Cell Reports, Volume *27*

Supplemental Information

Dissecting the Single-Cell Transcriptome Network Underlying Gastric Premalignant Lesions and Early Gastric Cancer

Peng Zhang, Mingran Yang, Yiding Zhang, Shuai Xiao, Xinxing Lai, Aidi Tan, Shiyu Du, and Shao Li

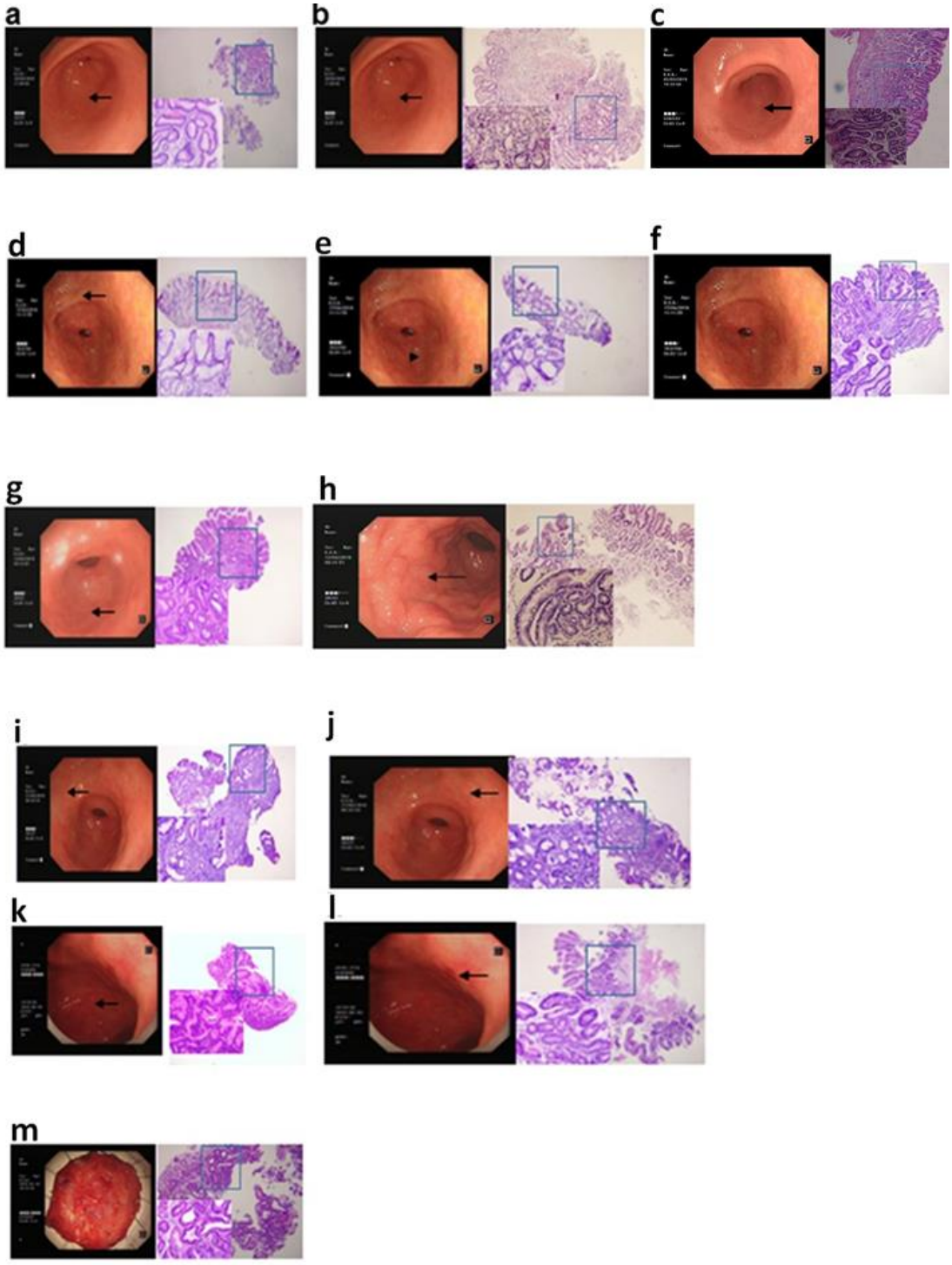

Figure S1. Gastroscopic images and the hematoxylin and eosin (H&E) staining of samples in this study, Related to **Figure 1**. Arrows show the sites of gastroscopic biopsies. a, NAG1 (P1); b, NAG2 (P2); c, NAG3 (P9); d, CAG1 (P3); e, CAG2 (P3); f, CAG3 (P4); g, IMW1 (P5); h, IMW2 (P6); i, IMS1 (P7); j, IMS2 (P7); k, IMS3 (P8); l, IMS4 (P8); m, EGC (P8).

