MDPI

Article

Single-Cell and Bulk RNA Sequencing Reveal Malignant Epithelial Cell Heterogeneity and Prognosis Signatures in Gastric Carcinoma

Zhihong Huang [1], Chao Wu [1], Xinkui Liu [1], Shan Lu [1], Leiming You [2], Fengying Guo [3], Antony Stalin [4], Jingyuan Zhang [1], Fanqin Zhang [1], Zhishan Wu [1], Yingying Tan [1], Xiaotian Fan [1], Jiaqi Huang [1], Yiyan Zhai [1], Rui Shi [1], Meilin Chen [1], Chunfang Wu [5,*], Huiying Li [6] and Jiarui Wu [1,*]

[1] Department of Clinical Pharmacology of Traditional Chinese Medicine, School of Chinese Materia Medica, Beijing University of Chinese Medicine, Beijing 100029, China
[2] School of Life Sciences, Beijing University of Chinese Medicine, Beijing 100029, China
[3] School of Management, Beijing University of Chinese Medicine, Beijing 100029, China
[4] Institute of Fundamental and Frontier Sciences, University of Electronic Science and Technology of China, Chengdu 610054, China
[5] Department of Operations, Beijing Zest Bridge Medical Technology Inc., Beijing 100176, China
[6] School of Biology, Beijing Forestry University, Beijing 100091, China
* Correspondence: wuchunfang@zeqiao.com (C.W.); wujiarui@bucm.edu.cn (J.W.)

Abstract: Gastric carcinoma (GC) heterogeneity represents a major barrier to accurate diagnosis and treatment. Here, we established a comprehensive single-cell transcriptional atlas to identify the cellular heterogeneity in malignant epithelial cells of GC using single-cell RNA sequencing (scRNA-seq). A total of 49,994 cells from nine patients with paired primary tumor and normal tissues were analyzed by multiple strategies. This study focused on the malignant epithelial cells, which were divided into three subtypes, including pit mucous cells, chief cells, and gastric and intestinal cells. The trajectory analysis results suggest that the differentiation of the three subtypes could be from the pit mucous cells to the chief cells and then to the gastric and intestinal cells. Lauren's histopathology of GC might originate from various subtypes of malignant epithelial cells. The functional enrichment analysis results show that the three subtypes focused on different biological processes (BP) and pathways related to tumor development. In addition, we generated and validated the prognostic signatures for predicting the OS in GC patients by combining the scRNA-seq and bulk RNA sequencing (bulk RNA-seq) datasets. Overall, our study provides a resource for understanding the heterogeneity of GC that will contribute to accurate diagnosis and prognosis.

Keywords: single-cell RNA sequencing; bulk RNA sequencing; gastric carcinoma; malignant epithelial cells; prognostic signatures

1. Introduction

Gastric carcinoma (GC) has a high incidence and mortality globally, and the most common histological type is stomach adenocarcinoma (STAD) [1,2]. GC is characterized by particular subtypes and clinical manifestations, which are the great challenges for clinical diagnosis and treatment [3,4]. Although advances have been made in identifying different molecular subtypes of GC through The Cancer Genome Atlas (TCGA) and Asian Cancer Research Group (ACRG), the prognosis improvement of patients is not significant [5,6]. GC is mainly the pathological change of gastric mucosa glandular epithelial cells, and the exact pathogeny is unclear. This study shows that *H. pylori* infection, a high salt diet, and a low intake of fruits and vegetables are risk factors for GC [7].

High-throughput sequencing technologies, such as bulk RNA sequencing (bulk RNA-seq) and single-cell RNA sequencing (scRNA-seq), have greatly accelerated the pace of

research into the molecular characteristics of human tumors. So far, a large number of omic-studies have described the GC's genetic and epigenetic heterogeneity [8–10]. Although bulk RNA-seq can provide data for mass gene expression profiling, it cannot distinguish the relationship between different cell lineage and cellular interaction [11]. The advent of scRNA-seq compensates for the shortcomings of bulk RNA-seq. It provides a method that can characterize the transcriptional status of thousands of cells simultaneously [12,13]. The integration of scRNA-seq and bulk RNA-seq is of great scientific importance for the study of tumor development and heterogeneity.

Here, we constructed a transcriptomic landscape of GC malignant epithelial cells by scRNA-seq, encompassing various subtypes of GC pathology subtypes. We also combined the results with bulk RNA-seq to find the prognostic signatures that can predict the overall survival (OS) of GC patients. The landscape of GC malignant epithelial cells was constructed to interpret the heterogeneity of GC and shed light on the clinical diagnosis and prognosis of GC in our study (Figure 1).
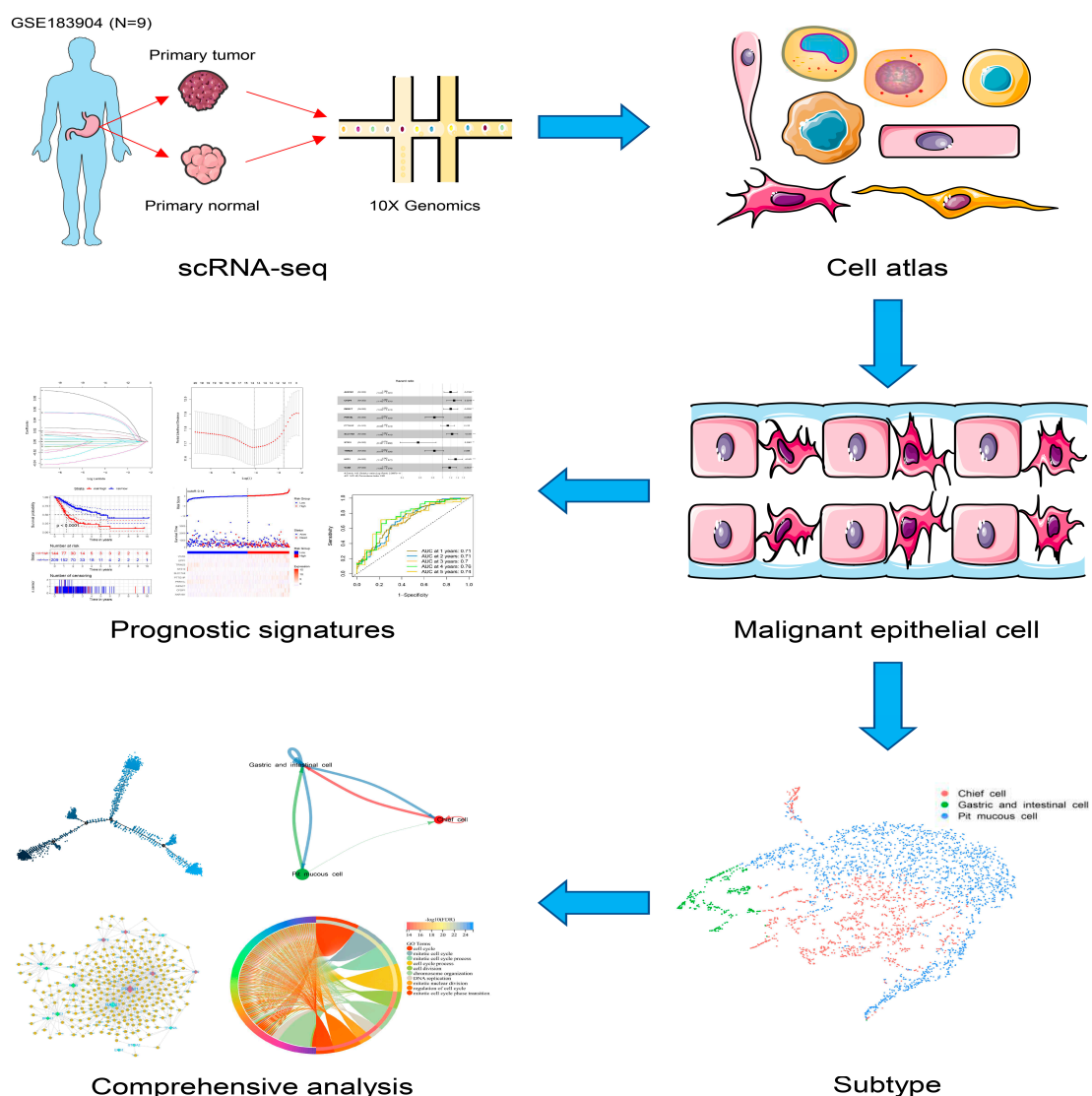

Figure 1. The workflow of the present study.

2. Methods

2.1. Data Source and Preprocessing

The GC scRNA-seq dataset GSE183904 was downloaded as required from the Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/gds, accessed on

24 March 2022). The nine patients with paired primary tumor and normal tissues were selected for analysis in the present study. Using the R package Seurat, all samples were analyzed for genes/features shared by more than three cells and for cells containing 500–6000 features. Cells with mitochondrial RNA content higher than 20% were excluded [14].

Public clinical data and gene expression information were obtained from the TCGA database (https://portal.gdc.cancer.gov/, accessed on 24 March 2022), the Genotype-Tissue Expression (GTEx) database (https://www.gtexportal.org/, accessed on 24 March 2022) and the GEO database [5]. The bulk RNA-seq dataset included TCGA-STAD, GTEx-stomach, GSE15459, GSE29272, GSE57303, GSE62254, and GSE66229. Differentially expressed genes (DEGs) were calculated using the R package limma between tumor and normal samples. Only genes with an expression fold change (FC) >2.0 or <−2.0 and a false discovery rate (FDR) < 0.05 were taken into subsequent analysis.

2.2. scRNA-Seq Normalization, Clustering and DEGs Analyzing

SCtransform normalization was carried out separately for each sample, and FindIntegrationAnchors was run to find anchor genes and integrate the data [15]. The integrated data were scaled, principal component analysis (PCA) was performed, and the data were visualized using the t-Distributed stochastic neighbor embedding (tSNE) method. A shared nearest neighbor (SNN) modularity optimization-based clustering algorithm with a resolution of 1.0 was set to identify cell clusters. The FindAllMarkers module was used to identify DEGs, and genes expressed in more than 25% of cells in each cluster were selected for cluster demarcation. The cell type markers were employed to construct cell atlas.

