Dynamic changes of the immune microenvironment in the development of gastric cancer caused by inflammation

Siyuan Gan,[1,8] Changfu Li,[2,8] Rui Hou,[3,4] Geng Tian,[3,4] Yuan Zhao,[5] Dan Ren,[6] Wenjing Zhou,[7] Fei Zhao,[1] Kebo Lv,[1] and Jialiang Yang[3,4]

[1]School of Mathematical Sciences, Ocean University of China, Qingdao, China; [2]Department of Digestive Internal Medicine, Daqing Longnan Hospital, The Fifth Affiliated Hospital of Qiqihar Medical College, Daqing 163000, China; [3]Geneis Beijing Co., Ltd, Beijing 100102, China; [4]Qingdao Geneis Institute of Big Data Mining and Precision Medicine, Qingdao 266000, China; [5]School of Electrical and Information Engineering, Anhui University of Technology, Ma'anshan, China; [6]Department of Pathology, Daqing Longnan Hospital, The Fifth Affiliated Hospital of Qiqihar Medical College, Daqing 163000, China; [7]Department of Oncology, Hiser Medical Center of Qingdao, No. 4, Renmin Road, Shibei District, Qingdao, China

Precancerous lesions typically precede gastric cancer (GC), but the molecular mechanisms underlying the transition from these lesions to GC remain unclear. Therefore, it is urgent to understand this transition from precancerous lesions to GC, which is crucial for the early diagnosis and treatment of GC. In this study, we merged multiple single-cell RNA sequencing datasets to investigate the molecular changes in distinct cell types associated with the progression of GC. First, we observed an increasing abundance of immune cells and a decrease in non-immune cells from non-atrophic gastritis to GC. Five immune cell types were significantly enriched in GC compared to precancerous lesions. Moreover, we found that the interleukin (IL)-17 signaling pathway and Th17 cell differentiation were significantly up-regulated in immune cell subsets during GC progression. Some genes in these processes were predominantly expressed at the GC stage, highlighting their potential as diagnostic markers. Furthermore, we validated our findings using bulk RNA sequencing data from The Cancer Genome Atlas and confirmed consistent immune cell changes during GC progression. Our study provides insights into the immune infiltration and signaling pathways involved in the development of GC, contributing to the development of early diagnosis and targeted treatment strategies for this malignancy.

INTRODUCTION

Nowadays, cancer is one of the serious threats to public health. The number of new cancer cases and deaths is increasing year by year.[1] There are many hypotheses about the cause of cancer, including smoking, radiation, viruses, cancer-causing chemicals (carcinogens), and chronic inflammation.[2,3] As early as 1863, Balkwill et al. discovered that tumor tissue contains a large amount of inflammatory cell infiltration and put forward the hypothesis that tumors are originated from chronic inflammation.[4] Inflammation is thought to be the body's local protective response to injury or infection.[5] In recent years, studies have found that chronic or excessive inflammation

can lead to cancer. More than 20% of cancers are caused by inflammation, such as gastritis for gastric cancer (GC),[6] pneumonia for lung cancer,[7] and colitis for colorectal cancer.[8] It is crucial to analyze the molecular mechanisms under this transformation to systematically study the transformation process from chronic inflammation into cancer and discover the important factors involved in this malignant transformation. Breakthroughs in this field will contribute to cancer prevention as well as early diagnosis and treatment.

GC is one of the most common cancers in clinical practice, with a high incidence and mortality worldwide, especially in Asia.[9] According to the World Health Organization Global Cancer Statistics Report 2020, it is estimated that the number of new patients with GC worldwide exceeds 1 million every year, and the main age of onset is between 60 and 84 years old.[10,11] The link between intestinal metaplasia (IM) and GC was revealed in 1938.[12] In 1955, Morson reported that GC occurred in IM.[13] Intestinal GC is preceded by precancerous lesions, including non-atrophic gastritis (NAG), chronic atrophic gastritis (CAG), and IM.[14] Patients with these precancerous lesions have an increased risk of developing GC.[14]

Identifying the genes and pathways that drive cancer formation has been the focus of large-scale genomics research. However, most research has focused on the overall characteristics of advanced gastric tumors, largely ignoring related studies on precancerous lesions.[8] Our understanding of the phenotypic changes and molecular drivers of the transition from precancerous lesions to GC remains inadequate.[8]
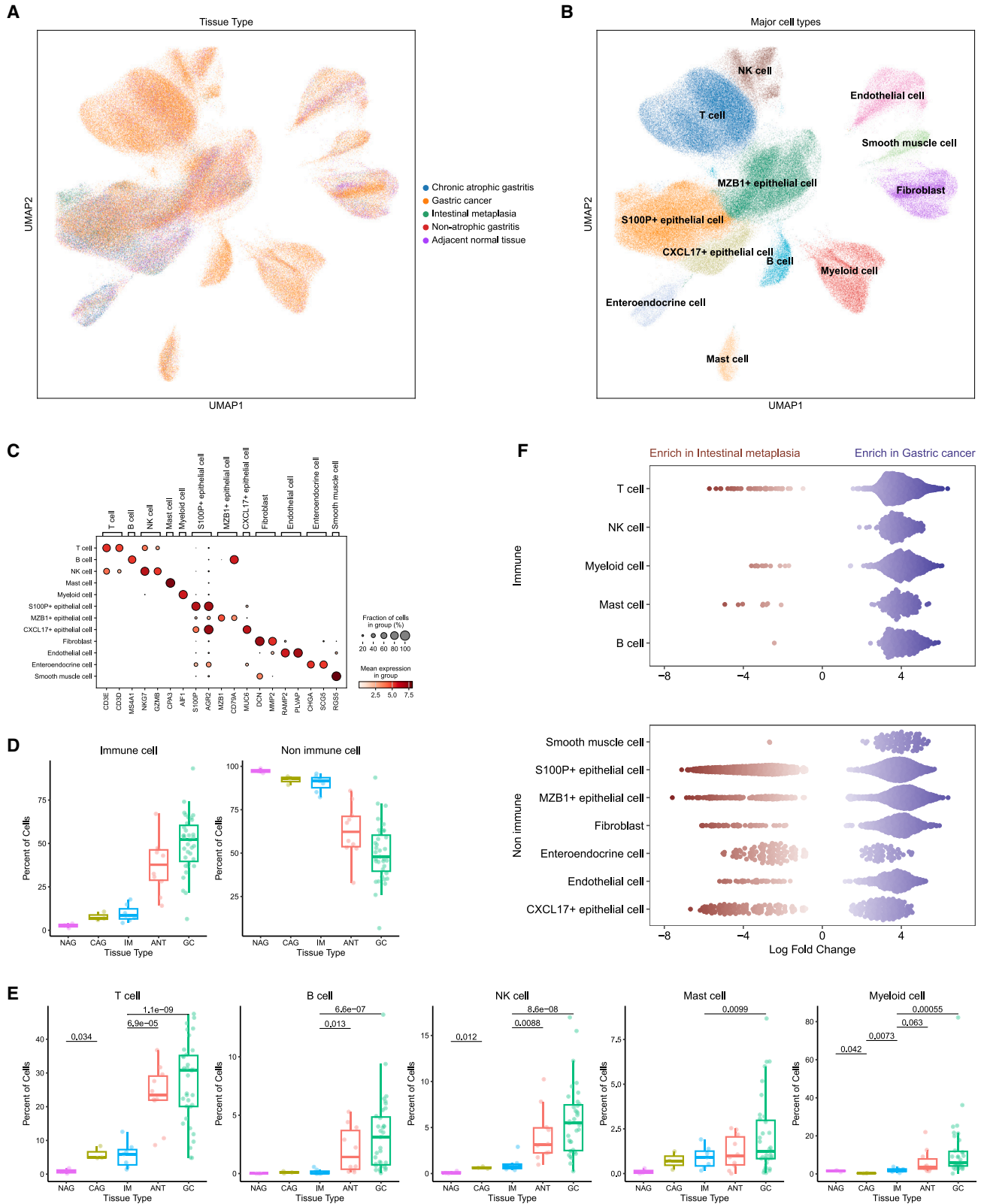
The identification of the biomarkers in this transformation process could help in the early diagnosis of GC so that most patients with GC can get a radical cure to improve their health.