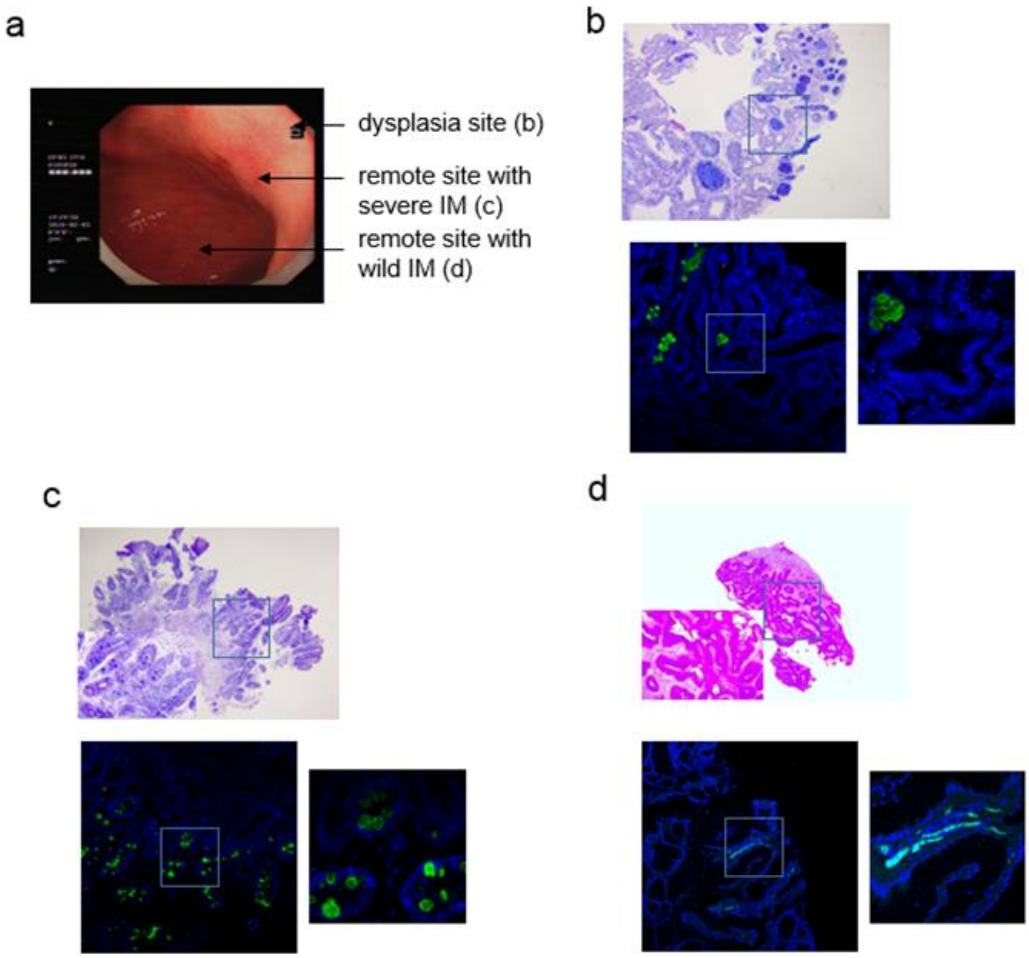

Figure S2. The AB-PAS staining and MUC2-based immunofluorescence (IF) staining for the three biopsies collecting at three distinct sites from the early-malignant patient (P8), Related **Figure 1**. a, The three sites of biopsies in the gastric antrum of P8, including the neoplastic site, the remote site with severe IM and the remote site with wild IM. b-d, the AB-PAS staining (upper) and Immunofluorescence staining (bottom) of gastric tissue sections at the neoplastic site (b) , remote site with severe IM (c) and remote site with wild site (d) (original magnification, 40x and 200x) .