2.3. Recognition of Malignant and Non-Malignant Epithelial Cells

To identify chromosomal copy number variations, InferCNV (https://github.com/broadinstitute/inferCNV, accessed on 28 March 2022) was used to explore scRNA-seq datasets from primary tumors compared with normal scRNA-seq datasets. Each epithelial cell in the present study was clustered, and a CNV score was calculated using the k-means clustering algorithm. The classes with a CNV score higher than the average was defined as a group of malignant epithelial cells, whereas the others were considered as a group of non-malignant epithelial cells. The malignant epithelial cells were also clustered and annotated as different subtypes based on cell type markers. DEGs between malignant and non-malignant epithelial cells and among different subtypes of malignant epithelial cells were identified using the FindAllMarkers module.

2.4. Functional Enrichment Analysis

Gene ontology (GO) enrichment analysis, Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis, and gene set enrichment analysis (GSEA) were performed to uncover the potential biological function of the different cells through the Database for Annotation, Visualization, and Integrated Discovery (DAVID, https://david.ncifcrf.gov/, accessed on 28 March 2022) database and the Molecular Signatures Database (MSigDB, https://www.gsea-msigdb.org/, accessed on 28 March 2022). The enrichment results were visualized using the R package ggplot2.

2.5. Trajectory Analysis

Pseudotime and cell trajectory analyses were used to investigate the differentiation trajectories and related genes in different cell clusters and to explain the molecular mechanism in the progression of GC. The R package Monocle was used to cluster cells and draw learning maps to compare trajectories among different subtypes of malignant epithelial cells [16,17].

2.6. Cell–Cell Communication Analysis

To explore underlying interactions between various subtypes in malignant epithelial cells, an analysis of cell–cell communication was conducted using the R package Cellchat, which simulates intercellular communication by binding ligands, receptors and their co-factors [18]. The enriched receptor-ligand interactions between the two cell types were deduced based on the receptor expression of one cell type and the corresponding ligand expression of another cell type.

2.7. Construction of Gene Regulatory Network

SCENIC is a tool for simultaneously reconstructing gene regulatory networks and identifying stable cell states from scRNA-seq data [19]. The gene regulatory network was derived based on co-expression and DNA motif analysis, and then the network activity in each cell was analyzed to identify the cell state. In order to determine the search space around the transcription start site, two gene-motif rankings (10 kb around the transcription start site or 500 bp upstream and 100 bp downstream of the transcription start site) were used as reference. Gene regulation of various malignant epithelial cells was constructed using the R package GENIE3, RcisTarget and AUCell. Then, the gene regulatory network was visualized using Cytoscape software.

2.8. Generation and Validation of the Prognostic Signatures

The bulk RNA-seq datasets GSE15459, GSE57303 and GSE66229 from the same platform were integrated for subsequent analysis. The DEGs from the scRNA-seq and the bulk RNA-seq datasets were screened related to the OS of TCGA-STAD patients using a univariate Cox analysis. Moreover, LASSO analysis was employed to choose the reliable predictors for multivariate Cox regression. Then, the prognostic signatures were used to create a polygenic risk score and divide the TCGA-STAD samples into low- or high-risk groups. The time-dependent receiver operating characteristic (ROC) curve was used to assess the predictive ability of the prognostic signatures. The GSE62254 dataset was used to validate the prognostic value of the prognostic signatures. All results were analyzed and visualized using the R package survival, survminer, rms and timeROC.

2.9. Statistical Analysis

R version 4.0.3 was used for statistical analysis (https://www.r-project.org/, accessed on 28 March 2022). The Wilcoxon-rank sum test was used to evaluate associations with continuous variables. Student's *t*-test was used to analyze significant differences among distinct groups. Kaplan–Meier curves with log-rank statistics were used to compare OS. $p < 0.05$ was considered to indicate statistical significance.

3. Results

3.1. A Single-Cell Transcriptome Atlas of GC

Cell populations and corresponding molecular characteristics in GC were investigated in the GSE183904 dataset, which includes nine patients with paired primary tumor and normal tissues (Supplementary Table S1). After the removal of low-quality cells based on the screening criteria in Methods 2.1, 49,994 cells were persisted for biological analysis (Figure 2A). The cells were divided into 33 clusters after normalization, integration and PCA. These clusters were noted as eight known cell types according to marker genes recorded in the literature. The details were as follows: (1) the epithelial cells highly expressing *CDH1*; (2) the endothelial cells highly expressing *PLVAP*; (3) the fibroblast cells highly expressing *FN1*; (4) the T cells highly expressing *CD8A*; (5) the B cells highly expressing *TNFRSF17*; (6) the macrophage cells highly expressing *CD163*; (7) the NK cells highly expressing *KLRD1*; (8) the mast cells highly expressing *KIT*. The proportion of each cell type varied in different samples (Figure 2B–E, Supplementary Table S2).
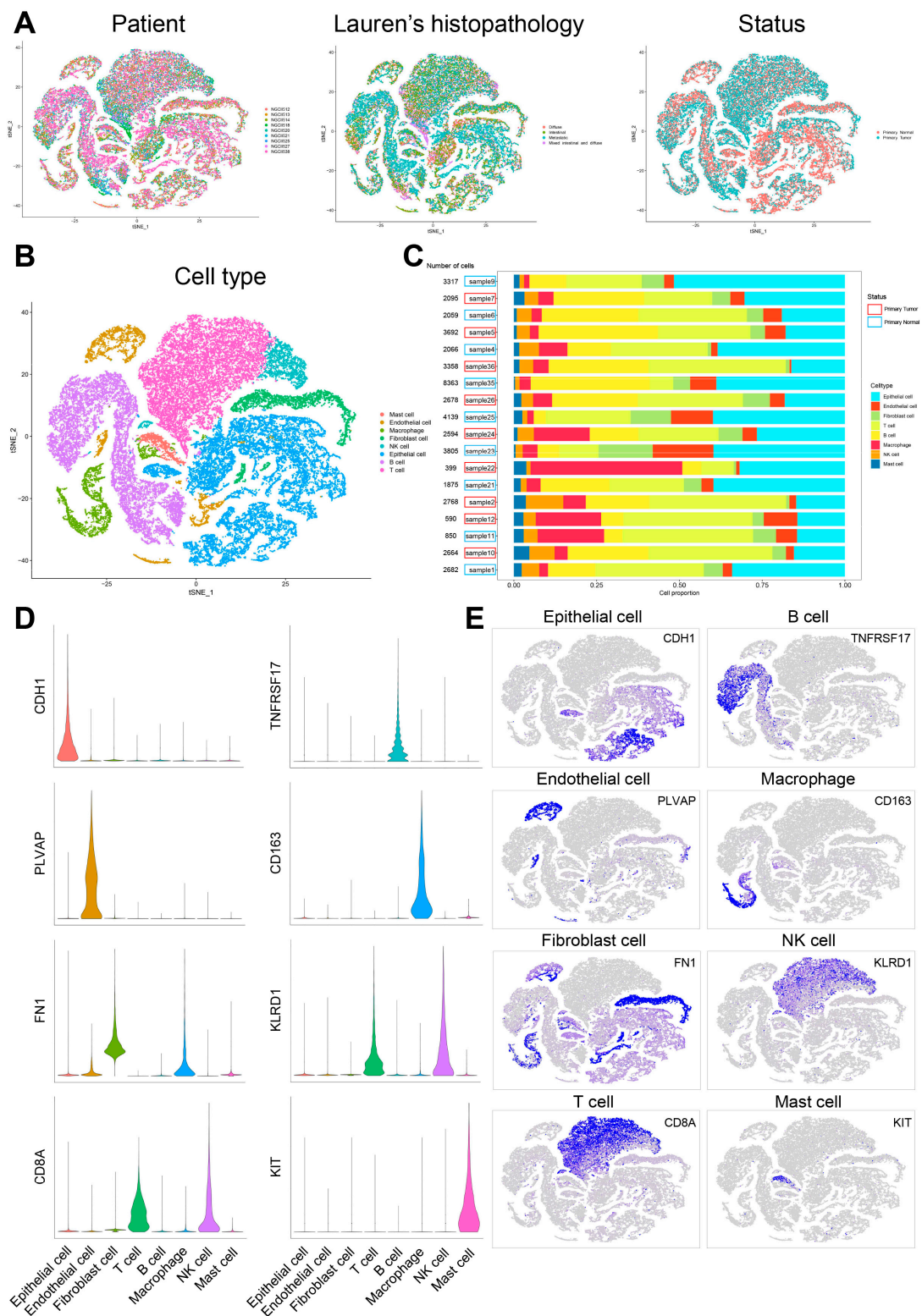
Figure 2. Cellular atlas of GC primary tumor and normal tissues. (**A**) tSNE plot for the 49,994 high-quality cells showing sample origin, Lauren's histopathology and status. (**B**) tSNE plot showing cell types for the 49,994 cells. (**C**) The proportion of each cell type in 18 samples. (**D**) Violin plots showing the expression distribution of marker genes in eight cell types. (**E**) tSNE plots showing the expression levels of marker genes for eight cell types.