In this study, we collected multiple single-cell RNA sequencing (scRNA-seq) datasets from the Gene Expression Omnibus (GEO). These include NAG, CAG, IM, GC tumor samples, and adjacent normal tissues (ANTs). By comparing the samples from gastritis to IM and then to GC, we found that the abundances of the five immune cell types vary identically. Meanwhile, all of them are significantly enriched in the GC. From IM to GC, interleukin (IL)-17 signaling pathway is up-regulated in the tumor, and its related genes are significantly differentially expressed in the tumor, such as *TNF*, *IL17RA*, *IKBKG*, *TAB2*, *IL1B*, and *CASP8*. These differentially expressed genes (DEGs) can be used as clinical diagnostic markers for the early diagnosis and treatment of GC. We have identified a set of genes that demonstrate up-regulation in immune cell subsets, suggesting their potential involvement in the disease process. While these genes show promise as biomarkers for the early diagnosis and treatment of GC, we acknowledge that the up-regulation of these genes is not exclusive to GC but is also observed in various immune-dependent disorders such as infections, allergies, and autoimmune diseases. This overlap in gene expression profiles complicates the specificity of these markers for GC and may lead to potential misdiagnosis in patients with concurrent diseases.

RESULTS

Immune modifications during gastric carcinogenesis

To identify markers of dynamic changes in the tumor immune microenvironment that are associated with the development and progression of GC from precancerous lesions, we obtained high-quality single-cell expression profiles of 211,056 high-quality cells from 58 samples (Figure S1A; materials and methods). These included 3 NAG samples, 3 CAG samples, 7 IM samples, 35 primary tumor samples, and 10 adjacent normal samples (Figure 1A). The tumor samples included one early GC sample, and the tumor samples ranged from stage I to stage IV (Figure S1B). In order to identify different cell populations based on the expression of marker genes, we performed dimension reduction and unsupervised cell clustering using the method implemented in the Scanpy software suite[15–17] and ultimately identified a total of 12 major cell clusters (Figure 1B; materials and methods). Based on the expression of known marker genes (Figure 1C), all cell types in the gastric samples including immune cells, epithelial cells, fibroblasts,

endothelial cells, enteroendocrine cells, and smooth muscle cells were revealed. Among them, immune cells consist of T cells, B cells, natural killer (NK) cells, mast cells, and myeloid cells. Epithelial cells have three subpopulations, S100P+ epithelial cells, MZB1+ epithelial cells, and CXCL17+ epithelial cells (materials and methods).

By tracking the changes in the proportion of these cell types in each sample, we found that the proportions of immune cell subsets gradually increase and the proportions of non-immune cell subsets gradually decrease from NAG to GC (Figure 1D). Immune cells are significantly enriched at GC and ANT stages (t test, $p < 0.05$). Next, we captured the changes in the relative abundances of the five immune cell types separately. Their relative abundances show basically the same upward trend along the development of GC. We then tested whether the abundance of every immune cell type is significantly different between adjacent stages and found that the abundances of all five immune cell types were significantly different between IM and GC stages (Figure 1E, t test, $p < 0.05$). We used another generalized linear model-based method (Milo) to test the difference in the abundances of cell types.[18] Milo's results show that, relative to IM, all five immune cell types we identified were enriched at the GC stage, which is consistent with the previous observations (Figure 1F). Taken together, these data suggest that the turning point in the trajectory of immune evasion of GC is at the IM stage.

Activated signaling pathways during GC progression

Since the turning point in the trajectory of immune evasion is at the IM stage, we next explored the altered biological processes during this turning point. The NEBULA algorithm[19] was used to obtain the DEGs within five immune cell subsets between IM and GC (Figure 2A; materials and methods). We then performed enrichment analysis on these DEGs within each immune cell subset separately and found that Th17 cell differentiation is significantly up-regulated at the GC stage compared with IM for all five immune cell subsets (Figures S2–S4; materials and methods). Th17 cells are a subset of effector T helper cells that produce IL-17 pro-inflammatory factors.[20] Besides, the IL-17 signaling pathway is significantly up-regulated in subsets of T cells, B cells, mast cells, and myeloid cells at the GC stage compared with IM (Figures S2–S4). IL-17 is an effective pro-inflammatory cytokine, which has been confirmed to be closely related to the formation, growth, and metastasis of tumors in a variety of cancer types.[21] IL-17 binds to its receptor, IL-17R, to form a heterodimeric receptor complex, which activates downstream pathways such as nuclear factor κB

Figure 1. Single-cell atlas of expression and changes in cellular composition during the development of GC

(A) Uniform manifold approximation and projection (UMAP) representations of all scRNA-seq cells colored by the original tissue types.

(B) UMAP representations and annotations of all cells.

(C) Dot plot showing the expression of the marker genes of each cell type.

(D) Boxplots of the cellular proportions of immune cells and non-immune cells in different samples. Each box represents data for a single tissue type. The boxplots and statistics are derived from 3 NAG samples, 3 CAG samples, 7 IM samples, 10 ANT samples, and 35 GC samples. Boxes represent the median, 25th percentile, and 75th percentile of the data, whiskers represent the highest and lowest values within 1.5 times the interquartile range of the boxplot, and all points are plotted.