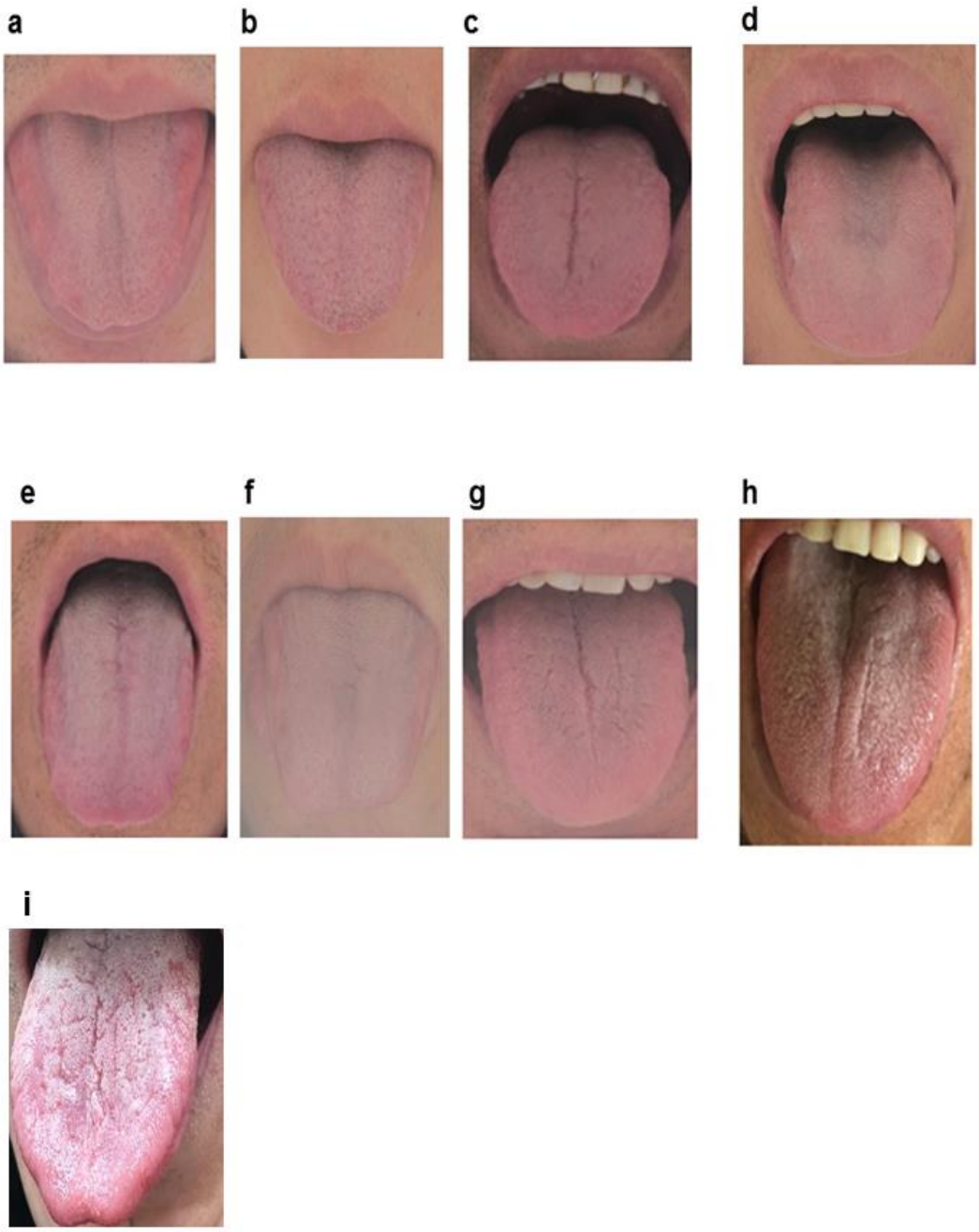

Figure S3. **Tongue images of patients in the study,** Related to **Figure 1**. a, P1; b, P2; c, P3; d, P4; e, P5; f, P6; g, P7; h, P8. i, P9.

