Single cell RNA-seq reveals immunosuppressive gastric stem-like cancer cells as a poor prognostic factor

Hong Kai Lee[1,†], Kok Siong Ang[1,†], Marion Chevrier[1,‡], Michelle Goh[1,‡], Jingjing Ling[1], Vivien Koh[2], Xiaomeng Zhang[1], Jia Chi Tan[1], Nicole Lee[1], Sergio Erdal Irac[1,3], Ze Ming Lim[1], Josephine Lum[1], Alicia Tay[1], Charles Antoine Dutertre[1], Carolyn Tang[4], Calista Wong[1], Mai Chan Lau[1], Chun Jye Lim[5], Samuel Wen Jin Chuah[5], Sherlly Lim[6], Joe Poh Sheng Yeong[6,7], Valerie Chew[5], Anis Larbi[1], Amit Singhal[1], Michael Poidinger[8], Matthew Chau Hsien Ng[9], Shanshan Wu Howland[1], Patrick Tan[10], Florent Ginhoux[1], Jimmy So[11], Wei Peng Yong[2,13], Jinmiao Chen[1,12,13*]

† These author contributed equally
‡ These author contributed equally
* Corresponding author

[1]Singapore Immunology Network (SIgN), Agency for Science, Technology and Research (A*STAR), Singapore 138648
[2]Department of Haematology-Oncology, National Cancer Institute of Singapore (NCIS), National University Hospital, Singapore 119074
[3]School of Veterinary Science, The University of Queensland, Gatton, QLD 4343, Australia
[4]Molecular, Cell and Developmental Biology, University of California Santa Cruz, California 95064, United States
[5]Translational Immunology Institute (TII), SingHealth-DukeNUS Academic Medical Centre, Singapore 169857
[6]Institute of Molecular Cell Biology (IMCB), Agency of Science Technology and Research (A*STAR), Singapore 138648
[7]Department of Anatomical Pathology, Singapore General Hospital, Singapore 169608
[8]Murdoch Children's Research Institute, Royal Children's Hospital, Parkville VIC 3052, Australia
[9]Division of Medical Oncology, National Cancer Centre Singapore, Singapore 169610
[10]Genome Institute of Singapore (GIS), Agency for Science, Technology and Research (A*STAR), Singapore 138672
[11]Department of Surgery, Yong Loo Lin School of Medicine, National University of Singapore, Singapore 117597
[12]Department of Microbiology and Immunology, Yong Loo Lin School of Medicine, NUS
[13]Senior author

Jinmiao Chen
8A Biomedical Grove
Immunos, #03-06 Singapore 138648
Email: chen_jinmiao@immunol.a-star.edu.sg
Tel: (+65) 6407-0395

Running head: Immunosuppressive gastric stem-like cancer cells

Declaration of Interests

The authors declared no conflict of interest.

SUMMARY

The tumor microenvironment is characterized by high cellular heterogeneity and a complex network of cell-cell communications. Analysis at the single cell level is required to dissect this complexity. Here we apply single-cell full-length transcriptome sequencing to analyze 3443 cells from paired normal-adjacent and tumor tissues of 15 gastric cancer patients with different histological profiles. Among the epithelial cells profiled, we discovered subsets of proliferative stem-like cells with upregulated pro-tumor pathways and that were associated with poorer prognosis. Stem-like cell enriched tissues also show distinct T cell populations with proportionally more TIM3$^+$ CD8$^+$ cells, while non-enriched tissues had proportionally more KLRC1$^+$ CD8$^+$ cells, offering opportunities for checkpoint inhibitor therapy. *In silico* predictions of ligand-receptor interactions revealed immune-suppressive interactions between these stem-like cells and immune cells via interacting pairs such as Nectin-2(CD112)-*TIGIT*, Galectin-9-*TIM3* and *CXCL16*-*CXCR6*. Using immunohistochemistry, we found CD8$^+$ T cells expressing TIGIT in close proximity to stem-like cells expressing NECTIN2.

(150 words)

Keywords: gastric cancer, single-cell RNA-sequencing, cancer stem-like cells, cell–cell interaction, tissue heterogeneity, immunity

INTRODUCTION

Gastric cancer (GC) is currently the fifth most common cancer in males and seventh most common in females (Bray et al., 2018). While GC incidence has declined in North America, and North and Western Europe, it remains prevalent in other parts of the world such as East Asia, Latin America, and Eastern Europe (Balakrishnan et al., 2017). Without proactive screening, it is usually discovered at late stages with poor survival prospects.

In current clinical practice, GC tumors are commonly categorized histologically using World Health Organization (WHO) and Lauren classification into intestinal and diffuse subtypes. Together with the tumor, node, metastasis (TNM) staging, these information is used to guide treatment and prognostication (Lauren, 1965). In recent years, the Cancer Genome Atlas (TGCA) project identified four molecular subtypes (Bass et al., 2014), while the Asian Cancer Research Group (ACRG) produced an alternative molecular classification which was demonstrated to be correlated with patient survival (Cristescu et al., 2015; Ichikawa et al., 2017). These molecular classifications rely on bulk tissue characterization techniques, ignoring intratumoral heterogeneity. On the other hand, single cell techniques can dissect the heterogeneity and discover rare cell types or cell states not found using bulk techniques (Zhang et al., 2019b). Characterizing the different cell sub-types and states present can lead to gene signatures for more fine-grained patient classifications based on tumor and immune cellular compositions. This approach is applicable towards developing personalized therapy formulations, including immunotherapies.

Immune checkpoint inhibitors (ICI) have emerged as an effective strategy against solid tumors (Nixon et al., 2018). Available therapies include cytotoxic T-lymphocyte antigen 4 (CTLA4) and programmed death 1 (PD-1)/PD-L1 inhibitors (Pardoll, 2012), and have seen success on multiple solid cancers such as melanoma and non-small cell lung cancer (Nixon et al., 2018). For cancers such as GC, the response rate is low with ICI monotherapy ranging from 10-15% (Fuchs et al., 2018; Janjigian et al., 2018; Kang et al., 2017; Kim et al., 2018; Le et al., 2015; Shitara et al., 2018; Yang et al., 2019). High response rates have been reported with deficient mismatch repair, microsatellite unstable and EBV positive GC, however these are uncommon subgroups. The commonly used immunohistochemical staining of PD-L1 is only weakly predictive for efficacy with single agent ICI (Kulangara et al., 2019). Therefore, better selection of biomarkers, new therapies and combinations are needed to extend the benefits of checkpoint therapy to more patients. Here, single cell methods provide a window in studying interactions between tumor and immune cells to shed light on immune evasion mechanisms (Nishida-Aoki and Gujral, 2019). Characterizing these interactions affords opportunities to develop new therapies, as well as rational design of combination strategies.

Recent advancements in single cell RNA sequencing (scRNA-seq) studies on GC such as Zhang *et al.*(2019b), showed distinct cell type compositions and gene expression heterogeneity between different stages of gastric pre-malignant lesions and early GC (Zhang et al., 2019b). Sathe *et al.* reported intra- and inter- tumor transcription heterogeneity among tumor and infiltrated immune cells, as well as ligand-receptor interactions between different cell types (Sathe et al., 2020). Wang *et al.* analyzed scRNA-seq of peritoneal carcinomatosis samples from GC patients and developed a panel of genes to stratify patients for survival (Wang et al., 2019). These studies employed the 10X Genomics Chromium system, which is highly popular due to its low cost per cell and the ability to assay large number of cells per run. However, it offers a relatively low gene coverage and does not sequence full-length transcripts (See et al., 2018). Thus, a more detailed landscape of GC tumor microenvironment is still crucially missing. In our study, we employed Smart-seq2 on single cells sorted by fluorescence-activated cell sorting (FACS). This approach offers greater gene coverage and full-length transcript sequencing, and the latter enabled us to more reliably capture point mutations and even reconstruct T cell receptors.

Here, using Smart-seq2 on FACS sorted single cells, we performed scRNA-seq analysis on 15 histologically classified GC patients (i.e. diffuse, intestinal, and mixed) with different disease stages. We constructed a GC cell resource containing both tumor and immune cells, and characterized the cell types present and receptor-ligand pairs responsible for tumor-immune interactions. Using our acquired data, we not only characterized the cell states present, but also linked various cell subsets to patient survival, and predicted cell-cell interactions that help to explain the observations. Within the tumor cell compartment, we identified cancer stem-like cells that were associated with poor patient survival. By analyzing the tumor-immune interaction landscape, we found putative mechanisms by which these cancer stem-like cells suppress the immune response. We also characterized the various immune cell subtypes and compared distinct immune microenvironments of stem-like cell enriched vs non-enriched patients.

RESULTS

Dissecting gastric cancer heterogeneity at single cell resolution

Paired gastric tissue samples (namely tumor and normal adjacent) were collected from 15 GC patients, with six patients diagnosed histologically as diffuse, four as intestinal, and five as mixed (**Table S1**). Using Smart-seq2 in conjunction with index sorting, we performed single-cell full-length transcriptome sequencing and obtained raw data of 3443 cells from all patient samples (**Figure 1A**). After quality filtering, 2878 cells were retained for downstream analysis with a median of 2342 genes detected per cell (**Figure S1A**). In this study, we adopted a multimodal bioinformatics approach, to provide a holistic view of individual cells and an in-depth dissection of cellular heterogeneity. We analyzed cell of origin, differentially expressed genes (DEGs) and enriched pathways, regulon activities, genetic variants, and developmental trajectories. We also studied how the tumor-immune interactions contribute towards immune suppression and tumor survival.

Using the Seurat package (**Material and Methods**), 12 cell clusters were identified, which are visualized in **Figure 1B** using the t-Distributed Stochastic Neighbor Embedding (tSNE) (Butler et al., 2018). Each cluster was manually annotated based on the expression of known markers (**Figure 1C; Table S2**). Expectedly, we found fibroblast and epithelial cells in the sample tissues, as well as different immune cells including natural killer (NK) cells, T cells, B cells, plasma cells, and macrophages. To compare our cell clusters with published data, we integrated our data with Zhang *et al.* (Zhang et al., 2019b). Visualization of the integrated data (**Figure S1G**) shows that the majority of our data, namely NK cells, T cells, mast cells, macrophages, B cells, plasma cells, fibroblast and epithelial cells, cluster closely with matching cell types in Zhang *et al.* This demonstrates the concordance between the two datasets and thus supports the validity of our annotation.

GC epithelial cell type compositions correlate with histological classification

The epithelial cells were highly heterogeneous and formed four clusters, of which two clusters each consisted primarily of cells from patients SC008 (92%) and SC005 (87%). These patients were diagnosed clinically as the sole cases of well-differentiated Stage IA and rare mucinous adenocarcinoma, respectively. The other two epithelial cell clusters, labelled as Epithelial 1 (E1) and Epithelial 2 (E2), contained cells mainly from the remaining 13 patients diagnosed with the common types of adenocarcinoma (Hu et al., 2012). We further probed the cell type composition of these clusters and identified their cell-of-origin by using the scHCL package (Han et al., 2020).

In terms of cell of origin, the E1 cluster of cells transcriptomically resembled stomach mucosal cells (98%) and stomach pit cells (2%), while the E2 cluster was predicted to contain an even mix of stomach- and intestinal-like cells, as well as other cells (45%, 31%, and 24%, respectively) (**Figure 1D**). In terms of cell type, the E2 cluster showed a mix of stomach mucosal (25%) and stomach pit (19%) cells, while most of the intestinal-like cells were predicted to be colonic goblet cells (8%), enterocytes (5%), and enterocyte progenitors (5%). This is in line with pathology findings of intestinal goblet and enterocyte cells in intestinal metaplasia, which usually leads to intestinal (i.e. E2) type of GC (Correa et al., 2010).

The epithelial cell cluster from patient SC008 was predicted to contain mostly stomach cells (98%), but the ratio of stomach mucosal to stomach pit cells was more similar (58% and 40%, respectively) compared to the E1 cluster. Clinically, SC008 was diagnosed as well-differentiated intestinal GC with an AJCC staging of IA. The early stage and well-differentiated tumor offers a plausible explanation for the divergence. The tumor cells may be early in their transition to assume a more intestinal morphology while their whole transcriptome profile remains comparatively closer to stomach cells. In the case of the SC005 epithelial cluster, it mainly consisted of intestinal-like cells (93%) with 52% being jejunum enterocytes and 41% traverse colon enterocytes. This is a much larger proportion of intestinal enterocytes than the E2 cell cluster that has a mix of cell types with stomach and intestinal origins. This variance from the broader E2 cluster may be attributed to the SC005 tumor being a rare type mucinous adenocarcinoma. In view of the disparate cell type compositions, we relabeled the two patient specific cell clusters as E1–SC008 and E2–SC005 to reflect the patient-specific compositions.

Based on the percentage of their cells assigned to the E1/E2 clusters, we divided the patients into E1-dominant and E2-dominant groups. This division corresponds well with the histological classification, where most E1-dominant patients were diagnosed as diffuse type (3 of 5), while E2-dominant patients were a mix of intestinal and mixed histology (3 diffuse, 3 intestinal, and 2 mixed). In particular, the mix of stomach and intestinal-like cells in the E2-dominant patient samples indicates the process of intestinal metaplasia. Taken together, our data highlight the heterogeneity in epithelial cell composition among GC patients which correlate with the commonly used histological findings.

Identification of stem-like epithelial cells associated with poorer survival

We next performed a deeper characterization of epithelial cell subtypes by selecting only epithelial cells and re-clustering them. We obtained six clusters with three derived from the formerly labelled E2 cells, while the remaining three corresponded to the earlier E1, E1–SC008, and E2–SC005 clusters (**Figures 2A and S2A; Table S3**). To identify and characterize the two new sub-clusters, we first employed gene regulatory network analysis using the SCENIC workflow (Aibar et al., 2017). The tSNE computed from the regulon activity scores revealed four clusters, with most of the E1 and E2 cells falling into Regulon 0 and 1 clusters, respectively (**Figures 2B and S2B**). A significant portion of the two new sub-clusters appeared in Regulon 2 (93% and 48%), which showed higher activity in stem-related regulons. These regulons include *CDX1* extended (Jones et al., 2015), *MYC* extended (Varlakhanova et al., 2010), *HDAC1* (Zimberlin et al., 2015), *HDAC2* extended (Zimberlin et al., 2015), *RAD21* extended (Nitzsche et al., 2011), and *E2F1* extended (Yeo et al., 2011), and *EZH2* extended (Shin et al., 2012) (**Figure 2C; Table S4**). In particular, *CDX1* has been implicated in gastric epithelial cell dedifferentiation and subsequent differentiation into intestinal cells (Fujii et al., 2012), which is a possible path of intestinal metaplasia undertaken by this subset of E2 cells. Thus, we provisionally labelled these two new sub-clusters as Stem-like Epithelial A (SEA) and Stem-like Epithelial B (SEB) to reflect their stem-likeness.

Next, we examined the expression levels of known stem cell markers. Checking for known markers of normal gastric stem cells such as *AQP5* (Tan et al., 2020), *LGR5*, *BHLHA15*, *RUNX1*, *SOX2*, and *TNFRSF19* (Phesse and Sansom, 2017), their expression among the epithelial cells was minimal. We also checked *STMN1* and *MKI67*, which have been recently reported to be upregulated by actively cycling stem cells at pit-isthmus-neck region of the gastric corpus epithelium (Han et al., 2019). We found *STMN1* to be upregulated in the Stem-like Epithelial (SE) cells, while *MKI67* was not. In addition, metaplastic stem-like markers *OLFM4*, *CD44*, and *SOX4* (Zhang et al., 2019b) was found up-regulated in the SE cells. Moreover, in the tSNE plot with integrated data from our samples and Zhang et al., our stem-like cells and their metaplastic stem-like cells overlap (**Figure S2C**). Of the reported common cancer stem cell markers such as *CD24*, *CD44*

(Jaggupilli and Elkord, 2012; Takaishi et al., 2009), *PROM1* (Saha et al., 2020), *ALDH1A1* (Tomita et al., 2016), and *ALDH1A3* (Tomita et al., 2016), the stem-like cells showed comparatively higher expression of *CD24*, *CD44*, *PROM1*, and *ALDH1A1* than other epithelial cells (**Table S5; Figure S2D)**.

To further confirm the stem-likeness of these cells, we employed Gene Set Enrichment Analysis (GSEA) (Subramanian et al., 2005) with the MSigDB Hallmark v7.0 database, Short-term hematopoietic stem cell (ST-HSC), Long-term HSC (LT-HSC) (Ivanova, 2002), and stomach LGR5-positive (GSE86603) (Leushacke et al., 2017) datasets to analyze for gene sets enriched in the Stem-like Epithelial (SE) cell populations. Many stem-cell-related pathways, such as KRAS Proto-Oncogene, GTPase (*KRAS*) Signaling, WNT Beta Catenin Signaling, Notch Signaling, Hedgehog Signaling, and targets, such as G2M Checkpoint, E2F, *MYC* Targets, ST-HSC, LT- HSC, and LGR5-positive stem cell signature genes, were found to be highly enriched in the SE cells (**Figure 2D**), adding support to their stem-likeness. All other significantly enriched pathways and targets associated with SE cells can be found in **Table S6**.

We also investigated the stem-likeness of these cells using the LandSCENT package, which offers a marker-free method to estimate cell differentiation potency (Teschendorff and Enver, 2017). The computed signaling entropy scores were the highest among the SEA cells, followed by the SEB cells. (**Figures 2E and 3D**). The higher differentiation potency compared to the E1 and E2 cells provides more support to the stem-like nature of the SE cells.

We then examined the proliferative potential of the SE cells by computing the average expression of the top 500 proliferation genes in the epithelial cells (Ref: GO:0008283; **Figures 2F and 3E**). Surprisingly, the SE cells showed higher average expression levels than other epithelial cells. We also compared the proliferative properties by computing the cell cycle gene expression levels using the Seurat package. The computed G2/M phase scores of the SEA cells were also significantly higher than all the other epithelial cells (**Figure 2G**). Overall, these results indicate that the SE cancer cells were not quiescent but instead were proliferative. This is uncommon for stem cells, as adult stem cells generally remain quiescent most of the time, though a population of actively cycling stem cells was recently found in the isthmus of gastric glands within the stomach corpus (Han et al., 2019; Takaishi et al., 2009).

Currently, stem-related gene upregulation is associated with poor cancer patient survival (Bekaii-Saab and El-Rayes, 2017; Camilo et al., 2014). Therefore, we analyzed how the SE cells in our data correlate to patient survival. We first classified the TCGA GC patients (Bass et al., 2014) by their transcriptomic similarity to the different epithelial cell clusters, and then used their survival status for survival analysis (**Figures 2H and S3**). The univariate survival analysis showed that the SEA- and SEB-associated patients had poorer survival than E1- or E2-like patients in pairwise comparisons ($p < 0.05$), with the exception of E1 vs SEA. Multi-variate survival analysis adjusted by the associated cancer stages and patient age demonstrated that SEA- and SEB-like patients had poorer survival than E2-like patients (**Figure 2I**). Together, these results affirmed the negative impact of cancer stem-like cells on patient survival.

Upregulated pathways in SE cells promote cancer aggressiveness

To further characterize the different clusters of epithelial cells in our data, we obtained the cluster-specific DEGs and performed pathway analysis using the Ingenuity Pathway Analysis (IPA, Qiagen, GmbH) software (**Figures 2J and 2K; Table S7**).

In the SE cells, pro-tumor and/or drug-resistance-related pathways were found to be significantly upregulated (**Figure 2K**). In the SE cells, pathways such as the EIF2 signaling (Ali et al., 2017), gluconeogenesis I (Grasmann et al., 2019), glycolysis I, oxidative phosphorylation (OXPHOS), and mTOR signaling were upregulated. These pathways show that the SE cells were metabolically active and this behavior is linked to tumor aggressiveness (Liu et al., 2019). The dual usage of aerobic and anaerobic pathways highlights the metabolic plasticity of cancer cells, and the need to target both as an anti-cancer strategy (Jia et al., 2019). mTOR signaling can also inhibit autophagy, which helps tumor cell survival (Saxton and Sabatini, 2017), and also mediate the TGF-β signaling pathway (shown as upregulated in **Figure 2D**) that enhances cancer stem-ness and drug resistance (Francipane and Lagasse, 2016; Katsuno et al., 2019).

We also noticed the Antigen Presentation pathway (de Charette et al., 2016) to be upregulated in the SE cells (**Figure 2K**). This was supported by GSEA analysis using gene ontology (GO), which showed the enrichment of GO gene sets, namely antigen processing and presentation of peptide antigen via MHC class I (de Charette et al., 2016), antigen processing and presentation of exogenous peptide antigen via MHC class I, and antigen processing and presentation of peptide or polysaccharide antigen via MHC class II (de Charette et al., 2016) (**Figure S2G; Table S6**). The upregulation of these pathways is expected for proliferating stem cells (Agudo et al., 2018), and the MHC I pathway upregulation in particular enables them to escape destruction by NK cells. Meanwhile, the accompanying downregulation of *TAP1* and *TAP2* to avoid neoantigen presentation may help the cells to evade CD8$^+$ T cells (Ling et al., 2017) (**Figure S4A**).

DEG analysis comparing SEA and SEB (**Table S8)** showed that SEA cells had higher activations on many of the activated pathways detected earlier, particularly in the EIF2 Signaling and OXPHOS pathways (**Figures 2L and 2M; Table S9**), as well as stem-cell related gene sets, namely ST−HSC, E2F, and MTORC1 signaling (Ren et al., 2017) (**Figure S2I; Table S6**). On the other hand, SUMOylation pathway, which is involved in genome stabilization and DNA damage response during cell proliferation, differentiation, senescence, and apoptosis (Gong et al., 2017), was found to be upregulated within the SEB population; this pathway was also reported to be upregulated in colorectal cancer (CRC) stem cells (Du et al., 2016). In the SEB cells, the senescence pathway (**Figure 2M**) was found to be activated, suggesting greater senescence among the SEB cells compared to SEA. Of note, SEA cells were mainly present in the tumor tissues while SEB were primarily in normal adjacent tissues (**Figures 2A and S2A**), suggesting that these two populations are distinct.

In the E1 cells, pro-tumor and/or drug-resistance-related pathways were also found to be significantly upregulated (**Figure 2K**). For example, the Rho family GTPases signaling pathway was upregulated. In particular, the small Rho GTPase *CDC42*, which can be initialized through the integrin signaling pathway (Wiesner et al., 2005), plays a critical role in cancer malignancy by regulating cytoskeletal and microtubule dynamics, migration/invasion, epithelial mesenchymal transformation (EMT), angiogenesis, and cell–cell and cell–extracellular matrix adhesions (Mahajan and Mahajan, 2013; Maldonado and Dharmawardhane, 2018; Svensmark and Brakebusch, 2019). We also found the integrin-linked kinase (*ILK*) signaling pathway upregulated, which had been previously found to be over-expressed in 47% of gastric cancer tumor tissues (Gu et al., 2014), and promotes metastasis (Wiesner et al., 2005) and multi-drug resistance (Zheng et al., 2019). Compared to other epithelial cells, the E1 cells showed a substantial reduction in the cadherin binding expression (**Figure S2H; Table S6**). Cadherins are important for cell-cell adhesion, and its dysregulation enables tumor cell migration (Novak et al., 1998; Wu et al., 1998). We also found the *IL6* signaling and the chemokine receptor type 4 (*CXCR4*) pathways to be upregulated in E1 cells. Activation of the *IL6* pathway has been shown to promote cancer progression by mediating EMT in GC cells (Wu et al., 2017; Zhang et al., 2017; Zhu et al., 2014). Meanwhile, the *CXCR4* signaling pathway is frequently observed to be upregulated in highly invasive GC and tumor EMT (Xiang et al., 2017). Taken together, the activation of these pathways help to explain the poorer prognosis of the E1-dominant or diffuse GC patients compared to E2-dominant or intestinal GC patients (Petrelli et al., 2017; Wang et al., 2019).

Tumor trajectories revealed waves of gene regulation and dynamic immune response

We employed the Monocle 3 (Cao et al., 2019) to study the developmental transitions among the epithelial cells. In particular, we were interested in the differences between E1 and E2 cellular development, and where the SEA and SEB cells fit in the trajectories. Monocle predicted separate trajectories for the E1 and E2 clusters, highlighting the separate developmental paths for different cancer types (**Figures 3A and 3B**). We assumed the starting points of the two developmental trajectories to be the cells with the lowest tumor mutation burden (TMB; **Figure 3C**) and mean expression of cancer genes (**Figure S5D**). Both the E1 and E2 paths showed increasing trends of entropy (**Figure 3D**), proliferating gene expression (**Figure 3E**), cell cycle gene expression (**Figure 3F**), as well as expression of metaplastic stem cells (MSCs) genes extracted from Zhang *et al.* (Zhang et al., 2019b) (**Figure S5E**). Along the E2 developmental path (**Figure S5F**), we also observed the change in tissue morphology from well differentiated to poorly differentiated, reflecting the process of cancer development.

The differentially expressed gene clusters along the developmental trajectories were further analyzed using IPA (**Figures 3G and 3H; Table S10**). Along the E1 cellular development path (**Figure 3G**), three different trends in gene expression were observed. The first trend showed gradual upregulation of pathways such as regulation of actin-based motility by Rho, RhoA signaling, Cdc42 signaling, and actin cytoskeleton signaling. These pathways regulate cell morphology and motility, and their dysregulation are linked to the EMT process. RhoA, in particular, is associated with malignancy in gastric cancer, and its upregulation has been observed in diffuse GC tumors (Nam et al., 2019). We also noticed other upregulated metabolic pathways including mTOR signaling, EIF2, and OXPHOS. In particular, mTOR regulates many key cellular processes such as metabolic activities crucial for cell growth, and has been found to be activated in GC tissues (Hibdon et al., 2019; Saxton and Sabatini, 2017). EIF2, a member of the eukaryotic initiation factor family, participates in regulating protein translation and its overexpression has been detected in gastric cancers (Ali et al., 2017). The second trend was the reverse of the first, being a downregulation of pathways. The pathways in this category include death receptor signaling, retinoic acid mediated apoptosis signaling, PEDF signaling, and IL-15 production. The PEDF protein is a commonly secreted protein with anti-tumorigenic, anti-angiogenic, and anti-metastatic properties (Becerra and Notario, 2013). Downregulating of such signaling pathway reduced its impact on tumor suppression. IL-15 stimulates CD8$^+$ T and NK cells, and suppression of its production aids tumor growth (Santana Carrero et al., 2019). In the third trend, pathway upregulation was observed towards the middle of the pseudotime ordering, which was then followed by downregulation. Most of the pathways in this group are associated with immune regulation and responses. Here we speculate that these pathways were triggered during the course of tumor development, and tumor cells that were successful in shutting down those pathways, were able to survive. As observed in this study, majority of the tumor cells along this developmental trajectory underwent this immune ‘shutdown’ event.

For the E2 cellular developmental path (**Figure 3H**), one group of pathways showed continued upregulation along the trajectory. Among them, the regulation of actin-based motility by Rho pathways, mTOR, EIF2, and OXPHOS pathways, which were shared with those with the similar trend in the E1 trajectory (**Figure 3G**). We also note the sirtuin signaling pathway appearing in this category for E2. Members of the sirtuin family have shown both oncogenic and tumor suppressive properties (Carafa et al., 2019). The upregulation of sirtuins may also be connected to oxidative stress in the tumor cells, as suggested by the accompanying upregulation of OXPHOS and mitochondrial dysfunction. The second group of pathways showed the opposite trend of steady downregulation. Here, the PEDF pathway is shared with E1. Another down regulated pathway was FGF signaling, which helps promoting motility and invasion by cancer cells (Zhang et al., 2019a). We also observed the downregulation of the tryptophan degradation to 2-amino-3-carboxymuconate semialdehyde, and serotonin and melatonin biosynthesis pathways. Melatonin is a free radical scavenger with powerful anti-oxidative properties, and its downregulation could help to explain the uptrend in mitochondrial dysfunction found in the first group (Tarocco et al., 2019). The third group of pathways in E2 showed a small upregulation before a greater downregulation to below basal level at the beginning. Similar to the third group in E1, most of these pathways are immune-associated and a number of them are shared, including granulocyte adhesion and diapedesis, agranulocyte adhesion and diapedesis, primary immunodeficiency signaling, and natural killer cell signaling. Shutting down these pathways during tumor development is likely to aid immune evasion and tumor cell survival.