(E) Boxplots of immune cell type as in (D). t tests were performed for the proportion of adjacent tissue types and the proportion between IM and GC. *p* values smaller than 0.05 are shown in the plot.

(F) Milo analysis of differential abundance changes between IM and GC samples. Neighborhoods that are significantly differentially abundant in IM samples are colored in red, and neighborhoods that are differentially abundant in GC samples are colored in blue.

A

B

C

D

E
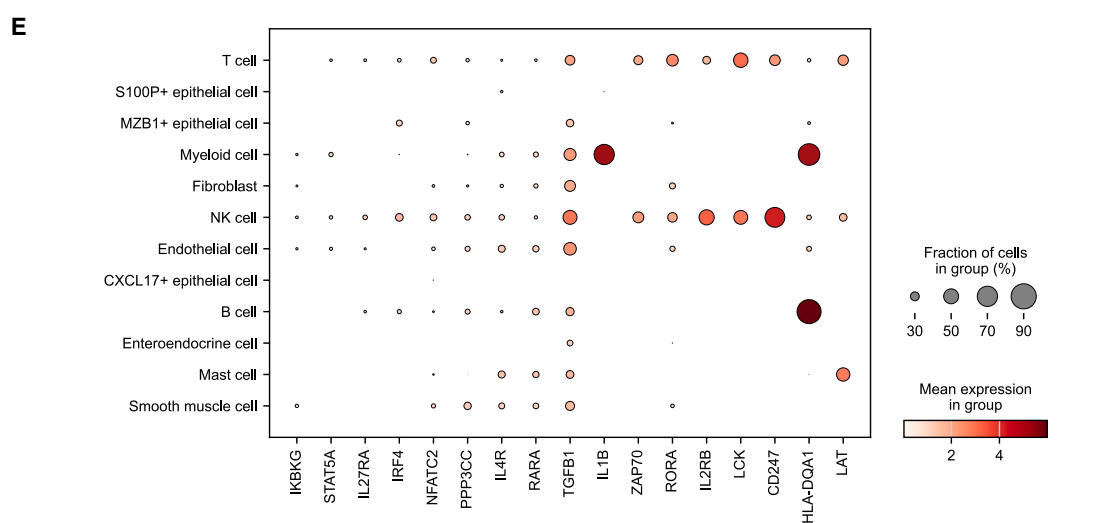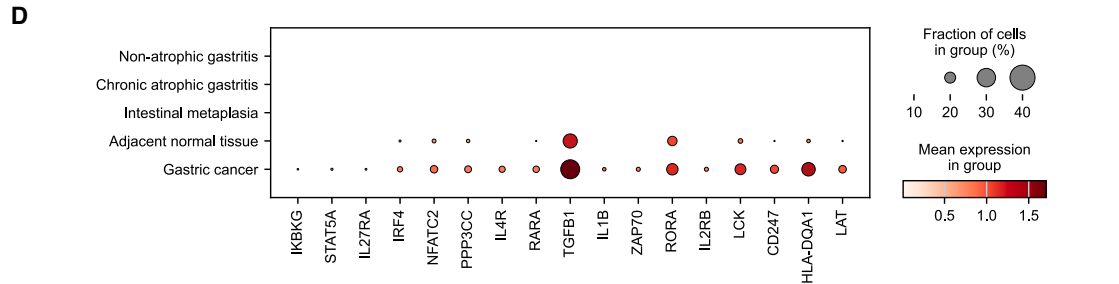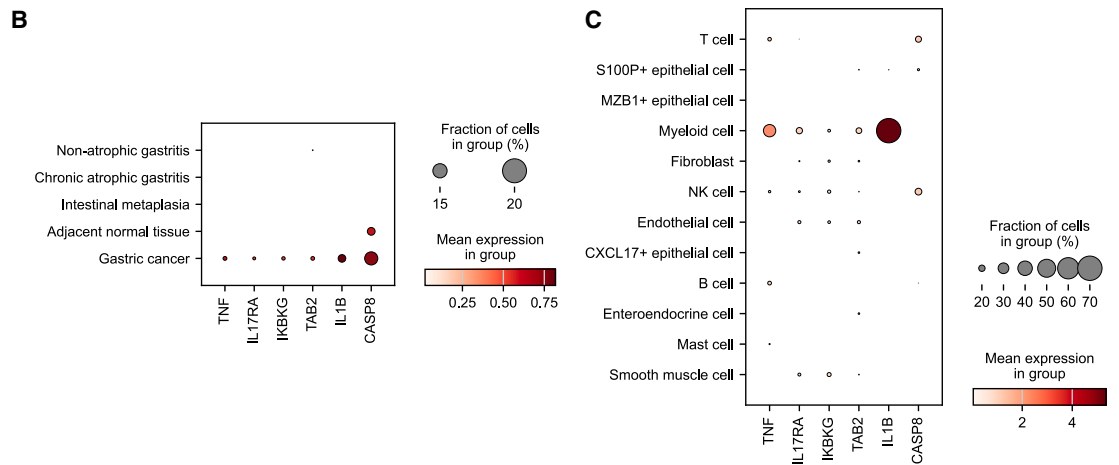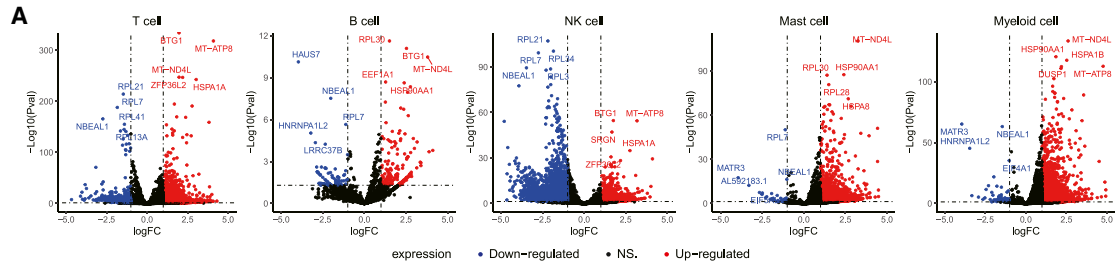
(legend on next page)

(NF-κB), Microtubule Affinity-Regulating Kinase (MAPKs), and CCAAT/Enhancer Binding Protein (C/EBPs) and induces the expression of antimicrobial peptides, cytokines, and chemokines.[21] This is consistent with our observation that the NF-κB signaling pathway and MAPK signaling pathway are significantly up-regulated in all subsets of immune cells during GC (Figures S2–S4).