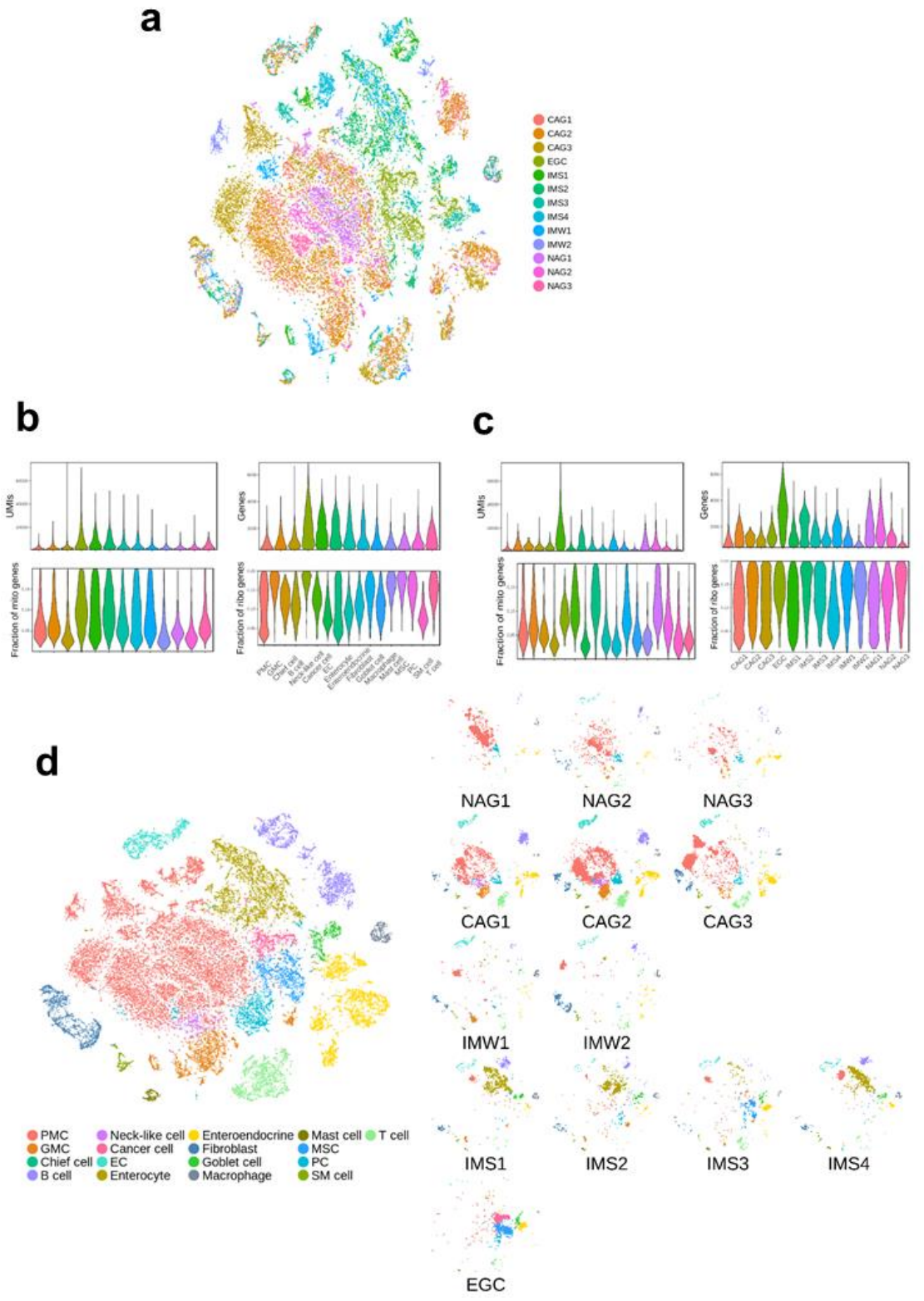

Figure S4. Consistency of cell capture and identification in biopsies from patients with different lesions, Related to **Figure 1.** a, The t-SNE plot show the cell distribution among multiple batches. b, Number of unique molecular identifiers (nUMI) and genes identified, and fraction of reads mapping to mitochondrial or ribosomal genes across identified cell types. c, nUMI and genes identified, and fraction of reads mapping to mitochondrial or ribosomal genes across patient samples. d, t-SNE plot as in Fig. 1b colored by cell types across all patients and then separated by sample.

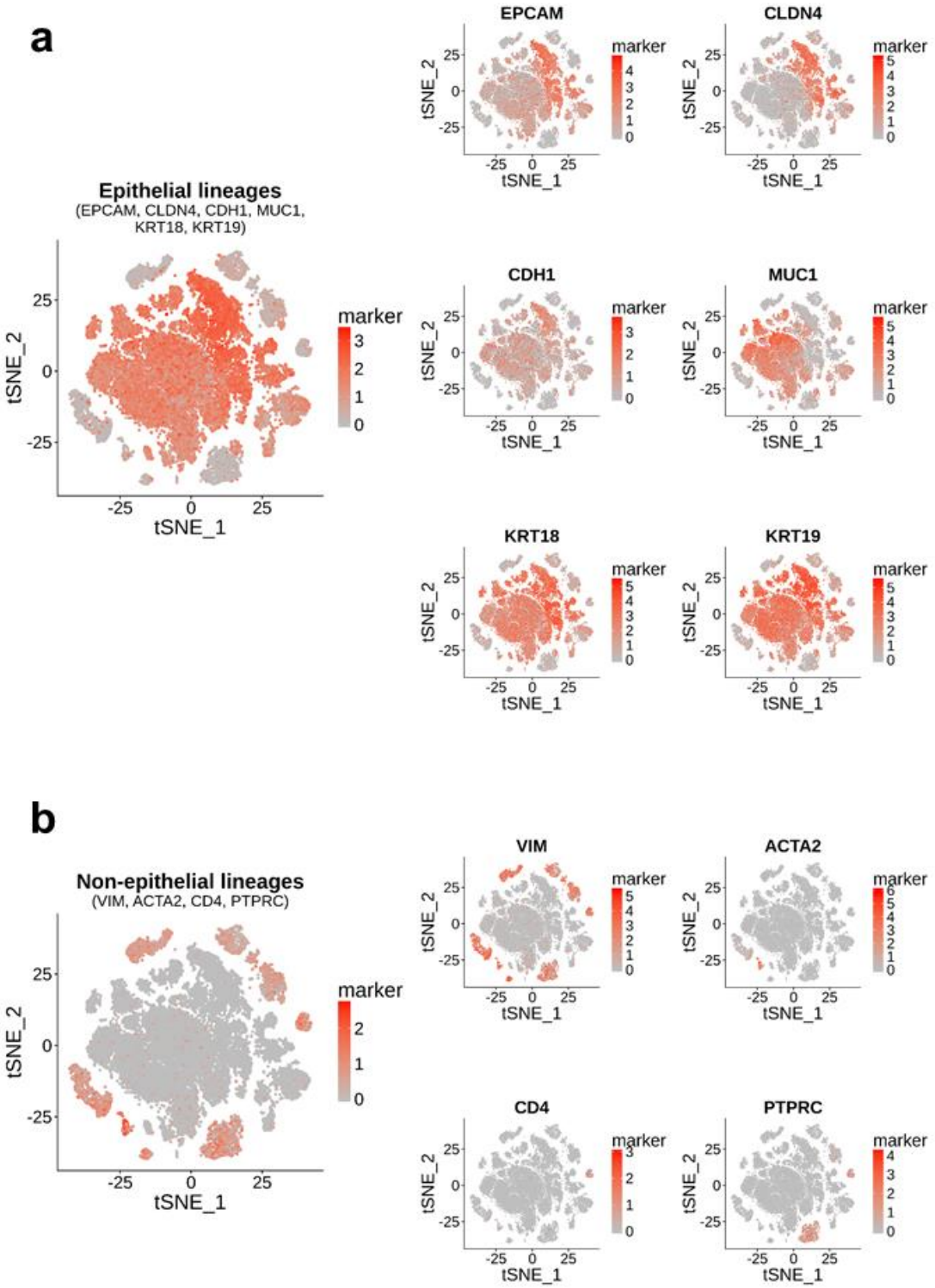

Figure S5. **The t-SNE plots show expression of epithelial cell lineages (a) and non-epithelial cell lineages (b) according to the mean (left) and individual (right) expression distribution of known markers**, Related to **Figure 1**. The known markers of gastrointestinal epithelial cell lineages include EPCAM, CLDN4, CDH1, MUC1, KRT18 and KRT19, while those of non-epithelial cell lineages include VIM, ACTA2, CD4 and PRPRC (CD45). The clear distinction between epithelial and non-epithelial cell lineages was observed.