Tumor stem-like cells harbor more genetic mutations

As antigen presentation-related pathways are upregulated in the SE cells, we sought to compare the mutation burdens of the SE cells versus others. The SEB population with a total cell count less than 30 was excluded from this analysis. The copy number variations (CNVs) of the epithelial cells at tumor sites were predicted using the inferCNV package (Patel et al., 2014) with normal-adjacent sites as reference (**Figures 4A-C**). We then used InferCNV's HMM prediction to highlight amplified and deleted regions. Only SEA cells were found to have any duplicated or deleted regions (**Figure 4D**). In particular, roughly half of the SEA cells show high levels of gene expression amplification (> 1.5x) in chromosome 7. Other chromosomes such as 1, 5, 6, 8, and 20 also showed gene expression amplification. We also investigated Tumor Mutation Burden (TMB) in terms of point mutations. Non-synonymous variations of the coding genes were obtained using the GATK-3-4 RNASeq germline SNPs and indels workflow (see **Material and Methods**), and then annotated using the snpEFF (Cingolani et al., 2012) software. Again, SEA cells in the normal-adjacent and tumor tissues carried the highest mutation burdens (**Figure 4E**). This result is in line with reports of mutation burden being positively correlated with stem-ness in cancers (Miranda et al., 2019). In the tumor tissues, E1 cells were found to have significantly higher mutation burdens than E2 cells. We also found that E1 cells from the tumor tissues carried higher mutation burdens than their corresponding populations from the normal-adjacent tissues.

(**Figure 4F,** *p*<0.005). Overall, the mutation burden is uneven between different cell subsets, and between tumor and normal adjacent sites.

Immune microenvironment and its association with stem-like cell abundance

The immune compartment is another key component of the tumor microenvironment. Here we characterized the different immune cell states present and how they correlate with stem-like cell abundance. We first investigated the correlations between the percentage of immune cells and epithelial cells using the FACS results. We unbiasedly quantitated the live/dead $CD45^+$ and $CD3^+$ cells present in the tissues, except the first 5 patients recruited in this study, (i.e. subjects SC001-SC005), where $CD3^+$ information was not available (**Figure S6**). We found a significant positive correlation between the proportion of SE cells and percentage of $CD45^+$ among the total live cells, within both normal-adjacent and tumor tissues ($p<0.05$; **Figure 5A**). This is in contrast with reported negative correlation between stem-ness and immune infiltration into solid tumors (Miranda et al., 2019). Indeed, we observed upregulated antigen presentation pathways in the SE cells (**Figures 2K and S2G**), which may have attracted immune cells into the tumor. This suggests, despite a “hot” immune environment, the tumor cells were able to avoid immune eradication.

We next extracted the NK and T cells derived from the previous global clustering for re-clustering (**Figure 1B**) and subtype identification. Based on the expression of T cell receptor (*TCR*) genes, *CD8A*/*CD8B*, *GNLY*, *IL2RA* and *FOXP3* genes, we identified $CD4^+$ T, $CD8^+$ T, regulatory T cells (Treg), and NK cell clusters (**Figures 5B and S7B**). The cell cluster that did not express TCR but expressed *GNLY* was labelled as NK cells, while the cluster that expressed TCR and *CD8A*/*CD8B* was labelled as $CD8^+$ T cells, and the cluster that expressed *IL2RA* and *FOXP3* was labelled as Treg cells (Lee and Lee, 2018). The remaining cluster was labelled as $CD4^+$ T cells.

We then re-clustered the CD4$^+$ T, CD8$^+$ T and NK cell clusters separately for subsets. The NK cells were sub-clustered into CD16$^+$ and CD16$^-$ NKs based on CD16 expression (**Figure 5C; Table S12**), with the CD16$^+$ sub-cluster labeled as CD16$^+$ NK cells. Upon activation, CD16$^+$ NKs lose CD16 expression and become mature CD16$^-$ NK cells (Romee et al., 2013; Tsukerman et al., 2014). The CD16$^+$ NK cells upregulated *CCL4*, *KLRF1*, *CRTAM*, *EOMES*, and *SH2D1A*, while the CD16$^-$ NK cells upregulated *IL32*, *TNFRSF18*, and *GZMA*.

Analysis of the CD4$^+$ T cells identified five sub-clusters. We found one cluster of central memory T (Tcm) cells marked by the expression of *SELL* and *CCR7* (Martin and Badovinac, 2018), two resident memory (Trm) clusters marked by the expression of *CD69*, *ITGAE*, and *CD44* (Baaten et al., 2010; Cibrián and Sánchez-Madrid, 2017; Duhen et al., 2018; Mackay et al., 2013), one cluster of CD4$^+$ T cells that overexpressed *CCL5* and calmodulin, and one cluster of exhausted CD4$^+$ T cells with downregulated TCR expression (**Figure 5D; Table S12**). The TCR downregulation has been reported in CD4$^+$ T cells during TCR activation, and is thought to de-sensitize them to avoid prolonged and excessive inflammatory response (Gallegos et al., 2016; Lauritsen et al., 1998; San José et al., 2000). The exhausted CD4$^+$ T cells also significantly downregulated the EIF2 signaling pathway, which is essential for expressing pro-inflammatory cytokines (Shrestha et al., 2012) (**Table S13**). These exhausted CD4$^+$ T cells may have lost their function as CD4$^+$ helper T cells in terms of antigen recognition and cytokine production. Among the two Trm clusters, one cluster overexpressed *CCL5*, which is linked to the recruitment of CD4$^+$ T into tumor site (Huffman et al., 2020; Quandt and Dorovini-Zis, 2004); the other Trm cluster overexpressed various T cell activation markers, such as *ICOS*, *CCR6*, and *CXCR6*, and was therefore labelled as activated CD4 Trm.

The CD8$^+$ T cells all appeared to be resident memory T cells (Trm) as they expressed *CD69*, *CD44*, and *ITGAE*, but not homing markers such as *CCR7* and *LEF1*

(**Figure S7C**) (Martin and Badovinac, 2018; Steinbach et al., 2018). Sub-clustering of these cells revealed 4 subsets (**Figure 5E; Table S12**), among which two cell subsets overexpressed IFN-γ, indicating that these cells were activated and functional. These two sets of activated Trm cells differed in their expression of heat shock proteins. One subset overexpressed protein chaperones such as *HSP70*, *HSP27*, *HSP47*, *DNAJ*s, as well as polyubiquitin precursors *UBB* and *UBC*. These cells might had been stressed during the sorting process and were therefore removed from the downstream analysis. We also found two subsets of exhausted cells that overexpressed *KLRC1* and *TIM3* respectively. The presence of these exhausted T cells is consistent with a number of previous reports (Chen et al., 2020; Wu et al., 2020a; Zheng et al., 2017). To investigate whether exhausted CD8$^+$ were more enriched in tumor site, we used IHC staining on patient's tumor and normal adjacent samples and found more CD8$^+$ *TIM3*$^+$ T cells in the patient's tumor tissue compared to normal tissue (**Figure S7F**).

In our patient cohort, immune cells were primarily found in the E2 dominant patients. Here we dissected the proportion of immune cell subtypes in samples from E2 dominant patients and compared immune cell composition in SE enriched vs SE non-enriched patients. We divided the immune cells by source, normal-adjacent and tumor samples that are SE enriched (≥ 20% SE cells) or non-enriched (< 20% SE cells) (**Table 1**). Among the CD4$^+$ T subtypes, there were different proportions in SE enriched and non-enriched samples, especially in tumor tissues. Notably, exhausted CD4$^+$ T and activated CD4$^+$ Trm cells make up a larger portion of CD4$^+$ cells in SE enriched tumor samples, while the reverse holds true for CD4$^+$ Tcm and CCL5$^+$ CD4$^+$ Trm cells. For CD8$^+$ T cells, TIM3$^+$ CD8$^+$ cells form a greater proportion in SE enriched tumor samples, while KLRC1$^+$ CD8$^+$ cells form a greater percentage in non-enriched normal-adjacent and tumor samples. This suggests that SE non-enriched patients may benefit from immunotherapy via *NKG2A* (*KLRC1*) blockade by Monalizumab, which is under clinical trials (André et al., 2018), while the SE enriched patients may benefit from anti-*TIM-3* checkpoint inhibitor *MBG453* instead. The observed T cell subset proportion differences suggest differential immune microenvironment between SE enriched and non-enriched patients, and the SE cells may be a contributing factor to those differences.

We also compared the proportions of immune cell subsets in normal-adjacent and tumor tissue. The NK populations in these two compartments appeared to be a mix of $CD16^+$ and $CD16^-$ NK cells, with $CD16^-$ forming the majority in both normal and tumor tissues, 85% and 68% respectively (**Table S14**). Among the $CD4^+$ T cells, $CD4^+$ Trm (activated and $CCL5^+$) constituted a greater proportion in normal tissue than in tumors (43% vs 17%), while the trend was reversed for $CD4^+$ Tcm (8% vs 33%). Among the $CD8^+$ cells in normal tissue, 29% were activated, while 71% were exhausted (*KLRC1*$^+$ and *TIM3*$^+$). In the tumor sites, the proportion of exhausted T cells was higher at 90% compared to activated (10%). As highlighted earlier, some patient subsets may benefit from immune blockade therapy to reactivate the exhausted T cells. Within the T cell population in tumors, $CD4^+$ cells constituted 48% (306 of 642), compared to 30% (132 of 438) in normal tissues. This higher proportion of $CD4^+$ cells in GC tumor had been previously observed (Yuan et al., 2017).

Utilizing the RNA-seq reads, we reconstructed the T cell receptor sequences using TraCeR (Stubbington et al., 2016). The assembled sequences were then analyzed to quantify clonal expansion among the CD4 and CD8 T cells (**Figure 5F**). Among the different CD8 subsets in tumor tissue, 47% - 56% of each subset are unexpanded singletons. This is broadly similar to the proportion of singletons observed in other different cancer tissues (Wu et al., 2020b). In general, no statistical difference in proportions of cells with different degrees of clonal expansion was found in all CD4 and CD8 subsets, when comparing between normal-adjacent and tumor tissues.

Signals from epithelial stem-like cells contribute to immune evasion

Tumor-immune cell interaction is a crucial aspect of how tumor cells avoid immune eradication (Andrews et al., 2019). Here we used CellPhoneDB (Efremova et al., 2020) to identify the expression of receptors/ligands in different immune and epithelial cell clusters, and to predict their interactions within normal and tumor tissue micro-environment settings (**Figure 6A**). We were especially interested in the contributions by the different epithelial cell clusters towards immune evasion. The detected ligand-receptor interactions were cross-validated with published reports on them as observed in gastric and other cancers, and organized into pro-tumor, anti-tumor, immune-inducing, and immune-suppressing categories (**Figure 6B; Table S15**). SEB and CCL5$^+$ CD4 Trm with total cell count less than 30 were excluded from this interactome analysis.

In general, we found more ligand-receptor interactions between epithelial and immune cells in the tumor tissues (total interaction count of 260), compared to the normal-adjacent tissues (total interaction count of 164) (**Figure 6A**). The highest number of interactions were identified between SEA and immune cells, with a total interaction count of 89 from the normal-adjacent tissues and 128 from the tumor tissues. Specifically, of the 128 interactions found in the tumor tissues, 16 (12.5%) pro-tumor, 5 (4%) anti-tumor, 28 (22%) immune-suppressing, and 13 (10%) immune-inducing interactions were detected (**Figure 6B**). Between the E1 and immune cells, 45 and 91 interaction pairs were identified in the normal-adjacent and tumor tissues, respectively, with more pro-tumor interactions detected from the tumor (19) compared to that from the normal-adjacent tissues (5). In contrast, the lowest ligand-receptor interaction counts were identified between E2 and immune cells, with total interaction count of 30 in the normal-adjacent tissues and 41 in the tumor tissues. Finally, neither immune-inducing nor –suppressing interactions were detected between the non-SE E2 and immune cells from both normal-adjacent and tumor tissues (**Figure 6B**).

Among the epithelial cells, the SEA cells were highly engaged in immune-inducing interactions (11 and 13 in normal-adjacent and tumor tissues respectively). In contrast, the

E1 cells only had 4 of such interactions in tumor tissues, while E2 had none. This may partly explain the positive correlation between the abundance of SE cells and CD45$^+$ immune cells in the normal-adjacent and tumor tissues (**Figure 5A**). However, the SE cells were engaged in more immune-suppressing than immune inducing interactions with the different immune cell types. The SEA cells had 28 immune-suppressing and 13 immune-inducing interactions in normal adjacent tissue, and 25 immune-suppressing versus 11 immune-inducing interactions in tumor tissue. While both inducing and suppressive interactions were present, the suppressive interactions were more numerous. In view of the poorer survival of patients that were transcriptomically similar to the SE cells, it is likely that these suppressive interactions were efficacious and contribute to the poor survival.

Delving into the interactions, four pro-tumor interactions, *LTBR-LTB*, *CXCL16-CXCR6*, *TNFRSF1B-TNF*, and *CEACAM6-CEACAM5*, were found only among the SEA cells' interactions with immune cells. In particular, the Lymphotoxin Beta Receptor (*LTBR*) bearing tumor cells, when interacting with the *LTB* protein on lymphocytes, can promote tumor angiogenesis and metastasis (Das et al., 2019; Hehlgans et al., 2002). In our data, we found this interaction to be specific to SEA tumor cells with CD4Tcm and Activated CD4Trm cells (**Figure 6B).** The *CXCR6-CXCL16* interaction was only observed between the T and SE cells, and is of interest due to potentially opposing roles; this interaction can positively or negatively regulate metastasis and invasion of cancer, arising from the two forms of *CXCL16* molecules. The trans-membranous *CXCL16* can induce an anti-tumor immune response, while the soluble *CXCL16* enhances tumor proliferation and migration (Gooden et al., 2014; Meijer et al., 2008). Another identified interacting pair, the Tumor Necrosis Factor Receptor Superfamily, Member 1B (*TNFRSF1B-TNFR2*) expressed by a tumor cell, when bound by the TNF, can result in TNF receptor-associated factor 2 recruitment and activate the nuclear factor-κB pathway that resists TNF-induced tumor cell death (Rothe et al., 1995). The activated *TNFRSF1B* was previously found to mediate Granulin Precursor (*GRN*) expression in CRC, which in turn regulates *MKi67* and Vascular Endothelial Growth Factor A expression, leading to tumor proliferation and angiogenesis (Yang et al., 2015). Lastly, the normal adjacent tissue compartment of our data contained the interacting pair of *CEACAM6* expressed by the SEA cells and *CEACAM5* on exhausted CD4T cells. The expression of CEA Cell Adhesion Molecule 6 (*CEACAM6*) was previously reported to positively correlate with poorer prognosis and metastasis in multiple cancer types, including CRC and GC (Blumenthal et al., 2007; Pandey et al., 2019; Ru et al., 2017).

Among the ligands in tumor-immune interactions, Galectin-9 (*LGALS9*), a carbohydrate-binding protein involved in T-cell homeostasis (Wiersma et al., 2013), was found to be highly expressed in SEA cells from both normal and tumor tissues, and E1 cells from tumor tissues (**Figures 6A and 6B**). Contrasting roles have been described for Galectin-9 on anti-tumor immunity. On one hand, *LGALS9* can enhance T helper 1 type antitumor immunity (Dai et al., 2005), while the binding of *LGALS9* to CD44 (*LGALS9-CD44*) also promotes T-regulatory cell differentiation (Wu et al., 2014). However, *LGALS9* binding to its receptor, T cell immunoglobulin- and mucin domain-containing-3 (*TIM3*) expressed on activated T cells (*LGALS9-TIM3*), causes T cell dysfunction and apoptosis (Li et al., 2012). *LGALS9* also inhibits NK cell chemotaxis to the tumor microenvironment (Wang et al., 2016) and participates in suppressing NK activities through its interaction with *TIM3*-bearing NK cells (Golden-Mason et al., 2013; KOMITA et al., 2015; Li et al., 2016; Vento-Tormo et al., 2018). In addition to *LGALS9-TIM3*, other ligand-receptor interactions such as *GRN-TNFRSF1B*, *HLA-C-KIR2DL3*, and *NECTIN2-TIGIT*, can also suppress NK cells, especially in tumor tissues (**Figure 6B**). These NK suppressive effects were likely to be intensified among SE and E1 cells that highly expressed the *GRN*, Major Histocompatibility Complex, Class I, C (HLA-C), and Nectin-2 proteins, especially in tumor tissues (**Figure S8**). These proteins can bind to the *TNFRSF1B* (Tang et al., 2011; Xu et al., 2014), Killer Cell Immunoglobulin Like Receptor, Two Ig Domains And Long Cytoplasmic Tail 3 (*KIR2DL3*) (Fadda et al., 2010), and T cell immunoglobulin and ITIM domain (*TIGIT*) inhibitory receptors (Golden-Mason et al., 2013; Gonçalves Silva et al., 2017; KOMITA et al., 2015), respectively, to effectively suppress the NK anti-tumor response. *TIGIT* is also known to be expressed on $CD8^+$ T cells, and its expression often marks T cell dysfunction when *NECTIN2* is bounded to the *TIGIT* receptor (Chauvin et al., 2015). Blocking the effect of *TIGIT* is being tested as a therapy for solid tumors (Harjunpää and Guillerey, 2020).

As the predictions were computed based on mRNA expression and without spatial information, we employed multiplex immunohistochemistry staining for validation. Among the interacting pairs predicted, the *NECTIN2-TIGIT* pair was of significant interest, due to the interaction being specific to SEA cells that were $STMN1^+$. The IHC panel includes EPCAM, STMN1, NECTIN2, TIGIT, and CD8. The image showed that $EPCAM^+$ $STMN1^+$ stem-like cancer cells could be detected in the tumor region (**Figures 6C and S9**). To further investigate the interactions between the stem-like cancer cells and the immune cells, we zoomed into a region where both cell types were in close proximity. We found the stem-like cancer cells were in close proximity with *TIGIT* expressing $CD8^+$ T cells. Notably, the stem-like cancer cells also expressed NECTIN2, which could suppress the nearby T cells via their TIGIT receptors.

DISCUSSION

While dietary changes and lower *H. pylori* incidence have helped to reduce the incidence of GC, it remains a disease with high incidence rates, particularly in East Asia. Immunotherapy has gained prominence in recent years as a promising approach in cancer therapy, but has not been a resounding success in treating GC. To harness the body's immune system against GC, a clear understanding of immune evasion by GC is needed. Here, we deployed single cell RNA-seq to characterize cancer and immune cells in the tumor environment of different GC types, and mapped out the predicted interactions between them.

Our scRNA-seq of GC tissues revealed a heterogeneous collection of epithelial cells of different types and origins. The predicted composition of each sample appeared to mirror the histological diagnosis. E1 cell dominant patients were largely diagnosed as diffuse, while E2 cell dominant patients' samples had the histological appearance of intestinal tissue. In the single-cell study on peritoneal metastasis of GC, Wang *et al.* (2019) stratified patients into two groups: gastric type wherein the peritoneal carcinomatosis cells resembled stomach cells, and GI-mixed type wherein the tissue contained cells that resembled intestinal and other non-gastric cells types from the GI tract (Wang et al., 2019). Based on available clinical data, the gastric type patients had poorer prognosis than GI-mixed patients. Similarly, our study also showed that shorter survival was found among the E1 dominant (diffuse type) patients. We then probed the differences between E1 and E2 cells. In the DEG analysis, E1 cells showed a number of pro-tumor and/or drug-resistance-related pathways to be upregulated. In particular, EMT-related pathways, including *CDC42*, *CXCR4*, *IL6,* and *ILK* signaling, were highly upregulated (Cheng et al., 2018; Mao et al., 2019; Sun et al., 2018; Wu et al., 2017; Zhang et al., 2017; Zhu et al., 2014). The enhanced motility and invasiveness of the cancer cells acquired through EMT is critical in the initiation of metastasis for cancer progression, and is strongly associated with poor disease prognosis in both colorectal (Loboda et al., 2011) and gastric cancers (Cristescu et al., 2015).

One of the key findings in this study is our discovery of stem-like epithelial (SE) cells which clustered close to the E2 cells. Initial characterization of these cells by regulon analysis showed upregulation of stem-related transcription factors. Though we did not find the upregulation of gastric stem cell markers such as *AQPR5*, *LGR5*, and *MKI167*, they instead had higher expressions of cancer stem cell markers like *CD24*, *CD44*, *Prominin 1*, and *ALDH1A1*. We also found that these SE cells had higher differentiation potency and proliferation gene expressions than other epithelial cells. Upregulated gluconeogenesis, glycolysis, and mTOR signaling pathways also indicated that the SE cells were metabolically active. In a recent report by Han et al (Han et al., 2019), actively cycling stem cells were found in the isthmus of gastric glands. These cells are responsible for pit-isthmus-neck region maintenance, and expressed *STMN1* and *MKI67*. Interestingly, our SE cells also upregulated *STMN1* but not *MKI67*. One possible avenue of inquiry is to determine if these gastric isthmus stem cells are possible progenitors of the SE cells. In another recent publication, Zhang *et al*. reported MSCs in their pre-GC lesion samples (Zhang et al., 2019b). These MSCs also showed upregulation of *OLFM4, CD44*, and *SOX4* like in the SE cells, and furthermore, data integration of the datasets showed that the SE cells and MSCs overlapped. Zhang *et al.* also speculated that their MSCs were the source of GC cells. However, the SE cells in this study had the greatest tumor mutation burden, and in our trajectory analysis, mainly appeared at the ending points of the predicted E1 and E2 developmental trajectories. It appears that the SE cells may be a product of the EMT process, as the up-regulation of Rho (Bendris et al., 2012), mTOR (Gulhati et al., 2011), and EIF2 (Feng et al., 2014) pathways was traced along the trajectories. Further investigation is needed to determine their similarities with other reported stem-like cells and the roles they might play in GC tumors.

We also found the SE cells to be significantly associated with poorer disease prognosis, compare to E1 and E2 cells. From our data, a number of possible explanations may be found. Firstly, the SE cells showed higher mutation burdens and expression of proliferation genes than other epithelial cells. Secondly, pro-tumor pathways were upregulated in the SE cells. Pathways such as EIF2 that controls protein translation, and the mTOR pathway that regulates cell growth; their upregulation contributes towards cell proliferation (Ali et al., 2017; Hibdon et al., 2019; Saxton and Sabatini, 2017). We also found the MHC I pathway to be upregulated, which enabled avoidance of NK cell killing, while the low expression of *TAP1* and *TAP2* reduced neoantigen processing for MHC presentation and thus improved survival against T cells. Nevertheless, we found that the proportion of stem-like cells in a normal-adjacent/tumor tissue to be positively correlated with the proportion of host immune cells in the tissue, suggesting that the antigen presentation pathways were not completely dysfunctional.

Analysis of the epithelial – immune cell interactions painted an interesting picture. SEA cells were more heavily involved in interactions than the E1 or E2 cells. In particular, the SEA cells were involved in more immune interactions, especially immune-suppressive interactions, in both normal and tumor micro-environment settings. Some other interactions include *LTBR-LTB* and *TNFRSF1B-TNF*, which have pro-tumor effects. The greater number of immune-suppressive interactions by SE cells is a possible explanation for their correlation with poor patient prognosis, despite their tendency to attract immune cells. Our analysis also highlighted interactions that are currently under investigation for immunotherapy. For instance, the blockade of the immune checkpoint mediated by *TIGIT*/Nectin-2 (Figueroa-Protti et al., 2019; Oshima et al., 2013; Stamm et al., 2018) and/or *Tim-3*/Galectin-9 (Chen et al., 2019; He et al., 2018) interactions are currently under investigation. The treatments can potentially re-activate NK and/or T cell activities in GC patients, especially SE enriched patients among the GI-mixed/E2-dominant patients.

Our single cell dissection of the immune compartment found the diverse range of immune cells present in both normal and tumor tissues. T cells, both cytotoxic and at various levels of exhaustion, formed the bulk of immune cells sequenced. Dissection of T cell subpopulations showed that a greater proportion of $CD8^+$ T cells in SE enriched tumors expressed the TIM3 marker, highlighting the opportunity for blockade therapy by *MBG453*. On the other hand, SE non-enriched patients show comparatively more $CD8^+$ T cells with the KLRC1 marker, which may benefit from immunotherapy via NKG2A (KLRC1) blockade by Monalizumab instead.

In conclusion, we dissected the heterogeneity of epithelial and immune cells in GC tumors at single cell level. We found stem-like epithelial cells that correlated with poor patient prognosis. These stem-like cells were more proliferative than other cancer cells, and were more engaged in cell-cell interactions with the immune system, both stimulatory and suppressive. This study provides an essential resource on gastric cancer tissue interacting with the immune compartment at the single cell level, and can help suggest new cell-specific tumor immune escape mechanisms for potential checkpoint inhibitor therapy development for GC.

MATERIALS & METHODS

LEAD CONTACT AND MATERIALS AVAILABILITY STATEMENT

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Jinmiao Chen (chen_jinmiao@immunol.a-star.edu.sg). Single cell RNA-seq data generated in this study have been deposited to GEO (GEO: GSE153745).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Ethics Statement

The research activities involved in this study were reviewed and approved by the local institutional ethics board (National Healthcare Group Domain Specific Review Board: B/2016/00059). All tissue samples were obtained with written informed consent from all participants.

Patient samples

Matching tumor-edge and normal tissues were collected prospectively from 15 GC patients who visited the National University Hospital in Singapore between July 2016 and May 2017. All tissue samples were obtained from surgical specimens immediately following removal from patients undergoing primary surgery, with no pre-operative irradiation or chemotherapy.

Briefly, six to eight surgical bites (each measuring approximately 5 mm in size) on each of the fresh gastric tumor and normal-adjacent tissues were collected from the patients during planned surgical resection. The bites were preferably obtained from the periphery of the resected tissue whenever possible, in order to capture region known to contain higher proportion of immune cells. Gastric adenocarcinoma was confirmed by histological examination and then classified according to Lauren's criteria into diffuse and intestinal types. Ethics-approved demographic and clinical characteristics were collected from all cases (**Table S1**).

METHOD DETAILS

Tissue dissociation and single cell sorting

All fresh tissues were first collected in cold FACS buffer (5% FBS and 2 mM EDTA in PBS) and then rinsed with cold HBSS (Sigma-Aldrich, Saint Louis, MO). After mincing, the tissues were digested for 1-2 hours at 37 °C in 300 U/ml collagenase (Sigma-Aldrich) and 100 U/ml hyaluronidase (Sigma-Aldrich). Thereafter, the cell suspension was filtered and the cells were pelleted at 300 x g for 5 minutes at 4 °C. Cells were resuspended and washed twice in cold FACS buffer. Cell viability was determined using trypan blue (Invitrogen, Carlsbad, CA). Cells were again pelleted and subsequently resuspended in a cold blocking buffer [composition withheld by SIgN]. An antibody cocktail was then added and incubated for 20 minutes on ice before FACS sorting. The stained cells were then sorted using the BD FACSAria III 3-Laser system and subjected to flow cytometry analysis. Stratified FACS-sorting was conducted to allow 1:1 ratio of EpCAM$^+$ and CD45$^+$ cells collection in the 96 well plates.