Since Th17 cell-associated activities are the common event in altered immune cells in GC, we further examined the expression of specific genes involved in the IL-17 signaling pathway and Th17 differentiation. In the IL-17 signaling pathway, some genes were only expressed at the GC stage or up-regulated at the GC stage, such as *TNF*, *IL17RA*, *IKBKG*, *TAB2*, *IL1B*, and *CASP8*, among others (Figure 2B). As shown in Figure 2C, *IL1B* is mainly expressed by myeloid cells, *CASP8* is mainly expressed by T and NK cells, and *TNF* is mainly expressed by myeloid cells, but *TNF* is also expressed by other immune cells. These results suggest that the IL-17 signaling pathways in GC require coordination between diverse immune cell types. Genes that contribute to Th17 cell differentiation are *IL1B*, *IL2RB*, *ZAP70*, *HLA-DQA1*, *LAT*, and *IL4R*. We found that many genes related to Th17 cell differentiation were expressed only at the GC stage or up-regulated at the GC stage, whereas none of these genes were found to be expressed in precancerous lesion samples (Figure 2D). Meanwhile, we found that *IL1B* is mainly expressed by myeloid cells. Additionally, some genes are mainly expressed by two or three immune cells, such as *ZAP70*, *IL2RB*, *LCK*, and *CD247* are mainly expressed by T and NK cells, *HLA-DQA1* is mainly expressed by myeloid cells and B cells, and *LAT* is mainly expressed by T, NK, and mast cells. In addition, some genes are expressed in all immune cells, for instance, *IL4R* and *RARA* (Figure 2E). These results further confirm that the IL-17 signaling pathway and Th17 cell differentiation in GC require the coordination between diverse immune cell types.

Immune infiltration of GC samples in TCGA

To track the changes in immune cell states along GC progression, we obtained the bulk RNA-seq data of GC from The Cancer Genome Atlas (TCGA)[22] and used the CIBERSORTx[23] platform to deconvolute these GC bulk RNA-seq data to infer the cellular proportion of each sample of the TCGA GC data. Due to the limited storage space on the CIBERSORTx platform (1,000 megabytes), we used the Metacell-2 algorithm to reduce the number of cells in the merged scRNA-seq dataset.[24] Finally, the derived metacell dataset contains 11,337 metacells (materials and methods). We identified 10 major cell clusters based on marker gene expression (Figures 3A and 3B; materials and methods). Although the granularity of metacell clusters is slightly different from those of single-cell clusters identified by single-cell expression profile, the major cell types in gastric tissues are all revealed, and the five immune cell types that we focus on in this study

are consistent. Since the metacell dataset is still bigger than 1,000 megabytes, we downsampled approximately 75% of the metacells multiple times to ensure that the deconvolution results are stable.

Using metacells as the reference gene expression profiles, we deconvolved the GC TCGA data and obtained the inferred proportion of each cell type in each sample (materials and methods). According to the clinical information, we grouped the samples by the stage of GC. As shown in Figure 3C, the proportions of five immune cells show an overall upward trend with the progression of GC, and the proportions of T cells, B cells, and myeloid cells showed significant differences in different stages. Next, we compared the changes in the proportion of each immune cell in the four tumor stages of GC in the single-cell data and TCGA data. We found that the proportion of immune cells in the two data showed a similar trend during tumor progression (Figure 3D). These data validate the changes in the abundance of immune cell subsets found in our single-cell data.

Signaling pathway analysis of subsets of T cells and myeloid cells

Next, we further subdivided T cells and myeloid cells into subtypes and examined changes in the abundance of these two immune cell subtypes as well as their differential gene expression. We used the Scanpy software suite to perform dimension reduction and unsupervised cell clustering for both T cells and myeloid cells.[15–17] We then annotated the subtypes based on the expression of known marker genes (materials and methods). Ultimately, three T cell subtypes were identified, including regulatory T cells, cytotoxic T cells, and naive T cells (Figure 4A). Meanwhile, five subtypes of myeloid cells were identified, including intermediate monocytes, monocytes/macrophages, non-classical monocytes, conventional dendritic cells (DCs), and plasmacytoid DCs (pDCs) (Figure 4B).

Next, we examined the changes in abundance of the eight immune cell subtypes obtained above and used the t test to determine the significance of differences in abundance between the adjacent stages. We found the abundance of T cell and myeloid cell subtypes gradually increased from NAG to GC (Figures 4C and 4D). The common feature among the changes in the abundance of the eight subtypes is that the difference in abundance between IM and GC was significant, and all eight subtypes were relatively significantly enriched in the GC and ANT stages (t test, $p < 0.05$). This is consistent with our previous conclusion that the turning point in the trajectory of immune evasion of GC is at the IM stage.

From IM to GC, we used the NEBULA algorithm to obtain DEGs within subsets of T cells and myeloid cells. Then, we examined the DEGs within each subset and performed Kyoto Encyclopedia of

---

Figure 2. Gene expression in GC development

(A) Volcano plots of the results of differential expression analysis in immune cells between IM and GC by using the NEBULA algorithm. Red dots represent genes up-regulated (log FC ≥ 1 and $p < 0.05$) in GC, and blue dots represent genes down-regulated (log FC ≤ −1 and $p < 0.05$) in GC.

(B and C) Dot plots of representative genes in the IL-17 signaling pathway mapped onto tissue types and cell types.