3.2. Classification of Malignant and Non-Malignant Epithelial Cells

First, the cell population annotated as epithelial cells in the dataset was accurately determined and clustered (Figure 3A). Next, InferCNV was employed to distinguish malignant and non-malignant epithelial cells according to the CNV score. The kemans classes with a CNV score higher than the average (0.00124) were defined as the malignant epithelial cells group, and the others were defined as the non-malignant epithelial cells group (Figure 3B–E, Figure 4A). The top five DEGs of malignant and non-malignant epithelial cells were as follows: (1) malignant epithelial cells highly expressing *CAPN8*, *CLDN4*, *CYP3A5*, *PHGR1*, and *PLEC*; (2) non-malignant epithelial cells highly expressing *IGFBP2*, *LIPF*, *PGA3*, *PGA4*, and *PGA5* (Figure 4B,C). Compared with non-malignant epithelial cells, malignant epithelial cells were enriched for signaling pathways such as TNF-α/NF-κB, KRAS and IL6/STAT3/JAK. Moreover, there were also cells enriched in the epithelial mesenchymal transition genes associated with cancer development and progression (Figure 5A, Supplementary Table S3).
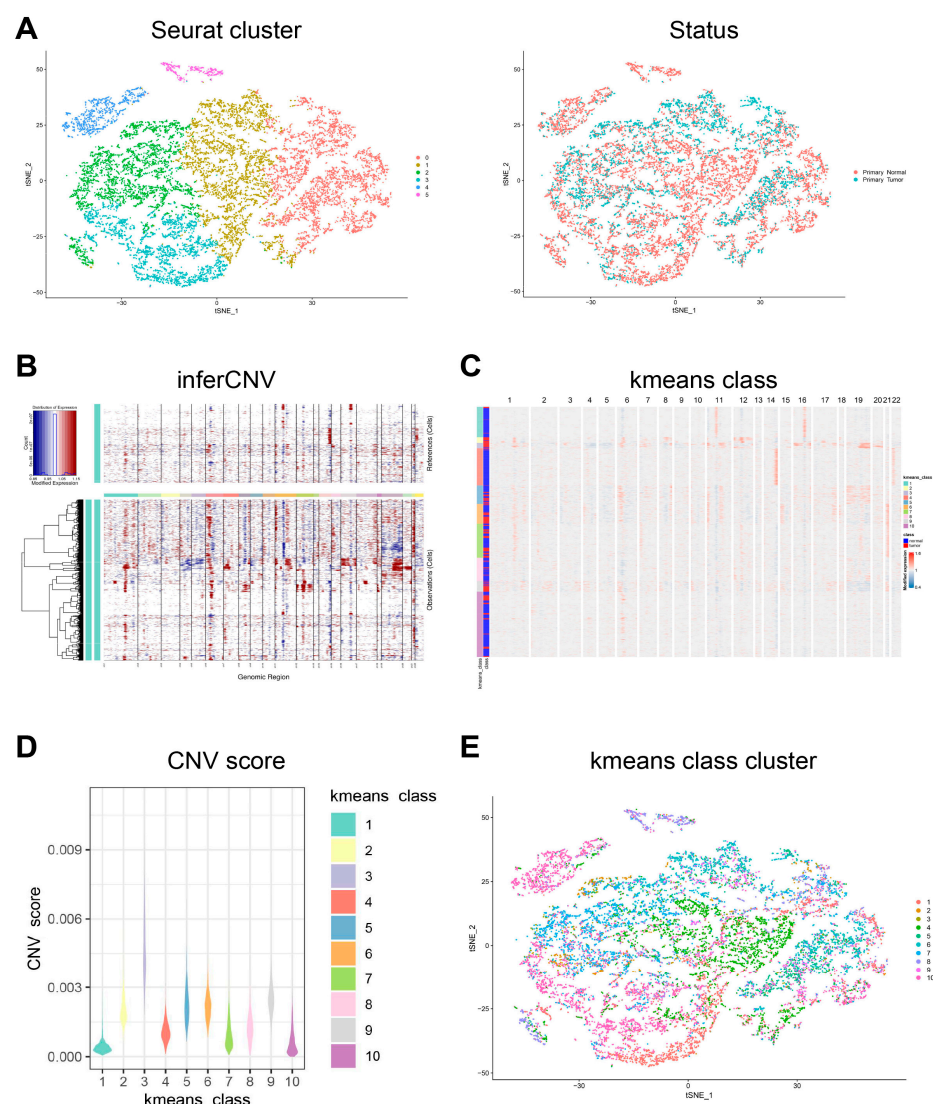

Figure 3. Copy-number variation analysis of epithelial cells. (**A**) tSNE plots showing the Seurat cluster and status of epithelial cells. (**B**) Hierarchical heatmap showing large-scale CNVs in the epithelial cells of a primary tumor and in normal tissues. (**C**) Hierarchical heatmap showing the results of k-means class clustering. (**D**) Violin plots showing CNV scores in the different k-means classes. (**E**) tSNE plot showing the epithelial cells in the different k-means class clusters.
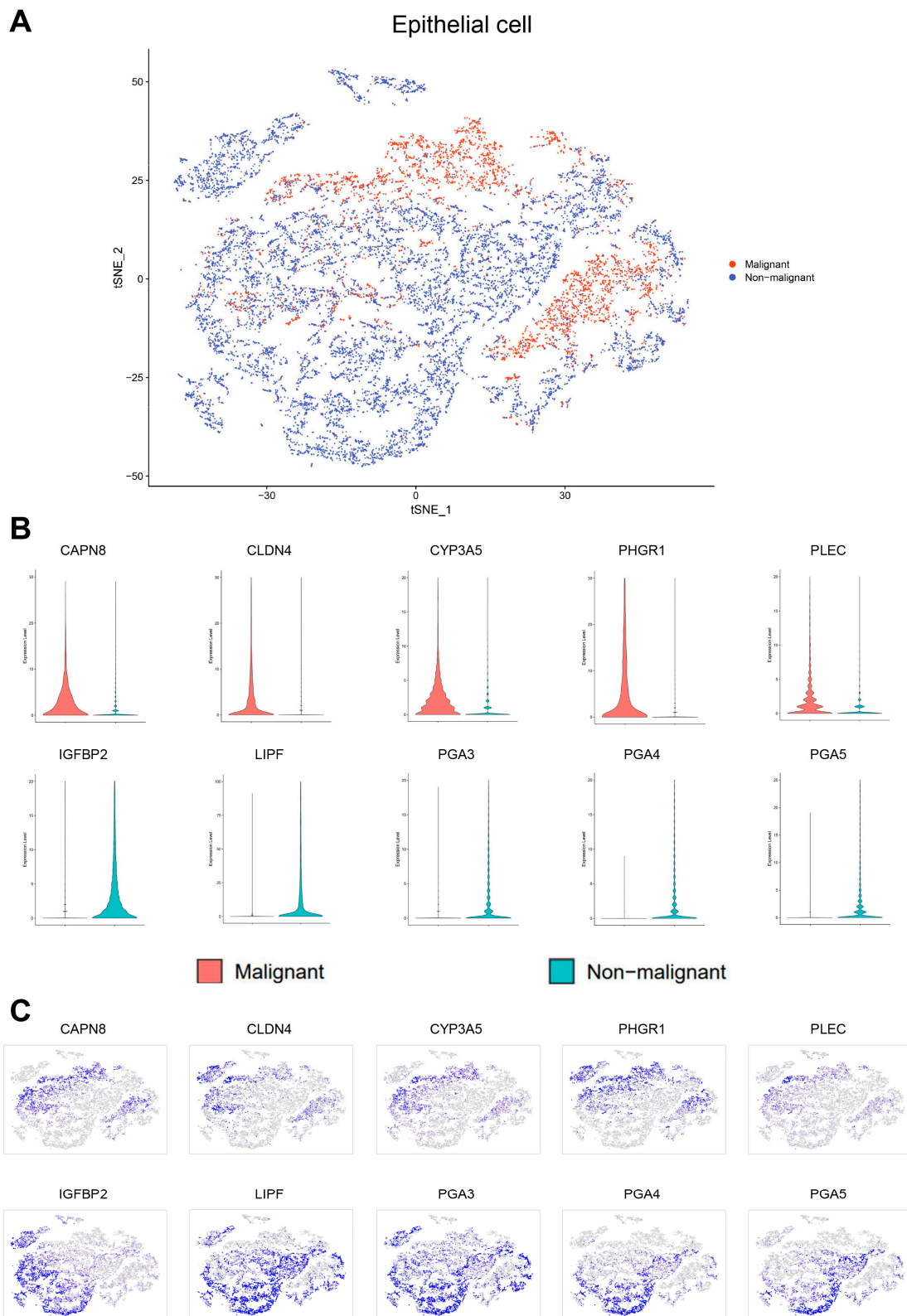
Figure 4. Classification of the epithelial cells as malignant or non-malignant. (**A**) tSNE plot showing the classification of malignant and non-malignant cells. (**B**) Violin plots showing the expression of ten representative genes with differential expression between malignant and non-malignant cells. (**C**) tSNE plots showing the expression of ten representative genes with differential expression.
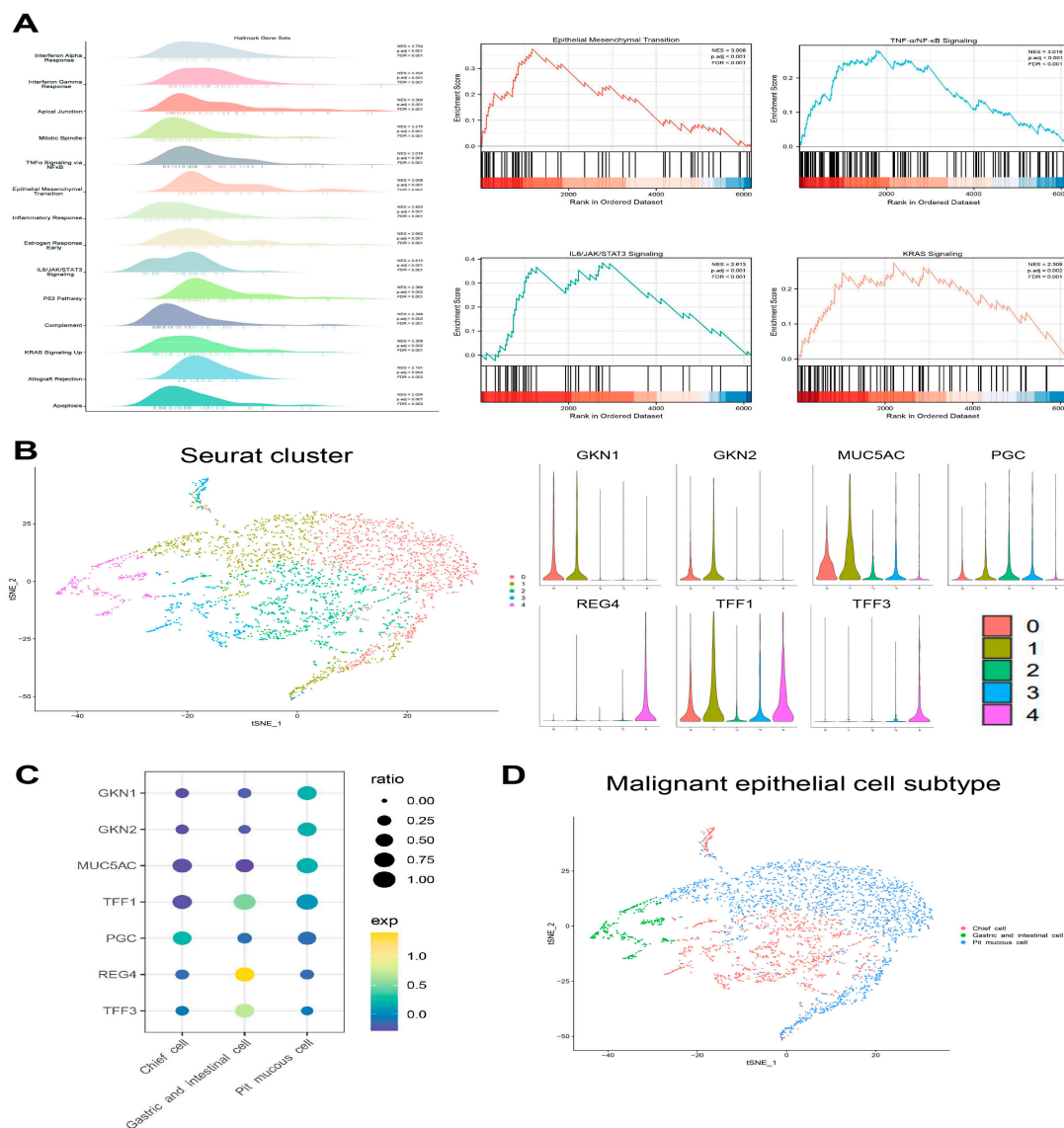
Figure 5. Cellular atlas of the malignant epithelial cell subtypes. (**A**) The GSEA results show the enrichment gene sets in malignant epithelial cells, and four of them were GC-associated gene sets. (**B**) tSNE plot showing the malignant epithelial cells seurat cluster, and Violin plots showing the expression distribution of subtype marker genes in five clusters. (**C**) Bubble plot showing the expression levels of marker genes in three subtypes. (**D**) tSNE plot showing the subtypes for malignant epithelial cells.