Single cell RNA sequencing

The single cells were index-sorted into 96-well plates and subjected to single cell RNA sequencing using the Smart-Seq2 protocol (Picelli et al., 2014) the following modifications: 1. 1mg/ml BSA Lysis buffer (Ambion® ThermoFisher Scientific, Waltham, MA, USA); 2. Use of

200 pg cDNA with 1/5 reaction of Illumina Nextera XT kit (Illumina, San Diego, CA, USA). The length distribution of the cDNA libraries was monitored using a DNA High Sensitivity Reagent Kit on the Perkin Elmer Labchip (Perkin Elmer, Waltham, MA, USA). All samples were subjected to an indexed paired-end sequencing run of 2x151 cycles on an Illumina HiSeq 2000 and HiSeq 4000 system for about 1 million reads per sample.

Single cell RNA raw data pre-processing

Reads from single-cell RNA-sequencing were mapped to the GRCh38 human reference genome GENCODE release 25 using RSEM version 1.3.0 (Dobin et al., 2013), and transcript per million reads (TPM) values were calculated by the same RSEM (Li and Dewey, 2014).

Seurat analysis and cell type identification

The Seurat version 3.1.3.9000 (Butler et al., 2018) was used for downstream analysis. Initial quality control was performed to remove cells with high mitochondrial activity (> 50%), retaining 2,878 cells likely to be viable. We then performed SCTransform of the gene expression values. This was followed by feature selection, where the top 12,000 highly variable genes were selected for further processing. Principal component analysis (PCA) was performed and the top 20 PCs were used to generate tSNE and UMAP as well as cell clustering. Unsupervised clustering was performed at resolution = 0.8 to cluster all the cells. To annotate the clusters, we computed the differentially expressed genes (DEGs) among the clusters using the LogNormalized values and the FindAllMarkers function with logfc.threshold = 0 and test.use = "bimod". Pathways enriched by DEGs were identified using Ingenuity Pathway Analysis (IPA) software. When analyzing the epithelial cells, the same process was applied except for the significant PCs, of which PCs 1 to 9 were used, and cell clusters were identified at resolution = 0.4.

To aid confirmation of our cell type annotation, we examined FACS data of 698 cells from obtained from Patient SC016, SC017 and SC020 (**Table S15**). All 616 except 7 (99%) of the annotated B, NK, plasma, and T cells were confirmed as immune cells by the expressed CD45 marker. Of 21 annotated B, 19 (90%) expressed CD19 marker. Of 106 annotated CD4T, 74 (72%) expressed CD3, 33 (31%) expressed CD4, 18 (17%) expressed CD8, 7 (6%) expressed CD56, and 9 (8%) co-expressed CD3 and CD56 markers. Of 15 annotated Treg, 10 (67%) expressed CD3, 6 (40%) expressed CD4, 1 (7%) expressed CD8, 2 (13%) expressed CD56, and 2 (13%) co-expressed CD3 and CD56 markers. Of 251 annotated CD8T, 219 (87%) expressed CD3, 120 (48%) expressed CD8, 29 (12%) expressed CD4, 6 (2%) expressed CD56, and 15 (6%) co-expressed CD3 and CD56 markers. Of 220 annotated NK, 50 (23%) expressed CD56, 14 (6%) expressed CD3, 3 (1%) expressed CD4, and 1 (0.5%) expressed CD8 markers. All 56 but 1 (98%) of the annotated epithelial cells expressed the epithelial cell surface marker EpCAM. Lastly, 23 (88%) of the 26 annotated fibroblast cells expressed the EpCAM marker.

Datasets integration with Zhang et al.

We used the Seurat package v3.1 for the integration of our SMART-Seq2 and Zhang *et al.* (Zhang et al., 2019b). data. Both datasets were first log normalized using the LogNormalize function, and 12000 most variable genes were then identified using the FindVariableGenes function. To integrate the datasets, cell anchors in the datasets were found using the FindIntegrationAnchors function and subsequently used for batch integration with the IntegrateData function. The integrated gene expression matrix was then scaled such that each gene had the same mean and standard deviation. To visualize the integrated data, dimension reduction with PCA followed by tSNE was performed and the output plotted. We annotated the cells in Zhang *et al.*’s data using the gene markers published in the accompanying supplementary file.

Cell of origin analysis

For cell of origin prediction, we used the scHCL version 0.1.1 https://github.com/ggjlab/scHCL/). The scHCL function was used without any custom settings. The prediction results were then processed manually to merge cell types with different markers.

Regulon analysis

AUC (Area Under the Curve) scores of regulon activities were calculated for all epithelial cells using the SCENIC (Single Cell rEgulatory Network Inference and Clustering) (Aibar et al., 2017) R package version 1.1.2-2 with default parameters. The SCENIC analysis returned AUC scores of 624 regulons which were further analyzed using the Seurat version 3.1.2. The LogNormalize and feature selection steps in Seurat pipeline were skipped. Original AUC scores of all 624 regulons were directly used for PCA analysis, and the top 10 PCs used to generate the tSNE plot. Four clusters of epithelial cells were identified using the top 10 PCs and resolution = 0.4. Differentially enriched regulons between clusters were identified using the Wilcoxon test.

Mouse-derived Lgr5-positive gene set

Using the GEO dataset GSE86603, differential gene expression analysis was performed by comparing three samples of uninjured mice Lgr5-GFPhigh corpus cells against three samples of uninjured mice Lgr5-GFPneg corpus cells (Leushacke et al., 2017). The six .cel files were first downloaded from the GEO database and pre-processed using the rma function in the oligo R package (version 1.46.0). Then, the mogene20sttranscriptcluster.db R package (version 8.7.0) was used to annotate the mouse data with their respective Entrez gene IDs and gene symbols. Probes that could not be mapped to a gene ID/symbol were removed and rows with the same gene ID were aggregated to produce mean expression values. Using the limma R package (version 3.38.3), differentially expressed genes (DEGs) between Lgr5-GFPneg and Lgr5-GFPhigh corpus cells were identified (Logfc > log2(1.5), p value < 0.05). Next, mouse DEGs were then mapped to their respective human homologues using the HomoloGene database (build 68) from NCBI. Of the 2954 human DEGs mapped, the top 300 DEGs (Logfc > 1.55, $p < 0.05$) were included as an Lgr5-positive gene set for the subsequent Gene Set Enrichment Analysis (GSEA).

Gene Set Enrichment Analysis

GSEA version 4.0.2 (Subramanian et al., 2005) was performed using the Hallmarks v7.1 and C5–GO Biological Processes v7.1 gene sets available from the MSigDB database, long-term Hematopoietic Stem Cell (LT-HSC), short-term HSC (ST-HSC) gene sets (Ivanova, 2002), and the Lgr5-positive gene sets derived above. All default settings of the GSEA software was used.

Single-cell entropy analysis

To quantify single-cell entropy, we used the LandSCENT version 0.99.3 (Teschendorff and Enver, 2017). First, we log-nomalized our scRNA-seq count table using the normalize function from scater package as recommended by the LandSCENT package. We then used the DoIntegPPI function to integrate the normalized data with the human protein-protein interaction network (net13Jun12.m) defined by the package. Finally, the CompSRana function was used to compute the signaling entropy rate and estimated differentiation potency.

Top 500 proliferation and cell cycle genes

We first extracted the expression levels of the human proliferation (GO:0008285, 1061 genes) and cell cycle (GO:0007049, 761 genes) genes, and summed their respective expression across all cells to rank them. The top 500 highly expressed genes were then selected and used to compute the sum of expression for each cell.

Epithelial-based classification of TCGA tissues

In order to classify the TCGA GC tissues based on the gene expression similarity with the epithelial clusters identified in this study, the TCGA and our SMART-Seq2 (including all epithelial and non-epithelial cells) datasets were integrated using the Seurat package (version 3.1.0) (Stuart et al., 2019). Adhering to the recommended integration workflow, we first normalized and scaled each dataset with the Seurat SCTransform normalization (Hafemeister and Satija, 2019), and selected 3000 HVGs as input to identify the 'anchors' for datasets integration. Upon data integration, a global clustering including both single cell and TCGA samples was performed to identify TCGA patients that clustered with the epithelial clusters. Then, a total of 125 epithelial-like TCGA samples were selected to re-cluster with only the SMART-Seq2 E1, E2, SEA, and SEB cells, to suggest E1-, E2-, SEA- and SEB-like tissues from the TCGA dataset.

Survival analysis of SEA and SEB like TCGA patients

Overall survival curves were estimated by the Kaplan-Meier method and compared using the log-rank (Mantel-Cox) test. Patients who were still alive at last follow-up were censored. For multivariate analysis, the Cox proportional hazards regression model was used to relate risk factors, considered simultaneously, to overall survival time. Tests of hypothesis were used to assess whether there are statistically significant associations between risk factors and time to event. For the age risk factor, which is a continuous variable, the cutp function from survMisc R package was used to determine the optimal cut-off point. The parameter estimates represent the change in the expected log of the hazard rate relative to a one-unit change in the risk factor, holding all other risk factor constant. They were generated using coxph function from survival R package and are shown with their p-values and their associate hazard rates along with their 95% confidence intervals. Two-tailed p-values that were less than 0.05 were considered to be statistically significant for all hypothesis testing conducted in this study.

Trajectory inference

First, we winsorized the TPM data of the epithelial cells with an upper limit of 97.5% on a per cluster basis. Next, the winsorized TPM was subjected to the Monocle 3 analysis workflow (Cao et al., 2019). The gene expression values in the TPM table were log-transformed and normalized by the size factor using the data pre-processing functions available in Monocle. We then performed PCA and projected the cells onto a 2-dimensional space encoding transcriptional state using the Uniform Manifold Approximation and Projection (UMAP) (McInnes et al., 2018) with the first 30 principal components and default parameters setting of n_neighbors = 15 and min_dist = 0.1. We then used the Leiden method (Traag et al., 2019) with a resolution equals to 0.1 to detect cell clusters. This resulted in four partitions, which were represented by the 4 ‘superclusters’ of the Leiden communities, found using the K-Nearest-Neighbor pruning method (Levine et al., 2015). For each partition, a cell trajectory was drawn atop the projection using the Monocle reversed graph embedding algorithm.

For the pseudotime ordering, we took the cells with the lowest TMB as the starting point for the two main partitions (one partition included Epithelial 1 cells, and the other included Epithelial 2 cells and Stem-like Epithelial cells). To find the genes that were differentially expressed along these two trajectories, we used the Monocle graph_test function to compute a measure of spatial autocorrelation, the Moran’s I. A statistic equals to +1 means that nearby cells on a trajectory have a perfect similar expression for the gene being tested, while 0 represents no correlation. We selected the genes with a Moran’s I higher than 0.2 for the Epithelial 1 partition (263 genes), and 0.25 for the Epithelial 2/Stem-like Epithelial partition where we removed the Stem-like Epithelial B cells (1106 genes). Then, we used the pheatmap function (R package) to visualize these DEGs, and clustered the genes using to the hierarchical clustering algorithm.

Tumor Mutation Burden

Non-synonymous variations of the coding genes were obtained using the GATK-3-4 RNASeq germline SNPs and indels workflow, using its default execution options and reference files stated in the reference json file (https://github.com/gatk-workflows/gatk3-4-rnaseq-germline-snps-indels). The variants found were annotated using the snpEFF version 4.3T software (Cingolani et al., 2012), with a genomic annotation database re-built using the same reference genome (GRCh37, Homo_sapiens_assembly19_1000genomes_decoy.fasta) and its annotation GTF (star.gencode.v19.transcripts.patched_contigs.gtf) as the GATK workflow. The total genome coverage with minimum 4 pass-filter reads were surveyed using the GATK CallableLoci module. The tumor mutation burden (TMB) was derived by normalizing the total pass-filter non-synonymous variants that had minimal depth of 4 reads (counts) with total genome coverage (per Mb) and total number of sequencing reads (per million reads).

Immune cell clustering

Only cells identified as T and NK cells in the global clustering were re-clustered for analysis using the Seurat package (v3.1.3). FACS data from Patient SC016, SC017 and SC020 were used to determine $CD4^+$ and $CD8^+$ T cells from $CD45^+$ cells and these information were used as the ground truth to guide separation among $CD4^+$ Ts, $CD8^+$ Ts and NK cells. Cell clusters that lacked TCR expression and expressed GNLY were labelled as NK cells; cell clusters that overexpressed CD8A and CD8B were labelled as $CD8^+$ T cells; the cell cluster that overexpressed CD4 and FOXP3 was labelled as Treg cells; all other cell clusters were labelled as $CD4^+$ T. After confirming with the FACS-based ground truth, one cluster was found to contain both $CD4^+$ Ts and $CD8^+$ Ts. This cluster was sub-clustered to separate the $CD4^+$ Ts and $CD8^+$ Ts. Following that, each of the $CD4^+$ T, $CD8^+$ T and NK cell clusters were further re-clustered separately to further identify sub-clusters within each cluster.

Reconstruction of T cell receptor

T cell receptor sequences were reconstructed for all cells using the TraCeR version 0.6.0 (Stubbington et al., 2016) assemble function with “-s Hsap --loci A B G D” as input parameters. Clonal type and size were summarized on a per patient basis using the TraCeR summarize function with “-s Hsap --loci A B G D” parameters. The clonal size was overlaid on a tSNE plot to visualize T cell expansion.

Cell-cell communication analysis

The CellPhoneDB 2 (Efremova et al., 2020) was used to analyze ligand and receptor expressions in cells to predict cell-cell communications. The TPM table of the sequenced cells was analyzed using the cellphonedb method with the default parameters, except that the minimum fraction of cells in a cluster expressing a gene was set to 0.3.

Statistical analyses

The two-tailed Wilcoxon rank-sum test was used to compare nonparametric continuous variables, respectively. The Spearman’s Rank correlation was used to evaluate the strength and correlation (negative or positive) of a relationship between two non-parametric continuous variables. Two-tailed *p* values that were less than 0.05 were considered to be statistically significant. The graph plotting and statistical analysis were performed using the ggplot2 and ggpubr in R statistics packages or the matplotlib and SciPy in Python libraries.

Multiplex immunohistochemistry

Multiplex immunohistochemistry was performed on SC004, SC007, and SC018 tumor slides using the Opal 7-colour kit (Akoya Biosciences, CA, US). Four micrometers of FFPE sections were dewaxed and rehydrated. The images were acquired using the Vectra 3.0 Automated Quantitative Pathology Imaging System (Perkin Elmer) with 4’,6-diamidino-2-phenylindole (DAPI) as the nuclear marker as previously described (Garnelo et al., 2017).. The antibodies used were anti-Human Stathmin (clone D1Y5A; Cell Signaling), anti-Human CD8 (clone C8/144B; DAKO), anti-Human TIGIT (clone 3106; Abcam), anti-Human Nectin-2 (clone D8D3F; Cell Signaling), and anti-Human EpCAM (clone VU-1D9; Invitrogen).

Author contribution

Jinmiao Chen conceptualized and supervised the study. Kok Siong Ang, Hong Kai Lee, Jingjing Ling and Jinmiao Chen wrote the manuscript. Jimmy So and Wei Peng Yong recruited patients, collected samples and provided critical insights. Vivien Koh dissociated the tissues. Calista Wong, Mai Chan Lau, Ze Ming Lim, Charles Antoine Dutertre, Anis Larbi, and Jinmiao Chen performed FACS staining and index sorting. Michelle Goh, Alicia Tay, Josephine Lum, and Shanshan Wu ran single-cell smart-seq2. Jinmiao Chen, Marion Chevrier, Hong Kai Lee, Kok Siong Ang, Xiaomeng Zhang, Jingjing Ling, Jia Chi Tan, Nicole Lee, and Carolyn Tang performed data analysis. Jia Chi Tan, Hong Kai Lee, Xiaomeng Zhang generated the figures. Sergio Erdal Irac, Chun Jye Lim, Sherlly Lim and Joe Poh Sheng Yeong performed IHC staining. Samuel Chuah Wen Jin and Valerie Chew performed gating of FACS data. Michael Poidinger, Amit Singhal, Matthew Ng, Shanshan Wu, Patrick Tan, and Florent Ginhoux provided scientific advices.

Acknowledgement

We would like to thank NMRC OFYIRG grant for funding this study; SIgN Immunomonitoring platform for running FACS and single-cell RNA-seq. SIgN Immunomonitoring platform is supported by a BMRC IAF 311006 grant and BMRC transition funds #H16/99/b0/011.

References

Agudo, J., Park, E.S., Rose, S.A., Alibo, E., Sweeney, R., Dhainaut, M., Kobayashi, K.S., Sachidanandam, R., Baccarini, A., Merad, M., et al. (2018). Quiescent Tissue Stem Cells Evade Immune Surveillance. Immunity *48*, 271-285.e5.

Aibar, S., González-Blas, C.B., Moerman, T., Huynh-Thu, V.A., Imrichova, H., Hulselmans, G., Rambow, F., Marine, J.-C., Geurts, P., Aerts, J., et al. (2017). SCENIC: single-cell regulatory network inference and clustering. Nat. Methods *14*, 1083–1086.

Ali, M.U., Ur Rahman, M.S., Jia, Z., and Jiang, C. (2017). Eukaryotic translation initiation factors and cancer. Tumor Biol. *39*, 101042831770980.

André, P., Denis, C., Soulas, C., Bourbon-Caillet, C., Lopez, J., Arnoux, T., Bléry, M., Bonnafous, C., Gauthier, L., Morel, A., et al. (2018). Anti-NKG2A mAb Is a Checkpoint Inhibitor that Promotes Anti-tumor Immunity by Unleashing Both T and NK Cells. Cell *175*, 1731-1743.e13.

Andrews, L.P., Yano, H., and Vignali, D.A.A. (2019). Inhibitory receptors and ligands beyond PD-1, PD-L1 and CTLA-4: breakthroughs or backups. Nat. Immunol. *20*, 1425–1434.

Baaten, B.J.G., Li, C.-R., and Bradley, L.M. (2010). Multifaceted regulation of T cells by CD44. Commun. Integr. Biol. *3*, 508–512.

Balakrishnan, M., George, R., Sharma, A., and Graham, D.Y. (2017). Changing Trends in Stomach Cancer Throughout the World. Curr. Gastroenterol. Rep. *19*, 36.

Bass, A.J., Thorsson, V., Shmulevich, I., Reynolds, S.M., Miller, M., Bernard, B., Hinoue, T., Laird, P.W., Curtis, C., Shen, H., et al. (2014). Comprehensive molecular characterization of gastric adenocarcinoma. Nature *513*, 202–209.

Becerra, S.P., and Notario, V. (2013). The effects of PEDF on cancer biology: mechanisms of action and therapeutic potential. Nat. Rev. Cancer *13*, 258–271.

Bekaii-Saab, T., and El-Rayes, B. (2017). Identifying and targeting cancer stem cells in the treatment of gastric cancer. Cancer *123*, 1303–1312.

Bendris, N., Arsic, N., Lemmers, B., and Blanchard, J.M. (2012). Cyclin A2, Rho GTPases and EMT. Small GTPases *3*, 225–228.

Blumenthal, R.D., Leon, E., Hansen, H.J., and Goldenberg, D.M. (2007). Expression patterns of CEACAM5 and CEACAM6 in primary and metastatic cancers. BMC Cancer *7*, 2.

Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R.L., Torre, L.A., and Jemal, A. (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA. Cancer J. Clin. *68*, 394–424.

Butler, A., Hoffman, P., Smibert, P., Papalexi, E., and Satija, R. (2018). Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat. Biotechnol. *36*, 411–420.

Camilo, V., Barros, R., Celestino, R., Castro, P., Vieira, J., Teixeira, M.R., Carneiro, F., Pinto-de-Sousa, J., David, L., and Almeida, R. (2014). Immunohistochemical molecular phenotypes of gastric cancer based on SOX2 and CDX2 predict patient outcome. BMC Cancer *14*, 753.

Cao, J., Spielmann, M., Qiu, X., Huang, X., Ibrahim, D.M., Hill, A.J., Zhang, F., Mundlos, S., Christiansen, L., Steemers, F.J., et al. (2019). The single-cell transcriptional landscape of mammalian organogenesis. Nature *566*, 496–502.

Carafa, V., Altucci, L., and Nebbioso, A. (2019). Dual Tumor Suppressor and Tumor Promoter Action of Sirtuins in Determining Malignant Phenotype. Front. Pharmacol. *10*.

de Charette, M., Marabelle, A., and Houot, R. (2016). Turning tumour cells into antigen presenting cells: The next step to improve cancer immunotherapy? Eur. J. Cancer *68*, 134–147.

Chauvin, J.-M., Pagliano, O., Fourcade, J., Sun, Z., Wang, H., Sander, C., Kirkwood, J.M., Chen, T.T., Maurer, M., Korman, A.J., et al. (2015). TIGIT and PD-1 impair tumor antigen–specific CD8+ T cells in melanoma patients. J. Clin. Invest. *125*, 2046–2058.

Chen, L., Wang, W., Koide, A., Bolen, J., Miller, G., and Koide, S. (2019). Abstract 1551: First in class immunotherapy targeting Galectin-9 promotes T-cell activation and anti-tumor response against pancreatic cancer and other solid tumors. In Immunology, (American Association for Cancer Research), pp. 1551–1551.

Chen, Y., Xin, Z., Huang, L., Zhao, L., Wang, S., Cheng, J., Wu, P., and Chai, Y. (2020). CD8+ T Cells Form the Predominant Subset of NKG2A+ Cells in Human Lung Cancer. Front. Immunol. *10*, 1–11.

Cheng, Y., Song, Y., Qu, J., Che, X., Song, N., Fan, Y., Wen, T., Xu, L., Gong, J., Wang, X., et al. (2018). The Chemokine Receptor CXCR4 and c-MET Cooperatively Promote Epithelial-Mesenchymal Transition in Gastric Cancer Cells. Transl. Oncol. *11*, 487–497.

Cibrián, D., and Sánchez-Madrid, F. (2017). CD69: from activation marker to metabolic gatekeeper. Eur. J. Immunol. *47*, 946–953.

Cingolani, P., Platts, A., Wang, L.L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X., and

Ruden, D.M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. Fly (Austin). *6*, 80–92.

Correa, P., Piazuelo, B.M., and Wilson, K.T. (2010). Pathology of Gastric Intestinal Metaplasia: Clinical Implications. Am. J. Gastroenterol. *105*, 493–498.

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. (2015). Molecular analysis of gastric cancer identifies subtypes associated with distinct clinical outcomes. Nat. Med. *21*, 449–456.

Dai, S.-Y., Nakagawa, R., Itoh, A., Murakami, H., Kashio, Y., Abe, H., Katoh, S., Kontani, K., Kihara, M., Zhang, S.-L., et al. (2005). Galectin-9 Induces Maturation of Human Monocyte-Derived Dendritic Cells. J. Immunol. *175*, 2974–2981.

Das, R., Coupar, J., Clavijo, P.E., Saleh, A., Cheng, T.-F., Yang, X., Chen, J., Van Waes, C., and Chen, Z. (2019). Lymphotoxin-β receptor-NIK signaling induces alternative RELB/NF-κB2 activation to promote metastatic gene expression and cell migration in head and neck cancer. Mol. Carcinog. *58*, 411–425.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: Ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21.

Du, L., Li, Y.-J., Fakih, M., Wiatrek, R.L., Duldulao, M., Chen, Z., Chu, P., Garcia-Aguilar, J., and Chen, Y. (2016). Role of SUMO activating enzyme in cancer stem cell maintenance and self-renewal. Nat. Commun. *7*, 12326.

Duhen, T., Duhen, R., Montler, R., Moses, J., Moudgil, T., de Miranda, N.F., Goodall, C.P., Blair, T.C., Fox, B.A., McDermott, J.E., et al. (2018). Co-expression of CD39 and CD103 identifies tumor-reactive CD8 T cells in human solid tumors. Nat. Commun. *9*, 2724.

Efremova, M., Vento-Tormo, M., Teichmann, S.A., and Vento-Tormo, R. (2020). CellPhoneDB: inferring cell–cell communication from combined expression of multi-subunit ligand–receptor complexes. Nat. Protoc. *15*, 1484–1506.

Fadda, L., Borhis, G., Ahmed, P., Cheent, K., Pageon, S. V., Cazaly, A., Stathopoulos, S., Middleton, D., Mulder, A., Claas, F.H.J., et al. (2010). Peptide antagonism as a mechanism for NK cell activation. Proc. Natl. Acad. Sci. *107*, 10160–10165.

Feng, Y. -x., Sokol, E.S., Del Vecchio, C.A., Sanduja, S., Claessen, J.H.L., Proia, T.A., Jin, D.X., Reinhardt, F., Ploegh, H.L., Wang, Q., et al. (2014). Epithelial-to-Mesenchymal Transition Activates PERK-eIF2 and Sensitizes Cells to Endoplasmic Reticulum Stress. Cancer Discov. *4*, 702–715.

Figueroa-Protti, L., Soto-Molinari, R., Calderón-Osorno, M., Mora, J., and Alpízar-Alpízar, W. (2019). Gastric Cancer in the Era of Immune Checkpoint Blockade. J. Oncol. *2019*, 1–11.

Francipane, M.G., and Lagasse, E. (2016). Therapeutic potential of mTOR inhibitors for targeting cancer stem cells. Br. J. Clin. Pharmacol. *82*, 1180–1188.

Fuchs, C.S., Doi, T., Jang, R.W., Muro, K., Satoh, T., Machado, M., Sun, W., Jalal, S.I., Shah, M.A., Metges, J.-P., et al. (2018). Safety and Efficacy of Pembrolizumab Monotherapy in Patients With Previously Treated Advanced Gastric and Gastroesophageal Junction Cancer. JAMA Oncol. *4*, e180013.

Fujii, Y., Yoshihashi, K., Suzuki, H., Tsutsumi, S., Mutoh, H., Maeda, S., Yamagata, Y., Seto, Y., Aburatani, H., and Hatakeyama, M. (2012). CDX1 confers intestinal phenotype on gastric epithelial cells via induction of stemness-associated reprogramming factors SALL4 and KLF5. Proc. Natl. Acad. Sci. *109*, 20584–20589.

Gallegos, A.M., Xiong, H., Leiner, I.M., Sušac, B., Glickman, M.S., Pamer, E.G., and van Heijst, J.W.J. (2016). Control of T cell antigen reactivity via programmed TCR downregulation. Nat. Immunol. *17*, 379–386.

Garnelo, M., Tan, A., Her, Z., Yeong, J., Lim, C.J., Chen, J., Lim, K.H., Weber, A., Chow, P., Chung, A., et al. (2017). Interaction between tumour-infiltrating B cells and T cells controls the progression of hepatocellular carcinoma. Gut *66*, 342–351.

Golden-Mason, L., McMahan, R.H., Strong, M., Reisdorph, R., Mahaffey, S., Palmer, B.E., Cheng, L., Kulesza, C., Hirashima, M., Niki, T., et al. (2013). Galectin-9 Functionally Impairs Natural Killer Cells in Humans and Mice. J. Virol. *87*, 4835–4845.

Gonçalves Silva, I., Yasinska, I.M., Sakhnevych, S.S., Fiedler, W., Wellbrock, J., Bardelli, M., Varani, L., Hussain, R., Siligardi, G., Ceccone, G., et al. (2017). The Tim-3-galectin-9

Secretory Pathway is Involved in the Immune Escape of Human Acute Myeloid Leukemia Cells. EBioMedicine *22*, 44–57.

Gong, L., Qi, R., and Li, D. (2017). Sumoylation Pathway as Potential Therapeutic Targets in Cancer. Curr. Mol. Med. *16*, 900–905.

Gooden, M.J.M., Wiersma, V.R., Boerma, A., Leffers, N., Boezen, H.M., ten Hoor, K.A., Hollema, H., Walenkamp, A.M.E., Daemen, T., Nijman, H.W., et al. (2014). Elevated serum CXCL16 is an independent predictor of poor survival in ovarian cancer and may reflect pro-metastatic ADAM protease activity. Br. J. Cancer *110*, 1535–1544.