(D and E) Dot plots of representative genes in Th17 cell differentiation mapped onto tissue types and cell types.
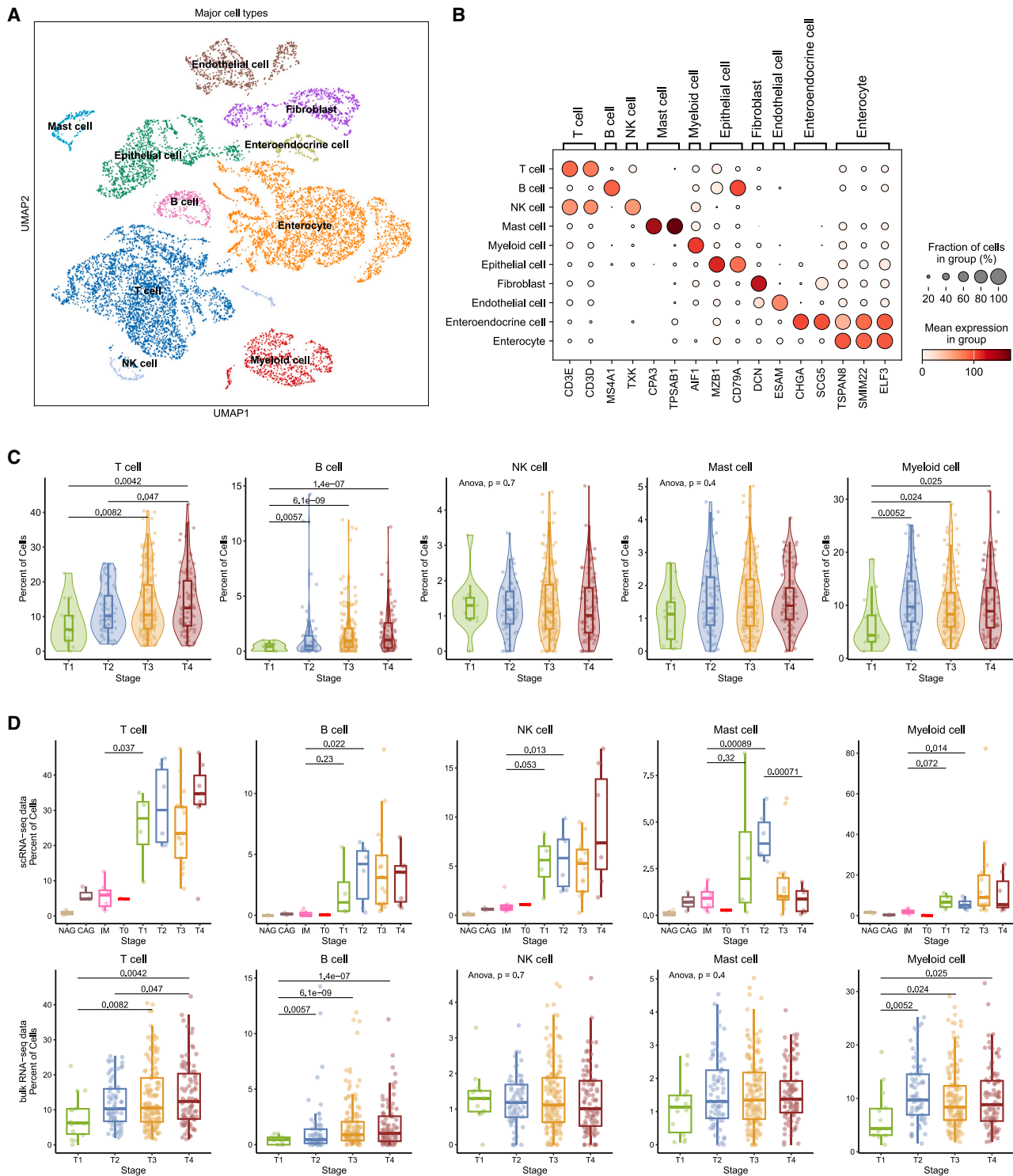
Figure 3. Immune cell composition of GC samples in TCGA

(A) UMAP representations and annotations of metacells.

(B) Dot plot of the expression of the marker genes for each cell type.

(C) Violin plots of the proportion of each immune cell type in different samples from bulk RNA-seq dataset. Each violin represents data for a single GC stage. t test *p* values between any two stages less than 0.05 are shown in the plot.

(legend continued on next page)

Genes and Genomes (KEGG) pathway enrichment analysis using these genes. In the up-regulated gene enrichment analysis, we found that compared to IM, the IL-17 signaling pathway was significantly up-regulated only in regulatory T cells, intermediate monocytes, and monocytes/macrophages during the GC stage. Additionally, we found that Th17 cell differentiation was significantly up-regulated only in regulatory T cells, cytotoxic T cells, naive T cells, intermediate monocytes, and pDCs during the GC stage. At the same time, we examined the expression of genes involved in the IL-17 signaling pathway that were up-regulated at the GC stage in regulatory T cells, intermediate monocytes, and monocytes/macrophages. We found that genes expressed or up-regulated only at the GC stage were the same as when T cells and myeloid cells were not divided. We also examined the expression of genes involved in Th17 cell differentiation that were up-regulated at the GC stage in regulatory T cells, cytotoxic T cells, naive T cells, intermediate monocytes, and pDCs. We found that genes expressed or up-regulated only at the GC stage were almost the same as when T cells and myeloid cells were not divided, and only *IL4R* was not found. Finally, we examined the expression of these genes in different cell types and did not find any genes that were specifically expressed only in a subset (Figures 4E and 4F).

In summary, these data and results suggest that only some, but not all, immune subsets are involved in the activities related to the IL-17 signaling pathway and Th17 cell differentiation. This also provides further evidence that the IL-17 signaling pathways in GC require coordination between diverse immune cell types.

DISCUSSION
Cancer is a growing threat to public health, with increasing incidence and mortality rates. Chronic inflammation has been recognized as a contributing factor to cancer development, and it has been observed that chronic or excessive inflammation can lead to a significant number of cancer types. In this study, we aimed to investigate the molecular mechanisms involved in the transformation of inflammation into GC and identify potential biomarkers for early diagnosis and treatment.

Our study revealed the immune modifications during gastric carcinogenesis. By analyzing scRNA-seq data from multiple samples of NAG, CAG, IM, GC at different stages, and ANTs, we identified distinct immune cell populations and their dynamic changes throughout the progression of GC. We observed a gradual increase in the proportion of immune cell subsets and a decrease in non-immune cell subsets from NAG to GC, indicating a shift toward immune cell enrichment in the tumor microenvironment. Furthermore, we found that the abundances of all five immune cell types (T cells, B cells, NK cells, mast cells, and myeloid cells) were significantly different between

IM and GC stages, with enrichment in the GC samples. This suggests that the turning point in the trajectory of immune evasion in GC occurs at the IM stage. These observations highlight the importance of understanding the immune cell dynamics and their functional changes during the progression of GC.

We then focused on the IL-17 signaling pathway, which plays a crucial role in inflammation and has been implicated in various cancers. Our analysis revealed the up-regulation of the IL-17 signaling pathway in subsets of T cells, B cells, mast cells, and myeloid cells at the GC stages compared to IM. The IL-17 signaling pathway is known to activate downstream pathways, such as NF-κB and MAPKs, and induce the expression of pro-inflammatory cytokines and chemokines. The significant up-regulation of the NF-κB and MAPK signaling pathways in immune cells during GC further supports the involvement of the IL-17 pathway in gastric carcinogenesis. Moreover, we identified several specific genes involved in the IL-17 signaling pathway and Th17 cell differentiation that were expressed or up-regulated in GC samples, including *TNF*, *IL17RA*, *IKBKG*, *TAB2*, *IL1B*, and *CASP8*. These genes may serve as potential clinical diagnostic markers for early detection and treatment of GC. The coordinated expression of specific genes involved in the IL-17 signaling pathway and Th17 cell differentiation was observed across diverse immune cell types, emphasizing the interplay and cooperation between different immune subsets in GC. Meanwhile, in-depth analysis of immune cell subtypes, specifically T cells and myeloid cells, revealed that only a portion of immune subsets are involved in the activities related to the IL-17 signaling pathway and Th17 cell differentiation.