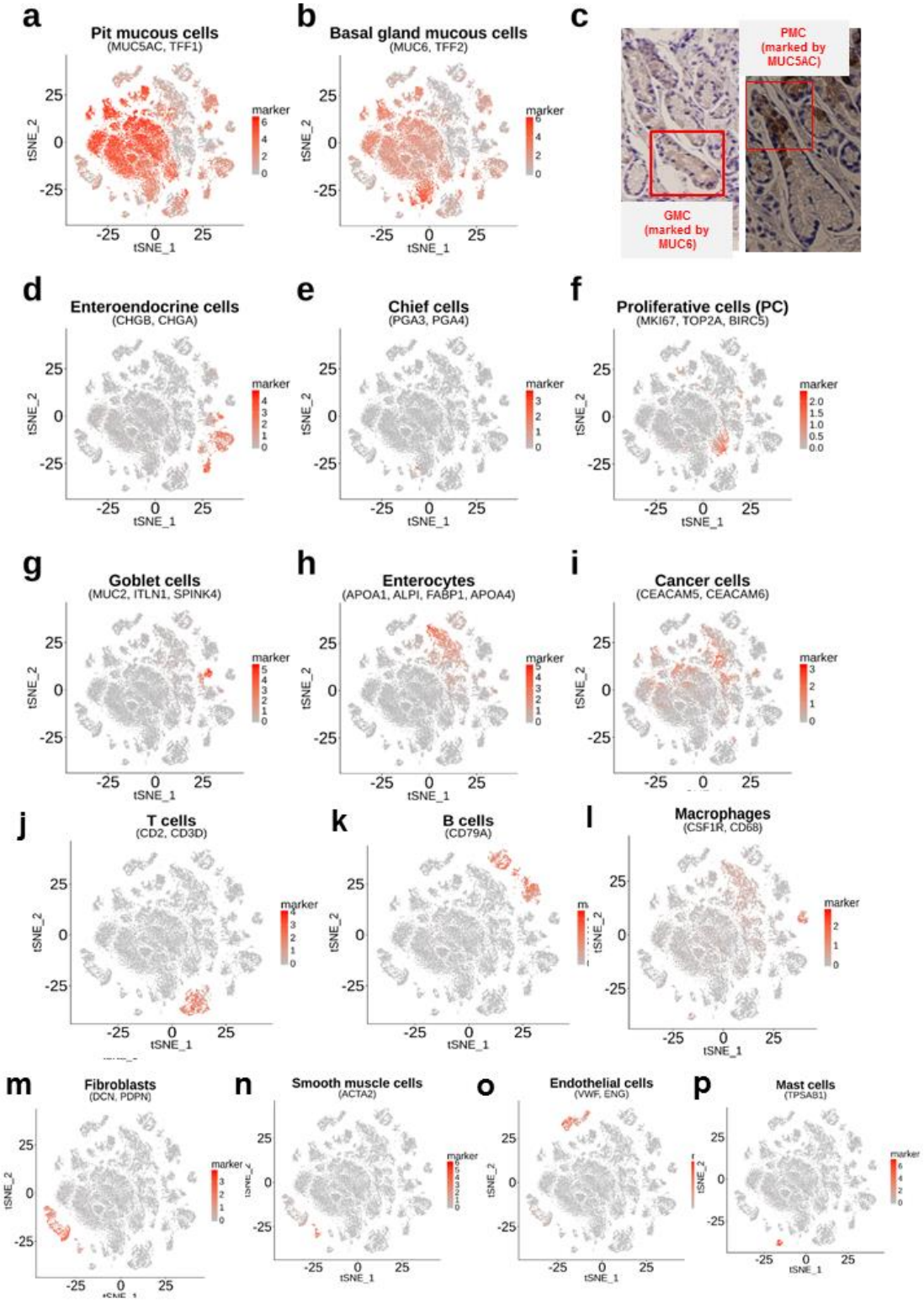

Figure S6. The t-SNE plots show expression of cell lineages, Related to **Figure 1**. It comprised Pit mucous cells (PMC, marked by MUC5AC and TFF1, 13026 cells, a), Basal gland mucous cells (GMC, marked by MUC6 and TFF2, 1886 cells, b), Enteroendocrine cells (marked by CHGB and CHGA, 2814 cells, d), Chief cells (marked by PGA3 and PGA4, 44 cells, e), Proliferative cells (PC, marked by MKI67, TOP2A and BIRC5, 1133 cells, f), Goblet cells (marked by MUC2, ITLN1 and SPINK4, 552 cells, g), Enterocytes (marked by APOA1, ALPI, FABP1 and APOA4, 3364 cells, h), Cancer cells

