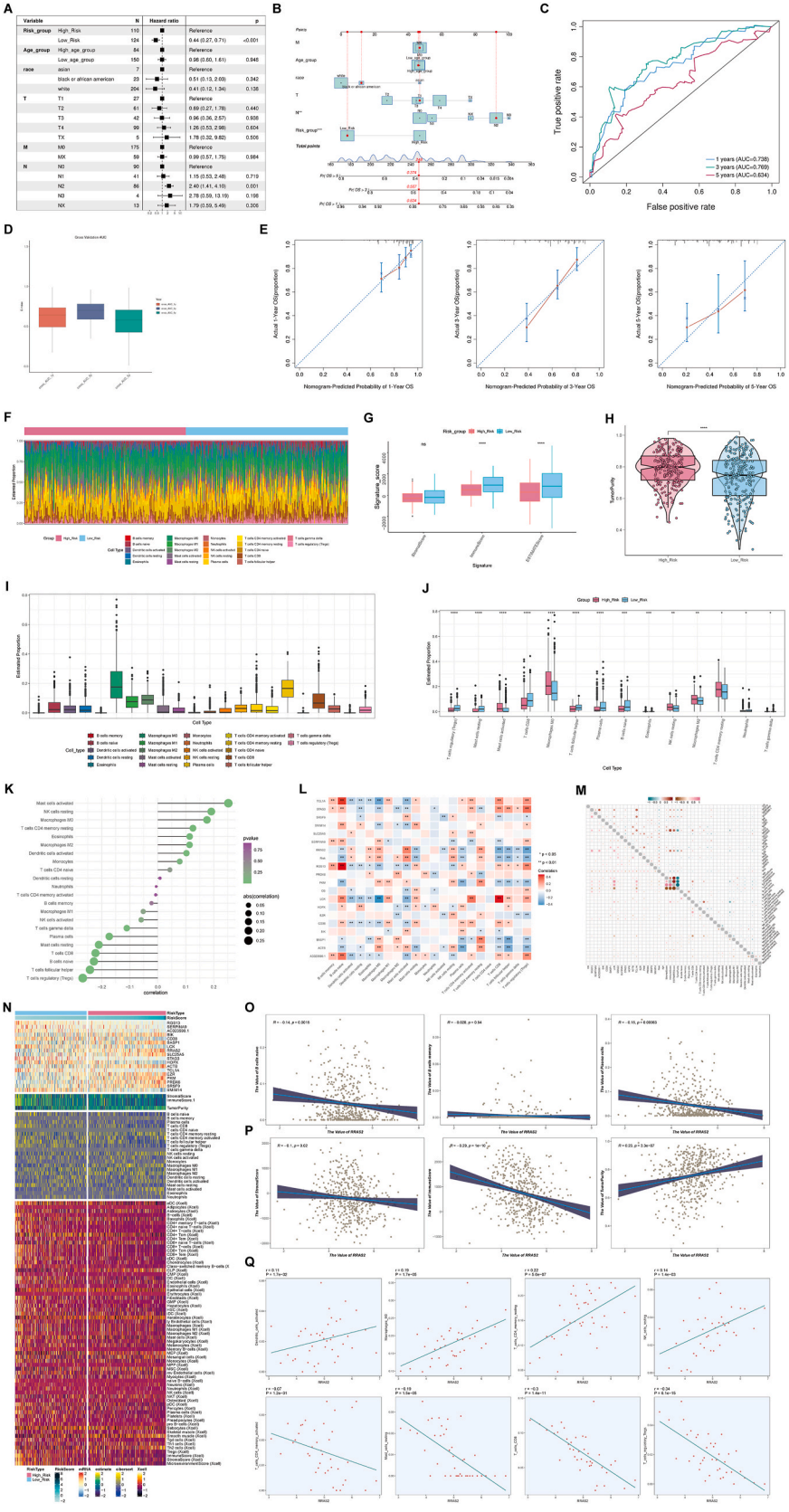
3.15. Immunofluorescence detection of MEF2B$^+$ B cells

Three tumor samples were used to explore the relationship between MEF2B and tumor microenvironment in HNSCC by mIHC (Fig. S3). B cells were stained with anti-CD79a, and CD8$^+$ T cells were stained with anti-CD8, and the tumor epithelium was stained with anti-EPCAM. It was revealed that CD8$^+$ T cells were less enriched in the MEF2B and CD79a double-positive area, suggesting that high-level of MEF2B in CD79a might correlated with lower immune response. Thus, we have investigated the tumor microenvironment feature of MEF2B + B cell in HNSCC through bioinformatics analysis and mIHC.

4. Discussion

HNSCC is one of the most common malignant tumors in the world, with a high metastatic rate and low survival rate, often accompanied by dysfunctions such as articulation and eating and disruption of facial appearance [37]. It was found that HNSCC exhibited immunosuppressive features [38]. Tumor-infiltrated B cells are an essential component of the tumor microenvironment (TME), and they play a fundamental role in the immunotherapeutic response as the primary effectors of humoral immunity [39]. However, in TME, B cells exert both pro- and anti-tumor effects. On the one hand, B cells exert their anti-tumor effects in three ways: 1. secreting tumor-specific antibodies, promoting T cell functions [40]; 2. secreting Granzyme B, interferon-γ (IFNγ), and tumor necrosis factor-related apoptosis-inducing ligand (TRAIL) to kill tumor cells directly [41–43]; 3. secreting immunoglobulins to mediate the processes associated with complement-dependent cytotoxicity (CDC), antibody-dependent cell-mediated cytotoxicity (ADCC), antibody-dependent cell-mediated phagocytosis (ADCP) [44]. On the other hand, B cells can also promote tumor progression through several mechanisms, such as TIM-1$^+$ Bregs. B cells can produce IL10, which inhibits the immune-killing effect of T cells [15]. B cells can also produce cytokines such as IL-1β to promote the growth and invasion of cancer cells [45]. CD20$^+$ B cells in TME are mainly associated with a good prognosis in different solid tumors, including melanoma, colorectal, gastric, and non-small cell lung cancer [46–49]. Distel et al. showed that CD20$^+$ tumor-infiltrating lymphocytes are prognostic indicators of favorable outcomes in early-stage HNSCC. Conversely, an inverse correlation emerged in advanced stages, implying a plasticity in B cells' functional and compositional attributes throughout the disease [50]. Consequently, B cells may be significant in the prognosis of HNSCC patients and may work as a predictive biomarker for immunotherapy [51].