Grasmann, G., Smolle, E., Olschewski, H., and Leithner, K. (2019). Gluconeogenesis in cancer cells – Repurposing of a starvation-induced metabolic pathway? Biochim. Biophys. Acta - Rev. Cancer *1872*, 24–36.

Gu, X., Xing, X., Yang, W., Hu, J., and Dai, D. (2014). High expression of integrin-linked kinase predicts aggressiveness and poor prognosis in patients with gastric cancer. Acta Histochem. *116*, 758–762.

Gulhati, P., Bowen, K.A., Liu, J., Stevens, P.D., Rychahou, P.G., Chen, M., Lee, E.Y., Weiss, H.L., O'Connor, K.L., Gao, T., et al. (2011). mTORC1 and mTORC2 Regulate EMT, Motility, and Metastasis of Colorectal Cancer via RhoA and Rac1 Signaling Pathways. Cancer Res. *71*, 3246–3256.

Hafemeister, C., and Satija, R. (2019). Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. Genome Biol. *20*, 296.

Han, S., Fink, J., Jörg, D.J., Lee, E., Yum, M.K., Chatzeli, L., Merker, S.R., Josserand, M., Trendafilova, T., Andersson-Rolf, A., et al. (2019). Defining the Identity and Dynamics of Adult Gastric Isthmus Stem Cells. Cell Stem Cell *25*, 342-356.e7.

Han, X., Zhou, Z., Fei, L., Sun, H., Wang, R., Chen, Y., Chen, H., Wang, J., Tang, H., Ge, W., et al. (2020). Construction of a human cell landscape at single-cell level. Nature.

Harjunpää, H., and Guillerey, C. (2020). TIGIT as an emerging immune checkpoint. Clin. Exp. Immunol. *200*, 108–119.

He, Y., Cao, J., Zhao, C., Li, X., Zhou, C., and Hirsch, F. (2018). TIM-3, a promising target for cancer immunotherapy. Onco. Targets. Ther. *Volume 11*, 7005–7009.

Hehlgans, T., Stoelcker, B., Stopfer, P., Müller, P., Cernaianu, G., Guba, M., Steinbauer, M., Nedospasov, S.A., Pfeffer, K., and Männel, D.N. (2002). Lymphotoxin-beta receptor immune interaction promotes tumor growth by inducing angiogenesis. Cancer Res. *62*, 4034–4040.

Hibdon, E.S., Razumilava, N., Keeley, T.M., Wong, G., Solanki, S., Shah, Y.M., and Samuelson, L.C. (2019). Notch and mTOR Signaling Pathways Promote Human Gastric Cancer Cell Proliferation. Neoplasia *21*, 702–712.

Hu, B., El Hajj, N., Sittler, S., Lammert, N., Barnes, R., and Meloni-Ehrig, A. (2012). Gastric cancer: Classification, histology and application of molecular pathology. J. Gastrointest. Oncol. *3*, 251–261.

Huffman, A.P., Lin, J.H., Kim, S.I., Byrne, K.T., and Vonderheide, R.H. (2020). CCL5 mediates CD40-driven CD4+ T cell tumor infiltration and immunity. JCI Insight *5*.

Ichikawa, H., Nagahashi, M., Shimada, Y., Hanyu, T., Ishikawa, T., Kameyama, H., Kobayashi, T., Sakata, J., Yabusaki, H., Nakagawa, S., et al. (2017). Actionable gene-based classification toward precision medicine in gastric cancer. Genome Med. *9*, 93.

Ivanova, N.B. (2002). A Stem Cell Molecular Signature. Science (80-.). *298*, 601–604.

Jaggupilli, A., and Elkord, E. (2012). Significance of CD44 and CD24 as Cancer Stem Cell Markers: An Enduring Ambiguity. Clin. Dev. Immunol. *2012*, 1–11.

Janjigian, Y.Y., Bendell, J., Calvo, E., Kim, J.W., Ascierto, P.A., Sharma, P., Ott, P.A., Peltola, K., Jaeger, D., Evans, J., et al. (2018). CheckMate-032 Study: Efficacy and Safety of Nivolumab and Nivolumab Plus Ipilimumab in Patients With Metastatic Esophagogastric Cancer. J. Clin. Oncol. *36*, 2836–2844.

Jia, D., Lu, M., Jung, K.H., Park, J.H., Yu, L., Onuchic, J.N., Kaipparettu, B.A., and Levine, H. (2019). Elucidating cancer metabolic plasticity by coupling gene regulation with metabolic pathways. Proc. Natl. Acad. Sci. *116*, 3909–3918.

Jones, M.F., Hara, T., Francis, P., Li, X.L., Bilke, S., Zhu, Y., Pineda, M., Subramanian, M., Bodmer, W.F., and Lal, A. (2015). The CDX1–microRNA-215 axis regulates colorectal cancer stem cell differentiation. Proc. Natl. Acad. Sci. *112*, E1550–E1558.

Kang, Y.-K., Boku, N., Satoh, T., Ryu, M.-H., Chao, Y., Kato, K., Chung, H.C., Chen, J.-S., Muro, K., Kang, W.K., et al. (2017). Nivolumab in patients with advanced gastric or gastro-oesophageal junction cancer refractory to, or intolerant of, at least two previous chemotherapy regimens (ONO-4538-12, ATTRACTION-2): a randomised, double-blind, placebo-controlled, phase 3 trial. Lancet *390*, 2461–2471.

Katsuno, Y., Meyer, D.S., Zhang, Z., Shokat, K.M., Akhurst, R.J., Miyazono, K., and Derynck, R. (2019). Chronic TGF-β exposure drives stabilized EMT, tumor stemness, and cancer drug resistance with vulnerability to bitopic mTOR inhibition. Sci. Signal. *12*.

Kim, S.T., Cristescu, R., Bass, A.J., Kim, K.-M., Odegaard, J.I., Kim, K., Liu, X.Q., Sher, X., Jung, H., Lee, M., et al. (2018). Comprehensive molecular characterization of clinical responses to PD-1 inhibition in metastatic gastric cancer. Nat. Med. *24*, 1449–1458.

KOMITA, H., KOIDO, S., HAYASHI, K., KAN, S., ITO, M., KAMATA, Y., SUZUKI, M., and HOMMA, S. (2015). Expression of immune checkpoint molecules of T cell immunoglobulin and mucin protein 3/galectin-9 for NK cell suppression in human gastrointestinal stromal tumors. Oncol. Rep. *34*, 2099–2105.

Kulangara, K., Zhang, N., Corigliano, E., Guerrero, L., Waldroup, S., Jaiswal, D., MS, M.J., Shah, S., Hanks, D., Wang, J., et al. (2019). Clinical Utility of the Combined Positive Score for Programmed Death Ligand-1 Expression and the Approval of Pembrolizumab for Treatment of Gastric Cancer. Arch. Pathol. Lab. Med. *143*, 330–337.

LAUREN, P. (1965). the Two Histological Main Types of Gastric Carcinoma: Diffuse and So-Called Intestinal-Type Carcinoma. an Attempt At a Histo-Clinical Classification. Acta Pathol. Microbiol. Scand. *64*, 31–49.

Lauritsen, J.P., Christensen, M.D., Dietrich, J., Kastrup, J., Odum, N., and Geisler, C. (1998). Two distinct pathways exist for down-regulation of the TCR. J. Immunol. *161*, 260–267.

Le, D.T., Uram, J.N., Wang, H., Bartlett, B.R., Kemberling, H., Eyring, A.D., Skora, A.D., Luber, B.S., Azad, N.S., Laheru, D., et al. (2015). PD-1 Blockade in Tumors with Mismatch-Repair Deficiency. N. Engl. J. Med. *372*, 2509–2520.

Lee, W., and Lee, G.R. (2018). Transcriptional regulation and development of regulatory T cells. Exp. Mol. Med. *50*, e456-10.

Leushacke, M., Tan, S.H., Wong, A., Swathi, Y., Hajamohideen, A., Tan, L.T., Goh, J., Wong, E., Denil, S.L.I.J., Murakami, K., et al. (2017). Lgr5-expressing chief cells drive epithelial regeneration and cancer in the oxyntic stomach. Nat. Cell Biol. *19*, 774–786.

Levine, J.H., Simonds, E.F., Bendall, S.C., Davis, K.L., Amir, E.D., Tadmor, M.D., Litvin, O., Fienberg, H.G., Jager, A., Zunder, E.R., et al. (2015). Data-Driven Phenotypic Dissection of AML Reveals Progenitor-like Cells that Correlate with Prognosis. Cell *162*, 184–197.

Li, B., and Dewey, C.N. (2014). RSEM: Accurate transcript quantification from RNA-seq data with or without a reference genome. Bioinforma. Impact Accurate Quantif. Proteomic Genet. Anal. Res. *12*, 41–74.

Li, H., Wu, K., Tao, K., Chen, L., Zheng, Q., Lu, X., Liu, J., Shi, L., Liu, C., Wang, G., et al. (2012). Tim-3/galectin-9 signaling pathway mediates T-cell dysfunction and predicts poor prognosis in patients with hepatitis B virus-associated hepatocellular carcinoma. Hepatology *56*, 1342–1351.

Li, Y.-H., Zhou, W.-H., Tao, Y., Wang, S.-C., Jiang, Y.-L., Zhang, D., Piao, H.-L., Fu, Q., Li, D.-J., and Du, M.-R. (2016). The Galectin-9/Tim-3 pathway is involved in the regulation of NK cell function at the maternal–fetal interface in early pregnancy. Cell. Mol. Immunol. *13*, 73–81.

Ling, A., Löfgren-Burström, A., Larsson, P., Li, X., Wikberg, M.L., Öberg, Å., Stenling, R., Edin, S., and Palmqvist, R. (2017). TAP1 down-regulation elicits immune escape and poor prognosis in colorectal cancer. Oncoimmunology *6*, e1356143.

Liu, Y., Zhang, Z., Wang, J., Chen, C., Tang, X., Zhu, J., and Liu, J. (2019). Metabolic reprogramming results in abnormal glycolysis in gastric cancer: a review. Onco. Targets. Ther. *Volume 12*, 1195–1204.

Loboda, A., Nebozhyn, M. V, Watters, J.W., Buser, C.A., Shaw, P.M., Huang, P.S., Van't Veer, L., Tollenaar, R.A., Jackson, D.B., Agrawal, D., et al. (2011). EMT is the dominant program in human colon cancer. BMC Med. Genomics *4*, 9.

Mackay, L.K., Rahimpour, A., Ma, J.Z., Collins, N., Stock, A.T., Hafon, M.-L., Vega-Ramos, J., Lauzurica, P., Mueller, S.N., Stefanovic, T., et al. (2013). The developmental pathway for CD103+CD8+ tissue-resident memory T cells of skin. Nat. Immunol. *14*, 1294–1301.

Mahajan, K., and Mahajan, N.P. (2013). ACK1 tyrosine kinase: Targeted inhibition to block cancer cell proliferation. Cancer Lett. *338*, 185–192.

Maldonado, M. del M., and Dharmawardhane, S. (2018). Targeting Rac and Cdc42 GTPases in Cancer. Cancer Res.

Mao, Q., Liang, X.-L., Wu, Y.-F., Pang, Y.-H., Zhao, X.-J., and Lu, Y.-X. (2019). ILK promotes survival and self-renewal of hypoxic MSCs via the activation of lncTCF7-Wnt pathway induced by IL-6/STAT3 signaling. Gene Ther. *26*, 165–176.

Martin, M.D., and Badovinac, V.P. (2018). Defining memory CD8 T cell. Front. Immunol. *9*, 1–10.

McInnes, L., Healy, J., and Melville, J. (2018). UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction.

Meijer, J., Ogink, J., Kreike, B., Nuyten, D., de Visser, K.E., and Roos, E. (2008). The Chemokine Receptor CXCR6 and Its Ligand CXCL16 Are Expressed in Carcinomas and Inhibit Proliferation. Cancer Res. *68*, 4701–4708.

Miranda, A., Hamilton, P.T., Zhang, A.W., Pattnaik, S., Becht, E., Mezheyeuski, A., Bruun, J., Micke, P., de Reynies, A., and Nelson, B.H. (2019). Cancer stemness, intratumoral heterogeneity, and immune response across cancers. Proc. Natl. Acad. Sci. *116*, 9020–9029.

Nam, S., Kim, J.H., and Lee, D.H. (2019). RHOA in Gastric Cancer: Functional Roles and Therapeutic Potential. Front. Genet. *10*.

Nishida-Aoki, N., and Gujral, T.S. (2019). Emerging approaches to study cell&#x2013;cell interactions in tumor microenvironment. Oncotarget *10*.

Nitzsche, A., Paszkowski-Rogacz, M., Matarese, F., Janssen-Megens, E.M., Hubner, N.C., Schulz, H., de Vries, I., Ding, L., Huebner, N., Mann, M., et al. (2011). RAD21 Cooperates with Pluripotency Transcription Factors in the Maintenance of Embryonic Stem Cell Identity. PLoS One *6*, e19470.

Nixon, N.A., Blais, N., Ernst, S., Kollmannsberger, C., Bebb, G., Butler, M., Smylie, M., and

Verma, S. (2018). Current landscape of immune therapy in treatment of solid tumours, with future opportunities and challenges. Curr. Oncol. *25*.

Novak, A., Hsu, S.-C., Leung-Hagesteijn, C., Radeva, G., Papkoff, J., Montesano, R., Roskelley, C., Grosschedl, R., and Dedhar, S. (1998). Cell adhesion and the integrin-linked kinase regulate the LEF-1 and -catenin signaling pathways. Proc. Natl. Acad. Sci. *95*, 4374–4379.

Oshima, T., Sato, S., Kato, J., Ito, Y., Watanabe, T., Tsuji, I., Hori, A., Kurokawa, T., and Kokubo, T. (2013). Nectin-2 is a potential target for antibody therapy of breast and ovarian cancers. Mol. Cancer *12*, 60.

Pandey, R., Zhou, M., Islam, S., Chen, B., Barker, N.K., Langlais, P., Srivastava, A., Luo, M., Cooke, L.S., Weterings, E., et al. (2019). Carcinoembryonic antigen cell adhesion molecule 6 (CEACAM6) in Pancreatic Ductal Adenocarcinoma (PDA): An integrative analysis of a novel therapeutic target. Sci. Rep. *9*, 18347.

Pardoll, D.M. (2012). The blockade of immune checkpoints in cancer immunotherapy. Nat. Rev. Cancer *12*, 252–264.

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. (2014). Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. Science (80-.). *344*, 1396–1401.

Petrelli, F., Berenato, R., Turati, L., Mennitto, A., Steccanella, F., Caporale, M., Dallera, P., de Braud, F., Pezzica, E., Di Bartolomeo, M., et al. (2017). Prognostic value of diffuse versus intestinal histotype in patients with gastric cancer: A systematic review and meta-analysis. J. Gastrointest. Oncol. *8*, 148–163.

Phesse, T.J., and Sansom, O.J. (2017). Lgr5 joins the club of gastric stem cell markers in the corpus. Nat. Cell Biol. *19*, 752–754.

Picelli, S., Faridani, O.R., Björklund, Å.K., Winberg, G., Sagasser, S., and Sandberg, R. (2014). Full-length RNA-seq from single cells using Smart-seq2. Nat. Protoc. *9*, 171–181.

Quandt, J., and Dorovini-Zis, K. (2004). The Beta Chemokines CCL4 and CCL5 Enhance Adhesion of Specific CD4+ T Cell Subsets to Human Brain Endothelial Cells. J. Neuropathol.

Exp. Neurol. *63*, 350–362.

Ren, R., Ocampo, A., Liu, G.-H., and Izpisua Belmonte, J.C. (2017). Regulation of Stem Cell Aging by Metabolism and Epigenetics. Cell Metab. *26*, 460–474.

Romee, R., Foley, B., Lenvik, T., Wang, Y., Zhang, B., Ankarlo, D., Luo, X., Cooley, S., Verneris, M., Walcheck, B., et al. (2013). NK cell CD16 surface expression and function is regulated by a disintegrin and metalloprotease-17 (ADAM17). Blood *121*, 3599–3608.

Rothe, M., Sarma, V., Dixit, V., and Goeddel, D. (1995). TRAF2-mediated activation of NF-kappa B by TNF receptor 2 and CD40. Science (80-.). *269*, 1424–1427.

Ru, G.-Q., Han, Y., Wang, W., Chen, Y., Wang, H.-J., Xu, W.-J., Ma, J., Ye, M., Chen, X., He, X.-L., et al. (2017). CEACAM6 is a prognostic biomarker and potential therapeutic target for gastric carcinoma. Oncotarget *8*, 83673–83683.

Saha, S.K., Islam, S.M.R., Kwak, K.-S., Rahman, M.S., and Cho, S.-G. (2020). PROM1 and PROM2 expression differentially modulates clinical prognosis of cancer: a multiomics analysis. Cancer Gene Ther. *27*, 147–167.

San José, E., Borroto, A., Niedergang, F., Alcover, A., and Alarcón, B. (2000). Triggering the TCR complex causes the downregulation of nonengaged receptors by a signal transduction-dependent mechanism. Immunity *12*, 161–170.

Santana Carrero, R.M., Beceren-Braun, F., Rivas, S.C., Hegde, S.M., Gangadharan, A., Plote, D., Pham, G., Anthony, S.M., and Schluns, K.S. (2019). IL-15 is a component of the inflammatory milieu in the tumor microenvironment promoting antitumor responses. Proc. Natl. Acad. Sci. *116*, 599–608.

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.

Saxton, R.A., and Sabatini, D.M. (2017). mTOR Signaling in Growth, Metabolism, and Disease. Cell *169*, 361–371.

See, P., Lum, J., Chen, J., and Ginhoux, F. (2018). A Single-Cell Sequencing Guide for Immunologists. Front. Immunol. *9*.

Shin, D.-M., Liu, R., Wu, W., Waigel, S.J., Zacharias, W., Ratajczak, M.Z., and Kucia, M. (2012). Global Gene Expression Analysis of Very Small Embryonic-Like Stem Cells Reveals that the Ezh2 -Dependent Bivalent Domain Mechanism Contributes to Their Pluripotent State. Stem Cells Dev. *21*, 1639–1652.

Shitara, K., Özgüroğlu, M., Bang, Y.-J., Di Bartolomeo, M., Mandalà, M., Ryu, M.-H., Fornaro, L., Olesiński, T., Caglevic, C., Chung, H.C., et al. (2018). Pembrolizumab versus paclitaxel for previously treated, advanced gastric or gastro-oesophageal junction cancer (KEYNOTE-061): a randomised, open-label, controlled, phase 3 trial. Lancet *392*, 123–133.

Shrestha, N., Bahnan, W., Wiley, D.J., Barber, G., Fields, K.A., and Schesser, K. (2012). Eukaryotic Initiation Factor 2 (eIF2) signaling regulates proinflammatory cytokine expression and bacterial invasion. J. Biol. Chem. *287*, 28738–28744.

Stamm, H., Klingler, F., Grossjohann, E.-M., Muschhammer, J., Vettorazzi, E., Heuser, M., Mock, U., Thol, F., Vohwinkel, G., Latuske, E., et al. (2018). Immune checkpoints PVR and PVRL2 are prognostic markers in AML and their blockade represents a new therapeutic option. Oncogene *37*, 5269–5280.

Steinbach, K., Vincenti, I., and Merkler, D. (2018). Resident-Memory T Cells in tissue-restricted immune responses: For better or worse? Front. Immunol. *9*, 1–15.

Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M., Hao, Y., Stoeckius, M., Smibert, P., and Satija, R. (2019). Comprehensive Integration of Single-Cell Data. Cell *177*, 1888-1902.e21.

Stubbington, M.J.T., Lönnberg, T., Proserpio, V., Clare, S., Speak, A.O., Dougan, G., and Teichmann, S.A. (2016). T cell fate and clonality inference from single-cell transcriptomes. Nat. Methods *13*, 329–332.

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., et al. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. *102*, 15545–15550.

Sun, L., Wang, Q., Chen, B., Zhao, Y., Shen, B., Wang, H., Xu, J., Zhu, M., Zhao, X., Xu, C., et al. (2018). Gastric cancer mesenchymal stem cells derived IL-8 induces PD-L1 expression in gastric cancer cells via STAT3/mTOR-c-Myc signal axis. Cell Death Dis. *9*, 928.

Svensmark, J.H., and Brakebusch, C. (2019). Rho GTPases in cancer: friend or foe? Oncogene *38*, 7447–7456.

Takaishi, S., Okumura, T., Tu, S., Wang, S.S.W., Shibata, W., Vigneshwaran, R., Gordon, S.A.K., Shimada, Y., and Wang, T.C. (2009). Identification of Gastric Cancer Stem Cells Using the Cell Surface Marker CD44. Stem Cells *27*, 1006–1020.

Tan, S.H., Swathi, Y., Tan, S., Goh, J., Seishima, R., Murakami, K., Oshima, M., Tsuji, T., Phuah, P., Tan, L.T., et al. (2020). AQP5 enriches for stem cells and cancer origins in the distal stomach. Nature *578*, 437–443.

Tang, W., Lu, Y., Tian, Q.-Y., Zhang, Y., Guo, F.-J., Liu, G.-Y., Syed, N.M., Lai, Y., Lin, E.A., Kong, L., et al. (2011). The Growth Factor Progranulin Binds to TNF Receptors and Is Therapeutic Against Inflammatory Arthritis in Mice. Science (80-.). *332*, 478–484.

Tarocco, A., Caroccia, N., Morciano, G., Wieckowski, M.R., Ancora, G., Garani, G., and Pinton, P. (2019). Melatonin as a master regulator of cell death and inflammation: molecular mechanisms and clinical implications for newborn care. Cell Death Dis. *10*, 317.

Teschendorff, A.E., and Enver, T. (2017). Single-cell entropy for accurate estimation of differentiation potency from a cell’s transcriptome. Nat. Commun. *8*, 15599.

Tomita, H., Tanaka, K., Tanaka, T., and Hara, A. (2016). Aldehyde dehydrogenase 1A1 in stem cells and cancer. Oncotarget *7*, 11018–11032.

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

Tsukerman, P., Stern-Ginossar, N., Yamin, R., Ophir, Y., Stanietsky, A.M.N., and Mandelboim, O. (2014). Expansion of CD16 positive and negative human NK cells in response to tumor stimulation. Eur. J. Immunol. *44*, 1517–1525.

Varlakhanova, N. V., Cotterman, R.F., DeVries, W.N., Morgan, J., Donahue, L.R., Murray, S., Knowles, B.B., and Knoepfler, P.S. (2010). myc maintains embryonic stem cell pluripotency and self-renewal. Differentiation *80*, 9–19.

Vento-Tormo, R., Efremova, M., Botting, R.A., Turco, M.Y., Vento-Tormo, M., Meyer, K.B., Park, J.E., Stephenson, E., Polański, K., Goncalves, A., et al. (2018). Single-cell reconstruction of the early maternal–fetal interface in humans. Nature *563*, 347–353.

Wang, R., Song, S., Harada, K., Han, G., Pizzi, M.P., Zhao, M., Tatlonghari, G., Zhang, S., Wang, Y., Zhao, S., et al. (2019). Cell-of-Origin Analysis of Metastatic Gastric Cancer Uncovers the Origin of Inherent Intratumor Heterogeneity and a Fundamental Prognostic Signature. BioRxiv 725390.

Wang, Y., Sun, J., Ma, C., Gao, W., Song, B., Xue, H., Chen, W., Chen, X., Zhang, Y., Shao, Q., et al. (2016). Reduced Expression of Galectin-9 Contributes to a Poor Outcome in Colon Cancer by Inhibiting NK Cell Chemotaxis Partially through the Rho/ROCK1 Signaling Pathway. PLoS One *11*, e0152599.

Wiersma, V.R., de Bruyn, M., Helfrich, W., and Bremer, E. (2013). Therapeutic potential of Galectin-9 in human disease. Med. Res. Rev. *33*, E102–E126.

Wiesner, S., Legate, K.R., and Fässler, R. (2005). Integrin-actin interactions. C. Cell. Mol. Life Sci. *62*, 1081–1099.

Wu, C., Keightley, S.Y., Leung-Hagesteijn, C., Radeva, G., Coppolino, M., Goicoechea, S., McDonald, J.A., and Dedhar, S. (1998). Integrin-linked Protein Kinase Regulates Fibronectin Matrix Assembly, E-cadherin Expression, and Tumorigenicity. J. Biol. Chem. *273*, 528–536.

Wu, C., Thalhamer, T., Franca, R.F., Xiao, S., Wang, C., Hotta, C., Zhu, C., Hirashima, M., Anderson, A.C., and Kuchroo, V.K. (2014). Galectin-9-CD44 Interaction Enhances Stability and Function of Adaptive Regulatory T Cells. Immunity *41*, 270–282.

Wu, T.D., Madireddi, S., de Almeida, P.E., Banchereau, R., Chen, Y.J.J., Chitre, A.S., Chiang, E.Y., Iftikhar, H., O’Gorman, W.E., Au-Yeung, A., et al. (2020a). Peripheral T cell expansion predicts tumour infiltration and clinical response. Nature *579*, 274–278.

Wu, T.D., Madireddi, S., de Almeida, P.E., Banchereau, R., Chen, Y.-J.J., Chitre, A.S., Chiang, E.Y., Iftikhar, H., O’Gorman, W.E., Au-Yeung, A., et al. (2020b). Peripheral T cell expansion predicts tumour infiltration and clinical response. Nature *579*, 274–278.

Wu, X., Tao, P., Zhou, Q., Li, J., Yu, Z., Wang, X., Li, J., Li, C., Yan, M., Zhu, Z., et al. (2017).

IL-6 secreted by cancer-associated fibroblasts promotes epithelial-mesenchymal transition and metastasis of gastric cancer via JAK2/STAT3 signaling pathway. Oncotarget *8*.

Xiang, Z., Zhou, Z.-J., Xia, G.-K., Zhang, X.-H., Wei, Z.-W., Zhu, J.-T., Yu, J., Chen, W., He, Y., Schwarz, R.E., et al. (2017). A positive crosstalk between CXCR4 and CXCR2 promotes gastric cancer metastasis. Oncogene *36*, 5122–5133.

Xu, F., Zhou, G., Han, S., Yuan, W., Chen, S., Fu, Z., Li, D., Zhang, H., Li, D., and Pang, D. (2014). Association of TNF-α, TNFRSF1A and TNFRSF1B Gene Polymorphisms with the Risk of Sporadic Breast Cancer in Northeast Chinese Han Women. PLoS One *9*, e101138.

Yang, D., Wang, L.-L., Dong, T.-T., Shen, Y.-H., Guo, X.-S., Liu, C.-Y., Liu, J., Zhang, P., Li, J., and Sun, Y.-P. (2015). Progranulin promotes colorectal cancer proliferation and angiogenesis through TNFR2/Akt and ERK signaling pathways. Am. J. Cancer Res. *5*, 3085–3097.

Yang, L., Wang, Y., and Wang, H. (2019). Use of immunotherapy in the treatment of gastric cancer (Review). Oncol. Lett.

Yeo, H.C., Beh, T.T., Quek, J.J.L., Koh, G., Chan, K.K.K., and Lee, D.-Y. (2011). Integrated Transcriptome and Binding Sites Analysis Implicates E2F in the Regulation of Self-Renewal in Human Pluripotent Stem Cells. PLoS One *6*, e27231.

Yuan, L., Xu, B., Yuan, P., Zhou, J., Qin, P., Han, L., Chen, G., Wang, Z., Run, Z., Zhao, P., et al. (2017). Tumor-infiltrating CD4+ T cells in patients with gastric cancer. Cancer Cell Int. *17*, 114.