3.3. Transcriptional Heterogeneity of Malignant Epithelial Cells

The malignant epithelial cells were accurately determined and clustered. Subsequently, the subtypes have been annotated based on marker genes in the literature. There were three subtypes as follows: (1) the pit mucous cells with high expression of *GKN1*, *GKN2*, *MUC5AC*, and *TFF1*; (2) the chief cells with high expression of *PGC*; (3) the gastric and intestinal cells with high expression of *TFF1*, *TFF3* and *REG4* (Figure 5B–D, Supplementary Table S2). The proportion of each subtype varied considerably among the different Lauren's classifications, and the top 20 DEGs of the three malignant epithelial cell subtypes were shown (Figure 6A,B). The GSEA results show that the pit mucous cells were mainly enriched in the pathways and gene sets correlated with cell proliferation, metastasis and invasion, cell differentiation and inflammatory response. The chief cells were mainly enriched in the pathways and gene sets related to cell proliferation, cell migration and invasion, cell cycle, cell differentiation, inflammatory response and cell survival. The gastric and intestinal cells

were enriched mainly in pathways and genes that were relevant to apoptosis, cell growth, cell survival and cell metabolism (Figure 6C, Supplementary Table S4).
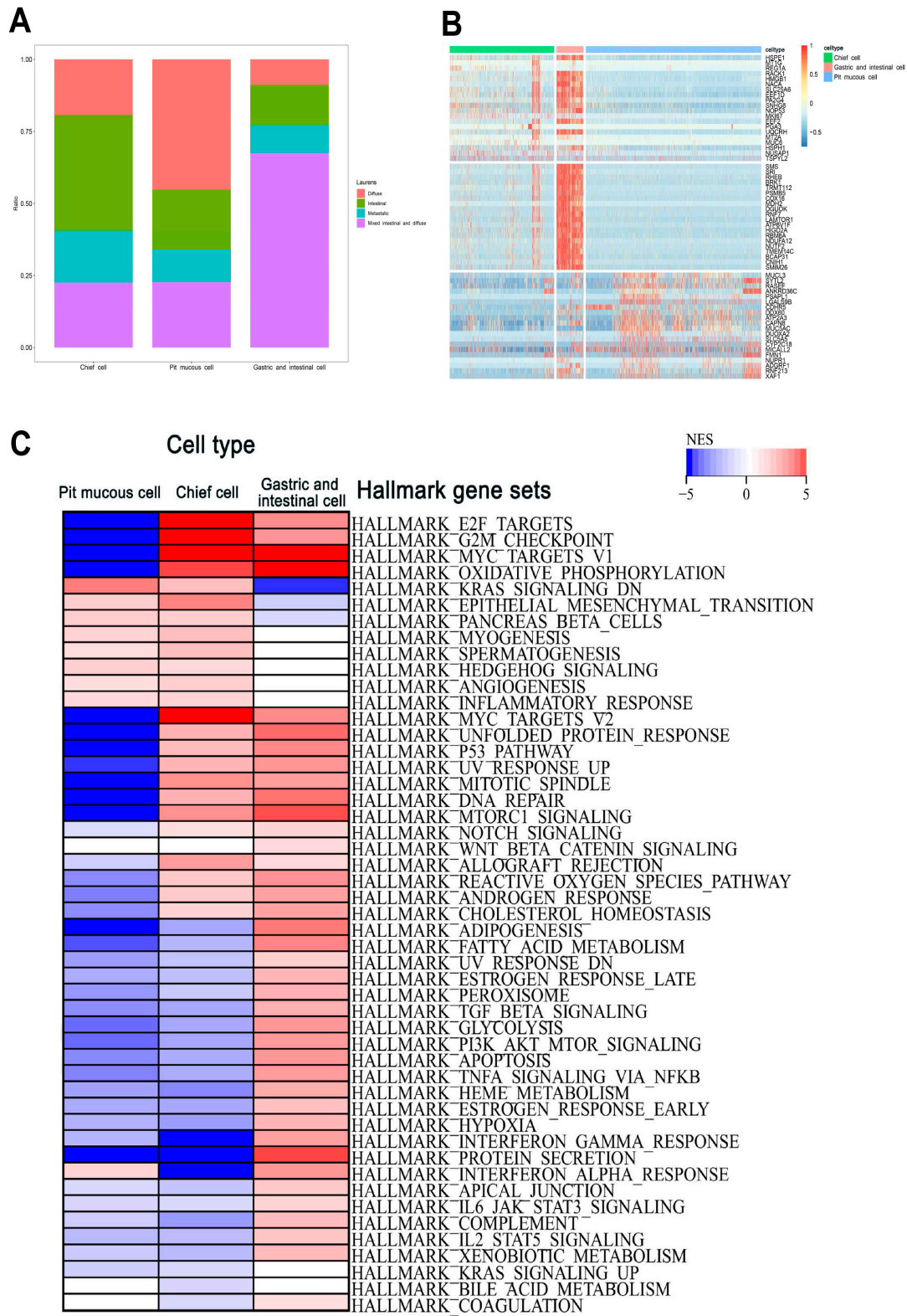

Figure 6. The heterogeneity of the malignant epithelial cell subtypes in GC. (**A**) The proportion of each subtype in Lauren's histopathology. (**B**) Heatmap showing the top 20 marker genes in three subtypes. (**C**) The GSEA results showing the enrichment gene sets in three subtypes.

3.4. Trajectory of Malignant Epithelial Cells

Trajectory analysis was performed to identify the differentiation heterogeneity of malignant epithelial cells within and between tumors. It was found that the underlying cell differentiation trajectories of malignant epithelial cells comprised nine states, and the contents of three subtypes differed in various states. The pit mucous cells differentiated into the chief cells and the gastric and intestinal cells (Figure 7A). In the whole potential differentiation trajectories of malignant epithelial cells, the activities of genes related to the biological processes (BP) and pathways of cell proliferation, cell migration and invasion, cell cycle and apoptosis were reduced. In contrast, the activities of genes associated with translation, enzyme activity and regulation of other cell biological processes and pathways were increased (Figure 7B, Supplementary Table S5).
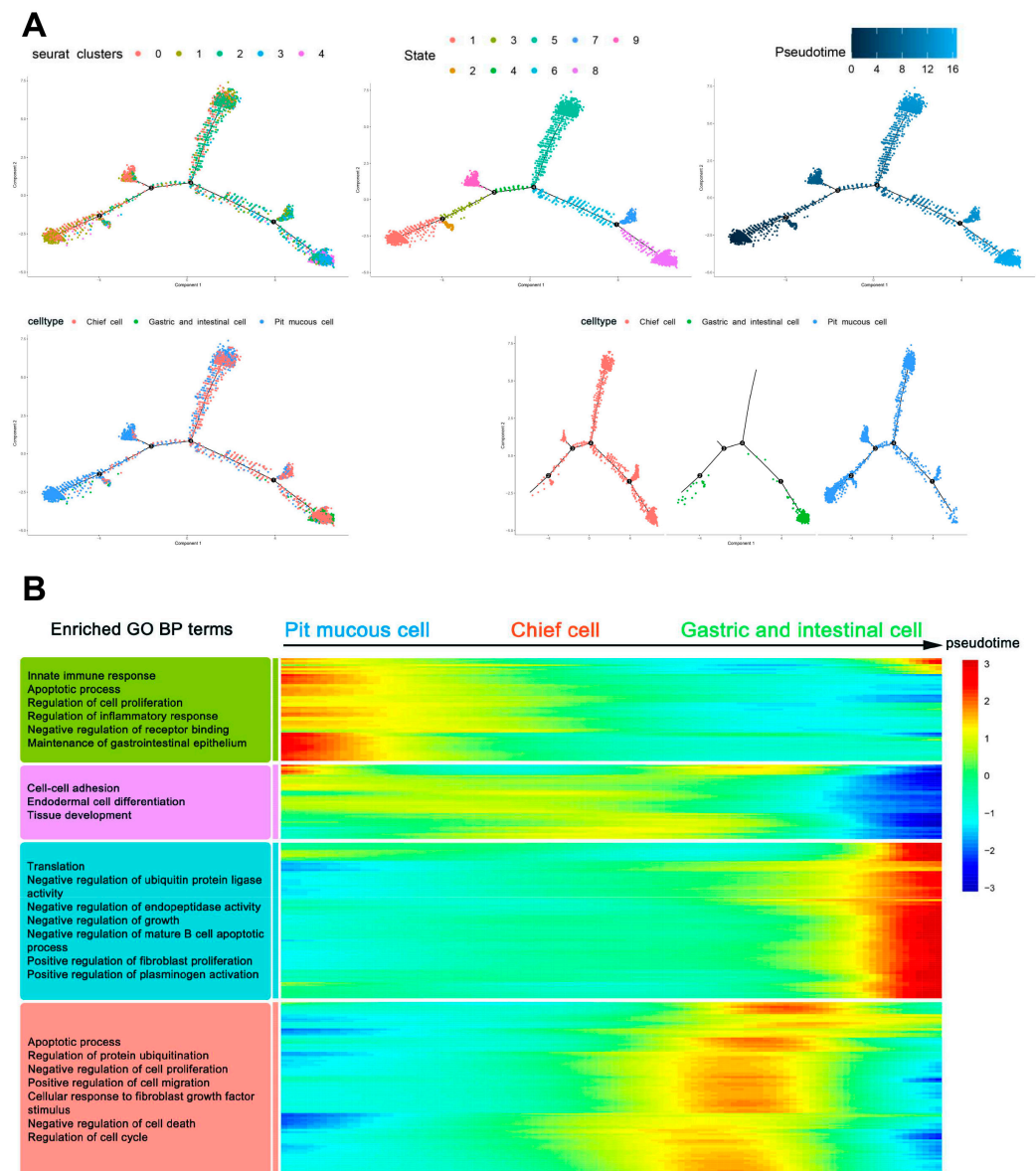