Historically, investigations into B cells have predominantly relied on Immunohistochemistry (IHC) or flow cytometry techniques, which inherently constrain the precise categorization of B cell populations. Typically, based on their differentiation progression, B cells can be stratified into three primary subgroups: naïve B cells, Memory B Cells (MBCs), and Antibody-Secreting Cells (ASCs), each manifesting distinctive phenotypic profiles and functions in TME [52]. Due to the classification, functionality, and spatial presence of different B cell subgroups in different tissues and the low proportion of B cells in specific tumors, B cells are also commonly classified as CD19$^+$CD20$^+$ B cells and CD138$^+$SDC1$^+$ PCs. The application of scRNA-seq provides higher resolution for identifying B cell subgroups. In our study, we analyzed the intratumoral B cell profiles within HNSCC and investigated the functional information of tumor-responsive B cells. From a longitudinal perspective, we used cluster analysis to divide B cells into NBs, GCBs, MBs, and PCs for



(caption on next page)

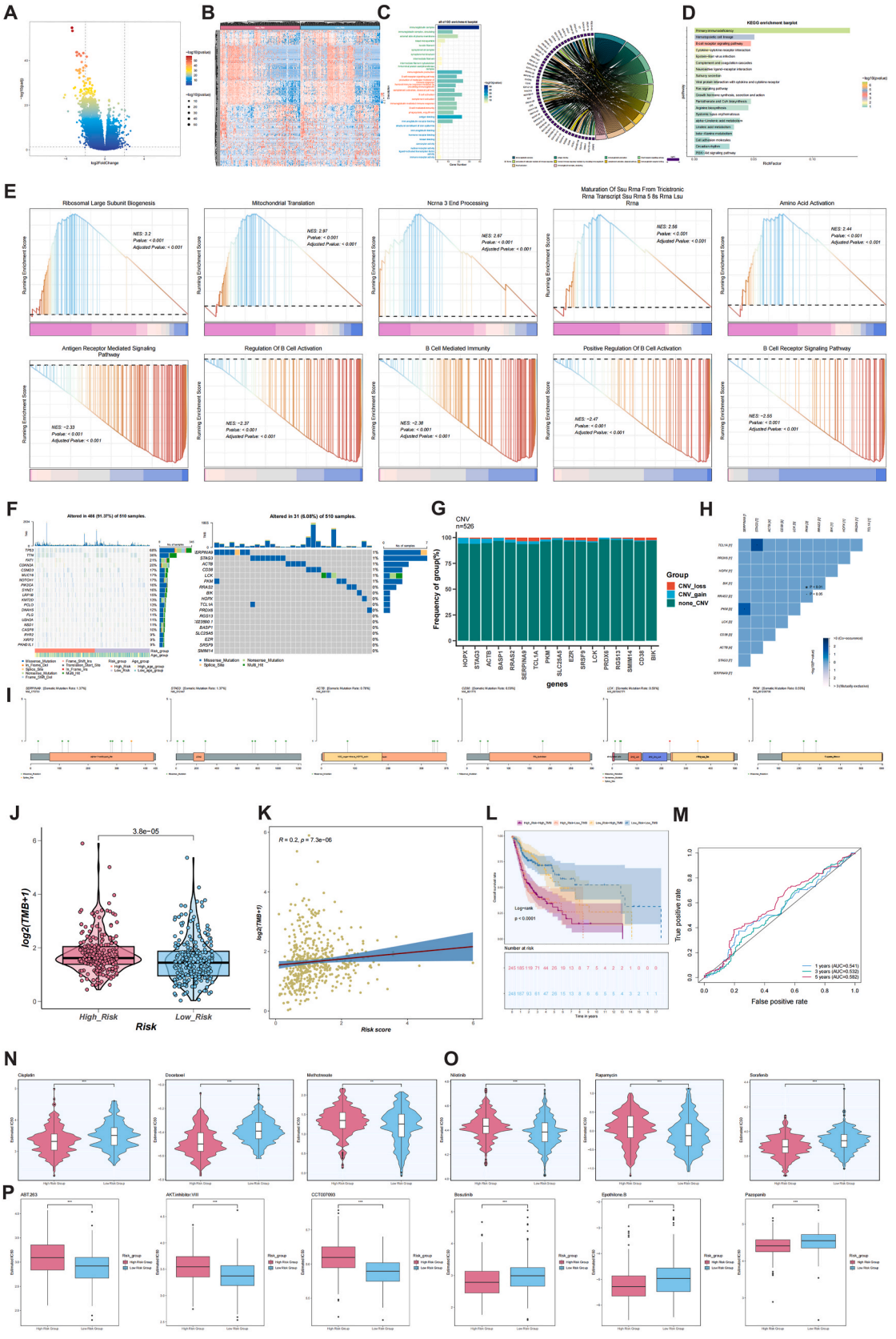
Fig. 12. | Construction and verification of Nomogram and analysis of immune cell infiltration. (A) Multivariate analyses of the prognosis model. (B) Nomogram plot constructed by combining the risk score and clinical features. (C–D) AUC values evaluate the prognostic performance of the risk score model at 1, 3 and 5 years. (E) Deviations of actual 1-year, 3-year, and 5-year survival probabilities. (F)A heat map of principal component analysis for the immune cell infiltration between high and low risk groups. (G–H) Differential analysis of estimate score, immune score, stromal score, and tumor purity between high and low risk groups. (I)A boxplot of immune cells in the two risk subgroups. (J)Differential analysis of immune cell infiltration distribution between high and low risk groups. (K)The correlation between risk score and immune cells. The horizontal axis represents the correlation with the risk score, the vertical axis represents the proportion of immune cells, and the size of the circle represents the absolute value of the correlation coefficient. The larger the circle, the higher the correlation. (L)The correlation between prognostic genes and immune cells. Red indicates positive correlation, blue indicates negative correlation, $^{*}P < 0.05$, $^{**}P < 0.01$. (M)An overview of the correlation of immune cells. (N)The distribution of prognostic genes, estimate score, immune score, stromal score, tumor purity and immune cells in high and low risk groups. (O) RRAS2 was positively correlated with NBs, MBs, PB. (P) RRAS2 was negatively correlated with stromal score, immune score and tumor purity. (Q) Analysis of the correlation between RRAS2 and immune cells.

the first time. High-throughput scRNA-Seq and pseudotime trajectory analysis showed that in the early stage of tumor-infiltrating B cell development in HNSCC, NBs and GCBs with high differentiation ability were primarily present. In contrast, in the later stage of development, a large number of MBs and a small number of PCs appeared. Nevertheless, the composition of B cells changes differently during the development of different tumors and promotes or inhibits tumor development. In advanced non-small cell lung cancer (NSCLC), a decrease in naive-like B cells is associated with a poor prognosis, whereas plasma-like B cells inhibit cancer cell growth in the early stage of NSCLC but promote cell growth in the advanced stage of NSCLC [53]. B cells isolated from early-stage colorectal cancer (CRC) exhibited a pre-B cell-like phenotype characterized by the expression of tumor suppressor molecules. In contrast, B cells derived from advanced CRC displayed a propensity for differentiation into plasma cells and predicted a poor prognosis [54]. Research has revealed that activated B cell populations exhibit a better prognosis, higher rates of immune cell infiltration, and significant levels of immune checkpoint expression, including elevated PD-L1 levels [55]. In addition, the composition of tumor-infiltrating B cells varies in many ways across the transversal dimensions of the tumor. The objective diversity of B cell subgroups in different tumors and the subjective individualized approach of single-cell analysis contribute to the diversity of B cell subgroup definitions. For example, a higher proportion of IgG plasma cells were found at the tumor sites than in the adjacent normal mucosal tissues in CRC [56]; however, in a study in NSCLC, there were more B cells in the adjacent tissues than in the tumor, which may be related to infiltration of inflammatory cells [53]. B cells show less infiltration in esophageal squamous cell carcinoma (ESCC), breast invasive carcinoma (BRCA), kidney cancer (KIRC), and hepatocellular carcinoma (HCC) than lung cancer (LC), colon adenocarcinoma (COAD), and skin cutaneous melanoma (SKCM) [57]. The above results can reveal the existence of specific B cell subgroups of TME in different tissues, thus revealing the immunogenic or immunosuppressive of TME.