Zhang, J., Zhou, Y., Huang, T., Wu, F., Pan, Y., Dong, Y., Wang, Y., Chan, A.K.Y., Liu, L., Kwan, J.S.H., et al. (2019a). FGF18, a prominent player in FGF signaling, promotes gastric tumorigenesis through autocrine manner and is negatively regulated by miR-590-5p. Oncogene *38*, 33–46.

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

Zhang, W., Gu, J., Chen, J., Zhang, P., Ji, R., Qian, H., Xu, W., and Zhang, X. (2017).

Interaction with neutrophils promotes gastric cancer cell migration and invasion by inducing epithelial-mesenchymal transition. Oncol. Rep. *38*, 2959–2966.

Zheng, C.-C., Hu, H.-F., Hong, P., Zhang, Q.-H., Xu, W.W., He, Q.-Y., and Li, B. (2019). Significance of integrin-linked kinase (ILK) in tumorigenesis and its potential implication as a biomarker and therapeutic target for human cancer. Am. J. Cancer Res. *9*, 186–197.

Zheng, C., Zheng, L., Yoo, J.K., Guo, H., Zhang, Y., Guo, X., Kang, B., Hu, R., Huang, J.Y., Zhang, Q., et al. (2017). Landscape of Infiltrating T Cells in Liver Cancer Revealed by Single-Cell Sequencing. Cell *169*, 1342-1356.e16.

Zhu, Q., Zhang, X., Zhang, L., Li, W., Wu, H., Yuan, X., Mao, F., Wang, M., Zhu, W., Qian, H., et al. (2014). The IL-6–STAT3 axis mediates a reciprocal crosstalk between cancer-derived mesenchymal stem cells and neutrophils to synergistically prompt gastric cancer progression. Cell Death Dis. *5*, e1295–e1295.

Zimberlin, C.D., Lancini, C., Sno, R., Rosekrans, S.L., McLean, C.M., Vlaming, H., van den Brink, G.R., Bots, M., Medema, J.P., and Dannenberg, J.-H. (2015). HDAC1 and HDAC2 collectively regulate intestinal stem cell homeostasis. FASEB J. *29*, 2070–2080.

Table

Table 1. Tabulation of NK, CD4T, CD8T and Treg immune subsets obtained from Epithelial-2-dominant patients (n=1021) according to tissues with >= 20% and < 20% Stem-like Epithelial (SE) of all epithelial cells, with respect to Normal-adjacent and Tumor tissues.

	NK (n=270)		**CD4 T (n=316)[a]**					**CD8 T (n=329)**			**Treg**
Tissue	**NK**	**CD16^{+} NK**	**CD4Tcm**	**CCL5^{+} CD4T**	**Exh. CD4T**	**Activated CD4Trm**	**CCL5^{+} CD4Trm**	**KLRC1^{+} CD8Trm**	**Activated CD8Trm**	**TIM3^{+} CD8Trm**	**Treg**
Normal-Adjacent											
SE cells >= 20% (%)	59 (82)	13 (18)	8 (22)	9 (25)	7 (19)	12 (33)	4 (N.A.)	55 (54)	30 (30)	16 (16)	4 (N.A.)
SE cells < 20% (%)	56 (90)	6 (10)	1 (2)	15 (27)	13 (24)	26 (47)	14 (N.A.)	70 (79)	12 (13)	7 (8)	29 (N.A.)
Fisher's Exact (*p*)[b]	0.2164		0.0022	1.0	0.7968	0.2013	N.A.	0.0007	<0.01	0.1193	N.A.
χ^2 (Statistic, *p*)	1.9216, 0.1657		10.3882, <0.05				N.A.	12.3273, <0.005			
Tumor											
SE cells >= 20% (%)	93 (77)	28 (23)	14 (26)	5 (9)	15 (28)	19 (36)	2 (N.A.)	45 (44)	16 (16)	42 (41)	25 (N.A.)
SE cells < 20% (%)	8 (53)	7 (47)	74 (43)	72 (42)	13 (8)	13 (8)	0 (N.A.)	26 (72)	5 (14)	5 (14)	28 (N.A.)
Fisher's Exact (*p*)[b]	0.0627		<0.05	<0.0001	0.0002	<0.0001	N.A.	<0.005	1.0	<0.005	N.A.
χ^2 (Statistic, *p*)	3.8648, <0.05		52.1159, <0.00001				N.A.	10.0032, <0.01			

[a] Excluded CCL5+ CD4Trm due to count of zero value;

[b] Target versus others in NK, CD4 T, and CD8 T cells, respectively.

Figure legends

Figure 1. Tumor and normal tissues collected from gastric cancer patients can be classified into Epithelial-1-dominant (Gastric-dominant) and Epithelial-2-dominant (Gastrointestinal-mixed) tissues. (A) A schema of overall study workflow. **(B)** tSNE overview of the 2878 cells that were selected (QC-passed) for subsequent analyses in this study. Each dot in the tSNE plot indicates a single cell. **(C)** Differential expression genes for cell-type clusters identified in **(B)**. Each row represents 1 cell. Each column represents a gene known for the respective cell-type cluster. **(D)** The closest matching organs/ cell-types of the epithelial cells obtained from each patient. The middle panel shows the closest matching organ/cell-type of the epithelial cells (row) by patient (column). The size of circle indicates the proportion of epithelial cells for each specific organ/cell-type. The circles are color coded by the closest matching organs and cell types, respectively. The histogram on the top shows the cumulative number of each organ/cell-type-specific cells, grouped by patient. The histogram on the right panel shows the count and proportion of organ/cell-type-specific cells found in the respective clusters identified in **(B),** including Epithelial 1, Epithelial 2, Epithelial 1 specific to Patient SC008, and Epithelial 2 with mostly collected from Patient SC005. The pie chart tracks show proportions of epithelial clusters found in tumor and/or normal tissues collected from each patient. The bottom annotation tracks show (from top to bottom): the Lauren's histological classification, number of days to death since first diagnosed, and AJCC stage since first diagnosed.

Figure 2. Identification of Stem-like Epithelials from Epithelial 1 and 2 clusters. (A) Six epithelial clusters identified using gene expression values, including Epithelial 1 (E1), E2, E1-SC008, E2-SC005, Stem-like Epithelial A (SEA), and SEB. Each dot in the uMAP plot indicates an epithelial cell. **(B)** Four epithelial clusters identified using regulon activity scores, including Regulon 0, 1, 2, and 3 clusters, each mostly corresponds to E1/E1-SC008, E2, SE, and E2-SC005, respectively. Each dot in the tSNE plot indicates an epithelial cell. **(C)** The heatmap shows differential regulon activities comparing Regulon 0 (E1/E1-SC008), 1 (E2), and 2 (ES) cells. **(D)** Gene Set Enrichment Analysis using genesets obtained from the

MSigDB Hallmark v7.0 database, Short-term hematopoietic stem cell (ST-HSC), Long-term HSC (LT-HSC), and stomach LGR5-positive (GSE86603) shows significant positive/negative enrichments of stemness-related genesets in the SE cells. FDR<25% for each comparison, by phenotype permutation test. **(E-G)** Comparisons of entropy score, mean expressions of proliferation genes, and G2M score between epithelial clusters. **(H)** Classification of patients based on epithelial clusters are associated with overall survival in the TCGA gastric cohort. The number of patients that are E1-, E2-, SEA- and SEB-like are 101, 76, 24, and 9, respectively. The p-values from the pairwise log-rank tests are Bonferroni-corrected. **(i)** Multivariate analysis of overall survival using Cox proportional hazards regression. **(j, k)** Differential Expression Genes (DEGs) comparing E1, E2, and SE, and biological pathways predicted for each group based on the DEGs. **(L, M)** DEGs comparing SEA and SEB, and biological pathways predicted for each group based on the DEGs.

Figure 3. Distinctive developmental trajectories and biological pathways between Epithelial 1 and 2 clusters. (A) Developmental trajectory reconstruction of epithelial cells using Monocle. Here, cells are colored by their epithelial cell clusters. **(B)** Developmental trajectory with cells colored by their pseudotime. **(C)** Developmental trajectory with cells colored by the log of their total mutation burden. Increasing mutation burden along the trajectories in both E1 and E2 trajectories. **(D)** Developmental trajectory with cells colored by their entropy value computed using the LandSCENT package, showing higher estimated differentiation potency among the Stem-like Epithelial (SE) cells. **(E)** Developmental trajectory with cells colored by the mean value of the top 500 proliferation associated genes (GO:0008283). Mean expression gradually increase along the E1 and E2 trajectories towards the end points. **(F)** Developmental trajectory with cells colored by the mean value of the top 500 cell cycle associated genes (GO:0007049) with gradually increasing mean expression in both E1 an E2 trajectories. **(G)** Differentially expressed genes along the E1 trajectory, with genes grouped in three clusters based on trend (row), each colored in orange, purple and pink, respectively. The cells are sorted from left to right (column) according to

pseudotime order depicted in the trajectories. The enriched biological pathways are annotated with negative $\log_{10}$ P-value. **(H)** Differentially expressed genes and associated pathways along the E2 trajectory presented in the same format as **(G)**.

Figure 4. Comparisons of DNA copy number variations and mutation burdens between Epithelial 1, 2, Stem-like Epithelial A (SEA), and SEB cells. Predicted copy number variations (CNVs) by comparing **(A)** Epithelial 1 (E1), **(B)** E2, and **(C)** Stem-like Epithelial (SEA) obtained from the Normal-adjacent (Reference) and Tumor tissues. **(D)** Level of the deletion or amplification event found in different chromosomal regions of Stem-like Epithelial A cells in tumor tissue (reference to normal tissue), filtered with a default posterior probability threshold of 0.5 that removed copy number variations (CNVs) that are not likely to be true CNV events. **(E)** SEA cells show the highest mutation burden per million reads among all epithelial cells, in both Normal-adjacent and Tumor tissues. **(F)** E1 cells from the tumor tissues carry statistical higher mutation burdens compared to E1 cells derived from the normal-adjacent tissues.

Figure 5. Dissection of immune heterogeneity and immune-signatures-based stratification for patient survival. (A) Significant positive correlation between proportion of Stem-like Epithelial cell (of all the epithelial cells) and proportion of CD45+ immune cell (of all live cells) was found in gastric Normal-adjacent and Tumor tissues. **(B)** Four major immune cell clusters identified, namely CD4T, CD8T, NK, and T-regulatory (Treg) cells. Each dot on the uMAP plot indicates a single cell. **(C)** Two NK sub-clusters identified by known NK markers. **(D)** Five $CD4^+$ T sub-clusters identified by known $CD4^+$ T markers. **(E)** Four $CD8^+$ Trm sub-clusters identified using gene expression values and the heatmap for known $CD8^+$ T markers. **(F)** Distribution of T cell receptor clonal expansion patterns in $CD4^+$ T and $CD8^+$ T cell subsets. Normalized stacked bar plot for each subset show the fraction of cells with different degrees of clonal expansion, with respect to Normal-adjacent and Tumor tissues. Cells with unsuccessful TCR assembly were excluded.

Figure 6. Interactions between epithelial clusters and immune subsets. (A) Overview of all ligand–receptor interactions between different epithelial clusters and immune subsets detected from tumor and normal tissues, respectively. **(B)** Dot plots of selected ligand–receptor interactions between epithelial cell clusters and immune subsets detected from tumor and normal tissues, respectively. The shade of data point's color reflects the means of the average expression level of interacting molecule 1 from cluster 1 and interacting molecule 2 from cluster 2. The negative $\log_{10}$ *p* values are indicated by circle size. **(C)** Multiplex immunohistochemistry staining of tumor tissue from Patient SC007 shows close proximity of the localized $CD8^+$ T cells expressing TIGIT and stem-like epithelial cells ($STMN1^+$) expressing NECTIN2.

Supplemental Figure legends

Figure S1. Metadata and clustering information of gastric cancer cells. **(A)** Violin plots shows quality control performed based on 3 criteria, number of RNA feature per cell (nFeature_RNA) > 200 and < 5000 and percentage of mitochondria RNA count per cell (percent.mt) < 50%. tSNE presentations with different metadata information of all cells collected: **(B)** Patient ID, **(C)** Site, **(D)** Lauren's histological classification, and **(E)** Pathological staging (T, Tumor size). **(F)** Heatmaps of differentially expressed genes across different clusters identified. Cluster 0,8, CD8T; 1, Fibroblast; 2,9, CD4T; 3, Epithelial 1; 4, NK; 5, B cell; 6, Epithelial 2; 7, Epithelial 1 – SC008; 10, Plasma cell; 11, Mast cell; 12, Epithelial 2 – SC005; 13, Macrophage. **(G)** Integration with dataset obtained from Zhang et al. and dashed lines represent areas of different cell types (green for MSC and red for epithelial and immune cells). GMC, antral basal gland mucous cell; MSC, metaplastic stem-like cell; NLC, neck-like cell; PMC, pit mucous cell.

Figure S2. Further analysis of epithelial cells. **(A)** tSNE projections of epithelial cells plotted with mRNA expression showing different metadata information (Patient ID, Site, Histological Site and Pathological Stage Tumor size). **(B)** uMAP presentation of epithelial cells using regulon activities with various metadata information (Patient ID, Site, Lauren's histological classification, and Pathological Stage T). **(C)** Violin plots of stemness-related genes expressed by different epithelial clusters. **(D)** Validation of SE cells through data integration with gastric cancer cells sequenced by 10x sequencing, SE cells locations were circled out. **(E)** tSNE plots with mean expression of signature markers for both cancer cell type and Metaplastic stem-like cell (MSC) type in epithelial cells. Gene Set Enrichment Analyses show significant **(F)** positive enrichments of antigen presentation-related genesets in Stem-like Epithelials (SE), **(G)** positive/negative enrichments of Epithelial-Mesenchymal-Transition-related genesets in Epithelial 1, **(H)** positive enrichments of stemness-related

genesets in Stem-like Epithelial A (SEA), **(I)** positive enrichment of glutamate secretion genesets in SEB cells. FDR<0.25 for all comparison, by phenotype permutation test.

Figure S3. Integration with TCGA data and survival analysis. **(A)** uMAP presentations show classification of TCGA data into different cell types according to clustering information obtained from our current study. Left to right: Site, Cell type information, Clusters after integration and cell type information for different clusters. **(B)** uMAP projections of epithelial cells and further classification of TCGA samples into different epithelial cell subtypes. Left to right: Site, Cell type information, Clusters after integration and cell type information for different clusters. **(C)** Survival analysis predicted with Kaplan-Meier estimator for one epithelial subtype against other epithelial subtypes. Left to right: Epithelial 1 (E1) vs E2 + Stem-like Epithelial A (SEA) + SEB, E2 vs E1 + SEA +SEB, SEA vs E1 + E2 + SEB, and SEB vs E1+ E2 + SEA. **(D)** Survival analysis predicted with Kaplan-Meier estimator for different age groups and stages. **(E)** Schematic presentation of a decision tree for stratifying TCGA patients using our epithelial cell subset signatures.

Figure S4. Comparison of MHC Class I- and II-related gene expressions between epithelial clusters. **(A)** Violin plots shows most of the MHC Class I-related genes were highly expressed by the Stem-like Epithelial A cluster, except for *TAP1* and *TAP2* genes. **(B)** Violin plots shows MHC Class II-related genes, including HLA-DRA, HLA-DRB1, HLA-E, HLA-F, and HLA-G, were highly expressed by the Stem-like Epithelial A cluster.

Figure S5. Multi-faceted epithelial cells along the Monocle 3 developmental trajectories. Epithelial cells colored by **(A)** the closest matching tissue predicted by the scHCL; **(B)** the closest matching cell type predicted by scHCL; **(C)** the site of tissue collection (Normal vs Tumor); **(D)** the mean of gastric cancer genes (Zhang *et al.*); **(E)** the mean of metaplastic stem cell (MSC) genes (Zhang *et al.*); **(F)** the tumor morphology, **(G)** the G2M score computed by the Seurat Cell Cycle Regression, **(H)** the mean of the Epithelial-

Mesenchymal-transition (EMT) up-regulated genes (Loboda *et al.*); **(I)** the mean of the EMT down-regulated genes (Loboda *et al.*); **(J)** the mean of the MicroSatellite Instability (MSI) genes (The Cancer Genome Atlas Research Network); **(K)** the patient sample ID.

Figure S6. FACS gating strategies. **(A)** FACS Panel 2 and **(B)** FACS Panel 3 used in this study.

Figure S7. Characterization of immune subsets. **(A)** Bubble plots shows significant positive correlation between proportion of Stem-like Epithelial cell (of all the epithelial cells) and proportion of CD3+ cell (of all CD45+ immune cells) was found in gastric normal-adjacent tissues (Blue bubbles), but not in tumor tissues (Red bubbles). **(B)** tSNE plots of all immune cells, colored by TCR group size, gene expressions of known markers for NK (*GNLY*), CD8 T (CD8A and CD8B), regulatory T (*IL2RA*), and CD4 T (*FOXP3*) cells. **(C)** High resident memory T-related gene expressions (including CD69, CD44, *ITGAE*), but low homing genes (CCR7 and LEF1) for all CD8 T cells identified in this study. **(D)** No significant difference in $CD45^+$ cells proportion of total live cells between Epithelial 1- and Epithelial 2-dominant patients, in both normal-adjacent and tumor tissues, respectively. **(E)** Comparison of $CD45^+$ cell proportions of total live cells between histological diffuse, intestinal, and mixed patients. The $CD45^+$ cell proportions in mixed patients were found to be significantly higher from the diffuse patients. **(F)** Immunohistochemistry staining of normal-adjacent and tumor tissues for an intestinal patient, SC019. $TIM3^+$ and $CD8^+$ cells were detected in both tissues.

Figure S8. Violin plots for expressions of interacting molecule from each epithelial cluster and interacting molecule from each immune subset involved in immune-suppressive interactions. The violins are plotted separately for normal-adjacent and tumor tissues.

Figure S9. Multiplex immunohistochemistry staining of gastric cancer tissue slides using EPCAM, CD8, TIGIT, STMN1, and NECTIN2 markers for Patients **(A)** SC004 and **(B)** SC018.

Supplemental Table legends

Table S1. Demographics and clinical information of 15 patients included in this study.

Table S2. Differential expression genes and the assigned celltypes of the 14 cell clusters identified from the 2878 cells sequenced in this study.

Table S3. Seurat clusters and celltypes/subtypes assigned for each cell (n=2878) identified in this study. N.A., Not applicable.

Table S4. Differential expression regulons between Stem-like Epithelial and Epithelial 1/2.

Table S5. Differential expression genes between Epithelial 1, Epithelial 2, and Stem-like Epithelial cells.

Table S6. Geneset Enrichment Analyses between epithelial clusters, using genesets obtained from the MSigDB Hallmark v7.0, MSigDB C5 - GO Biological Processes databases, Short-term hematopoietic stem cell (ST-HSC), Long-term HSC (LT-HSC), and stomach LGR5-positive (GSE86603).

Table S7. Biological pathways activated or down-regulated in the Epithelial 1 (E1), E2, Stem-like Epithelial cell populations.

Table S8. Differential expression genes between Stem-like Epithelial A (SEA) and SEB.

Table S9. Biological pathways activated or down-regulated in the Stem-like Epithelial A (SEA) and SEB cell populations

Table S10. Biological pathways that involve Clusters 1, 2 and 3 genes in the respective Epithelial 1 and Epithelial 2 trajectories predicted by Monocle 3.

Table S11. Spearman's Rank correlations of Epithelial 1, 2, and Stem-like Epithelial cell proportions with %CD45$^+$ of total live cells and %CD3$^+$ of total CD45$^+$ cells, in Normal-adjacent and Tumor tissues, respectively.

Table S12. Differential expression genes between immune subsets identified in each NK, CD4T and CD8T population.

Table S13. Biological pathways activated or down-regulated in the Exhausted CD4T cell population.

Table S14. Tabulation of NK, CD4T, CD8T and Treg immune subsets (n = 1390) according to Normal-adjacent and Tumor tissues.

Table S15. Comparison of single cell RNA-Seq-based and FACS-marker-based cell-type predictions. Only 698 cells (obtained from Patient SC016, SC017 and SC020) with FACS data available from Panel 3 testing are evaluated. Immune cell groups were characterized by CD45+; Epithelial cell groups by EpCAM+; T cells by CD45+CD3+; NK cells by CD45+CD56+ ; B cells and plasma cells by CD45+CD19+; Monocytes by CD45+CD3-

CD19-CD56-CD14+; DCs by CD45+CD3-CD19-CD56-CD14-HLA-DR+; Ungated CD45+CD3-CD19-CD56-CD14-HLA-DR-.

Figure 1

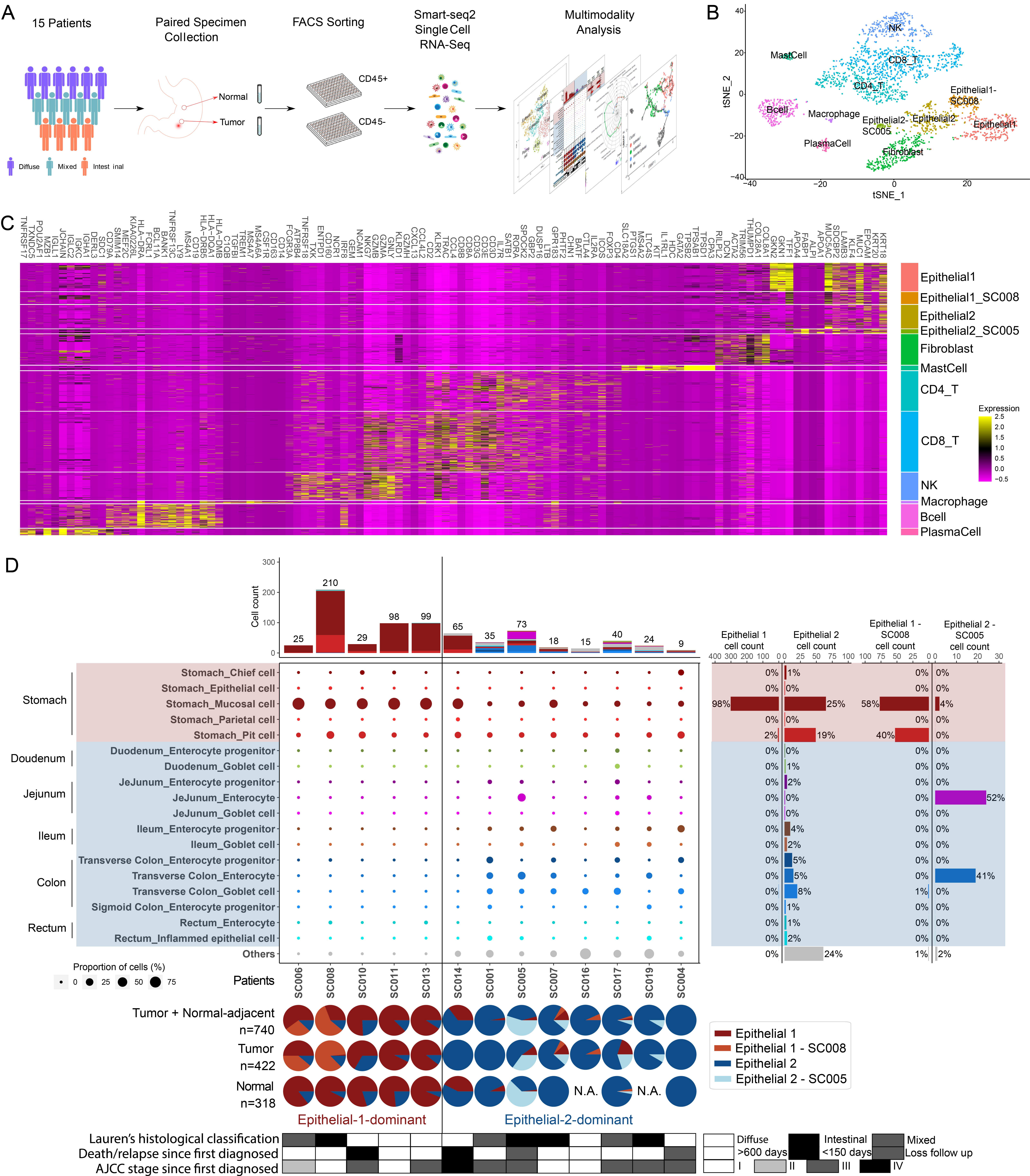

A
15 Patients
Diffuse
Mixed
Intest inal
Paired Specimen Collection
Normal
Tumor
FACS Sorting
CD45+
CD45-
Smart-seq2 Single Cell RNA-Seq
Multimodality Analysis
B
NK
CD8_T
MastCell
CD4_T
Epithelial1-SC008
Bcell
Macrophage
Epithelial2-SC005
Epithelial2
Epithelial1
PlasmaCell
Fibroblast
tSNE_2
tSNE_1
40
20
0
−20
−40
−40
−20
0
20
C
TNFRSF17
TXNDC5
POU2AF1
MZB1
IGLL1
JCHAIN
IGLC2
IGKC
IGHA1
DERL3
SDC1
CD79A
SMIM14
MEF2C
KIAA0226L
HLA-DRA
FCRL1
BCL11A
BANK1
TNFRSF13C
LY9
MS4A1
CD19
HLA-DRB5
HLA-DQA1
HLA-DMB
C1QB
TGFBI
TREM1
MS4A7
MS4A6A
CSF1R
CD163
CD14
FCGR3A
ATP8B4
TNFRSF18
TXK
ENTPD1
CD160
NCR1
IRF8
GEM
NCAM1
NKG7
GZMB
GZMA
GNLY
KLRD1
GZMH
CXCL13
CCL4L2
CD2
KLRK1
TRAC
CD8B
CD8A
CD3G
CD3E
CD3D
IL7R
SATB1
RORA
SPOCK2
GBP2
DUSP16
LTB
GPR183
PHTF2
CHN1
BATF
CTLA4
IL2RA
ICOS
FOXP3
CD4
SLC18A2
PTGS1
MS4A2
LTC4S
KIT
IL1RL1
HDC
GATA2
TPSB2
TPSAB1
TPSD1
CPA3
RILPL2
DCN
ACTA2
TRIM36
THUMPD1
COL28A1
COL8A1
GKN2
GKN1
TFF1
APOA4
FABP1
ALPI
APOA1
MUC5AC
SDCBP2
LAMB3
KLF4
MUC1
EPCAM
KRT20
KRT18
Epithelial1
Epithelial1_SC008
Epithelial2
Epithelial2_SC005
Fibroblast
MastCell
CD4_T
CD8_T
NK
Macrophage
Bcell
PlasmaCell
Expression
2.5
2.0
1.5
1.0
0.5
0.0
−0.5
D
Cell count
300
200
100
0
25
210
29
98
99
65
35
73
18
15
40
24
9
Epithelial 1 cell count
400 300 200 100 0
Epithelial 2 cell count
0 25 50 75 100
Epithelial 1 - SC008 cell count
100 75 50 25 0
Epithelial 2 - SC005 cell count
0 10 20 30
Stomach
Doudenum
Jejunum
Ileum
Colon
Rectum
Stomach_Chief cell
Stomach_Epithelial cell
Stomach_Mucosal cell
Stomach_Parietal cell
Stomach_Pit cell
Duodenum_Enterocyte progenitor
Duodenum_Goblet cell
JeJunum_Enterocyte progenitor
JeJunum_Enterocyte
JeJunum_Goblet cell
Ileum_Enterocyte progenitor
Ileum_Goblet cell
Transverse Colon_Enterocyte progenitor
Transverse Colon_Enterocyte
Transverse Colon_Goblet cell
Sigmoid Colon_Enterocyte progenitor
Rectum_Enterocyte
Rectum_Inflammed epithelial cell
Others
0% 1% 0% 0%
0% 0% 0% 0%
98% 25% 58% 4%
0% 0% 0% 0%
2% 19% 40% 0%
0% 0% 0% 0%
0% 1% 0% 0%
0% 2% 0% 0%
0% 0% 0% 52%
0% 0% 0% 0%
0% 4% 0% 0%
0% 2% 0% 0%
0% 5% 0% 0%
0% 5% 0% 41%
0% 8% 1% 0%
0% 1% 0% 0%
0% 1% 0% 0%
0% 2% 0% 0%
0% 24% 1% 2%
Proportion of cells (%)
0
25
50
75
Patients
SC006
SC008
SC010
SC011
SC013
SC014
SC001
SC005
SC007
SC016
SC017
SC019
SC004
Tumor + Normal-adjacent n=740
Tumor n=422
Normal n=318
N.A.
N.A.
Epithelial 1
Epithelial 1 - SC008
Epithelial 2
Epithelial 2 - SC005
Epithelial-1-dominant
Epithelial-2-dominant
Lauren’s histological classification
Death/relapse since first diagnosed
AJCC stage since first diagnosed
Diffuse
Intestinal
Mixed
>600 days
<150 days
Loss follow up
I
II
III
IV