(marked by CEACAM5 and CEACAM6, 774 cells, i), and non-epithelial cell lineages, including T cells (marked by CD2 and CD3D, 1618 cells, j), B cells (marked by and CD79A, 2158 cells, k), Macrophages (marked by CSF1R and CD68, 397 cells, l), Fibroblasts (marked by DCN and PDPN, 1381 cells, m), Smooth muscle cells (marked by and ACTA2, 279 cells, n), Endothelial cells (marked by VWF and ENG, 994 cells, o), Mast cells (marked by and TPSAB1, 227 cells, p). c) The immunohistochemistry staining of gastric glands with MUC6 (left) and MUC5AC (right) to clarify the GMC and PMC, respectively.

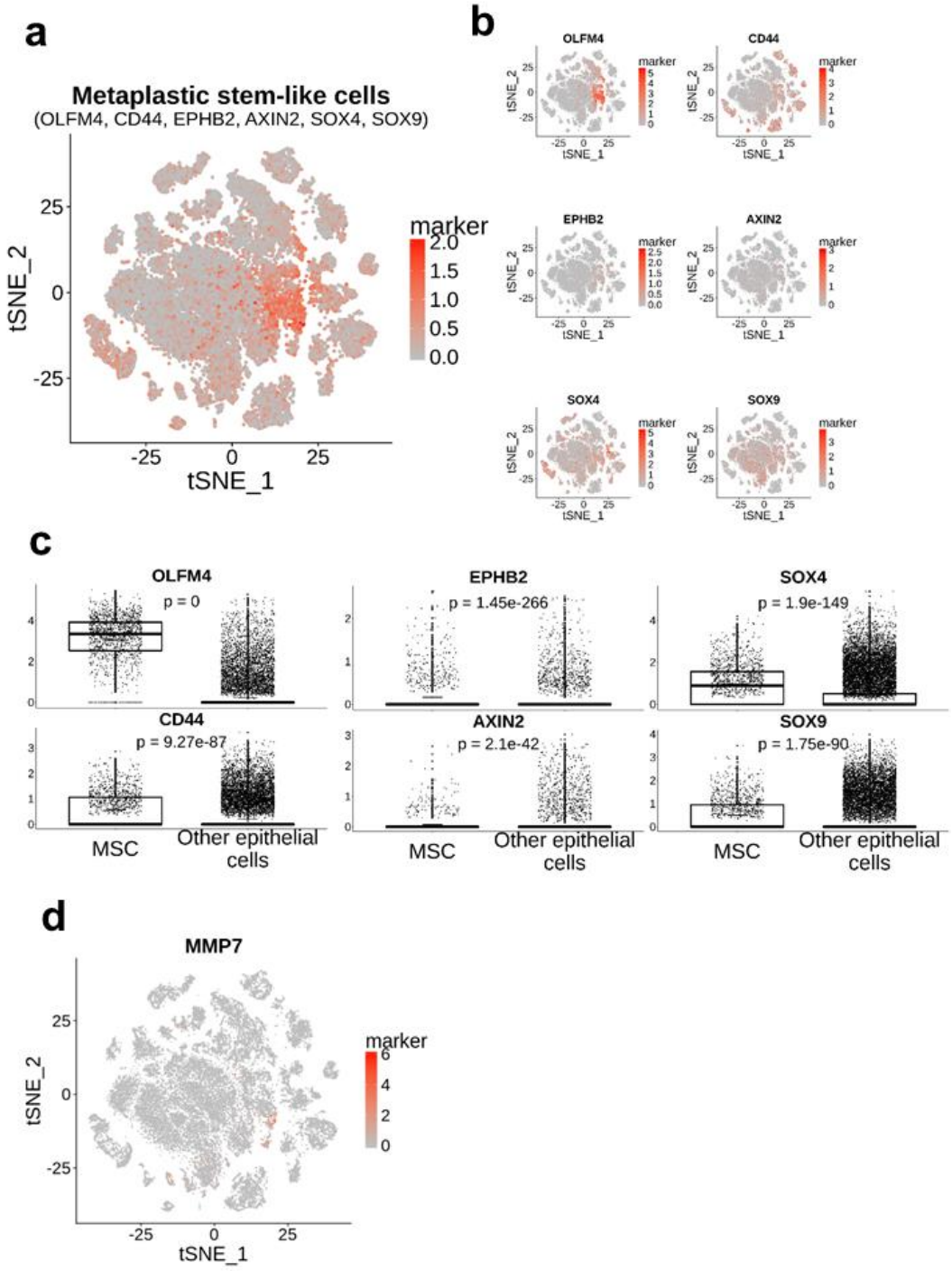

Figure S7. **The t-SNE plots show expression of metaplastic stem-like cells (MSCs),** Related to **Figure 1.** It was shown the mean (a) and individual (b) expression distribution of several stem cell-related markers, including OLFM4, CD44, EPHB2, AXIN2, SOX4 and SOX9. c. The boxplot for comparing the expression level of these stem cell-related markers in MSCs with that of markers in other cells.