The germinal center (GC) is located in the follicular zone of the secondary lymphoid organs. A study has generated a blueprint for B cells, categorizing them into two distinct developmental pathways: extrafollicular and germinal center. The data indicates a balance between the extrafollicular and germinal center responses in tumor-infiltrating B cells, suggesting that humoral immunity may be utilized for B cell-targeting immunotherapy [58]. This particular structure aggregates B cells during the humoral immune response [59]. Antigen-specific B cells accumulate in this remarkable structure and subsequently expand to produce high-frequency mutations in somatic cells, undergo antibody-type switching, and eventually turn into high-affinity B cells against the antigen, a specific means of targeting foreign invading antigens [60,61]. B cells undergo somatic high-frequency mutations along with high-affinity clones directed against antigens, and the clones produce immunoglobulin molecules with higher affinity for antigens [62,63]. Increased immunoglobulin affinity also requires antibody-type conversion. This process that requires the involvement of an enzyme called activation-induced cytidine deaminase (AID), which is highly expressed in B cells in the germinal center and is an essential enzyme in regulating the immune response in the germinal center, as well as being involved in the process of high-frequency mutation in somatic cells described above [64,65]. AID-mediated high-frequency mutations and type conversions make B-cell mutation rates much more likely, implying the potential for generating uncontrolled cancer cell clones and that germinal centers are critical sites for lymphocytic cancer production [66]. The immune response to GC depends on follicular helper T cells (T_{fh} cells), found in the follicular zone of secondary lymphoid organs, and co-localizes with GCBs to assist in B cell differentiation and clonal selection [67,68]. T_{fh} cells secrete cytokines and membrane surface molecules in direct contact with the GCBs, transmitting signals to the GCBs to help them differentiate into MBs or high-affinity PCs [69,70]. The recruitment of suitable Tfh cells into the GC during the humoral immune response is vital for the maturation of antibody affinity and the generation of an effective humoral immune response.

This paper finds that molecules on GCBs enable B-cell action on tumor cells by expressing the high-affinity ligands CD99 and SEMA4, which are involved in synaptic junctions to transmit signals for intercellular communication. SEMA4 is a member of the Semaphorin family of molecules, named for the fact that family members all have a sema structural domain consisting of 7 β-folded sheets in their molecular structures, consists of 8 subfamilies, and initially associated with axon guidance and neural development [71–73]. Because of its functional specificity, the immune system is most commonly characterized by migratory movements of cells. The most crucial function of Semaphorin signaling is its involvement in the migratory process of immune cells.

Dendritic cells (DCs) require SEMA4A for T-cell activation, and Th1 cells also precisely express SEMA4A, promoting the differentiation of activated T cells into Th1 subgroups [74]. SEMA4D is also abundantly expressed on activated T cells and has been reported to stimulate the production of high-affinity antibodies by humoral immunity [75,76]. Semaphorin involves several physiological processes such as cardiac development [77], angiogenesis and regeneration [78,79], osteoclastogenesis [80], and immune regulation [81]. It also engages in many pathological human disorders, including tumorigenesis and metastasis [82], neurogenic disorders [83], and immune disorders [84]. A growing amount of evidence suggests that Semaphorin plays an essential role in tumor development and



(caption on next page)

Fig. 13. | (A–B) DEGs analysis between high and low risk groups. (C–E) Functional enrichment analysis between the high and low risk groups of HNSCC patients. (C) Results of the GO enrichment analysis. (D) Results of the KEGG enrichment analysis. (E) Enrichment plots from the gene set enrichment analysis. Biological processes enriched in the high-risk group of HNSCC patients. Signaling pathways enriched in the low-risk HNSCC samples. (F) The waterfall map showed the distribution of the top 20 genes of TMB in the high and low risk groups. The horizontal coordinates represent the different samples and the vertical coordinates represent the genes. The different colors in the note below represent different types of mutations, the upper bars represent the probability of mutation in each sample, and the right bars represent the proportion of different types of mutations in each gene. (G) The CNV amplifcation, deletion or absence of data of risk genes. (H) Analysis of the correlation between gene and gene mutation. Mutually exclusive: if one gene is mutated, there is a high probability that no mutation will occur in the corresponding gene. Co−occurrence: two genes mutate at the same time. (I) The sample numbers, type, and location of mutations in each gene. The figure notes below represent the different mutation types. The dots in the figure represent the samples in which the mutation occurred, and the connected vertical lines indicate the location of the mutation. (J) TMB in high and low risk groups. (K) Correlation analysis between risk scores and TMB ($P < 0.001$). (L) Survival analysis of high risk-high TMB, high risk-low TMB, low risk-high TMB, low risk-low TMB. (M) AUC values evaluate the prognostic performance of the TMB model at 1, 3 and 5 years. (N–P) Chemotherapy and targeted drugs sensitivity of high- and low-risk subgroups.

holds promise as a potential therapeutic target in various tumors.

CD99 is a 32 kDa transmembrane protein encoded by the pseudoautosomal gene *MIC2* [85,86]. CD99 is well known to participate in several crucial biological processes, including cell adhesion [87,88], migration [89,90], apoptosis [89,91], T cell differentiation [92,93], inflammation [94], and immune responses [95], which suggests that CD99 plays a vital role in the peripheral immune response and cell differentiation. CD99 exerts dual effects on the clinical outcomes of patients afflicted with solid tumors. On the one hand, CD99 may act as a tumor suppressor in certain tumors. In patients with NSCLC, increased levels of CD99 stromal expression are associated with better prognosis [96]. In osteosarcoma, CD99 expression is down-regulated in the majority of patient samples. In contrast, overexpression of CD99 significantly negatively impacts osteosarcoma cell growth and survival, eliminating tumorigenicity and metastatic capacity [97]. On the other hand, CD99 can also be an oncogene for tumor migration and invasion. CD99 is strongly correlated with low survival and poor prognosis in gastric adenocarcinoma [98]. Overexpression of CD99 significantly increased migration and invasiveness of glioma cells and decreased survival [99,100]. High expression of CD99 is a distinguishing feature of Ewing sarcoma (EWS). As a result, CD99 stands out as the most frequently used diagnostic marker and therapeutic target for EWS [101]. Our study found that tumor cells are the only target cells of CD99 and SEMA4 signaling pathway, and the source of CD99 and SEMA4 ligands is also only GCBs, so it is speculated that GCBs may also be able to promote the development of HNSCC through CD99 and SEMA4 signaling pathway. Hence, CD99 and SEMA4 might serve as robust biomarkers for HNSCC and a promising target for tumor therapy.