Figure 2

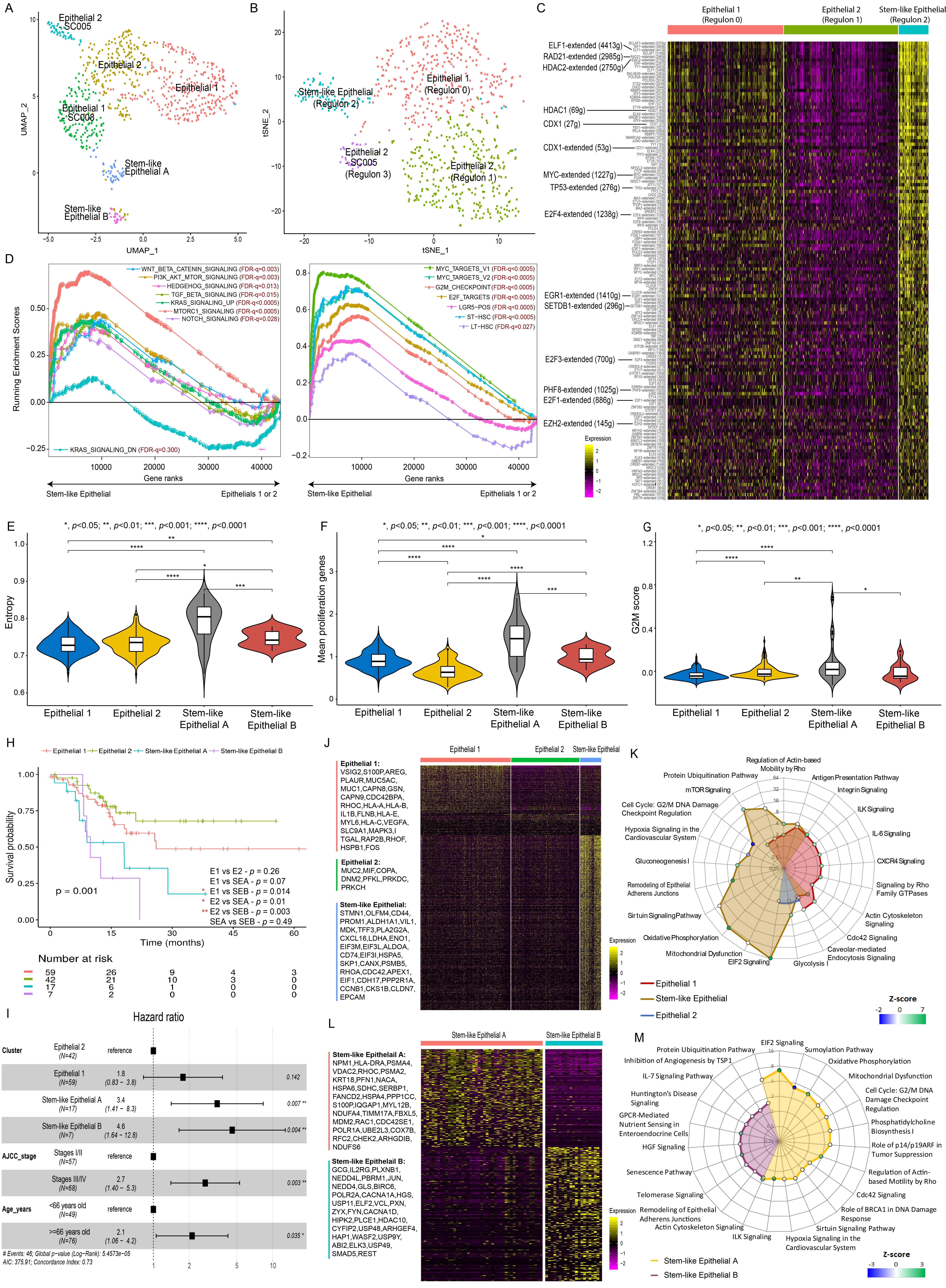

A
Epithelial 2 SC005
Epithelial 2
Epithelial 1
Epithelial 1 SC008
Stem-like Epithelial A
Stem-like Epithelial B
UMAP_1
UMAP_2
B
Stem-like Epithelial (Regulon 2)
Epithelial 1 (Regulon 0)
Epithelial 2 SC005 (Regulon 3)
Epithelial 2 (Regulon 1)
tSNE_1
tSNE_2
C
Epithelial 1 (Regulon 0)
Epithelial 2 (Regulon 1)
Stem-like Epithelial (Regulon 2)
ELF1-extended (4413g)
RAD21-extended (2985g)
HDAC2-extended (2750g)
HDAC1 (69g)
CDX1 (27g)
CDX1-extended (53g)
MYC-extended (1227g)
TP53-extended (276g)
E2F4-extended (1238g)
EGR1-extended (1410g)
SETDB1-extended (296g)
E2F3-extended (700g)
PHF8-extended (1025g)
E2F1-extended (886g)
EZH2-extended (145g)
Expression
D
WNT_BETA_CATENIN_SIGNALING (FDR-q=0.003)
PI3K_AKT_MTOR_SIGNALING (FDR-q=0.003)
HEDGEHOG_SIGNALING (FDR-q=0.013)
TGF_BETA_SIGNALING (FDR-q=0.015)
KRAS_SIGNALING_UP (FDR-q<0.0005)
MTORC1_SIGNALING (FDR-q<0.0005)
NOTCH_SIGNALING (FDR-q=0.028)
KRAS_SIGNALING_DN (FDR-q=0.300)
Running Enrichment Scores
Gene ranks
Stem-like Epithelial
Epithelials 1 or 2
MYC_TARGETS_V1 (FDR-q<0.0005)
MYC_TARGETS_V2 (FDR-q<0.0005)
G2M_CHECKPOINT (FDR-q<0.0005)
E2F_TARGETS (FDR-q<0.0005)
LGR5-POS (FDR-q<0.0005)
ST-HSC (FDR-q<0.0005)
LT-HSC (FDR-q=0.027)
Gene ranks
Stem-like Epithelial
Epithelials 1 or 2
E
*, p<0.05; **, p<0.01; ***, p<0.001; ****, p<0.0001
Entropy
Epithelial 1
Epithelial 2
Stem-like Epithelial A
Stem-like Epithelial B
F
*, p<0.05; **, p<0.01; ***, p<0.001; ****, p<0.0001
Mean proliferation genes
Epithelial 1
Epithelial 2
Stem-like Epithelial A
Stem-like Epithelial B
G
*, p<0.05; **, p<0.01; ***, p<0.001; ****, p<0.0001
G2M score
Epithelial 1
Epithelial 2
Stem-like Epithelial A
Stem-like Epithelial B
H
Epithelial 1
Epithelial 2
Stem-like Epithelial A
Stem-like Epithelial B
Survival probability
p = 0.001
E1 vs E2 - p = 0.26
E1 vs SEA - p = 0.07
* E1 vs SEB - p = 0.014
* E2 vs SEA - p = 0.01
** E2 vs SEB - p = 0.003
SEA vs SEB - p = 0.49
Time (months)
Number at risk
59 26 9 4 3
42 21 10 3 0
17 6 1 0 0
7 2 0 0 0
I
Hazard ratio
Cluster
Epithelial 2 (N=42) reference
Epithelial 1 (N=59) 1.8 (0.83 - 3.8) 0.142
Stem-like Epithelial A (N=17) 3.4 (1.41 - 8.3) 0.007 **
Stem-like Epithelial B (N=7) 4.6 (1.64 - 12.8) 0.004 **
AJCC_stage
Stages I/II (N=57) reference
Stages III/IV (N=68) 2.7 (1.40 - 5.3) 0.003 **
Age_years
<66 years old (N=49) reference
>=66 years old (N=76) 2.1 (1.06 - 4.2) 0.035 *
# Events: 46; Global p-value (Log-Rank): 5.4573e-05
AIC: 375.91; Concordance Index: 0.73
J
Epithelial 1
Epithelial 2
Stem-like Epithelial
Epithelial 1:
VSIG2,S100P,AREG,PLAUR,MUC5AC,MUC1,CAPN8,GSN,CAPN9,CDC42BPA,RHOC,HLA-A,HLA-B,IL1B,FLNB,HLA-E,MYL6,HLA-C,VEGFA,SLC9A1,MAPK3,ITGAL,RAP2B,RHOF,HSPB1,FOS
Epithelial 2:
MUC2,MIF,COPA,DNM2,PFKL,PRKDC,PRKCH
Stem-like Epithelial:
STMN1,OLFM4,CD44,PROM1,ALDH1A1,VIL1,MDK,TFF3,PLA2G2A,CXCL16,LDHA,ENO1,EIF3M,EIF3L,ALDOA,CD74,EIF3I,HSPA5,SKP1,CANX,PSMB5,RHOA,CDC42,APEX1,EIF1,CDH17,PPP2R1A,CCNB1,CKS1B,CLDN7,EPCAM
Expression
K
Regulation of Actin-based Mobility by Rho
Antigen Presentation Pathway
Integrin Signaling
ILK Signaling
IL-6 Signaling
CXCR4 Signaling
Signaling by Rho Family GTPases
Actin Cytoskeleton Signaling
Cdc42 Signaling
Caveolar-mediated Endocytosis Signaling
Glycolysis I
EIF2 Signaling
Mitochondrial Dysfunction
Oxidative Phosphorylation
Sirtuin Signaling Pathway
Remodeling of Epithelial Adherens Junctions
Gluconeogenesis I
Hypoxia Signaling in the Cardiovascular System
Cell Cycle: G2/M DNA Damage Checkpoint Regulation
mTOR Signaling
Protein Ubiquitination Pathway
Epithelial 1
Stem-like Epithelial
Epithelial 2
Z-score
L
Stem-like Epithelial A
Stem-like Epithelial B
Stem-like Epithelail A:
NPM1,HLA-DRA,PSMA4,VDAC2,RHOC,PSMA2,KRT18,PFN1,NACA,HSPA6,SDHC,SERBP1,FANCD2,HSPA4,PPP1CC,S100P,IQGAP1,MYL12B,NDUFA4,TIMM17A,FBXL5,MDM2,RAC1,CDC42SE1,POLR1A,UBE2L3,COX7B,RFC2,CHEK2,ARHGDIB,NDUFS6
Stem-like Epithelail B:
GCG,IL2RG,PLXNB1,NEDD4L,PBRM1,JUN,NEDD4,GLS,BIRC6,POLR2A,CACNA1A,HGS,USP11,ELF2,VCL,PXN,ZYX,FYN,CACNA1D,HIPK2,PLCE1,HDAC10,CYFIP2,USP48,ARHGEF4,HAP1,WASF2,USP9Y,ABI2,ELK3,USP49,SMAD5,REST
Expression
M
EIF2 Signaling
Sumoylation Pathway
Oxidative Phosphorylation
Mitochondrial Dysfunction
Cell Cycle: G2/M DNA Damage Checkpoint Regulation
Phosphatidylcholine Biosynthesis I
Role of p14/p19ARF in Tumor Suppression
Regulation of Actin-based Motility by Rho
Cdc42 Signaling
Role of BRCA1 in DNA Damage Response
Sirtuin Signaling Pathway
Hypoxia Signaling in the Cardiovascular System
ILK Signaling
Actin Cytoskeleton Signaling
Remodeling of Epithelial Adherens Junctions
Telomerase Signaling
Senescence Pathway
HGF Signaling
GPCR-Mediated Nutrient Sensing in Enteroendocrine Cells
Huntington's Disease Signaling
IL-7 Signaling Pathway
Inhibition of Angiogenesis by TSP1
Protein Ubiquitination Pathway
Stem-like Epithelial A
Stem-like Epithelial B
Z-score

Figure 3

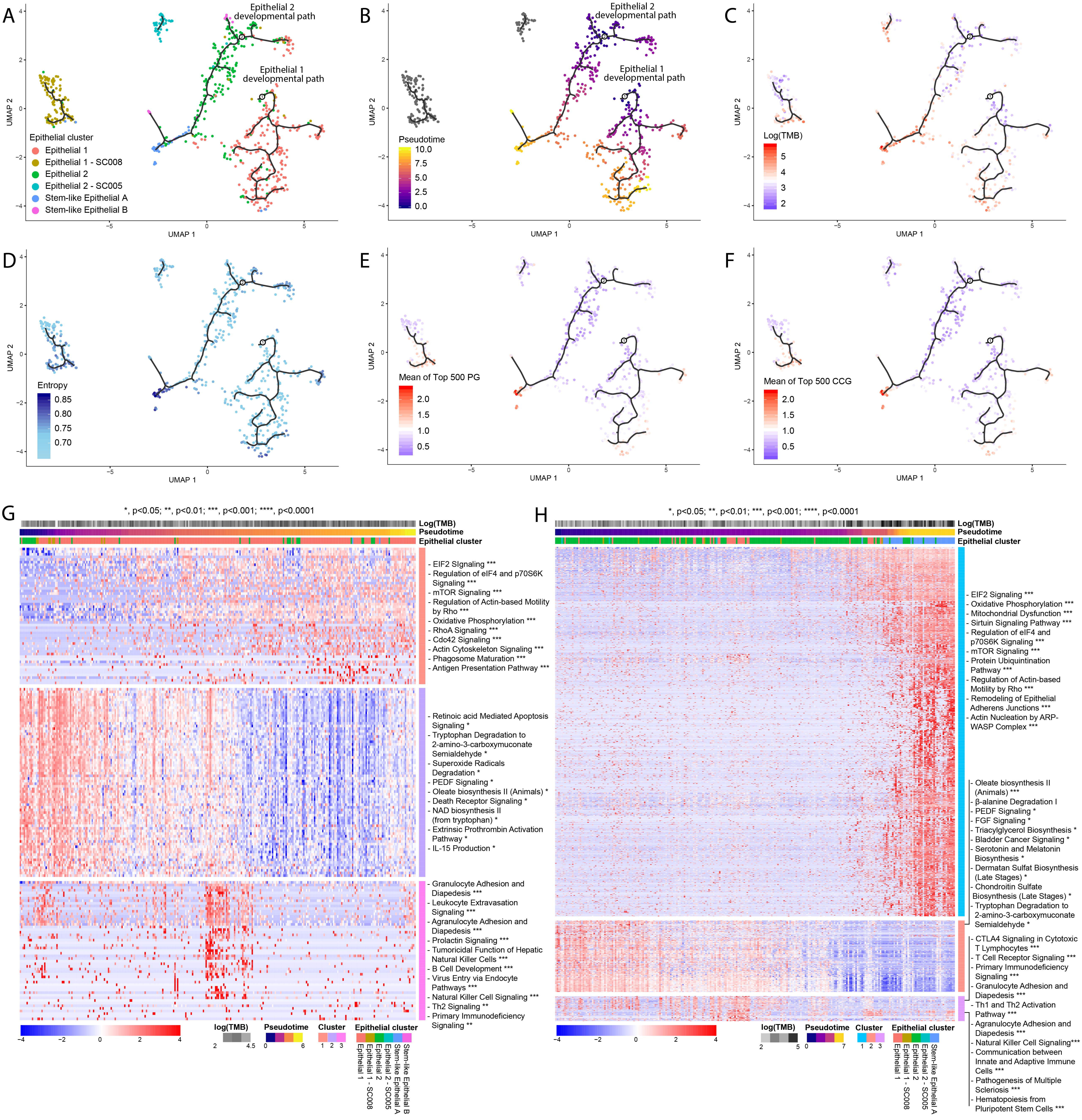
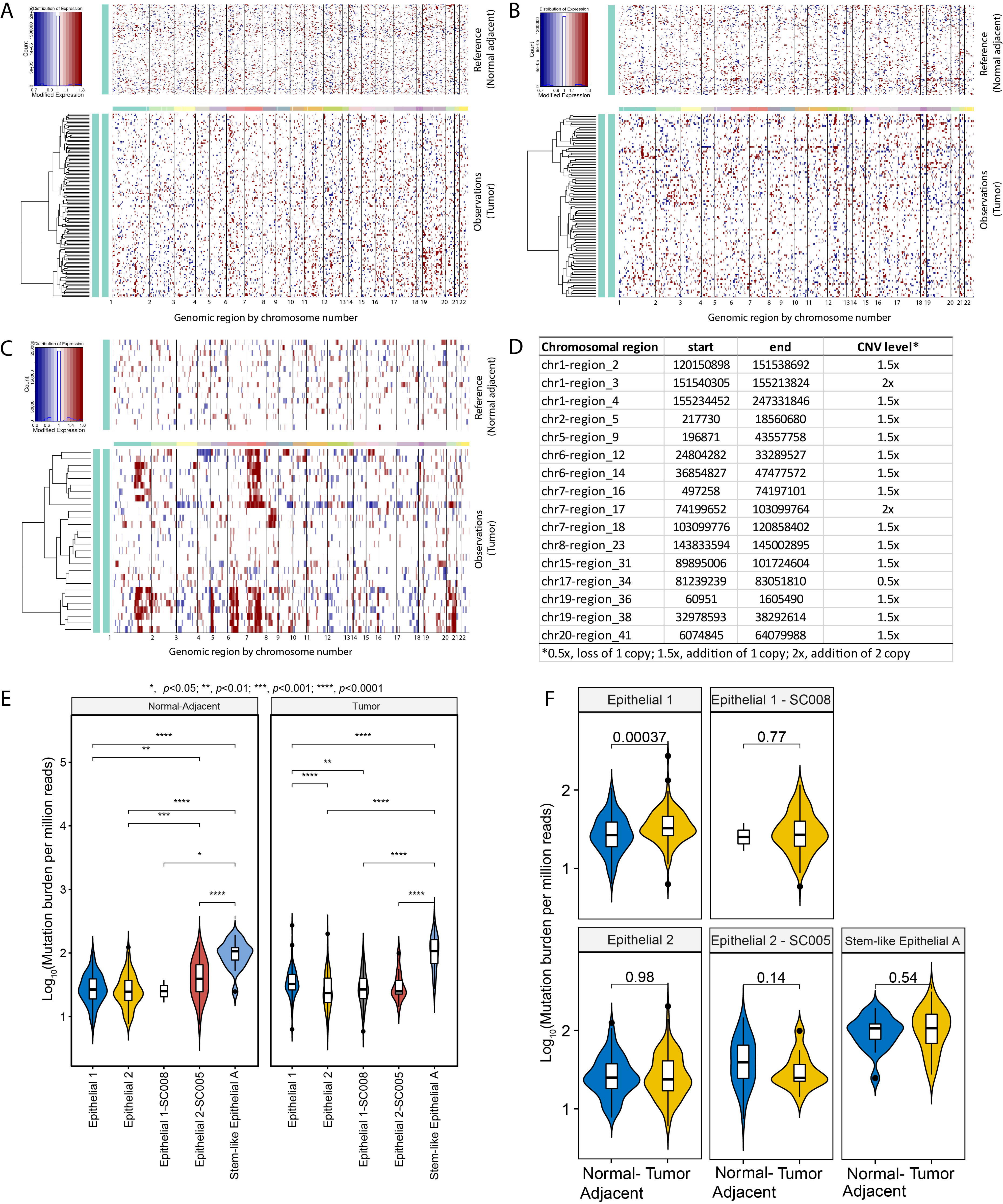

A
Epithelial 2 developmental path
Epithelial 1 developmental path
Epithelial cluster
Epithelial 1
Epithelial 1 - SC008
Epithelial 2
Epithelial 2 - SC005
Stem-like Epithelial A
Stem-like Epithelial B
UMAP 1
UMAP 2
B
Pseudotime
10.0
7.5
5.0
2.5
0.0
C
Log(TMB)
5
4
3
2
D
Entropy
0.85
0.80
0.75
0.70
E
Mean of Top 500 PG
2.0
1.5
1.0
0.5
F
Mean of Top 500 CCG
2.0
1.5
1.0
0.5
-5
0
5
4
2
0
-2
-4
G
*, p<0.05; **, p<0.01; ***, p<0.001; ****, p<0.0001
Log(TMB)
Pseudotime
Epithelial cluster
- EIF2 SIgnaling ***
- Regulation of eIF4 and p70S6K Signaling ***
- mTOR Signaling ***
- Regulation of Actin-based Motility by Rho ***
- Oxidative Phosphorylation ***
- RhoA Signaling ***
- Cdc42 Signaling ***
- Actin Cytoskeleton Signaling ***
- Phagosome Maturation ***
- Antigen Presentation Pathway ***
- Retinoic acid Mediated Apoptosis Signaling *
- Tryptophan Degradation to 2-amino-3-carboxymuconate Semialdehyde *
- Superoxide Radicals Degradation *
- PEDF Signaling *
- Oleate biosynthesis II (Animals) *
- Death Receptor Signaling *
- NAD biosynthesis II (from tryptophan) *
- Extrinsic Prothrombin Activation Pathway *
- IL-15 Production *
- Granulocyte Adhesion and Diapedesis ***
- Leukocyte Extravasation Signaling ***
- Agranulocyte Adhesion and Diapedesis ***
- Prolactin Signaling ***
- Tumoricidal Function of Hepatic Natural Killer Cells ***
- B Cell Development ***
- Virus Entry via Endocyte Pathways ***
- Natural Killer Cell Signaling ***
- Th2 Signaling **
- Primary Immunodeficiency Signaling **
-4
-2
0
2
4
log(TMB)
2
4.5
Pseudotime
0
6
Cluster
1
2
3
Epithelial cluster
Epithelial 1
Epithelial 1 - SC008
Epithelial 2
Epithelial 2 - SC005
Stem-like Epithelial A
Stem-like Epithelial B
H
*, p<0.05; **, p<0.01; ***, p<0.001; ****, p<0.0001
Log(TMB)
Pseudotime
Epithelial cluster
- EIF2 Signaling ***
- Oxidative Phosphorylation ***
- Mitochondrial Dysfunction ***
- Sirtuin Signaling Pathway ***
- Regulation of eIF4 and p70S6K Signaling ***
- mTOR Signaling ***
- Protein Ubiquintination Pathway ***
- Regulation of Actin-based Motility by Rho ***
- Remodeling of Epithelial Adherens Junctions ***
- Actin Nucleation by ARP-WASP Complex ***
- Oleate biosynthesis II (Animals) ***
- β-alanine Degradation I
- PEDF Signaling *
- FGF Signaling *
- Triacylglycerol Biosynthesis *
- Bladder Cancer Signaling *
- Serotonin and Melatonin Biosynthesis *
- Dermatan Sulfat Biosynthesis (Late Stages) *
- Chondroitin Sulfate Biosynthesis (Late Stages) *
- Tryptophan Degradation to 2-amino-3-carboxymuconate Semialdehyde *
- CTLA4 Signaling in Cytotoxic T Lymphocytes ***
- T Cell Receptor Signaling ***
- Primary Immunodeficiency Signaling ***
- Granulocyte Adhesion and Diapedesis ***
- Th1 and Th2 Activation Pathway ***
- Agranulocyte Adhesion and Diapedesis ***
- Natural Killer Cell Signaling***
- Communication between Innate and Adaptive Immune Cells ***
- Pathogenesis of Multiple Scleriosis ***
- Hematopoiesis from Pluripotent Stem Cells ***
-4
-2
0
2
4
log(TMB)
2
5
Pseudotime
0
7
Cluster
1
2
3
Epithelial cluster
Epithelial 1
Epithelial 1 - SC008
Epithelial 2
Epithelial 2 - SC005
Stem-like Epithelial A

Figure 4



Chromosomal region	start	end	CNV level*
chr1-region_2	120150898	151538692	1.5x
chr1-region_3	151540305	155213824	2x
chr1-region_4	155234452	247331846	1.5x
chr2-region_5	217730	18560680	1.5x
chr5-region_9	196871	43557758	1.5x
chr6-region_12	24804282	33289527	1.5x
chr6-region_14	36854827	47477572	1.5x
chr7-region_16	497258	74197101	1.5x
chr7-region_17	74199652	103099764	2x
chr7-region_18	103099776	120858402	1.5x
chr8-region_23	143833594	145002895	1.5x
chr15-region_31	89895006	101724604	1.5x
chr17-region_34	81239239	83051810	0.5x
chr19-region_36	60951	1605490	1.5x
chr19-region_38	32978593	38292614	1.5x
chr20-region_41	6074845	64079988	1.5x