Figure 7. Trajectory analysis of malignant epithelial cells. (**A**) Trajectory plots showing the differentiation of the three subtypes. (**B**) Heatmap showing the scaled expression of dynamic genes along the pseudotime. Rows of the heatmap represent genes that show dynamic changes along the pseudotime, and these genes were clustered into four groups according to their expression pattern along the pseudotime. The annotated GO:BP terms in each cluster were provided.

3.5. Intercellular Communication in Malignant Epithelial Cells

Cellchat was used to identify ligand–receptor pairs and molecular interactions among the three subtypes of malignant epithelial cells. The results show that multiple ligand–receptor-mediated cell interactions existed mainly in the MK signaling pathway (MDK-SDC4, MDK-SDC1, MDK-NCL, MDK-LRP1, MDK-ITGA6+ITGB1 and MDK-ITGA4+ITGB1) and in the MIF signaling pathway (MIF-CD74+CXCR4, MIF-CD74+CXCR2 and MIF-CD74+CD44) (Figure 8A–C, Supplementary Table S6).
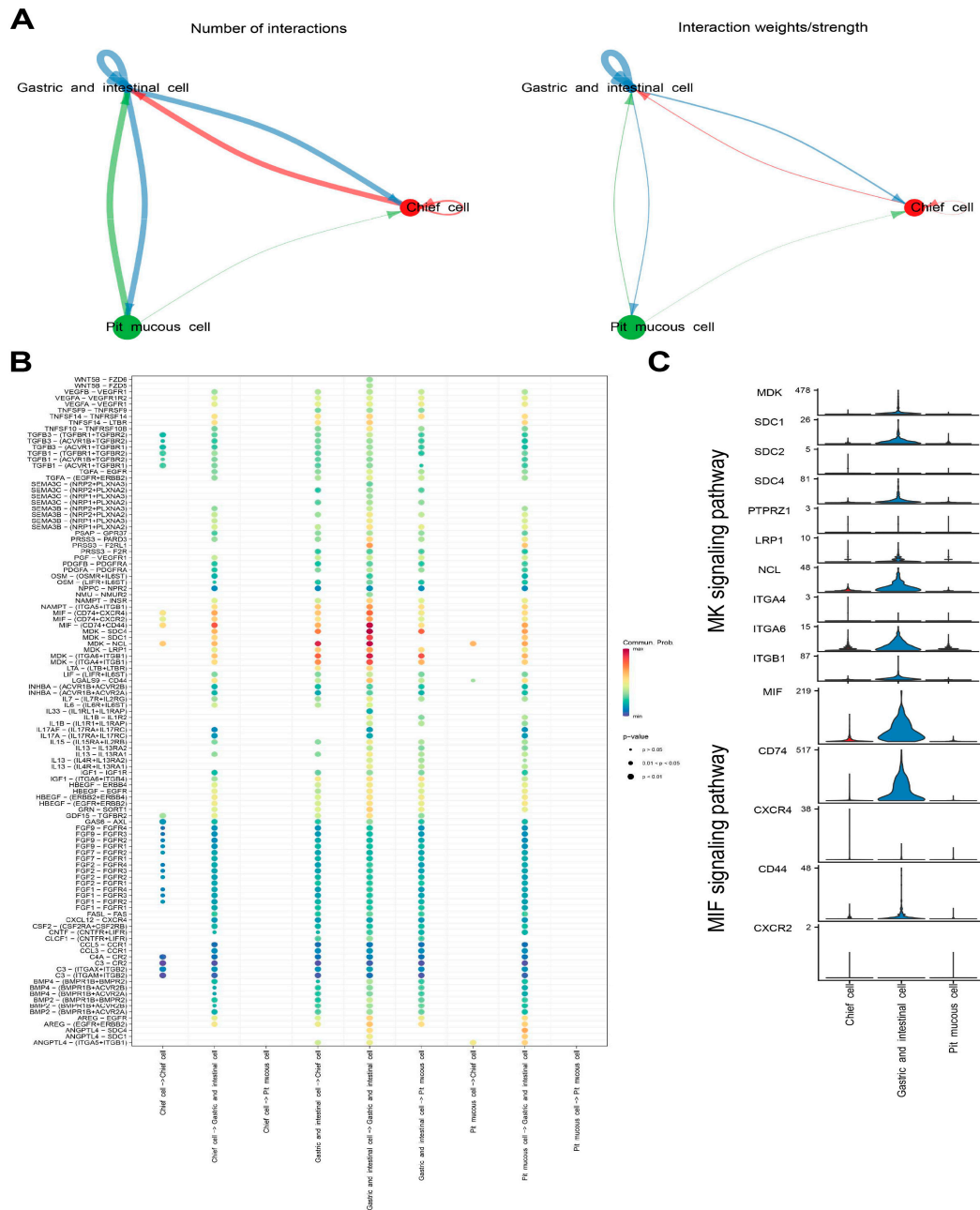

Figure 8. Cell–cell communication analysis of malignant epithelial cells. (**A**) Capacity for intercellular communication among the three malignant epithelial cell subtypes. The size of the circle is proportional to the number of cells. The cells that emit arrows express ligands and the cells that arrows point to express receptors. The more ligand–receptor pairs, the thicker the line. (**B**) Bubble plot showing the intercellular interaction relationship mediated by ligand–receptor pairs. (**C**) Violin plots showing the expression of representative genes in the MK and MIF signaling pathway.

3.6. Construction of Gene Regulatory Network in Malignant Epithelial Cells

The enrichment results of transcription factor (TF) in three subtypes of malignant epithelial cells were displayed, and the top 5 TFs were selected for further study (Figure 9A–C, Supplementary Table S7). The top 5 TF-gene regulatory networks were visualized using Cytoscape software (Figure 10A). The enrichment analysis results show that the common genes regulated by the top 5 TFs in three subtypes of malignant epithelial cells were mainly related to cell proliferation, signal transduction, and metabolism. The special genes regulated by the top 5 TFs in pit mucous cells were enriched in pathways and biological processes associated with cell proliferation, signal transduction and metabolism. The special genes regulated by the top 5 TFs in the chief cells were involved in the cell cycle and cell survival. The special genes regulated by the top 5 TFs in the gastric and intestinal cells involved the cell migration, invasion, and translation process (Figure 10B, Supplementary Table S8).
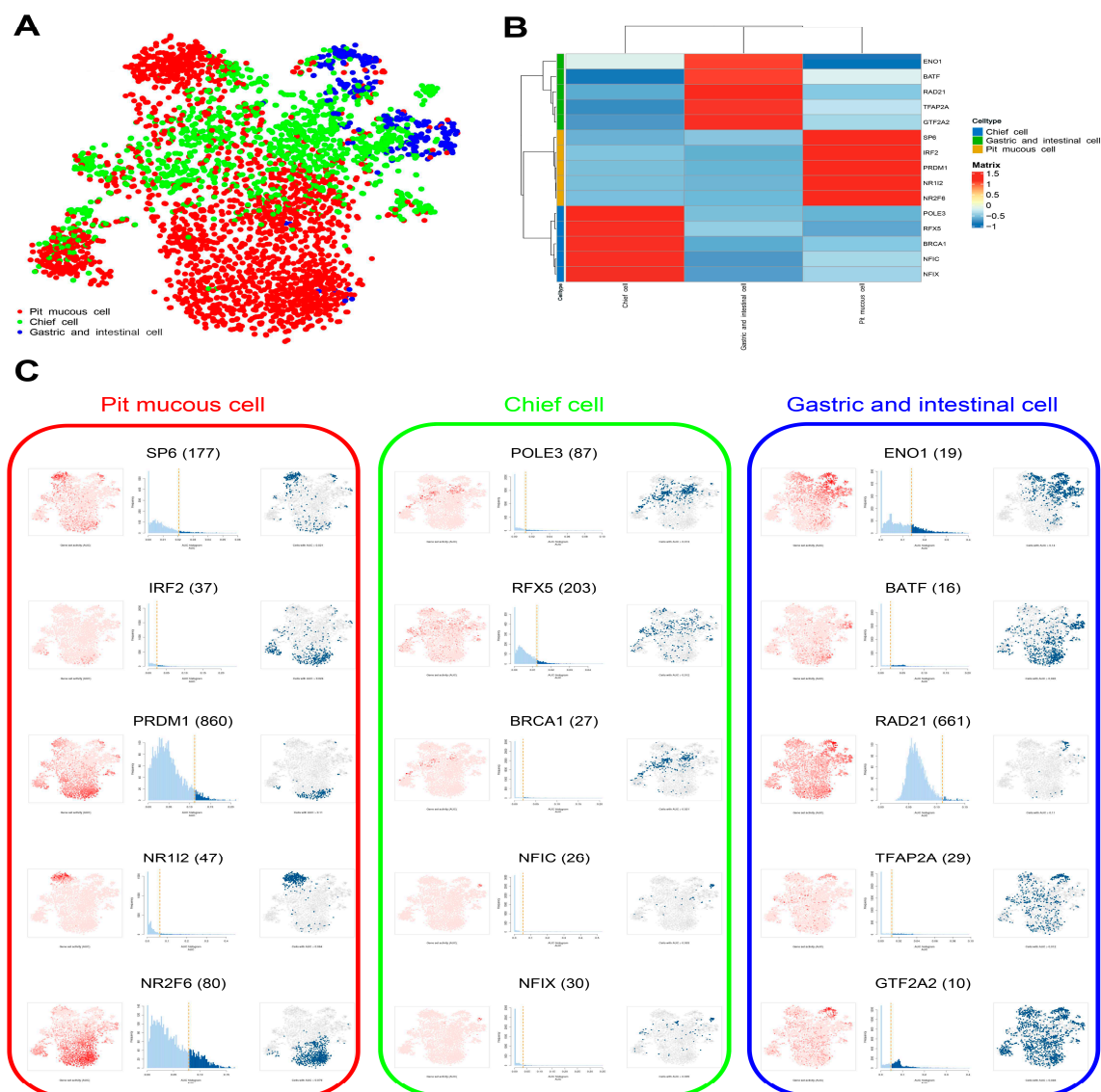