To validate our findings, we analyzed bulk RNA-seq data from TCGA and confirmed the changes in immune cell proportions observed in our scRNA-seq data. The similarities in immune cell dynamics between the two datasets further support the reliability of our findings.

In conclusion, our study provides comprehensive insights into the immune modifications and signaling pathways involved in gastric carcinogenesis. The identification of immune cell dynamics and the up-regulation of the IL-17 signaling pathway during GC progression highlight potential biomarkers for early diagnosis and treatment. Some genes involved in the IL-17 pathway and Th17 cell differentiation serve as potential drug targets for clinical applications. Further research and validation of these findings could contribute to improved cancer prevention, early detection, and therapeutic strategies for GC.

MATERIALS AND METHODS
Collection of scRNA-seq and bulk RNA-seq data
scRNA-seq data were obtained from the GEO database under GEO: GSE183904[9] and GSE134520,[14] and from the Database of Genotypes

---

(D) Boxplots of the proportion of each immune cell type in different samples from scRNA-seq data (the first row) and bulk RNA-seq data (the second row). Each box represents data for a single tissue type or GC stage. The boxplots and statistics are derived from 3 NAG samples, 3 CAG samples, 7 IM samples, 1 early GC samples, 4 stage I GC samples, 6 stage II GC samples, 17 stage III GC samples, and 7 stage IV GC samples in the scRNA-seq data and 14 stage I GC samples, 70 stage II GC samples, 157 stage III GC samples, and 95 stage IV GC samples in the bulk RNA-seq data. T0, early GC; T1–T4, stage I–IV GC.
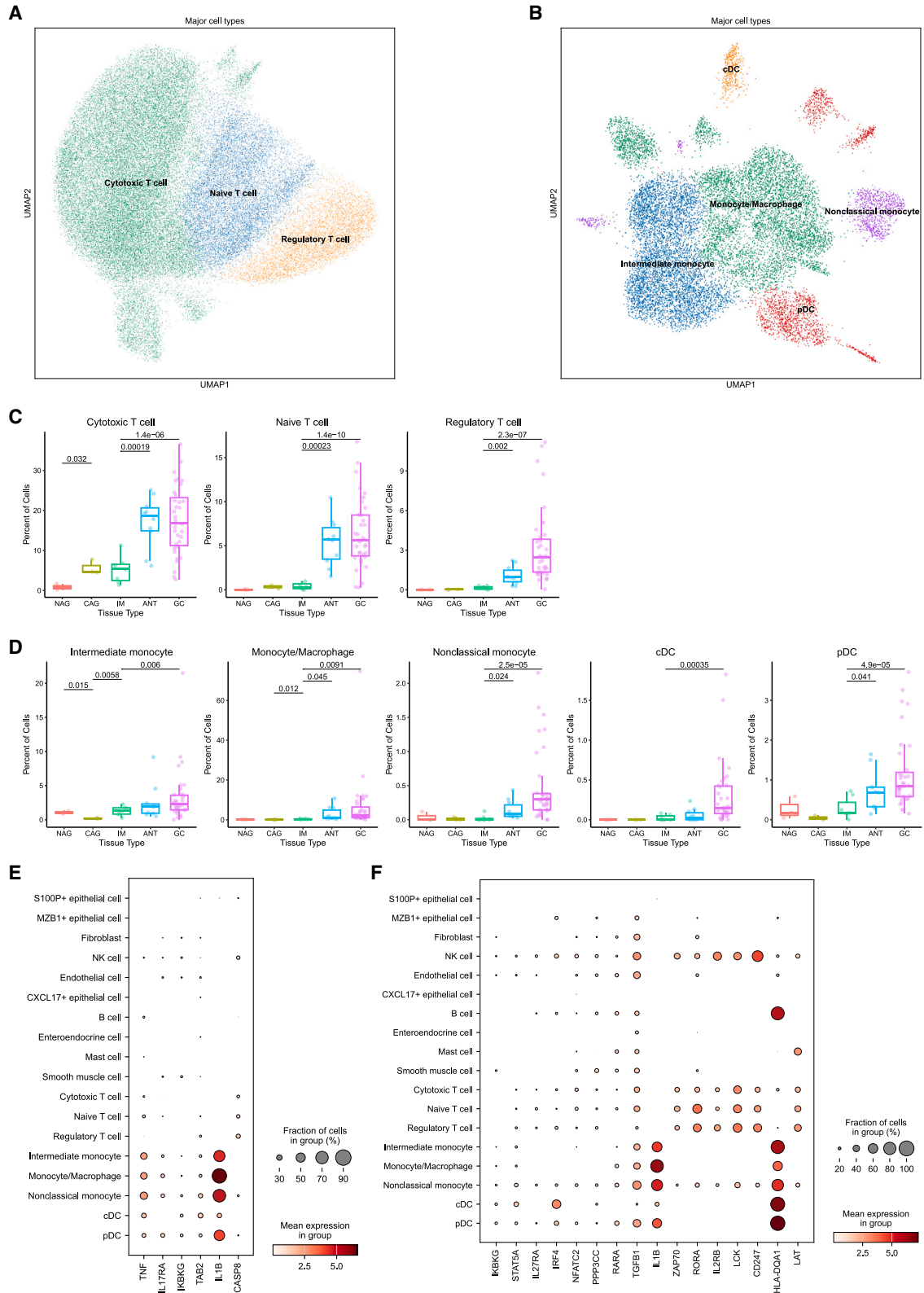
(legend on next page)

and Phenotypes (dbGaP) under the identifier phs001818.v1.p1.[25] In total, 58 samples from 43 patients were selected for analysis. These included 3 NAG samples, 3 CAG samples, 7 IM samples, 35 primary tumor samples, and 10 adjacent normal samples.

Bulk RNA-seq data of patients with GC were downloaded from TCGA database (https://portal.gdc.cancer.gov/).