*0.5x, loss of 1 copy; 1.5x, addition of 1 copy; 2x, addition of 2 copy

Figure 5

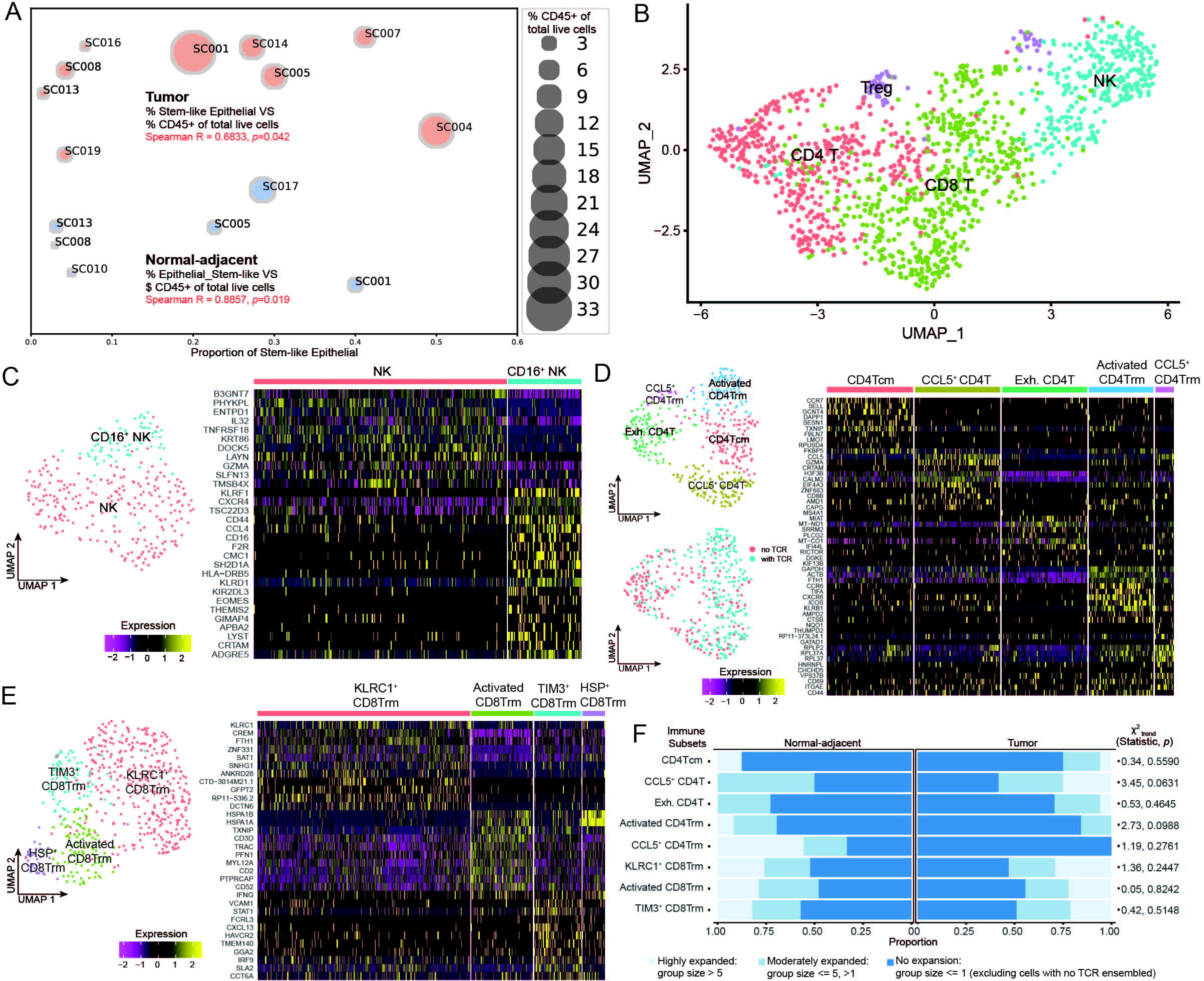
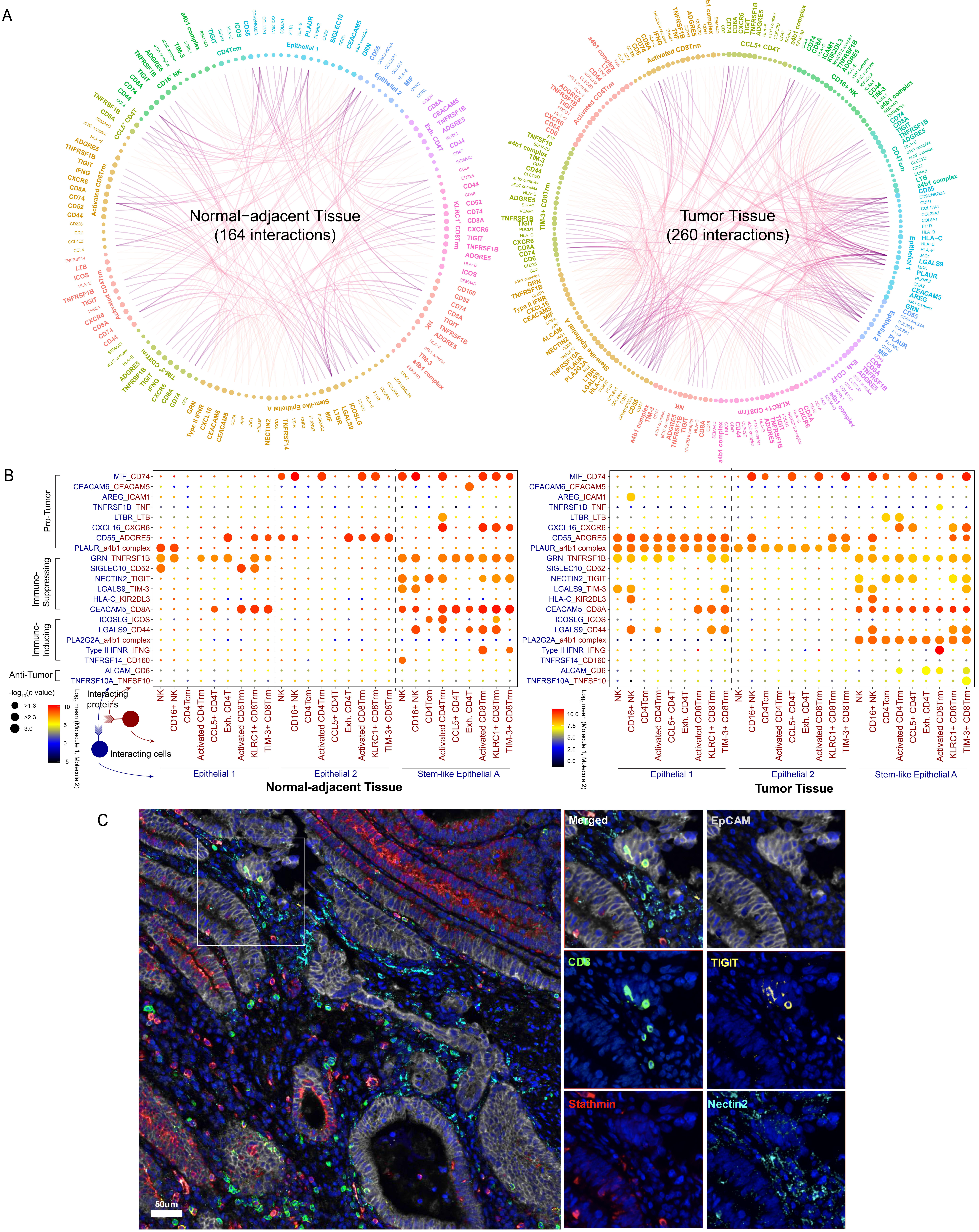

A
% CD45+ of total live cells
3
6
9
12
15
18
21
24
27
30
33
SC016
SC001
SC014
SC007
SC008
SC005
SC013
SC004
SC019
SC017
SC013
SC005
SC008
SC010
SC001
Tumor
% Stem-like Epithelial VS
% CD45+ of total live cells
Spearman R = 0.6833, p=0.042
Normal-adjacent
% Epithelial_Stem-like VS
$ CD45+ of total live cells
Spearman R = 0.8857, p=0.019
0.0
0.1
0.2
0.3
0.4
0.5
0.6
Proportion of Stem-like Epithelial
B
Treg
NK
CD4 T
CD8 T
UMAP_2
UMAP_1
2.5
0.0
−2.5
−6
−3
0
3
6
C
CD16⁺ NK
NK
UMAP 2
UMAP 1
Expression
−2
−1
0
1
2
NK
CD16⁺ NK
B3GNT7
PHYKPL
ENTPD1
IL32
TNFRSF18
KRT86
DOCK5
LAYN
GZMA
SLFN13
TMSB4X
KLRF1
CXCR4
TSC22D3
CD44
CCL4
CD16
F2R
CMC1
SH2D1A
HLA-DRB5
KLRD1
KIR2DL3
EOMES
THEMIS2
GIMAP4
APBA2
LYST
CRTAM
ADGRE5
D
CCL5⁺ CD4Trm
Activated CD4Trm
Exh. CD4T
CD4Tcm
CCL5⁺ CD4T
UMAP 2
UMAP 1
no TCR
with TCR
UMAP 2
UMAP 1
Expression
−2
−1
0
1
2
CD4Tcm
CCL5⁺ CD4T
Exh. CD4T
Activated CD4Trm
CCL5⁺ CD4Trm
CCR7
SELL
GCNT4
DAPP1
SESN1
TXNIP
FBLN7
LMO7
RPUSD4
FKBP5
CCL5
GZMA
CRTAM
H3F3B
CALM2
EIF4A3
ZNF683
CD8B
AMD1
CAPG
MS4A1
MIAT
MT−ND1
SRRM2
PLCG2
MT−CO1
IFI44L
RICTOR
DGKE
KIF13B
GAPDH
ACTB
FTH1
CCR6
TIFA
CXCR6
ICOS
KLRB1
AMPD2
CTSB
NQO1
THUMPD2
RP11−373L24.1
GATAD1
RPLP2
RPL37A
RPL37
HNRNPL
CHCHD5
VPS37B
CD69
ITGAE
CD44
E
TIM3⁺ CD8Trm
KLRC1⁺ CD8Trm
HSP⁺ CD8Trm
Activated CD8Trm
UMAP 2
UMAP 1
Expression
−2
−1
0
1
2
KLRC1⁺ CD8Trm
Activated CD8Trm
TIM3⁺ CD8Trm
HSP⁺ CD8Trm
KLRC1
CREM
FTH1
ZNF331
SAT1
SNHG1
ANKRD28
CTD−3014M21.1
GFPT2
RP11−53I6.2
DCTN6
HSPA1B
HSPA1A
TXNIP
CD3D
TRAC
PFN1
MYL12A
CD2
PTPRCAP
CD52
IFNG
VCAM1
STAT1
FCRL3
CXCL13
HAVCR2
TMEM140
GGA2
IRF9
SLA2
CCT6A
F
Immune Subsets
Normal-adjacent
Tumor
χ^2_{trend} (Statistic, p)
CD4Tcm
CCL5⁺ CD4T
Exh. CD4T
Activated CD4Trm
CCL5⁺ CD4Trm
KLRC1⁺ CD8Trm
Activated CD8Trm
TIM3⁺ CD8Trm
0.34, 0.5590
3.45, 0.0631
0.53, 0.4645
2.73, 0.0988
1.19, 0.2761
1.36, 0.2447
0.05, 0.8242
0.42, 0.5148
1.00
0.75
0.50
0.25
0.00
0.25
0.50
0.75
1.00
Proportion
Highly expanded: group size > 5
Moderately expanded: group size <= 5, >1
No expansion: group size <= 1 (excluding cells with no TCR ensembled)

Figure 6


A
Normal−adjacent Tissue
(164 interactions)
Tumor Tissue
(260 interactions)
Epithelial 1
Epithelial 2
Stem-like Epithelial A
CD4Tcm
CD16+ NK
CCL5+ CD4T
Activated CD8Trm
Activated CD4Trm
TIM-3+ CD8Trm
Exh. CD4T
KLRC1+ CD8Trm
NK
B
Pro-Tumor
Immuno-Suppressing
Immuno-Inducing
Anti-Tumor
MIF_CD74
CEACAM6_CEACAM5
AREG_ICAM1
TNFRSF1B_TNF
LTBR_LTB
CXCL16_CXCR6
CD55_ADGRE5
PLAUR_a4b1 complex
GRN_TNFRSF1B
SIGLEC10_CD52
NECTIN2_TIGIT
LGALS9_TIM-3
HLA-C_KIR2DL3
CEACAM5_CD8A
ICOSLG_ICOS
LGALS9_CD44
PLA2G2A_a4b1 complex
Type II IFNR_IFNG
TNFRSF14_CD160
ALCAM_CD6
TNFRSF10A_TNFSF10
-log10(p value)
>1.3
>2.3
3.0
Log2 mean (Molecule 1, Molecule 2)
10
5
0
-5
10.0
7.5
5.0
2.5
0.0
Interacting proteins
Interacting cells
NK
CD16+ NK
CD4Tcm
Activated CD4Trm
CCL5+ CD4T
Exh. CD4T
Activated CD8Trm
KLRC1+ CD8Trm
TIM-3+ CD8Trm
Epithelial 1
Epithelial 2
Stem-like Epithelial A
Normal-adjacent Tissue
Tumor Tissue
C
Merged
EpCAM
CD8
TIGIT
Stathmin
Nectin2
50um

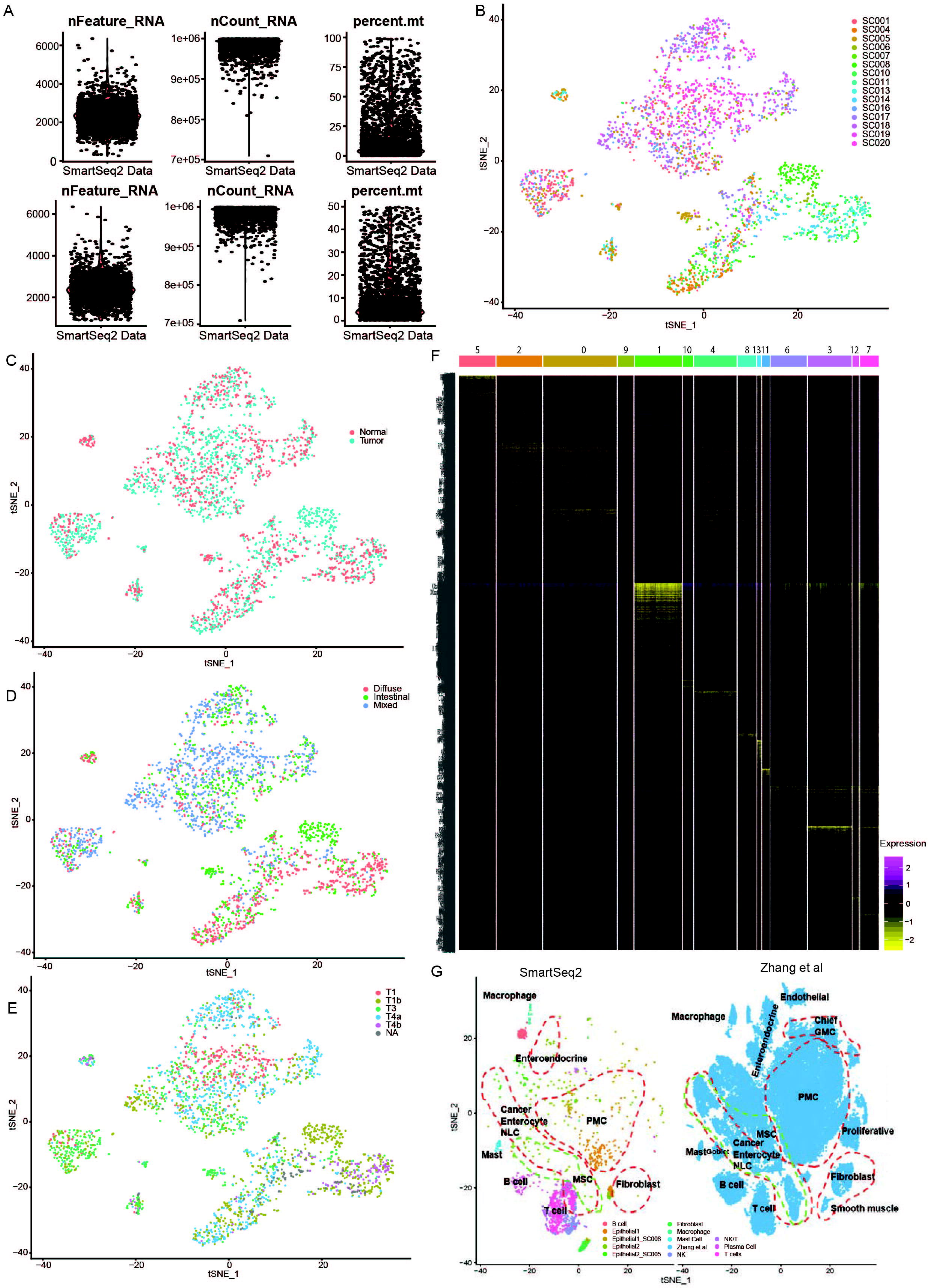


Figure S1. Metadata and clustering information of gastric cancer cells. **(A)** Violin plots shows quality control performed based on 3 criteria, number of RNA feature per cell (nFeature_RNA) > 200 and < 5000 and percentage of mitochondria RNA count per cell (percent.mt) < 50%. tSNE presentations with different metadata information of all cells collected: **(B)** Patient ID, **(C)** Site, **(D)** Lauren's histological classification, and **(E)** Pathological staging (T, Tumor size). **(F)** Heatmaps of differentially expressed genes across different clusters identified. Cluster 0,8, CD8T; 1, Fibroblast; 2,9, CD4T; 3, Epithelial 1; 4, NK; 5, B cell; 6, Epithelial 2; 7, Epithelial 1 – SC008; 10, Plasma cell; 11, Mast cell; 12, Epithelial 2 – SC005; 13, Macrophage. **(G)** Integration with dataset obtained from Zhang et al. and dashed lines represent areas of different cell types (green for MSC and red for epithelial and immune cells). GMC, antral basal gland mucous cell; MSC, metaplastic stem-like cell; NLC, neck-like cell; PMC, pit mucous cell.

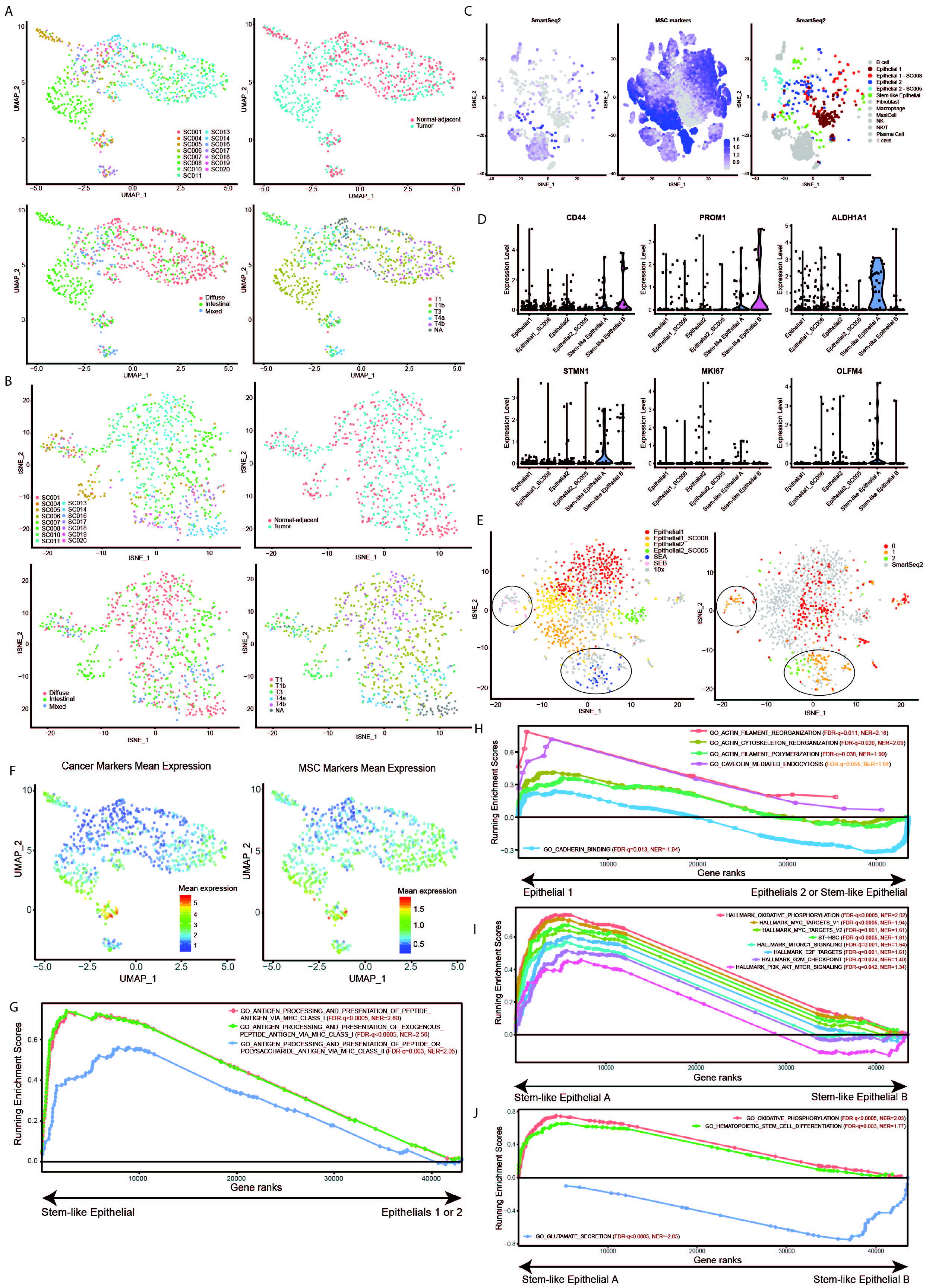
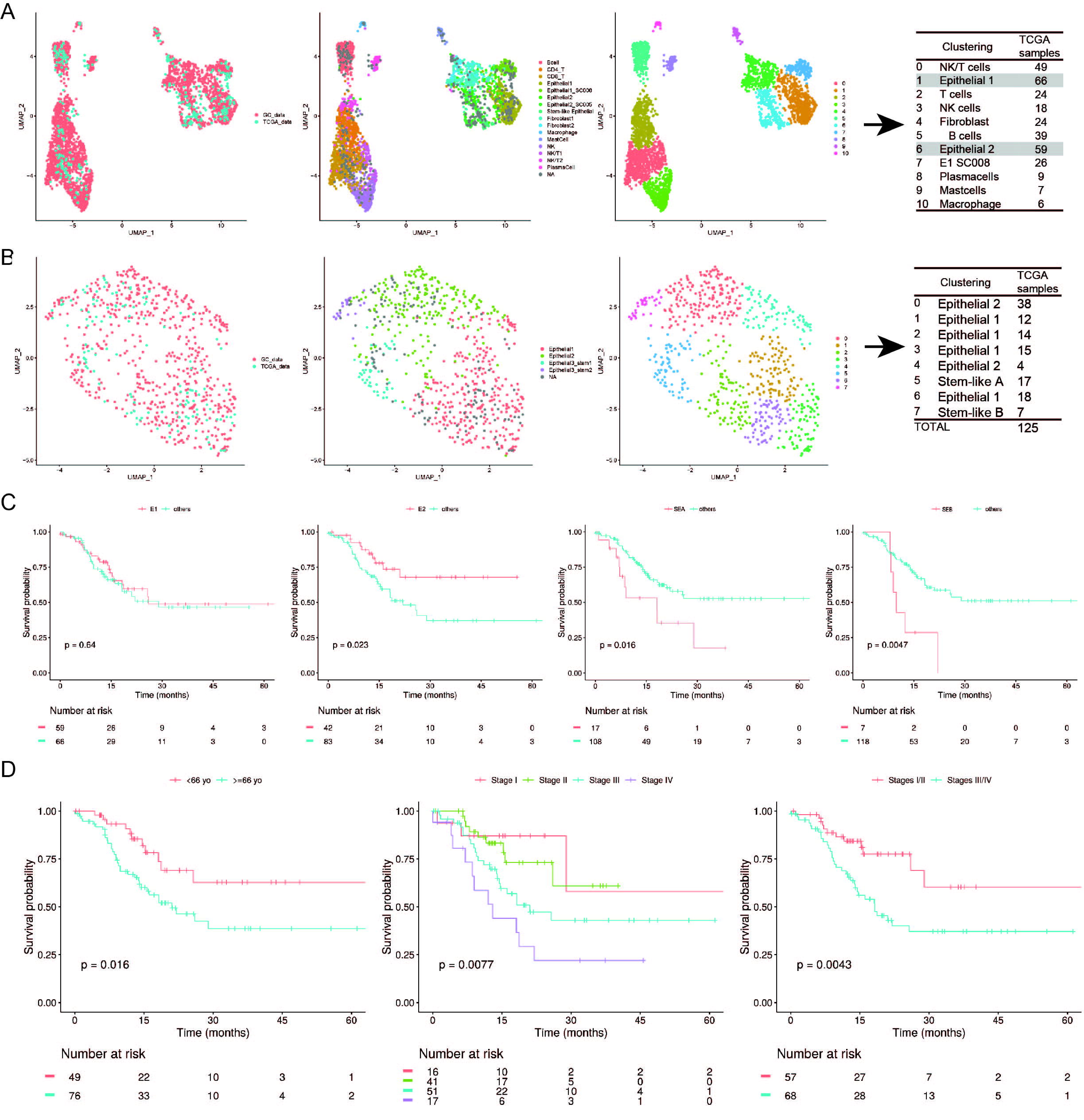


Figure S2. Further analysis of epithelial cells. **(A)** tSNE projections of epithelial cells plotted with mRNA expression showing different metadata information (Patient ID, Site, Histological Site and Pathological Stage Tumor size). **(B)** uMAP presentation of epithelial cells using regulon activities with various metadata information (Patient ID, Site, Lauren's histological classification, and Pathological Stage T). **(C)** Violin plots of stemness-related genes expressed by different epithelial clusters. **(D)** Validation of SE cells through data integration with gastric cancer cells sequenced by 10x sequencing, SE cells locations were circled out. **(E)** tSNE plots with mean expression of signature markers for both cancer cell type and Metaplastic stem-like cell (MSC) type in epithelial cells. Gene Set Enrichment Analyses show significant **(F)** positive enrichments of antigen presentation-related genesets in Stem-like Epithelials (SE), **(G)** positive/ negative enrichments of Epithelial-Mesenchymal-Transition-related genesets in Epithelial 1, **(H)** positive enrichments of stemness-related genesets in Stem-like Epithelial A (SEA), **(I)** positive enrichment of glutamate secretion genesets in SEB cells. FDR<0.25 for all comparison, by phenotype permutation test.



	Clustering	TCGA samples
0	NK/T cells	49
1	Epithelial 1	66
2	T cells	24
3	NK cells	18
4	Fibroblast	24
5	B cells	39
6	Epithelial 2	59
7	E1 SC008	26
8	Plasmacells	9
9	Mastcells	7
10	Macrophage	6

	Clustering	TCGA samples
0	Epithelial 2	38
1	Epithelial 1	12
2	Epithelial 1	14
3	Epithelial 1	15
4	Epithelial 2	4
5	Stem-like A	17
6	Epithelial 1	18
7	Stem-like B	7
TOTAL		125

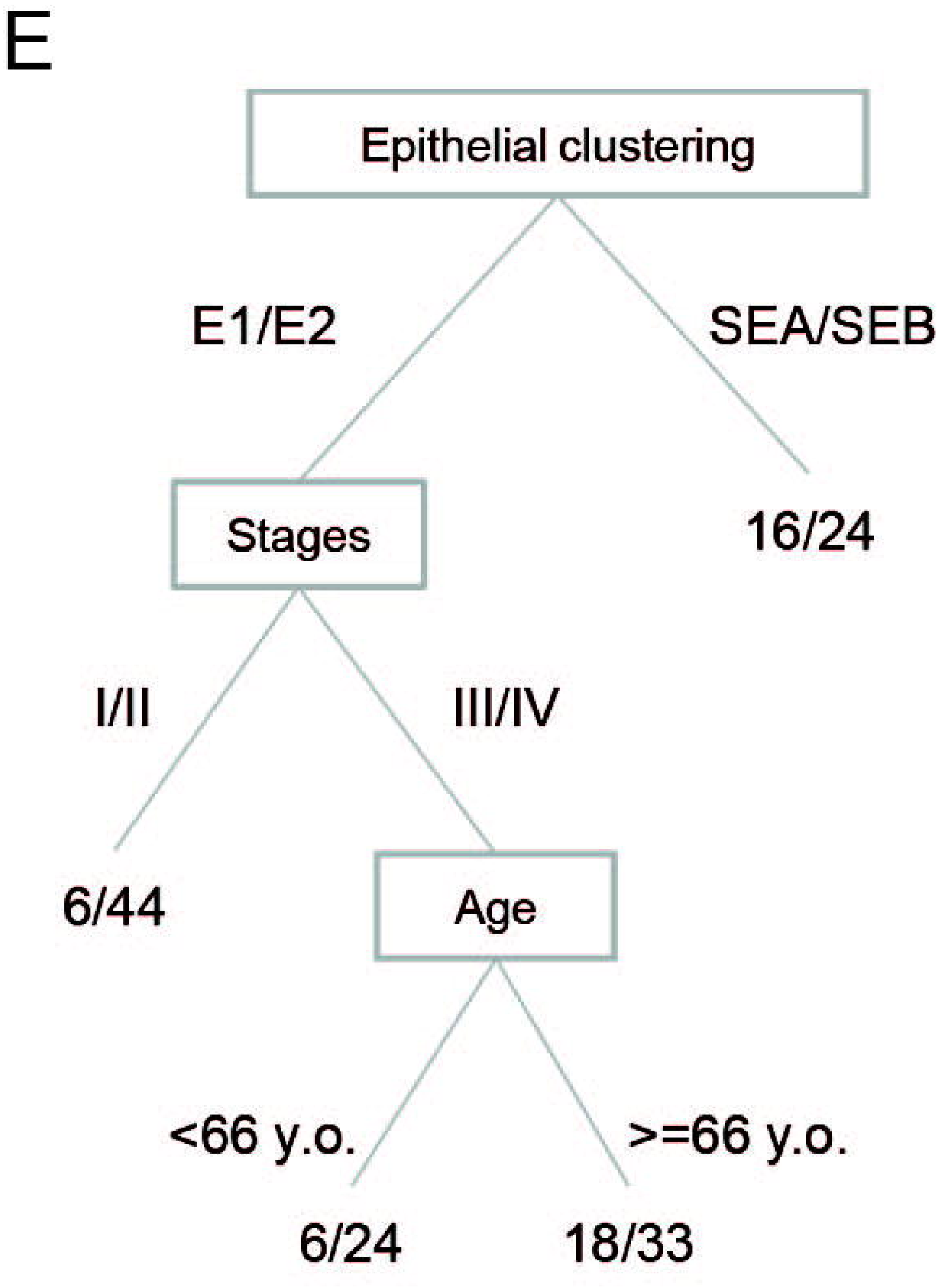


Figure S3. Integration with TCGA data and survival analysis. **(A)** uMAP presentations show classification of TCGA data into different cell types according to clustering information obtained from our current study. Left to right: Site, Cell type information, Clusters after integration and cell type information for different clusters. **(B)** uMAP projections of epithelial cells and further classification of TCGA samples into different epithelial cell subtypes. Left to right: Site, Cell type information, Clusters after integration and cell type information for different clusters. **(C)** Survival analysis predicted with Kaplan-Meier estimator for one epithelial subtype against other epithelial subtypes. Left to right: Epithelial 1 (E1) vs E2 + Stem-like Epithelial A (SEA) + SEB, E2 vs E1 + SEA +SEB, SEA vs E1 + E2 + SEB, and SEB vs E1+ E2 + SEA. **(D)** Survival analysis predicted with Kaplan-Meier estimator for different age groups and stages. **(E)** Schematic presentation of a decision tree for stratifying TCGA patients using our epithelial cell subset signatures.