Figure 9. The SCENIC analysis predicted the TF. (**A**) tSNE plot showing the TF in three subtypes of malignant epithelial cells. (**B**) Heatmap showing the top 5 TF in three subtypes of malignant epithelial cells. (**C**) tSNE plots and histograms show the top 5 TF activities in three subtypes of malignant epithelial cells.
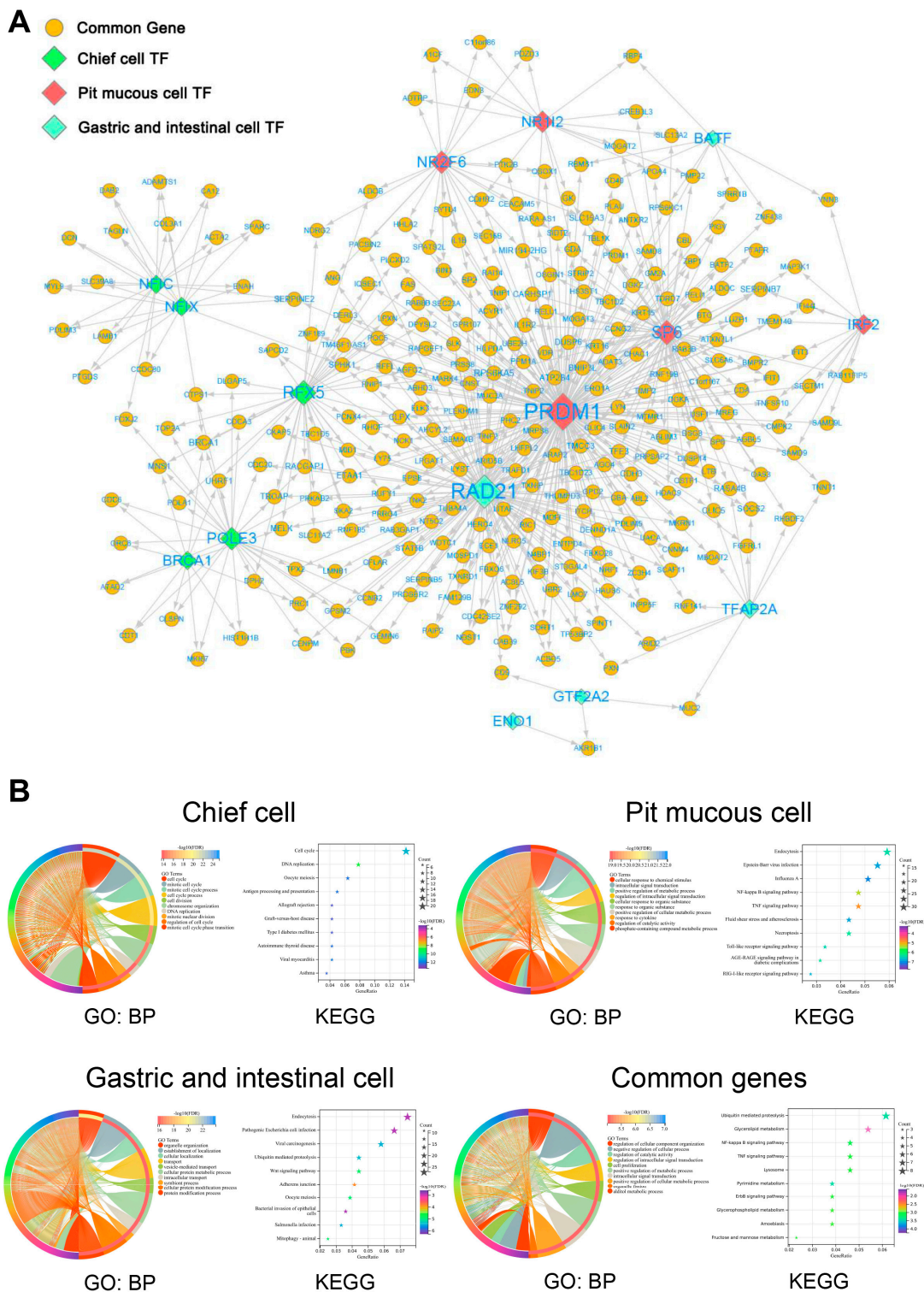
Figure 10. The TF-gene regulatory network of malignant epithelial cells. (**A**) Network showing the top 5 TF and their regulated genes in three subtypes of malignant epithelial cells. (**B**) Chord and bubble plots show the functional enrichment results of the top 5 TF regulated genes in three subtypes of malignant epithelial cells.

3.7. Bulk RNA-Seq Data Analysis

Differential gene expression analysis of all bulk RNA-seq datasets included in the present study was used to identify the DEGs between GC tumors and normal tissues. Then, the intersection of the results was taken to extract the common up-regulated and

down-regulated genes for functional enrichment analysis (Figure 11A,B). The down-regulated genes were mainly enriched in the biological processes and pathways related to metabolism and normal gastric function. In addition, the up-regulated genes were involved in the behavior of malignant tumor cells, such as cell migration and invasion (Figure 11C, Supplementary Table S9).
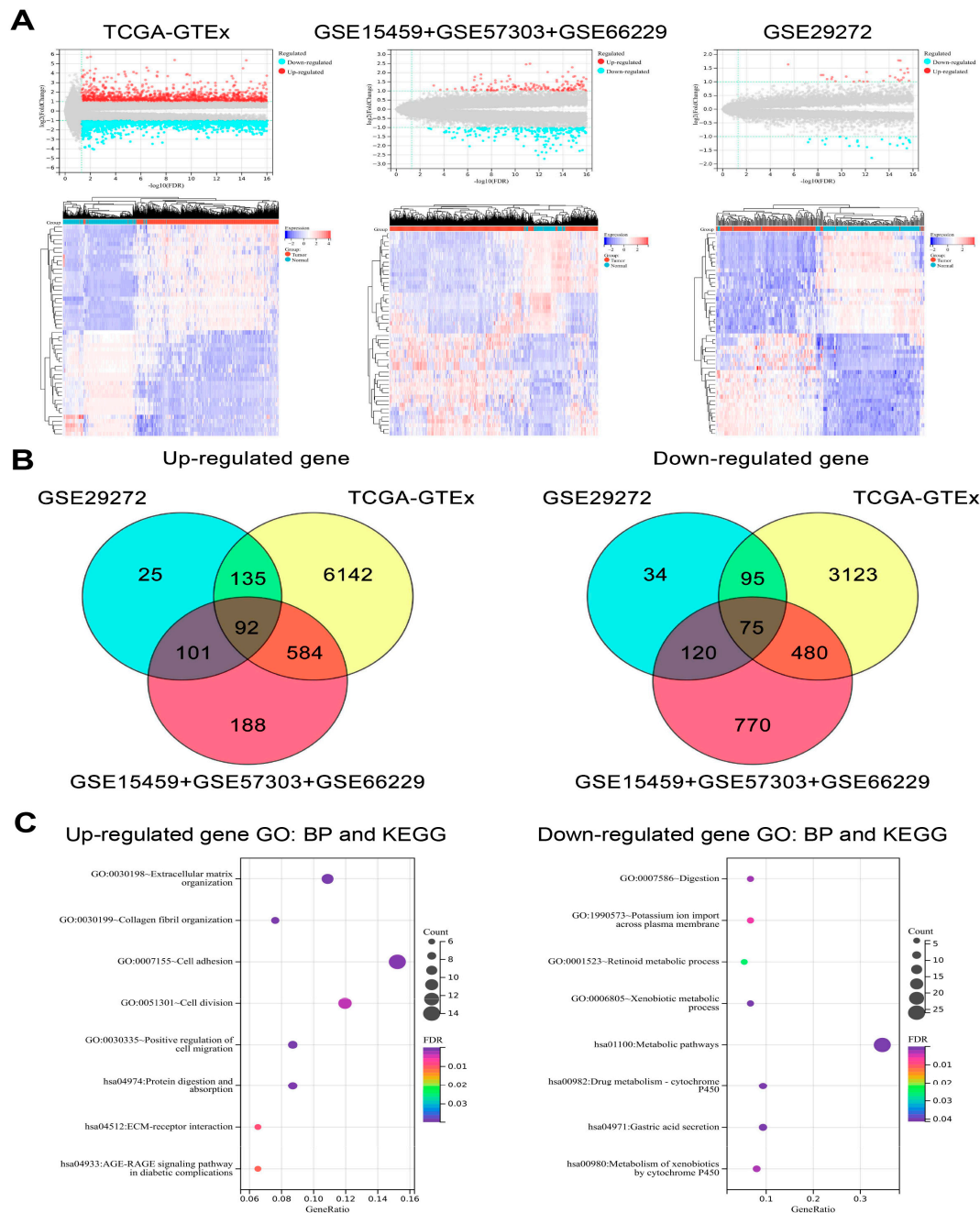

Figure 11. The analysis of the bulk RNA-seq datasets. (**A**) Volcano plots and heatmaps show the DEGs between the tumor and normal tissues in bulk RNA-seq datasets. (**B**) Venn plots showing the common up-regulated and down-regulated genes in different bulk RNA-seq datasets. (**C**) Bubble plots show the functional enrichment results of the up-regulated and down-regulated genes.