Quality control and preprocessing of scRNA-seq data

We performed quality control on each dataset using the method implemented in the Scanpy software suite[17] before merging them. Genes detected in fewer than three cells and cells expressing less than 200 detected genes were filtered out and excluded from downstream analysis. Next, cells would be labeled as poor-quality ones if they presented with one of the following conditions: (1) the number of Unique Molecular Identifier (UMIs) is lower than 600 or larger than 120,000, (2) the number of expressed genes is lower than 400 or larger than 7,000, or (3) 20% or more of the UMIs were mapped to mitochondrial genes.[14] Then, we continued to utilize the functions in the Scanpy to normalize the gene expression level to UMIs Per Million (UPM) and log transform the single-cell gene expression data. Finally, after merging these datasets, we obtained 211,056 high-quality cells, and 15,569 genes were kept in the merged dataset, which was then used for subsequent analysis.

Dimension reduction, unsupervised clustering, and annotation of scRNA-seq data

Dimension reduction, unsupervised clustering, and annotation of scRNA-seq data were done using functions in the Scanpy software package (v.1.9.1).

We identified a subset of highly variable genes (HVGs) in the preprocessed expression matrix and then scaled gene expression values before performing dimension reduction and clustering on them. The selection of HVGs in scRNA-seq data was implemented by the "pp.highly_variable_genes" function in the Scanpy package. The valid value of average expression was set to the range from 0.05 to 5, dispersion was set to no less than 0.5, and the default values were used for other parameters. Finally, the 4,000 HVGs were selected for further analysis.

Next, we used the "pp.pca" function to perform the principal-component analysis on the single-cell expression matrix with genes restricted to HVGs and retained the top 50 principal components for subsequent analysis. Using these data, we performed Harmony and BBKNN to construct the K-Nearest Neighbors (KNN) graph of high-quality cells. In the same graph, we used the "tl.umap" function to project cells to 2D space for visualization. Meanwhile, we utilized the "tl.leiden" function to conduct an unsupervised graph-based clus-

tering algorithm called Leiden to cluster cells. Due to the large number of cells in our study, we set the parameter resolution as 4. In this way, a total of 12 major cell clusters were identified.

We annotated these cell clusters using the expression of known marker genes. Marker genes include *CD3E* and *CD3D* for T cells; *MS4A1* for B cells; *NKG7* and *GZMB* for NK cells; *CPA3* for mast cells; *AIF1* for myeloid cells; *S100P* and *AGR2* for S100P+ epithelial cells; *MZB1* and *CD79A* for MZB1+ epithelial cells; *MUC6* for CXCL17+ epithelial cells; *DCN* and *MMP2* for fibroblasts; *RAMP2* and *PLVAP* for endothelial cells; *CHGA* and *SCG5* for enteroendocrine cells; and *RGS5* for smooth muscle cells.

Differential gene expression analysis

Differential gene expression analysis was performed using the tool NEBULA, which performs differential expression based on a negative binomial mixed model for two annotated cell groups.[19] In R (v.4.2.2), we performed differential gene expression analysis by using the NEBULA software package (v.1.2.2) to find DEGs in each immune cell type between IM and GC states. If the log-transformed fold change (log FC) is ≥ 1 and the p value is <0.05, then the gene was defined as up-regulated. If the log FC is ≤ -1 and the p is <0.05, then the corresponding gene was defined as down-regulated.

Gene set enrichment analysis

We performed the gene set enrichment analysis of DEGs by using the online platform WebGestalt (http://www.webgestalt.org/), with which the enriched KEGG pathways were derived, and we defined the pathways with a false discovery rate ≤ 0.05 as the significantly enriched pathways.

Metacell-2 for scRNA-seq data

We loaded the Metacells package (v.0.8.0)[24] into python (v.3.9.12) and ran the Metacell-2 algorithm on scRNA-seq data with default parameters. In the end, we obtained 11,337 metacells for subsequent analysis.

Annotation of metacells

We applied the same methods for dimension reduction, clustering, and annotation to metacells. First, we extracted the gene expression profiles of metacells of the top 4,000 HVGs. Then we ran "pp.pca" and "tl.umap" functions on this expression matrix in turn to achieve the purpose of dimension reduction. Finally, we used the "tl.leiden" function to cluster metacells and set the parameter resolution as 0.05. In this way, a total of 10 metacell clusters were identified.

We annotated these metacell clusters by using the expression of known marker genes. Marker genes included *CD3E* and *CD3D* for

Figure 4. The composition of subtypes of T cells and myeloid cells

(A and B) UMAP representations and annotations of subtypes in T cells and myeloid cells.

(C and D) Boxplots of the proportion of each T cell subtype and each myeloid cell subtype in different samples.

(E) Dot plots of representative genes in the IL-17 signaling pathway mapped onto cell types.

(F) Dot plots of representative genes in Th17 cell differentiation mapped onto cell types.

T cells; *MS4A1* for B cells; *TXK* for NK cells; *CPA3* and *TPSAB1* for mast cells; *AIF1* for myeloid cells; *MZB1* and *CD79A* for epithelial cells; *DCN* for fibroblasts; *ESAM* for endothelial cells; *CHGA* and *SCG5* for enteroendocrine cells; and *TSPAN8*, *SMIM22*, and *ELF3* for enterocytes.

Bulk RNA-seq data analysis
We used the online platform CIBERSORTx (https://cibersortx.stanford.edu/) to estimate the composition of various cell populations in bulk RNA-seq data. Signature gene matrices were created using the expression profiles of metacells as the reference gene expression profiles. We ran the "Impute Cell Fractions" module with default parameters and enabled S-mode batch correction. In this way, we obtained the proportions of each cell type for each bulk RNA-seq sample.

Division of T cells and myeloid cells
First, we extracted the expression matrices of the top 4,000 HVGs of T cells and myeloid cells, respectively. Then, we used the "pp.pca" and "tl.umap" functions to reduce the dimensions of the two expression matrices. Finally, we used the "tl.leiden" function to re-cluster T cells and myeloid cells, respectively, and set the parameter resolution as 0.25 for both. By this means, three subtypes in T cells and five subtypes myeloid cells were identified.

We annotated these subtypes by using the expression of known marker genes. Marker genes included *GZMK* and *CD8* for cytotoxic T cells; *CCR7* for naive T cells; *FOXP3* for regulatory T cells; *FCGR3A* for intermediate monocytes; *FCN1* and *VCAN* for monocytes/macrophages; *IFITF1* for non-classical monocytes; *IDO1* and *CLIC2* for conventional dendritic cells; and *PLD4* and *IRF8* for pDCs.

DATA AND CODE AVAILABILITY
This study did not generate new unique reagents. Raw data of this study are available on the GEO with GEO: GSE183904 and GSE134520 and the dbGaP with the identifier phs001818.v1.p1.

SUPPLEMENTAL INFORMATION
Supplemental information can be found online at https://doi.org/10.1016/j.omton.2024.200849.

ACKNOWLEDGMENTS
Ethical approval and consent to participate were waived since we used only publicly available data and materials in this study. There is no funding support for this study.