By integrating risk scores and clinical characteristics, we created a risk score model based on 18 GCBs risk genes, and the independent prognostic value of this model was subsequently validated. We used CIBERSORT, ssGSEA, and Xcell to analyze immune cell infiltration in HNSCC samples and to distinguish between different immune cell subgroups in the immune microenvironment, which have different prognostic implications. Compared with patients in the high-risk group, the low-risk group had a higher immune score and a greater abundance of tumor-infiltrating immune cells (TIIC), which resulted in a better prognosis. The low-risk group had more $CD8^+$ T cell infiltration, and the high-risk group had more macrophage M2 infiltration. $CD8^+$ T cells are the central effector cells in anti-tumor immunity by directly lysing tumor cells by releasing multiple cytokines and enzymes [102,103]. In the breast cancer tumor microenvironment interacting with tumor cells, the higher the number of $CD8^+$ T cells, the better the regulation of anti-tumor activity [104]. $CD8^+$ T cells are closely linked to the survival of tumor patients, with patients with higher $CD8^+$ T cell counts having a higher 5-year survival rate [105] and a better prognosis [106,107]. M2 macrophages can suppress tumor immunity by secreting the immunosuppressive factors IL-10 and TGF-β, and at the same time, secrete EGF to promote neovascularisation within the tumor, which promotes tumor invasion and metastasis [108,109]. Increased infiltration of M2 macrophages in breast cancers indicates a poor prognosis [110,111]. These findings are consistent with our results in HNSCC. In summary, the ratio of immune effector cells to suppressor cells, the location of immune cell infiltration, and the degree of immune cell activation in TME also affect the prognosis of patients and their clinical immunotherapy response.

There are some limitations to this study. Firstly, the data in this study came from part of the open single-cell cohort, which may not be representative because of the small sample size of B-cell studies in HNSCC. Secondly, there is a lack of studies on the association of B cells with treatment efficacy in HNSCC, which needs to be analyzed and studied in the future by further expanding the sample size. Thirdly, some studies related to assessing B cells on prognosis and efficacy in HNSCC are still controversial. These differences may be because B cell subgroups are highly functionally plastic and heterogeneous, so the composition and localization of tumor-infiltrating B cells vary. Hence, future validation in clinical trials is necessary.

In conclusion, we report that GCBs are B-cell-specific prognostic biomarkers for survival in HNSCC. The emergence and development of HNSCC are dependent on somatic cell gene mutations and are closely connected with the tumor microenvironment. All anti-tumor or pro-tumor factors form a network in the tumor microenvironment that determines tumor initiation, progression, and outcome. Therefore, we can further apply B cells in diagnosis and therapeutic evaluation and develop targeted therapies against the tumor immune microenvironment. These results underscore a compelling rationale for employing B-cell markers in patient stratification and exploring innovative therapeutics targeting B-cell subgroups.

Disclosure of funding

This study was funded by National Natural Science Foundation of China (82301158), China Postdoctoral Science Foundation

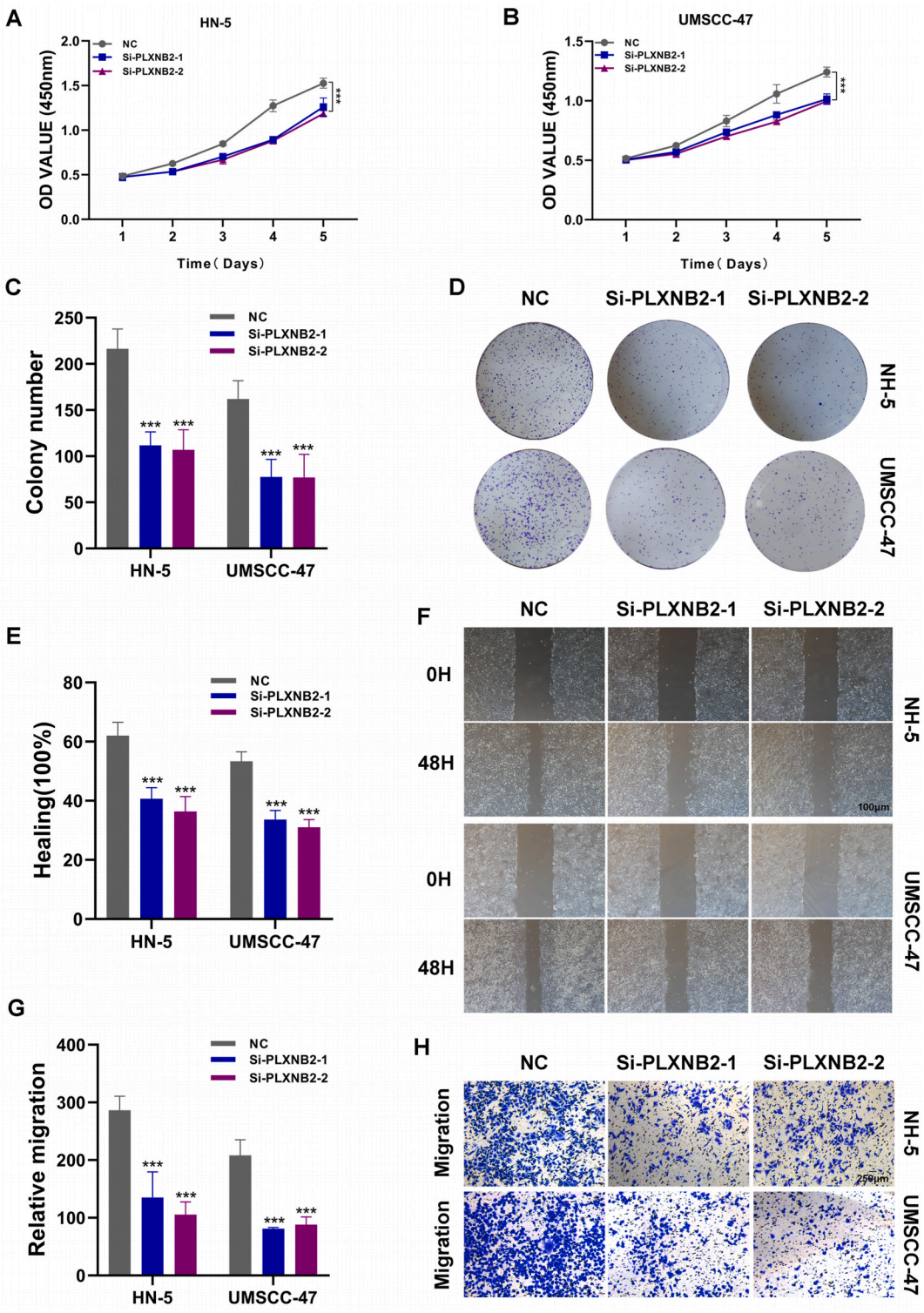


Fig. 14. | PLXNB2 significantly affects the proliferation and migration of head and neck cancer cell lines. (A, B) CCK-8 experiment. After PLXNB2 knockdown, the proliferation ability of HN-5 and UMSCC-47 cell lines decreased significantly. (C, D) Plate cloning experiment. After PLXNB2 knockdown, the colony-forming ability of HN-5 and UMSCC-47 cell lines was significantly decreased. (E, F) wound healing test. After PLXNB2 knockdown, the migration ability of HN-5 and UMSCC-47 cell lines decreased significantly. (G, H) Transwell experiment. After PLXNB2 knockdown, the migration and invasion ability of HN-5 and UMSCC-47 cell lines were significantly reduced. (*P < 0.05; **P < 0.01; ***P < 0.001).