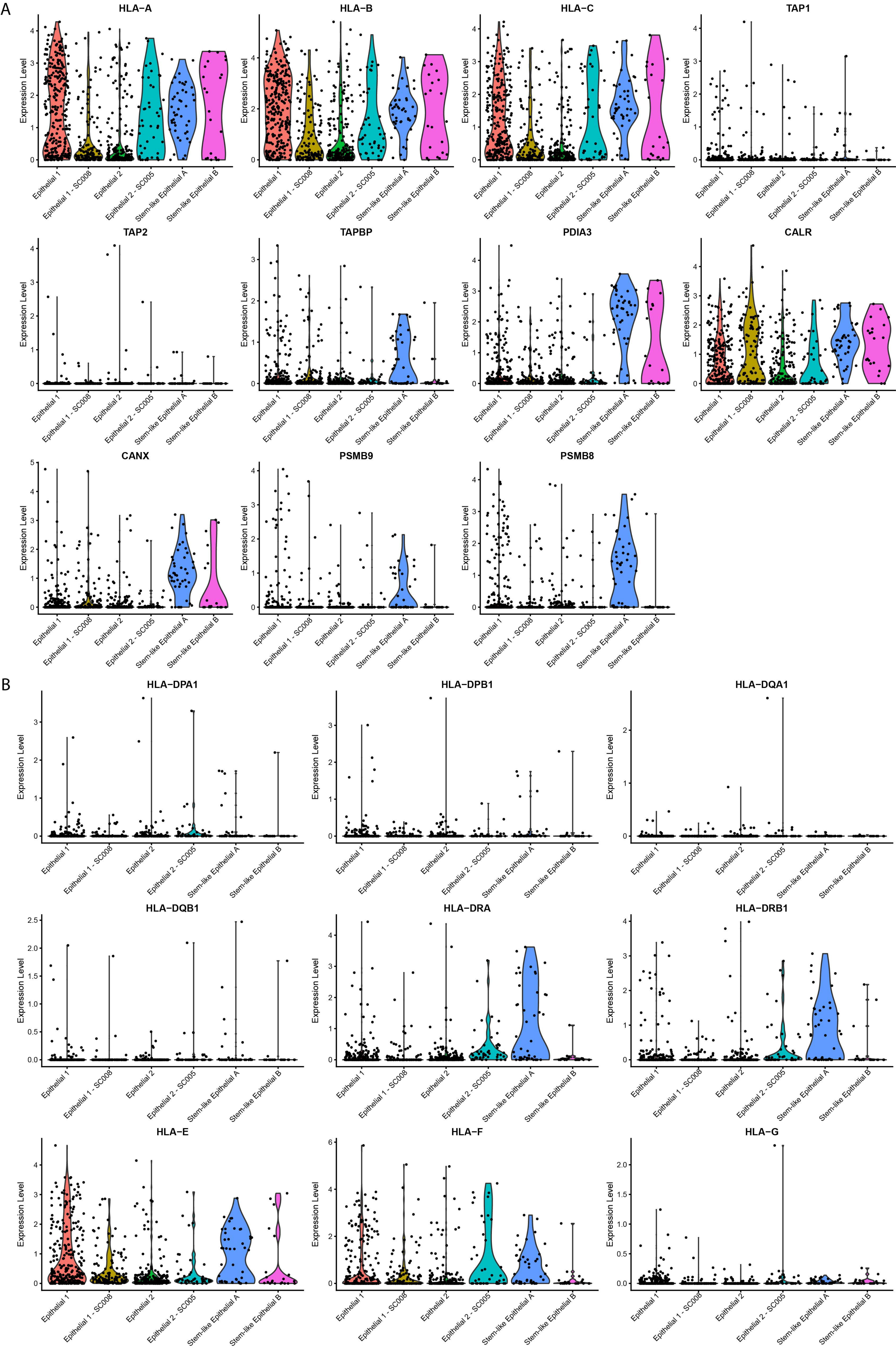


Figure S4. Comparison of MHC Class I- and II-related gene expressions between epithelial clusters. **(A)** Violin plots shows most of the MHC Class I-related genes were highly expressed by the Stem-like Epithelial A cluster, except for TAP1 and TAP2 genes. **(B)** Violin plots shows MHC Class II-related genes, including HLA-DRA, HLA-DRB1, HLA-E, HLA-F, and HLA-G, were highly expressed by the Stem-like Epithelial A cluster.

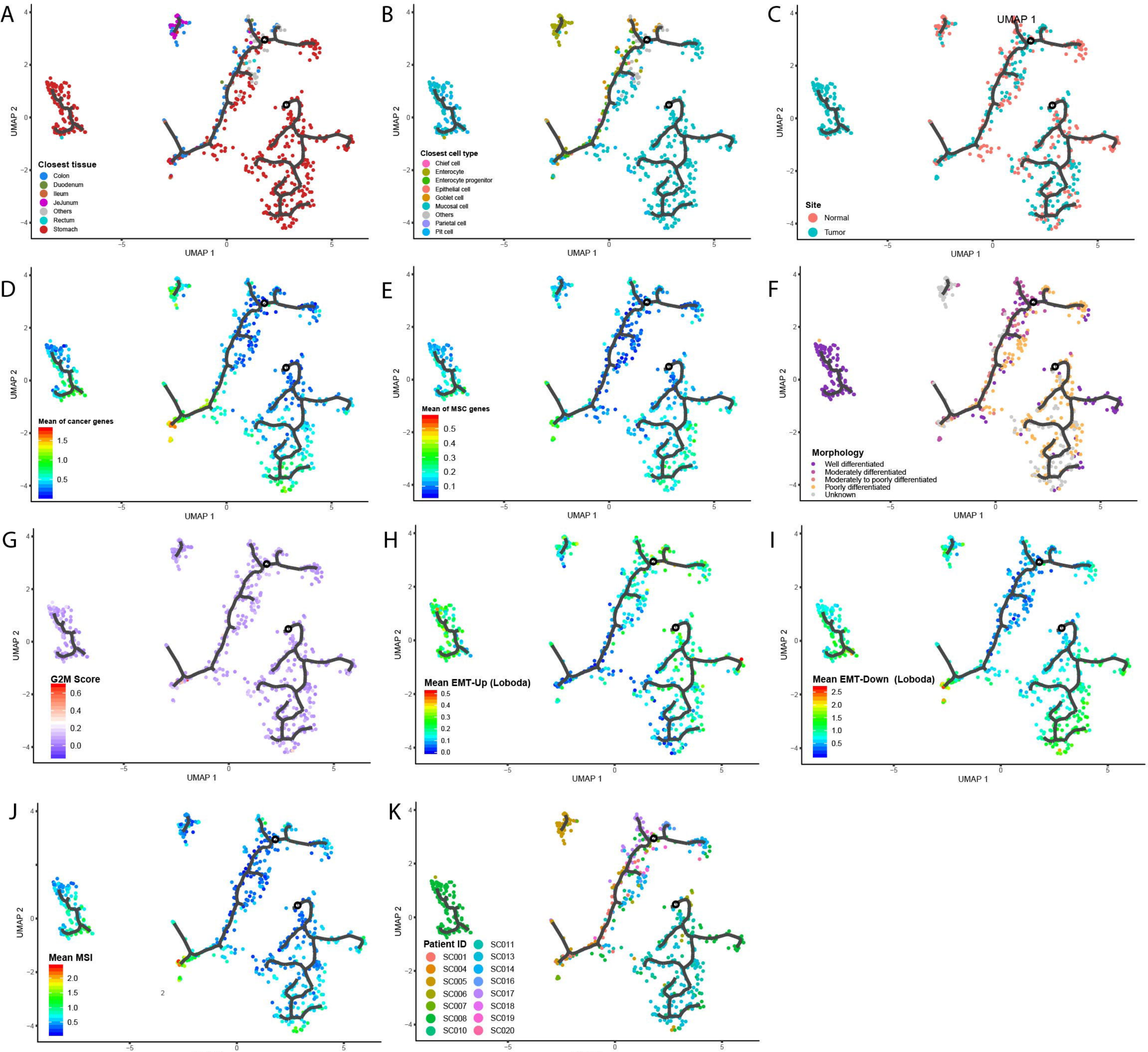


Figure S5. Multi-faceted epithelial cells along the Monocle 3 developmental trajectories. Epithelial cells colored by **(A)** the closest matching tissue predicted by the scHCL; **(B)** the closest matching cell type predicted by scHCL; **(C)** the site of tissue collection (Normal vs Tumor); **(D)** the mean of gastric cancer genes (Zhang *et al.*); **(E)** the mean of metaplastic stem cell (MSC) genes (Zhang *et al.*); **(F)** the tumor morphology, **(G)** the G2M score computed by the Seurat Cell Cycle Regression, **(H)** the mean of the Epithelial-Mesenchymal-transition (EMT) up-regulated genes (Loboda *et al.*); **(I)** the mean of the EMT down-regulated genes (Loboda *et al.*); **(J)** the mean of the MicroSatellite Instability (MSI) genes (The Cancer Genome Atlas Research Network); **(K)** the patient sample ID.

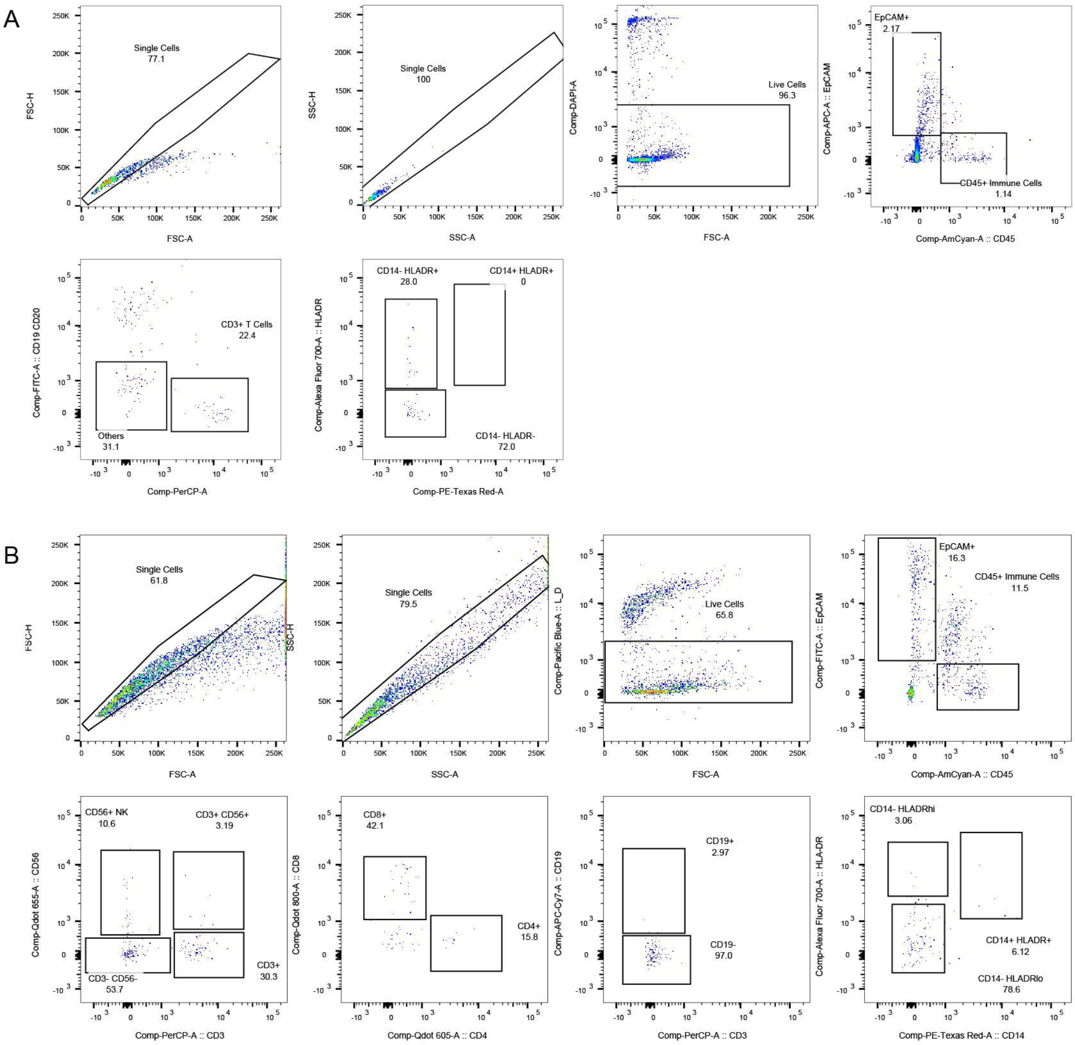


Figure S6. FACS gating strategies. **(A)** FACS Panel 2 and **(B)** FACS Panel 3 used in this study.

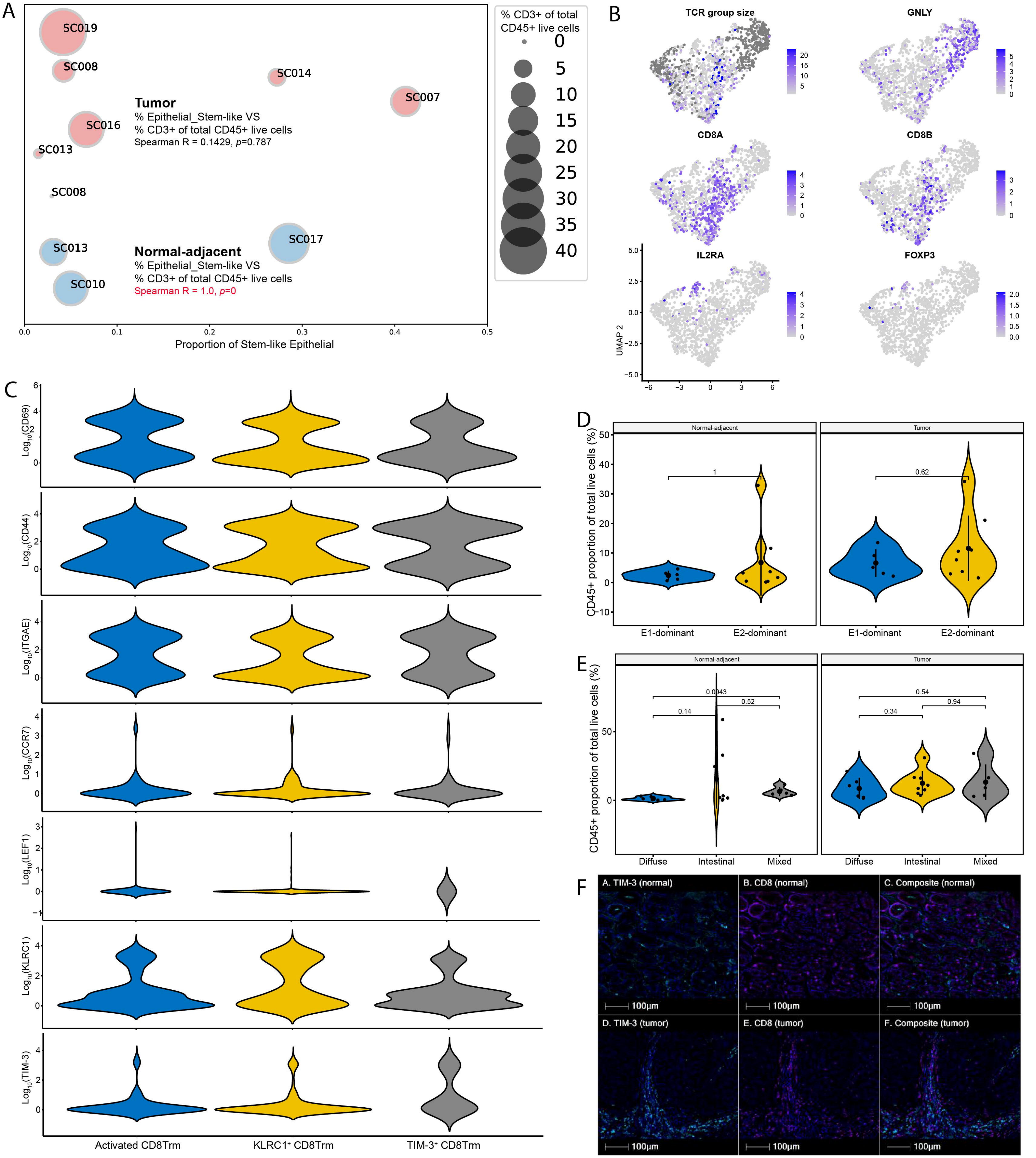


Figure S7. Characterization of immune subsets. **(A)** Bubble plots shows significant positive correlation between proportion of Stem-like Epithelial cell (of all the epithelial cells) and proportion of CD3+ cell (of all CD45+ immune cells) was found in gastric normal-adjacent tissues (Blue bubbles), but not in tumor tissues (Red bubbles). **(B)** tSNE plots of all immune cells, colored by TCR group size, gene expressions of known markers for NK (*GNLY*), CD8 T (CD8A and CD8B), regulatory T (*IL2RA*), and CD4 T (*FOXP3*) cells. **(C)** High resident memory T-related gene expressions (including CD69, CD44, *ITGAE*), but low homing genes (*CCR7* and *LEF1*) for all CD8 T cells identified in this study. **(D)** No significant difference in CD45+ cells proportion of total live cells between Epithelial 1- and Epithelial 2-dominant patients, in both normal-adjacent and tumor tissues, respectively. **(E)** Comparison of CD45+ cell proportions of total live cells between histological diffuse, intestinal, and mixed patients. The CD45+ cell proportions in mixed patients were found to be significantly higher from the diffuse patients. **(F)** Immunohistochemistry staining of normal-adjacent and tumor tissues for an intestinal patient, SC019. TIM3+ and CD8+ cells were detected in both tissues.

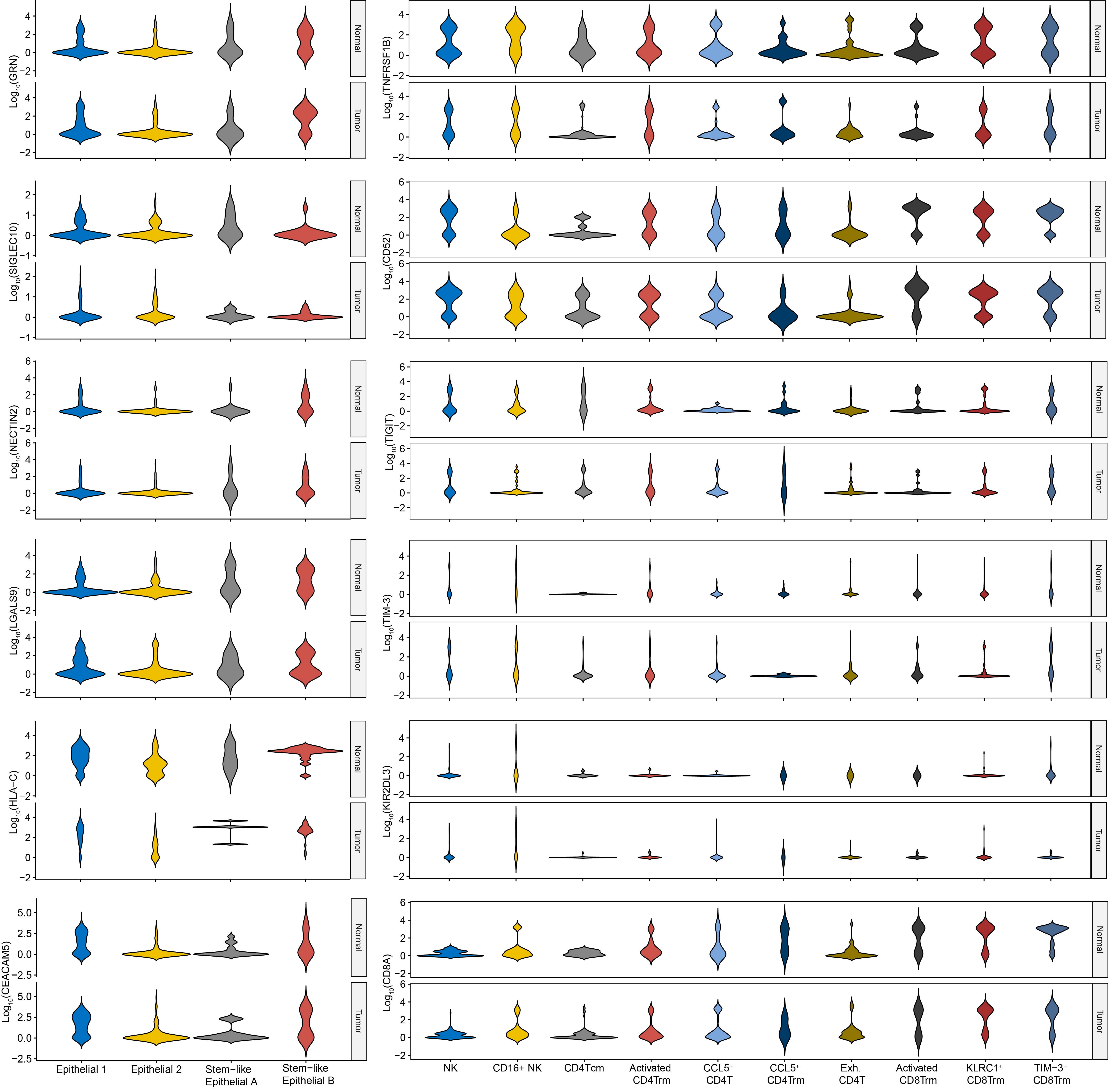


Figure S8. Violin plots for expressions of interacting molecule from each epithelial cluster and interacting molecule from each immune subset involved in immune-suppressive interactions. The violins are plotted separately for normal-adjacent and tumor tissues.

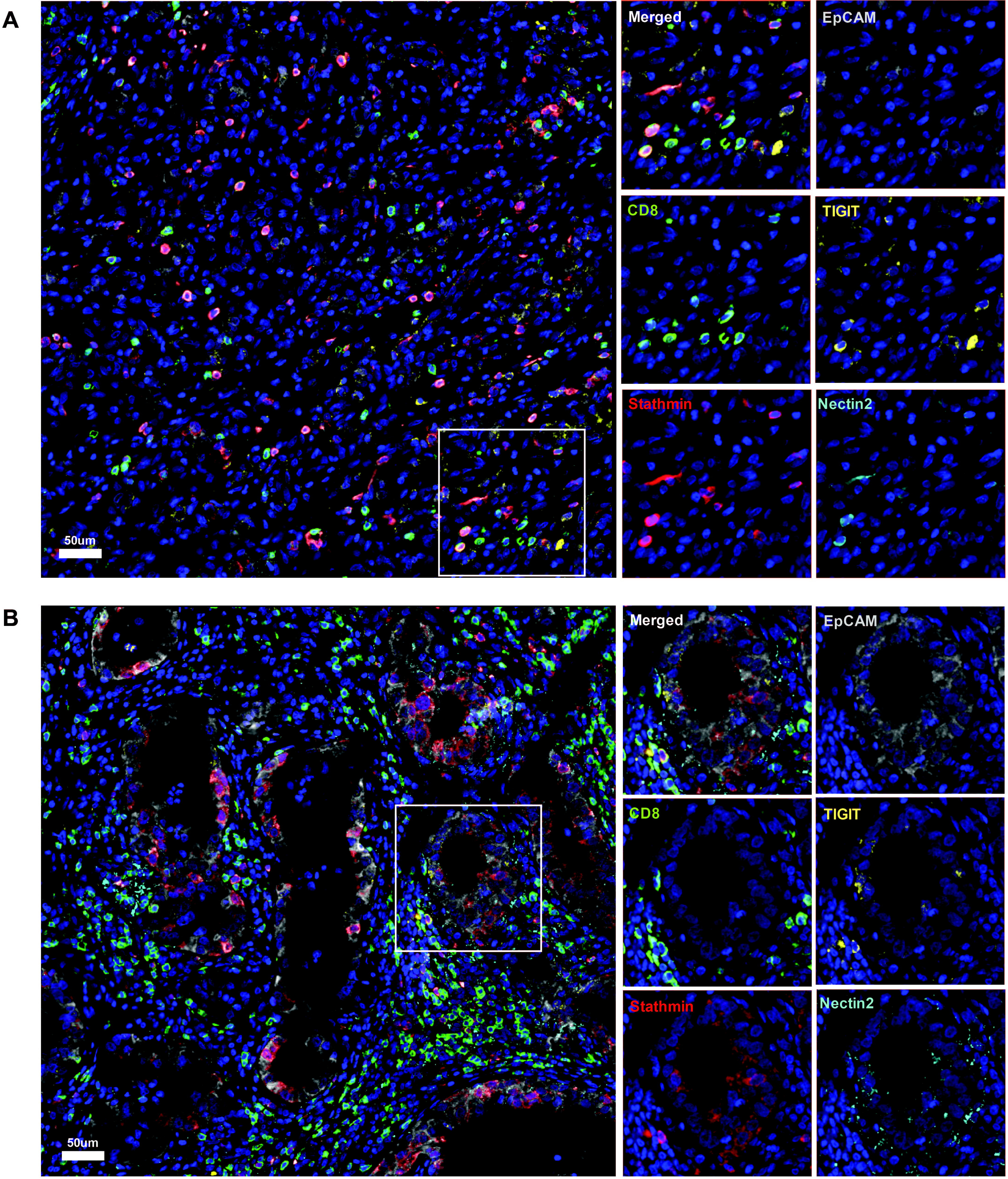

A
Merged
EpCAM
CD8
TIGIT
Stathmin
Nectin2
50um
B
Merged
EpCAM
CD8
TIGIT
Stathmin
Nectin2
50um

Figure S9. Multiplex immunohistochemistry staining of gastric cancer tissue slides using EPCAM, CD8, TIGIT, STMN1, and NECTIN2 markers for Patients **(A)** SC004 and **(B)** SC018.