3.8. Generation and Validation of the Prognostic Signatures in Malignant Epithelial Cells

The up-regulated genes from bulk RNA-seq and DEGs in malignant epithelial cells from scRNA-seq were integrated to construct a polygenic risk score based on prognostic signatures for predicting prognosis in GC. Ten genes (*AKR1B1, CFDP1, IMPACT, PRR15L,*

PTTG1IP, *SLC17A9*, *STX10*, *TRIM25*, *UPP1* and *VCAN*) were selected to calculate the polygenic risk score through univariable, multivariable Cox and LASSO regression. GC patients were divided into low-risk or high-risk groups, and the prognosis of patients in the high-risk group was poor in the training cohort (TCGA-STAD) and test cohort (GSE62254). The ROC curve showed that the ten prognostic signatures performed well in predicting the OS of GC patients, with an area under the ROC curve > 0.7 in the training cohort and around 0.6 in the test cohort (Figure 12A,B). Prognostic signatures in three subtypes of malignant epithelial cells were also generated and validated with prognostic values for GC patients (Supplementary Figures S1–S3).
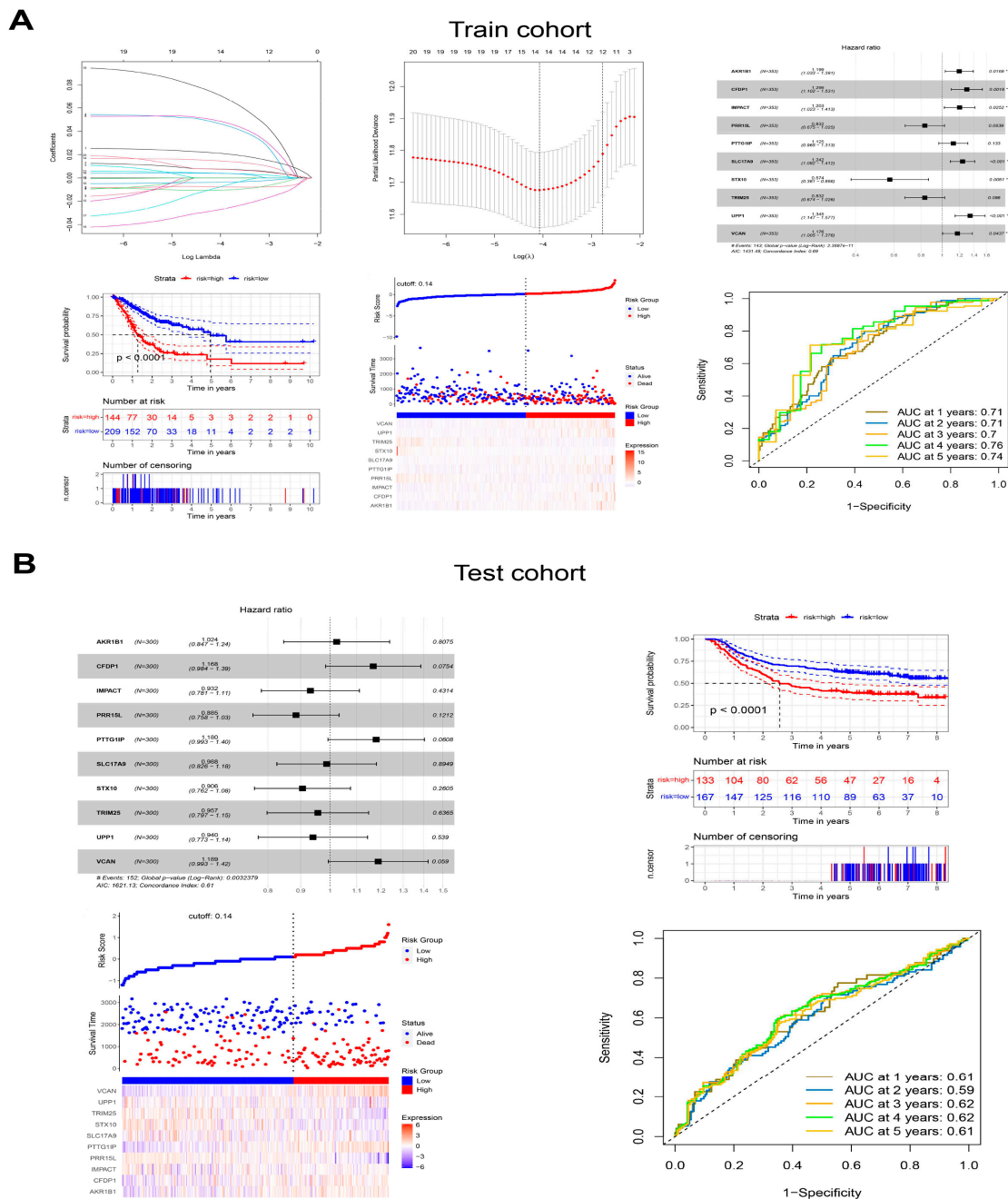

Figure 12. Generation and validation of the prognostic signatures in malignant epithelial cells. (**A**) The generation of the prognostic signatures in the training cohort. (**B**) The validation of the prognostic signatures in the test cohort. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

4. Discussion

GC is a heterogeneous disease influenced by many factors, presenting a number of difficulties for clinical diagnosis and individualized treatment. Burgeoning scRNA-seq has been widely applied to explore tumor heterogeneity, including the analysis of tumor development, drug resistance programs, cell–cell communication and immune infiltration patterns [20–22]. In the present study, we used the technology to construct a comprehensive landscape of malignant epithelial cells from GC at single-cell resolution.

Focusing on malignant epithelial cells, we identified three subtypes (the pit mucous cells, the chief cells, and the gastric and intestinal cells) with different transcriptome characteristics. The pit mucous cells and the chief cells were mainly derived from diffuse and intestinal GC patients, respectively. The gastric and intestinal cells had the largest proportion in GC patients with mixed intestinal and diffuse conditions. This suggests that the different Lauren's histopathology of GC might originate from various subtypes of malignant epithelial cells. We can use the marker genes of different subtypes of malignant epithelial cells to help determine the Lauren's classification. The results of the functional enrichment analysis show that they had different activities in the signaling pathways and biological processes associated with tumor development. The pit mucous cells and the chief cells were mainly enriched in cell proliferation, differentiation, migration and invasion, while the gastric and intestinal cells focused on cell survival and metabolism. The trajectory analysis results suggest that the differentiation of malignant epithelial cells could occur from pit mucous cells to chief cells and then to gastric and intestinal cells. These findings suggest that the tumor cells of diffuse-type GC were poorly differentiated and the cases with intestinal histology had a different degree of differentiation [23]. Zhang et al. found that the GC with a low differentiation degree was more aggressive, which was consistent with our findings [24].

Moreover, the cell–cell interactions indicated that the different subtypes of malignant epithelial cells communicate with each other to a high degree. The cytokines mainly mediated the intercellular communication among the three subtypes of malignant epithelial cells in the MK and MIF signaling pathway, which are involved in the inflammatory response, tumor growth, and cell migration, which have been poorly studied in GC [25–27]. The gene regulatory network in malignant epithelial cells showed that TF and its activities varied in different subtypes of malignant epithelial cells. Thus, the three subtypes of malignant epithelial cells were enriched in the different biological processes and pathways associated with malignant biological behavior. The above results suggest that intratumoral and intertumoral heterogeneity is a fundamental property of GC malignant epithelial cells.

In this study, we integrated the scRNA-seq and bulk RNA-seq datasets to find the prognostic signatures for predicting OS in GC patients. The prognostic signatures appeared to be essential for GC because they performed well in the train and test cohorts and could be used for GC clinical screening. In addition, we also generated and validated the prognostic signatures for different subtypes of malignant epithelial cells. They had prognostic value for the patients with different malignant epithelial cells in combination with Lauren's histopathology in clinical use. We could predict the progression of patients by detecting the expression of these prognostic signatures.

The results of this study may be used in clinical diagnosis and treatment in the future. The cell atlas and marker genes of GC malignant epithelial cells can be applied to identify Lauren's histopathology of GC patients by combining diagnostic methods such as endoscopy. Developing personalized therapy for malignant epithelial cell heterogeneity to improve clinical efficacy and the prognostic signatures can be used to test clinical application value and design the targeted drugs. At the same time, this study also had some limitations. First, the number of patients included in the study was small, and it was impossible to determine whether there was the same result in large-scale samples. Second, all results of this study were not validated in experiments and clinical practice to ensure they can be extended.

5. Conclusions

In summary, using scRNA-seq, we constructed a single-cell transcriptome atlas of GC malignant epithelial cells, encompassing several GC Lauren's histopathological subtypes. Various methods analyzed the subtypes of malignant epithelial cells, and the heterogeneity in terms of functional enrichment, cell differentiation trajectory, and intercellular communication was preliminarily clarified. Furthermore, we combined the scRNA-seq with the bulk RNA-seq datasets to find the prognostic signatures for predicting the OS of GC patients. All the results of the present study need to be further validated in experimental and clinical practice.