(2023M742332), Foundation of National Center for Translational Medicine (Shanghai) SHU Branch (SUITM-202304) and Shanghai Post-doctoral Excellence Incentive Program (2023787).

Ethics approval and consent to participate

Ethical approval (PJ2023-12-45) for this study was obtained from the Ethics Committee of the First Affiliated Hospital of Anhui Medical University. As the other data used in this study are publicly available, no ethical approval was required.

Consent for publication

All authors have read the manuscript and agree to publish.

Availability of data and material

The raw data for this study were generated at the corresponding archives, further inquiries can be directed to the corresponding authors.

Funding

This study was funded by National Natural Science Foundation of China (82301158), China Postdoctoral Science Foundation (2023M742332), Foundation of National Center for Translational Medicine (Shanghai) SHU Branch (SUITM-202304) and Shanghai Post-doctoral Excellence Incentive Program (2023787).

CRediT authorship contribution statement

Li Lin: Writing – review & editing. **Jiani Zou:** Writing – original draft. **Shengbin Pei:** Validation. **Wenyi Huang:** Investigation. **Yichi Zhang:** Supervision. **Zhijie Zhao:** Writing – review & editing, Formal analysis, Data curation. **Yantao Ding:** Writing – review & editing. **Can Xiao:** Writing – review & editing, Writing – original draft, Supervision.

Declaration of competing interest

None.

Abbreviation

(HNSCC) Head and neck squamous carcinoma
(OS) Overall survival
(TIME) Tumor immune microenvironment
(bulk RNA-seq) Bulk RNA-sequencing
(scRNA-seq) Single-cell RNA-sequencing
(HNSC) Head and Neck Cancer
(T_{fh} cells) Follicular helper T cells
(BCR) B cell receptor
(PBL) Peripheral blood
(DEGs) Differentially expressed genes
(NBs) Naive B cells
(GCBs) Germinal center B cells
(MBs) Memory B cells
(PCs) Plasma cells
(GEO) Gene Expression Omnibus
(TMB) Tumor mutational burden
(CT) Computed Tomography
(MRI) Magnetic Resonance Imaging
(TIME) Tumor immune microenvironment
(Bregs) Regulatory B cell subgroups
GCBs
(LASSO) Least Absolute Shrinkage and Selection Operator
(AUC) Area Under the ROC Curve
(TME) Tumor microenvironment
(Tregs) T cells regulatory
(GO) Gene Ontology

(KEGG) Kyoto Encyclopedia of Genes and Genomes
(IFNγ) Interferon-γ
(TRAIL) Tumor necrosis factor-related apoptosis-inducing ligand
(CDC) Complement-dependent cytotoxicity
(ADCC) Antibody-dependent cell-mediated cytotoxicity
(ADCP) Antibody-dependent cell-mediated phagocytosis
(IHC) Immunohistochemistry
(ASCs) Antibody-Secreting Cells
(NSCLC) Non-small cell lung cancer
(CRC) Colorectal cancer
(ESCC) Esophageal squamous cell carcinoma
(BRCA) Breast invasive carcinoma
(KIRC) Kidney cancer
(HCC) hepatocellular carcinoma
(LC) Lung cancer
(COAD) Colon adenocarcinoma
(SKCM) Skin cutaneous melanoma
(GC) Germinal center
(AID) Activation-induced cytidine deaminase
(DCs) Dendritic cells
(EWS)
Ewing sarcoma

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2024.e37726.

References

[1] C.R. Leemans, P. Snijders, R.H. Brakenhoff, The molecular landscape of head and neck cancer, Nat. Rev. Cancer 18 (2018) 269–282.
[2] A. Wyss, M. Hashibe, S.C. Chuang, et al., Cigarette, cigar, and pipe smoking and the risk of head and neck cancers: pooled analysis in the International Head and Neck Cancer Epidemiology Consortium, Am. J. Epidemiol. 178 (2013) 679–690.
[3] J.W. Carlisle, C.E. Steuer, T.K. Owonikoko, et al., An update on the immune landscape in lung and head and neck cancers, CA A Cancer J. Clin. 70 (2020) 505–517.
[4] D.C. Travassos, D. Fernandes, E. Massucato, et al., Squamous cell carcinoma antigen as a prognostic marker and its correlation with clinicopathological features in head and neck squamous cell carcinoma: systematic review and meta-analysis, J. Oral Pathol. Med. 47 (2018) 3–10.
[5] S.S. Sawant, S.M. Zingde, M.M. Vaidya, Cytokeratin fragments in the serum: their utility for the management of oral cancer, Oral Oncol. 44 (2008) 722–732.
[6] A. Kulasinghe, L. Kenny, C. Punyadeera, Circulating tumour cell PD-L1 test for head and neck cancers, Oral Oncol. 75 (2017) 6–7.
[7] R.M. Nagler, M. Barak, M. Peled, et al., Early diagnosis and treatment monitoring roles of tumor markers Cyfra 21-1 and TPS in oral squamous cell carcinoma, CANCER-AM CANCER SOC 85 (1999) 1018–1025.
[8] M.B. Amin, F.L. Greene, S.B. Edge, et al., The Eighth Edition AJCC Cancer Staging Manual: continuing to build a bridge from a population-based to a more "personalized" approach to cancer staging, CA A Cancer J. Clin. 67 (2017) 93–99.
[9] D. DE Paz, C.K. Young, H.T. Chien, et al., Prognostic roles of SCC antigen, CRP and CYFRA 21-1 in oral cavity squamous cell carcinoma, Anticancer Res. 39 (2019) 2025–2033.
[10] T. Fu, L.J. Dai, S.Y. Wu, et al., Spatial architecture of the immune microenvironment orchestrates tumor immunity and therapeutic response, J. Hematol. Oncol. 14 (2021) 98.
[11] T. Tang, X. Huang, G. Zhang, et al., Advantages of targeting the tumor immune microenvironment over blocking immune checkpoint in cancer immunotherapy, SIGNAL TRANSDUCT TAR 6 (2021) 72.
[12] C.M. Laumont, A.C. Banville, M. Gilardi, et al., Tumour-infiltrating B cells: immunological mechanisms, clinical impact and therapeutic opportunities, Nat. Rev. Cancer 22 (2022) 414–430.
[13] S.S. Kim, W.A. Sumner, S. Miyauchi, et al., Role of B Cells in responses to checkpoint blockade immunotherapy and overall survival of cancer patients, Clin. Cancer Res. 27 (2021) 6075–6082.
[14] O. Demaria, S. Cornen, M. Daeron, et al., Harnessing innate immunity in cancer therapy, NATURE 574 (2019) 45–56.
[15] S.S. Wang, W. Liu, D. Ly, et al., Tumor-infiltrating B cells: their role and application in anti-tumor immunity in lung cancer, Cell. Mol. Immunol. 16 (2019) 6–18.
[16] J.R. Conejo-Garcia, S. Biswas, R. Chaurio, et al., Neglected no more: B cell-mediated anti-tumor immunity, Semin. Immunol. 65 (2023) 101707.
[17] J. Tang, D. Ramis-Cabrer, V. Curull, et al., B cells and tertiary lymphoid structures influence survival in lung cancer patients with resectable tumors, Cancers 12 (2020).
[18] D. Pretscher, L.V. Distel, G.G. Grabenbauer, et al., Distribution of immune cells in head and neck cancer: CD8+ T-cells and CD20+ B-cells in metastatic lymph nodes are associated with favourable outcome in patients with oro- and hypopharyngeal carcinoma, BMC Cancer 9 (2009) 292.
[19] A. Lechner, H.A. Schlosser, M. Thelen, et al., Tumor-associated B cells and humoral immune response in head and neck squamous cell carcinoma, OncoImmunology 8 (2019) 1535293.
[20] O. Wood, J. Woo, G. Seumois, et al., Gene expression analysis of TIL rich HPV-driven head and neck tumors reveals a distinct B-cell signature when compared to HPV independent tumors, Oncotarget 7 (2016) 56781–56797.
[21] P.H. Li, X.Y. Kong, Y.Z. He, et al., Recent developments in application of single-cell RNA sequencing in the tumour immune microenvironment and cancer therapy, MILITARY MED RES 9 (2022) 52.
[22] S. Sun, J. Zhu, Y. Ma, et al., Accuracy, robustness and scalability of dimensionality reduction methods for single-cell RNA-seq analysis, Genome Biol. 20 (2019) 269.