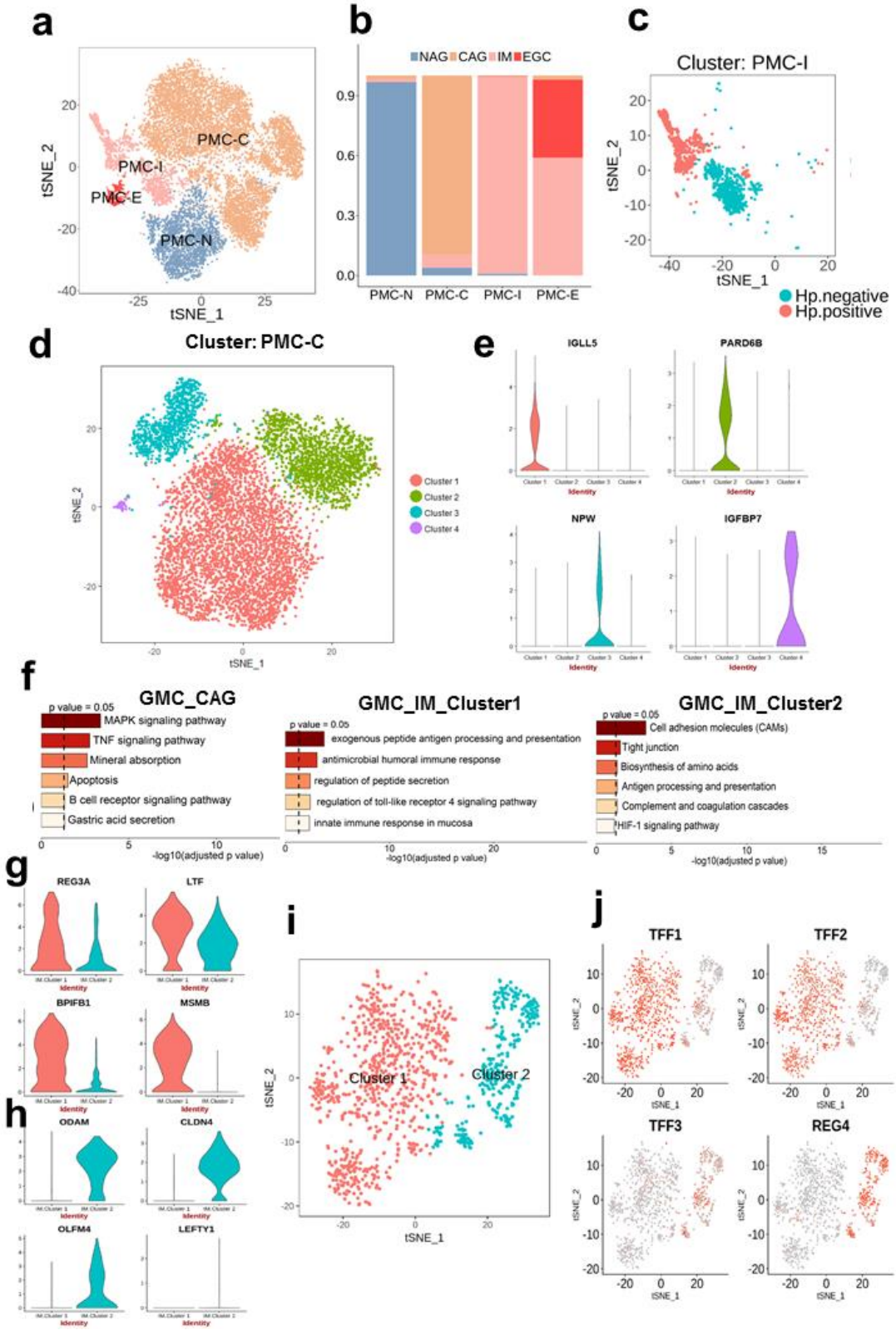

Figure S8. **The heterogeneity within the gastric epithelial cell lineages,** Related to **Figure 1 and Figure 3**. a. Re-clustering the PMC cell cluster into four main sub-clusters, which were annotated as PMC-N, PMC-C, PMC-I and PMC-E, respectively. b. The distribution of lesions in which cells in each sub-cluster are from. c. The PMC-I sub-cluster comprises of two isolated subtypes, of which one

represented *H. pylori* infection positive while the other represented *H. pylori* infection negative. d. The PMC-C sub-cluster comprises of four isolated subtypes (Cluster 1, Cluster 2, Cluster 3, and Cluster 4, respectively). e. Violin plots display the distribution of expression of markers of PMC-C subtypes, of which the Cluster 0 characterizing with inflammation response-related gene IGLL5, the Cluster 1 characterizing with cell polarity-related gene PARD6B, the Cluster 3 characterizing with neupeptide-related gene NPW, and the Cluster 4 characterizing with prostacyclin-related gene IGFBP7. f. The most enriched pathways for up-regulated genes in the GMCs from CAG samples (left), sub-cluster 1 of GMCs from IM samples (middle) and sub-cluster 2 of GMCs from IM samples (right). g. Violin plots display the distribution of expression of markers of the GMC_IM_Cluster 1, including REG3A, LTF, BPIFB1 and MSMB. h. Violin plots display the distribution of expression of markers of the GMC_IM_Cluster 2, including ODAM, CLDN4, OLFM4 and LEFTY1. i. Re-clustering the PC cell cluster into two main sub-clusters, which are annotated as Cluster 1 and Cluster 2, respectively. j. t-SNE plot shows the expression of the putative markers for the two PC sub-clusters (Cluster 1: TFF1 and TFF2 ; Cluster 2: TFF3 and REG4).