AUTHOR CONTRIBUTIONS
S.G. and C.L. conceived of the study, conducted the statistical analysis, and drafted the manuscript. R.H., G.T., Y.Z., D.R., F.Z., K.L., and J.Y. revised the manuscript and designed the study. All authors have reviewed and approved the final version of the text.

DECLARATION OF INTERESTS
The authors declare the following financial interests/personal relationships, which may be considered as potential competing interests: C.L. and D.R. are employed by Daqing Longnan Hospital, The Fifth Affiliated Hospital of Qiqihar Medical College; R.H. and J.Y. are employed by Geneis Beijing Co., Ltd; and G.T. is the chairman of Geneis Beijing Co., Ltd.

REFERENCES
1. Siegel, R.L., Miller, K.D., Fuchs, H.E., and Jemal, A. (2022). Cancer statistics, 2022. CA. Cancer J. Clin. *72*, 7–33.

2. Paduch, R. (2015). Theories of cancer origin. Eur. J. Cancer Prev. *24*, 57–67.

3. Afify, S.M., Hassan, G., Seno, A., and Seno, M. (2022). Cancer-inducing niche: the force of chronic inflammation. Br. J. Cancer *127*, 193–201.

4. Balkwill, F., and Mantovani, A. (2001). Inflammation and cancer: back to Virchow? Lancet *357*, 539–545.

5. Medzhitov, R. (2021). The spectrum of inflammatory responses. Science *374*, 1070–1075.

6. Tan, P., and Yeoh, K.G. (2015). Genetics and Molecular Pathogenesis of Gastric Adenocarcinoma. Gastroenterology *149*, 1153–1162.e3.

7. Liu, Y., Liu, L., Zhou, Y., Zhou, P., Yan, Q., Chen, X., Ding, S., and Zhu, F. (2019). CKLF1 Enhances Inflammation-Mediated Carcinogenesis and Prevents Doxorubicin-Induced Apoptosis via IL6/STAT3 Signaling in HCC. Clin. Cancer Res. *25*, 4141–4154.

8. Becker, W.R., Nevins, S.A., Chen, D.C., Chiu, R., Horning, A.M., Guha, T.K., Laquindanum, R., Mills, M., Chaib, H., Ladabaum, U., et al. (2022). Single-cell analyses define a continuum of cell state and composition changes in the malignant transformation of polyps to colorectal cancer. Nat. Genet. *54*, 985–995.

9. Kumar, V., Ramnarayanan, K., Sundar, R., Padmanabhan, N., Srivastava, S., Koiwa, M., Yasuda, T., Koh, V., Huang, K.K., Tay, S.T., et al. (2022). Single-Cell Atlas of Lineage States, Tumor Microenvironment, and Subtype-Specific Expression Programs in Gastric Cancer. Cancer Discov. *12*, 670–691.

10. Sung, H., Ferlay, J., Siegel, R.L., Laversanne, M., Soerjomataram, I., Jemal, A., and Bray, F. (2021). Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA. Cancer J. Clin. *71*, 209–249.

11. Zou, X., Jia, M., Wang, X., and Zhi, X. (2021). Interpretation of the World Cancer Report 2020. Chin. J. Clin. Thor. Cardiovasc. Surg. *28*, 11–18.

12. Correa, P., and Piazuelo, M.B. (2012). The gastric precancerous cascade. J. Dig. Dis. *13*, 2–9.

13. Morson, B.C. (1955). Carcinoma arising from areas of intestinal metaplasia in the gastric mucosa. Br. J. Cancer *9*, 377–385.

14. Zhang, P., Yang, M., Zhang, Y., Xiao, S., Lai, X., Tan, A., Du, S., and Li, S. (2019). Dissecting the Single-Cell Transcriptome Network Underlying Gastric Premalignant Lesions and Early Gastric Cancer. Cell Rep. *27*, 1934–1947.e5.

15. Becht, E., McInnes, L., Healy, J., Dutertre, C.A., Kwok, I.W.H., Ng, L.G., Ginhoux, F., and Newell, E.W. (2018). Dimensionality reduction for visualizing single-cell data using UMAP. Nat. Biotechnol. *37*, 38–44.

16. Traag, V.A., Waltman, L., and van Eck, N.J. (2019). From Louvain to Leiden: guaranteeing well-connected communities. Sci. Rep. *9*, 5233.

17. Wolf, F.A., Angerer, P., and Theis, F.J. (2018). SCANPY: large-scale single-cell gene expression data analysis. Genome Biol. *19*, 15.

18. Dann, E., Henderson, N.C., Teichmann, S.A., Morgan, M.D., and Marioni, J.C. (2022). Differential abundance testing on single-cell data using k-nearest neighbor graphs. Nat. Biotechnol. *40*, 245–253.

19. He, L., Davila-Velderrain, J., Sumida, T.S., Hafler, D.A., Kellis, M., and Kulminski, A.M. (2021). NEBULA is a fast negative binomial mixed model for differential or co-expression analysis of large-scale multi-subject single-cell data. Commun. Biol. *4*, 629.

20. Park, H., Li, Z., Yang, X.O., Chang, S.H., Nurieva, R., Wang, Y.H., Wang, Y., Hood, L., Zhu, Z., Tian, Q., and Dong, C. (2005). A distinct lineage of CD4 T cells regulates tissue inflammation by producing interleukin 17. Nat. Immunol. *6*, 1133–1141.

21. Zhao, J., Chen, X., Herjan, T., and Li, X. (2020). The role of interleukin-17 in tumor development and progression. J. Exp. Med. *217*, e20190297.

22. Cancer Genome Atlas Research Network, Weinstein, J.N., Collisson, E.A., Mills, G.B., Shaw, K.R.M., Ozenberger, B.A., Ellrott, K., Shmulevich, I., Sander, C., and Stuart, J.M. (2013). The Cancer Genome Atlas Pan-Cancer analysis project. Nat. Genet. *45*, 1113–1120.

23. 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. (2019). Determining cell type abundance and expression from bulk tissues with digital cytometry. Nat. Biotechnol. *37*, 773–782.

24. Ben-Kiki, O., Bercovich, A., Lifshitz, A., and Tanay, A. (2022). Metacell-2: a divide-and-conquer metacell algorithm for scalable scRNA-seq analysis. Genome Biol. *23*, 100.

25. Sathe, A., Grimes, S.M., Lau, B.T., Chen, J., Suarez, C., Huang, R.J., Poultsides, G., and Ji, H.P. (2020). Single-Cell Genomic Characterization Reveals the Cellular Reprogramming of the Gastric Tumor Microenvironment. Clin. Cancer Res. *26*, 2640–2653.