[23] X. Qiu, Q. Mao, Y. Tang, et al., Reversed graph embedding resolves complex single-cell trajectories, Nat. Methods 14 (2017) 979–982.
[24] J. Qi, H. Sun, Y. Zhang, et al., Single-cell and spatial analysis reveal interaction of FAP(+) fibroblasts and SPP1(+) macrophages in colorectal cancer, Nat. Commun. 13 (2022) 1742.
[25] P. Charoentong, F. Finotello, M. Angelova, et al., Pan-cancer immunogenomic analyses reveal genotype-immunophenotype relationships and predictors of response to checkpoint blockade, Cell Rep. 18 (2017) 248–262.
[26] A.M. Newman, C.L. Liu, M.R. Green, et al., Robust enumeration of cell subsets from tissue expression profiles, Nat. Methods 12 (2015) 453–457.
[27] K. Yoshihara, M. Shahmoradgoli, E. Martinez, et al., Inferring tumour purity and stromal and immune cell admixture from expression data, Nat. Commun. 4 (2013) 2612.
[28] M. Ashburner, C.A. Ball, J.A. Blake, et al., Gene ontology: tool for the unification of biology. The Gene Ontology Consortium, Nat. Genet. 25 (2000) 25–29.
[29] M. Kanehisa, M. Furumichi, Y. Sato, et al., KEGG: integrating viruses and cellular organisms, Nucleic Acids Res. 49 (2021) D545–D551.
[30] A. Subramanian, P. Tamayo, V.K. Mootha, et al., Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles, P NATL ACAD SCI USA 102 (2005) 15545–15550.
[31] D. Maeser, R.F. Gruener, R.S. Huang, oncoPredict: an R package for predicting in vivo or cancer patient drug response and biomarkers from cell line screening data, BRIEF BIOINFORM 22 (2021).
[32] P. Geeleher, N. Cox, R.S. Huang, pRRophetic: an R package for prediction of clinical chemotherapeutic response from tumor gene expression levels, PLoS One 9 (2014) e107468.
[33] A. Cerutti, M. Cols, I. Puga, Marginal zone B cells: virtues of innate-like antibody-producing lymphocytes, Nat. Rev. Immunol. 13 (2013) 118–132.
[34] S. Gu, L. Qian, Y. Zhang, et al., Significance of intratumoral infiltration of B cells in cancer immunotherapy: from a single cell perspective, BBA-REV CANCER 1876 (2021) 188632.
[35] K. Starska-Kowarska, The role of different immunocompetent cell populations in the pathogenesis of head and neck cancer-regulatory mechanisms of pro- and anti-cancer activity and their impact on immunotherapy, Cancers 15 (2023).
[36] A.R. Cillo, C. Kurten, T. Tabib, et al., Immune landscape of viral- and carcinogen-driven head and neck cancer, Immunity 52 (2020) 183–199.
[37] F. Bray, J. Ferlay, I. Soerjomataram, et al., Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries, CA A Cancer J. Clin. 68 (2018) 394–424.
[38] R. Mandal, Y. Senbabaoglu, A. Desrichard, et al., The head and neck cancer immune landscape and its immunotherapeutic implications, JCI Insight 1 (2016) e89829.
[39] Z.N. Willsmore, R.J. Harris, S. Crescioli, et al., B cells in patients with melanoma: implications for treatment with checkpoint inhibitor antibodies, Front. Immunol. 11 (2020) 622442.
[40] D. Ghosh, W. Jiang, D. Mukhopadhyay, et al., New insights into B cells as antigen presenting cells, Curr. Opin. Immunol. 70 (2021) 129–137.
[41] A.M. Gocher, C.J. Workman, D. Vignali, Interferon-gamma: teammate or opponent in the tumour microenvironment? Nat. Rev. Immunol. 22 (2022) 158–172.
[42] W.L. Xu, R.L. Wang, Z. Liu, et al., Granzyme B-producing B cells function as a feedback loop for T helper cells in liver transplant recipients with acute rejection, Inflammation 44 (2021) 2270–2278.
[43] G.V. Sharonov, E.O. Serebrovskaya, D.V. Yuzhakova, et al., B cells, plasma cells and antibody repertoires in the tumour microenvironment, Nat. Rev. Immunol. 20 (2020) 294–307.
[44] S. Herter, F. Herting, O. Mundigl, et al., Preclinical activity of the type II CD20 antibody GA101 (obinutuzumab) compared with rituximab and ofatumumab in vitro and in xenograft models, MOL CANCER THER 12 (2013) 2031–2042.
[45] G.J. Yuen, E. Demissie, S. Pillai, B lymphocytes and cancer: a love-hate relationship, TRENDS CANCER 2 (2016) 747–757.
[46] A. Hennequin, V. Derangere, R. Boidot, et al., Tumor infiltration by Tbet+ effector T cells and CD20+ B cells is associated with survival in gastric cancer patients, OncoImmunology 5 (2016) e1054598.