Supplementary Materials: The following are available online at https://www.mdpi.com/article/10.3390/cells11162550/s1, Figure S1: Generation and validation of the prognostic signatures in the pit mucous cells. * $p < 0.05$, ** $p < 0.01$. Figure S2: Generation and validation of the prognostic signatures in the chief cells. (A) The gen-eration of the prognostic signatures in the train cohort. (B) The validation of the prognostic sig-natures in the test cohort. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Figure S3: Generation and validation of the prognostic signatures in the gastric and intestinal cells. (A) The generation of the prognostic signatures in the train cohort. (B) The validation of the prognostic signatures in the test cohort. * $p < 0.05$. Table S2: List of cell-type markers from literature used for annotating cell-states. Table S3: The GSEA results of malignant vs non-malignant epithelial cells. Table S4.1: The GSEA results of the pit mucous cells. Table S4.2: The GSEA results of the chief cells. Table S4.3: The GSEA results of the gastric and intestinal cells. Table S5.1: The highly variable genes in different clusters. Table S5.2: The enrichment results of GO biological processes and KEGG pathways in green cluster. Table S5.3: The enrichment results of GO biological processes and KEGG pathways in purple cluster. Table S5.4: The enrichment results of GO biological processes and KEGG pathways in blue cluster. Table S5.5: The enrichment results of GO biological processes and KEGG pathways in red cluster. Table S6: Ligand-receptor pairs among the three subtypes of malignant epithelial cells. Table S7.1: Motif enrichment in the gene regulatory network of malignant epithelial cells. Table S7.2: Regulons and targets information in the gene regulatory network of malignant epithelial cells. Table S8.1: The GO and KEGG enrichment results of the common genes regulated by the top5 TF in three subtypes of malignant epithelial cells. Table S8.2: The GO and KEGG enrichment results of the special genes regulated by the top5 TF in the pit mucous cells. Table S8.3: The GO and KEGG enrichment results of the special genes regulated by the top5 TF in the chief cells. Table S8.4: The GO and KEGG enrichment results of the special genes regulated by the top5 TF in the gastric and intestinal cells. Table S9.1: The GO biological process and KEGG pathway enrichment results of up-regulated genes. Table S9.2: The GO biological process and KEGG pathway enrichment results of down-regulated genes.

Author Contributions: Conceptualization, data curation, investigation, writing—review and editing, Z.H., J.W., C.W. (Chunfang Wu), H.L. and A.S.; methodology, X.L., L.Y. and F.G.; formal analysis, J.Z., F.Z., Z.W., Y.T. and X.F.; validation, C.W. (Chao Wu) and S.L.; visualization, J.H., Y.Z., R.S. and M.C.; supervision, funding acquisition, J.W. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary Material.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

BP, biological process; DEGs, differentially expressed genes; GC, gastric carcinoma; GO, gene ontology; GSEA, gene set enrichment analysis; KEGG, Kyoto Encyclopedia of Genes and Genomes; OS, overall survival; PCA, principal component analysis; ROC, receiver operating characteristic; SNN, shared nearest neighbor; STAD, stomach adenocarcinoma; TF, transcription factor; tSNE, t-distributed stochastic neighbor embedding.

References

1. Sung, H.; Ferlay, J.; Siegel, R.L.; Laversanne, M.; Soerjomataram, I.; Jemal, A.; Bray, F. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J. Clin.* **2021**, *71*, 209–249. [CrossRef]
2. Etemadi, A.; Safiri, S.; Sepanlou, S.G.; Ikuta, K.; Bisignano, C.; Shakeri, R.; Amani, M.; Fitzmaurice, C.; Nixon, M.; Abbasi, N.; et al. The global, regional, and national burden of stomach cancer in 195 countries, 1990–2017: A systematic analysis for the Global Burden of Disease study 2017. *Lancet Gastroenterol. Hepatol.* **2020**, *5*, 42–54. [CrossRef]
3. Hudler, P. Challenges of deciphering gastric cancer heterogeneity. *World J. Gastroenterol.* **2015**, *21*, 10510–10527. [CrossRef]
4. Tan, P.; Yeoh, K.G. Genetics and Molecular Pathogenesis of Gastric Adenocarcinoma. *Gastroenterology* **2015**, *149*, 1153–1162. [CrossRef]
5. Comprehensive molecular characterization of gastric adenocarcinoma. *Nature* **2014**, *513*, 202–209. [CrossRef]
6. Cristescu, R.; Lee, J.; Nebozhyn, M.; Kim, K.M.; Ting, J.C.; Wong, S.S.; Liu, J.; Yue, Y.G.; Wang, J.; Yu, K.; et al. Molecular analysis of gastric cancer identifies subtypes associated with distinct clinical outcomes. *Nat. Med.* **2015**, *21*, 449–456. [CrossRef]
7. Smyth, E.C.; Nilsson, M.; Grabsch, H.I.; van Grieken, N.C.; Lordick, F. Gastric cancer. *Lancet* **2020**, *396*, 635–648. [CrossRef]
8. Wang, R.; Dang, M.; Harada, K.; Han, G.; Wang, F.; Pool, P.M.; Zhao, M.; Tatlonghari, G.; Zhang, S.; Hao, D.; et al. Single-cell dissection of intratumoral heterogeneity and lineage diversity in metastatic gastric adenocarcinoma. *Nat. Med.* **2021**, *27*, 141–151. [CrossRef]
9. Lei, Z.; Tan, I.B.; Das, K.; Deng, N.; Zouridis, H.; Pattison, S.; Chua, C.; Feng, Z.; Guan, Y.K.; Ooi, C.H.; et al. Identification of molecular subtypes of gastric cancer with different responses to PI3-kinase inhibitors and 5-fluorouracil. *Gastroenterology* **2013**, *145*, 554–565. [CrossRef]
10. Kumar, V.; Ramnarayanan, K.; Sundar, R.; Padmanabhan, N.; Srivastava, S.; Koiwa, M.; Yasuda, T.; Koh, V.; Huang, K.K.; Tay, S.T.; et al. Single-Cell Atlas of Lineage States, Tumor Microenvironment, and Subtype-Specific Expression Programs in Gastric Cancer. *Cancer Discov.* **2022**, *12*, 670–691. [CrossRef]
11. Newman, A.M.; Steen, C.B.; Liu, C.L.; Gentles, A.J.; Chaudhuri, A.A.; Scherer, F.; Khodadoust, M.S.; Esfahani, M.S.; Luca, B.A.; Steiner, D.; et al. Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nat. Biotechnol.* **2019**, *37*, 773–782. [CrossRef] [PubMed]
12. Baslan, T.; Hicks, J. Unravelling biology and shifting paradigms in cancer with single-cell sequencing. *Nat. Rev. Cancer* **2017**, *17*, 557–569. [CrossRef] [PubMed]
13. Wen, L.; Tang, F. Single-cell sequencing in stem cell biology. *Genome Biol.* **2016**, *17*, 71. [CrossRef] [PubMed]
14. Butler, A.; Hoffman, P.; Smibert, P.; Papalexi, E.; Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **2018**, *36*, 411–420. [CrossRef]
15. Stuart, T.; Butler, A.; Hoffman, P.; Hafemeister, C.; Papalexi, E.; Mauck, W.R.; Hao, Y.; Stoeckius, M.; Smibert, P.; Satija, R. Comprehensive Integration of Single-Cell Data. *Cell* **2019**, *177*, 1888–1902. [CrossRef]
16. Trapnell, C.; Cacchiarelli, D.; Grimsby, J.; Pokharel, P.; Li, S.; Morse, M.; Lennon, N.J.; Livak, K.J.; Mikkelsen, T.S.; Rinn, J.L. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* **2014**, *32*, 381–386. [CrossRef]
17. Qiu, X.; Mao, Q.; Tang, Y.; Wang, L.; Chawla, R.; Pliner, H.A.; Trapnell, C. Reversed graph embedding resolves complex single-cell trajectories. *Nat. Methods* **2017**, *14*, 979–982. [CrossRef]
18. Jin, S.; Guerrero-Juarez, C.F.; Zhang, L.; Chang, I.; Ramos, R.; Kuan, C.H.; Myung, P.; Plikus, M.V.; Nie, Q. Inference and analysis of cell-cell communication using CellChat. *Nat. Commun.* **2021**, *12*, 1088. [CrossRef]
19. Aibar, S.; Gonzalez-Blas, C.B.; Moerman, T.; Huynh-Thu, V.A.; Imrichova, H.; Hulselmans, G.; Rambow, F.; Marine, J.C.; Geurts, P.; Aerts, J.; et al. SCENIC: Single-cell regulatory network inference and clustering. *Nat. Methods* **2017**, *14*, 1083–1086. [CrossRef]
20. Patel, A.P.; Tirosh, I.; Trombetta, J.J.; Shalek, A.K.; Gillespie, S.M.; Wakimoto, H.; Cahill, D.P.; Nahed, B.V.; Curry, W.T.; Martuza, R.L.; et al. Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science* **2014**, *344*, 1396–1401. [CrossRef]
21. Ren, X.; Kang, B.; Zhang, Z. Understanding tumor ecosystems by single-cell sequencing: Promises and limitations. *Genome Biol.* **2018**, *19*, 211. [CrossRef] [PubMed]
22. Puram, S.V.; Tirosh, I.; Parikh, A.S.; Patel, A.P.; Yizhak, K.; Gillespie, S.; Rodman, C.; Luo, C.L.; Mroz, E.A.; Emerick, K.S.; et al. Single-Cell Transcriptomic Analysis of Primary and Metastatic Tumor Ecosystems in Head and Neck Cancer. *Cell* **2017**, *171*, 1611–1624. [CrossRef]
23. Baniak, N.; Senger, J.L.; Ahmed, S.; Kanthan, S.C.; Kanthan, R. Gastric biomarkers: A global review. *World J. Surg. Oncol.* **2016**, *14*, 212. [CrossRef]

24. Zhang, M.; Hu, S.; Min, M.; Ni, Y.; Lu, Z.; Sun, X.; Wu, J.; Liu, B.; Ying, X.; Liu, Y. Dissecting transcriptional heterogeneity in primary gastric adenocarcinoma by single cell RNA sequencing. *Gut* **2021**, *70*, 464–475. [CrossRef]

25. Sorrelle, N.; Dominguez, A.; Brekken, R.A. From top to bottom: Midkine and pleiotrophin as emerging players in immune regulation. *J. Leukoc. Biol.* **2017**, *102*, 277–286. [CrossRef] [PubMed]

26. Klasen, C.; Ohl, K.; Sternkopf, M.; Shachar, I.; Schmitz, C.; Heussen, N.; Hobeika, E.; Levit-Zerdoun, E.; Tenbrock, K.; Reth, M.; et al. MIF promotes B cell chemotaxis through the receptors CXCR4 and CD74 and ZAP-70 signaling. *J. Immunol.* **2014**, *192*, 5273–5284. [CrossRef]

27. Hughes, C.E.; Nibbs, R. A guide to chemokines and their receptors. *FEBS J.* **2018**, *285*, 2944–2971. [CrossRef]