[47] J. Berntsson, B. Nodin, J. Eberhard, et al., Prognostic impact of tumour-infiltrating B cells and plasma cells in colorectal cancer, Int. J. Cancer 139 (2016) 1129–1139.
[48] A. Ladanyi, J. Kiss, A. Mohos, et al., Prognostic impact of B-cell density in cutaneous melanoma, CANCER IMMUNOL IMMUN 60 (2011) 1729–1738.
[49] K.I. Al-Shibli, T. Donnem, S. Al-Saad, et al., Prognostic effect of epithelial and stromal lymphocyte infiltration in non-small cell lung cancer, Clin. Cancer Res. 14 (2008) 5220–5227.
[50] L.V. Distel, R. Fickenscher, K. Dietel, et al., Tumour infiltrating lymphocytes in squamous cell carcinoma of the oro- and hypopharynx: prognostic impact may depend on type of treatment and stage of disease, Oral Oncol. 45 (2009) e167–e174.
[51] M. Wouters, B.H. Nelson, Prognostic significance of tumor-infiltrating B cells and plasma cells in human cancer, Clin. Cancer Res. 24 (2018) 6125–6135.
[52] S. Gu, L. Qian, Y. Zhang, et al., Significance of intratumoral infiltration of B cells in cancer immunotherapy: from a single cell perspective, BBA-REV CANCER 1876 (2021) 188632.
[53] J. Chen, Y. Tan, F. Sun, et al., Single-cell transcriptome and antigen-immunoglobin analysis reveals the diversity of B cells in non-small cell lung cancer, Genome Biol. 21 (2020) 152.
[54] W. Wang, Y. Zhong, Z. Zhuang, et al., Multiregion single-cell sequencing reveals the transcriptional landscape of the immune microenvironment of colorectal cancer, Clin. Transl. Med. 11 (2021) e253.
[55] K. Li, C. Zhang, R. Zhou, et al., Single cell analysis unveils B cell-dominated immune subtypes in HNSCC for enhanced prognostic and therapeutic stratification, Int. J. Oral Sci. 16 (2024) 29.
[56] J. Xia, Z. Xie, G. Niu, et al., Single-cell landscape and clinical outcomes of infiltrating B cells in colorectal cancer, Immunology 168 (2023) 135–151.
[57] Y. Zheng, Z. Chen, Y. Han, et al., Immune suppressive landscape in the human esophageal squamous cell carcinoma microenvironment, Nat. Commun. 11 (2020) 6268.
[58] J. Ma, Y. Wu, L. Ma, et al., A blueprint for tumor-infiltrating B cells across human cancers, Science 384 (2024) eadj4857.
[59] G.D. Victora, M.C. Nussenzweig, Germinal centers, Annu. Rev. Immunol. 40 (2022) 413–442.
[60] C. Berek, A. Berger, M. Apel, Maturation of the immune response in germinal centers, Cell 67 (1991) 1121–1129.
[61] J. Jacob, G. Kelsoe, K. Rajewsky, et al., Intraclonal generation of antibody mutants in germinal centres, NATURE 354 (1991) 389–392.
[62] M.G. Weigert, I.M. Cesari, S.J. Yonkovich, et al., Variability in the lambda light chain sequences of mouse antibody, NATURE 228 (1970) 1045–1047.
[63] D. McKean, K. Huppi, M. Bell, et al., Generation of antibody diversity in the immune response of BALB/c mice to influenza virus hemagglutinin, P NATL ACAD SCI USA 81 (1984) 3180–3184.
[64] M. Muramatsu, K. Kinoshita, S. Fagarasan, et al., Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme, Cell 102 (2000) 553–563.
[65] R. Pavri, M.C. Nussenzweig, AID targeting in antibody diversity, Adv. Immunol. 110 (2011) 1–26.
[66] A. Nussenzweig, M.C. Nussenzweig, Origin of chromosomal translocations in lymphoid cancer, Cell 141 (2010) 27–38.
[67] D. Breitfeld, L. Ohl, E. Kremmer, et al., Follicular B helper T cells express CXC chemokine receptor 5, localize to B cell follicles, and support immunoglobulin production, J. Exp. Med. 192 (2000) 1545–1552.
[68] P. Schaerli, K. Willimann, A.B. Lang, et al., CXC chemokine receptor 5 expression defines follicular homing T cells with B cell helper function, J. Exp. Med. 192 (2000) 1553–1562.
[69] K. Luthje, A. Kallies, Y. Shimohakamada, et al., The development and fate of follicular helper T cells defined by an IL-21 reporter mouse, Nat. Immunol. 13 (2012) 491–498.
[70] D. Eto, C. Lao, D. DiToro, et al., IL-21 and IL-6 are critical for different aspects of B cell immunity and redundantly induce optimal follicular helper CD4 T cell (Tfh) differentiation, PLoS One 6 (2011) e17739.
[71] A.L. Kolodkin, D.J. Matthes, C.S. Goodman, The semaphorin genes encode a family of transmembrane and secreted growth cone guidance molecules, Cell 75 (1993) 1389–1399.

[72] P.K. Hota, M. Buck, Plexin structures are coming: opportunities for multilevel investigations of semaphorin guidance receptors, their cell signaling mechanisms, and functions, Cell. Mol. Life Sci. 69 (2012) 3765–3805.
[73] Y. Luo, D. Raible, J.A. Raper, Collapsin: a protein in brain that induces the collapse and paralysis of neuronal growth cones, Cell 75 (1993) 217–227.
[74] A. Kumanogoh, T. Shikina, K. Suzuki, et al., Nonredundant roles of Sema4A in the immune system: defective T cell priming and Th1/Th2 regulation in Sema4A-deficient mice, Immunity 22 (2005) 305–316.
[75] A. Kumanogoh, C. Watanabe, I. Lee, et al., Identification of CD72 as a lymphocyte receptor for the class IV semaphorin CD100: a novel mechanism for regulating B cell signaling, Immunity 13 (2000) 621–631.
[76] C. Gu, Y. Yoshida, J. Livet, et al., Semaphorin 3E and plexin-D1 control vascular pattern independently of neuropilins, Science 307 (2005) 265–268.
[77] A. Casazza, V. Finisguerra, L. Capparuccia, et al., Sema3E-Plexin D1 signaling drives human cancer cell invasiveness and metastatic spreading in mice, J. Clin. Invest. 120 (2010) 2684–2698.
[78] F. Maione, S. Capano, D. Regano, et al., Semaphorin 3A overcomes cancer hypoxia and metastatic dissemination induced by antiangiogenic treatment in mice, J. Clin. Invest. 122 (2012) 1832–1848.
[79] F. Maione, F. Molla, C. Meda, et al., Semaphorin 3A is an endogenous angiogenesis inhibitor that blocks tumor growth and normalizes tumor vasculature in transgenic mouse models, J. Clin. Invest. 119 (2009) 3356–3372.
[80] M. Hayashi, T. Nakashima, M. Taniguchi, et al., Osteoprotection by semaphorin 3A, NATURE 485 (2012) 69–74.
[81] G.M. Delgoffe, S.R. Woo, M.E. Turnis, et al., Stability and function of regulatory T cells is maintained by a neuropilin-1-semaphorin-4a axis, NATURE 501 (2013) 252–256.
[82] L. Tamagnone, Emerging role of semaphorins as major regulatory signals and potential therapeutic targets, in: Cancer. CANCER CELL, vol. 22, 2012, pp. 145–152.
[83] S. Nojima, T. Toyofuku, H. Kamao, et al., A point mutation in Semaphorin 4A associates with defective endosomal sorting and causes retinal degeneration, Nat. Commun. 4 (2013) 1406.
[84] T. Okuno, Y. Nakatsuji, A. Kumanogoh, The role of immune semaphorins in multiple sclerosis, FEBS Lett. 585 (2011) 3829–3835.
[85] P.N. Goodfellow, B. Pym, C. Pritchard, et al., MIC2: a human pseudoautosomal gene, PHILOS T R SOC B 322 (1988) 145–154.
[86] M. Pasello, M.C. Manara, K. Scotlandi, CD99 at the crossroads of physiology and pathology, J CELL COMMUN SIGNAL 12 (2018) 55–68.
[87] V. Cerisano, Y. Aalto, S. Perdichizzi, et al., Molecular mechanisms of CD99-induced caspase-independent cell death and cell-cell adhesion in Ewing's sarcoma cells: actin and zyxin as key intracellular mediators, Oncogene 23 (2004) 5664–5674.
[88] W. Kasinrerk, N. Tokrasinwit, S. Moonsom, et al., CD99 monoclonal antibody induce homotypic adhesion of Jurkat cells through protein tyrosine kinase and protein kinase C-dependent pathway, Immunol. Lett. 71 (2000) 33–41.
[89] A. Bremond, O. Meynet, K. Mahiddine, et al., Regulation of HLA class I surface expression requires CD99 and p230/golgin-245 interaction, Blood 113 (2009) 347–357.
[90] S.S. Yoon, K.I. Jung, Y.L. Choi, et al., Engagement of CD99 triggers the exocytic transport of ganglioside GM1 and the reorganization of actin cytoskeleton, FEBS Lett. 540 (2003) 217–222.
[91] G. Bernard, D. Zoccola, M. Deckert, et al., The E2 molecule (CD99) specifically triggers homotypic aggregation of CD4+ CD8+ thymocytes, J. Immunol. 154 (1995) 26–32.
[92] G. Bernard, D. Zoccola, M. Deckert, et al., The E2 molecule (CD99) specifically triggers homotypic aggregation of CD4+ CD8+ thymocytes, J. Immunol. 154 (1995) 26–32.
[93] G. Bernard, J.P. Breittmayer, M. de Matteis, et al., Apoptosis of immature thymocytes mediated by E2/CD99, J. Immunol. 158 (1997) 2543–2550.
[94] E.M. Dufour, A. Deroche, Y. Bae, et al., CD99 is essential for leukocyte diapedesis in vivo, Cell Commun. Adhes. 15 (2008) 351–363.
[95] R.L. Watson, J. Buck, L.R. Levin, et al., Endothelial CD99 signals through soluble adenylyl cyclase and PKA to regulate leukocyte transendothelial migration, J. Exp. Med. 212 (2015) 1021–1041.
[96] K. Edlund, C. Lindskog, A. Saito, et al., CD99 is a novel prognostic stromal marker in non-small cell lung cancer, Int. J. Cancer 131 (2012) 2264–2273.
[97] M.C. Manara, G. Bernard, P.L. Lollini, et al., CD99 acts as an oncosuppressor in osteosarcoma, Mol. Biol. Cell 17 (2006) 1910–1921.
[98] J.H. Lee, S.H. Kim, L.H. Wang, et al., Clinical significance of CD99 down-regulation in gastric adenocarcinoma, Clin. Cancer Res. 13 (2007) 2584–2591.
[99] H.J. Seol, J.H. Chang, J. Yamamoto, et al., Overexpression of CD99 increases the migration and invasiveness of human malignant glioma cells, Genes Cancer 3 (2012) 535–549.
[100] L.C. Cardoso, R. Soares, T.S. Laurentino, et al., CD99 expression in glioblastoma molecular subtypes and role in migration and invasion, Int. J. Mol. Sci. 20 (2019).
[101] N. Riggi, M.L. Suva, I. Stamenkovic, Ewing's sarcoma, NEW ENGL J MED 384 (2021) 154–164.
[102] A. Ocana, L. Diez-Gonzalez, E. Adrover, et al., Tumor-infiltrating lymphocytes in breast cancer: ready for prime time? J. Clin. Oncol. 33 (2015) 1298–1299.
[103] C. Gu-Trantien, S. Loi, S. Garaud, et al., CD4(+) follicular helper T cell infiltration predicts breast cancer survival, J. Clin. Invest. 123 (2013) 2873–2892.
[104] W.H. Fridman, F. Pages, C. Sautes-Fridman, et al., The immune contexture in human tumours: impact on clinical outcome, Nat. Rev. Cancer 12 (2012) 298–306.
[105] S.M. Mahmoud, E.C. Paish, D.G. Powe, et al., Tumor-infiltrating CD8+ lymphocytes predict clinical outcome in breast cancer, J. Clin. Oncol. 29 (2011) 1949–1955.
[106] A. Rody, U. Holtrich, L. Pusztai, et al., T-cell metagene predicts a favorable prognosis in estrogen receptor-negative and HER2-positive breast cancers, Breast Cancer Res. 11 (2009) R15.
[107] D.G. DeNardo, J.B. Barreto, P. Andreu, et al., CD4(+) T cells regulate pulmonary metastasis of mammary carcinomas by enhancing protumor properties of macrophages, Cancer Cell 16 (2009) 91–102.
[108] I. Shihab, B.A. Khalil, N.M. Elemam, et al., Understanding the role of innate immune cells and identifying genes in breast cancer microenvironment, Cancers 12 (2020).
[109] A. Mantovani, A. Sica, Macrophages, innate immunity and cancer: balance, tolerance, and diversity, Curr. Opin. Immunol. 22 (2010) 231–237.
[110] S.D. Heys, K.N. Stewart, E.J. McKenzie, et al., Characterisation of tumour-infiltrating macrophages: impact on response and survival in patients receiving primary chemotherapy for breast cancer, Breast Cancer Res. Treat. 135 (2012) 539–548.
[111] I. Shabo, O. Stal, H. Olsson, et al., Breast cancer expression of CD163, a macrophage scavenger receptor, is related to early distant recurrence and reduced patient survival, Int. J. Cancer 123 (2008) 780–786.