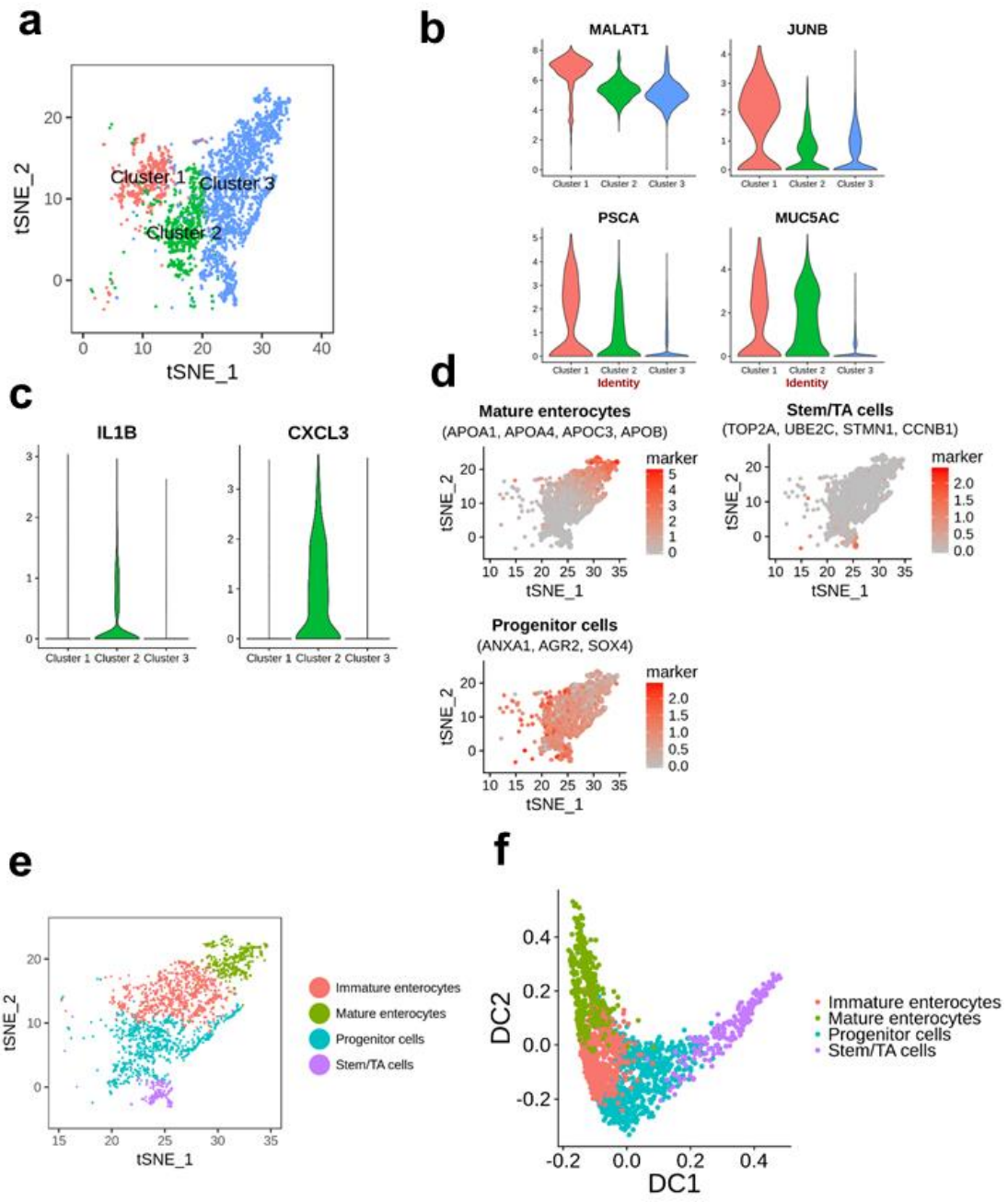

Figure S9. **The heterogeneity within the enterocyte cluster**, Related to **Figure 1**. a. Re-clustering the enterocyte into three main sub-clusters, which are annotated as Cluster 1, Cluster 2 and Cluster 3, respectively. b. c. The t-SNE plot shows the expression of the putative markers for the enterocyte subtype Cluster 1 (b) and Cluster 2 (c). d. t-SNE plot shows the expression levels of stem/TA (OLFM4, STMN1, CCNB1 and TOP2A), enterocyte progenitor (AGR2, AXNA1, and SOX4) and mature enterocyte markers (APOA1, APOA3, APOC3 and APOB). e. tSNE plot for enterocytes subsets representing representing distinct stages of maturation. f. the putative trajectory that indicates the differentiation of enterocyte from stem/TA to progenitor to immature enterocyte and toward mature enterocyte. It was inferred by using diffusion mapping.

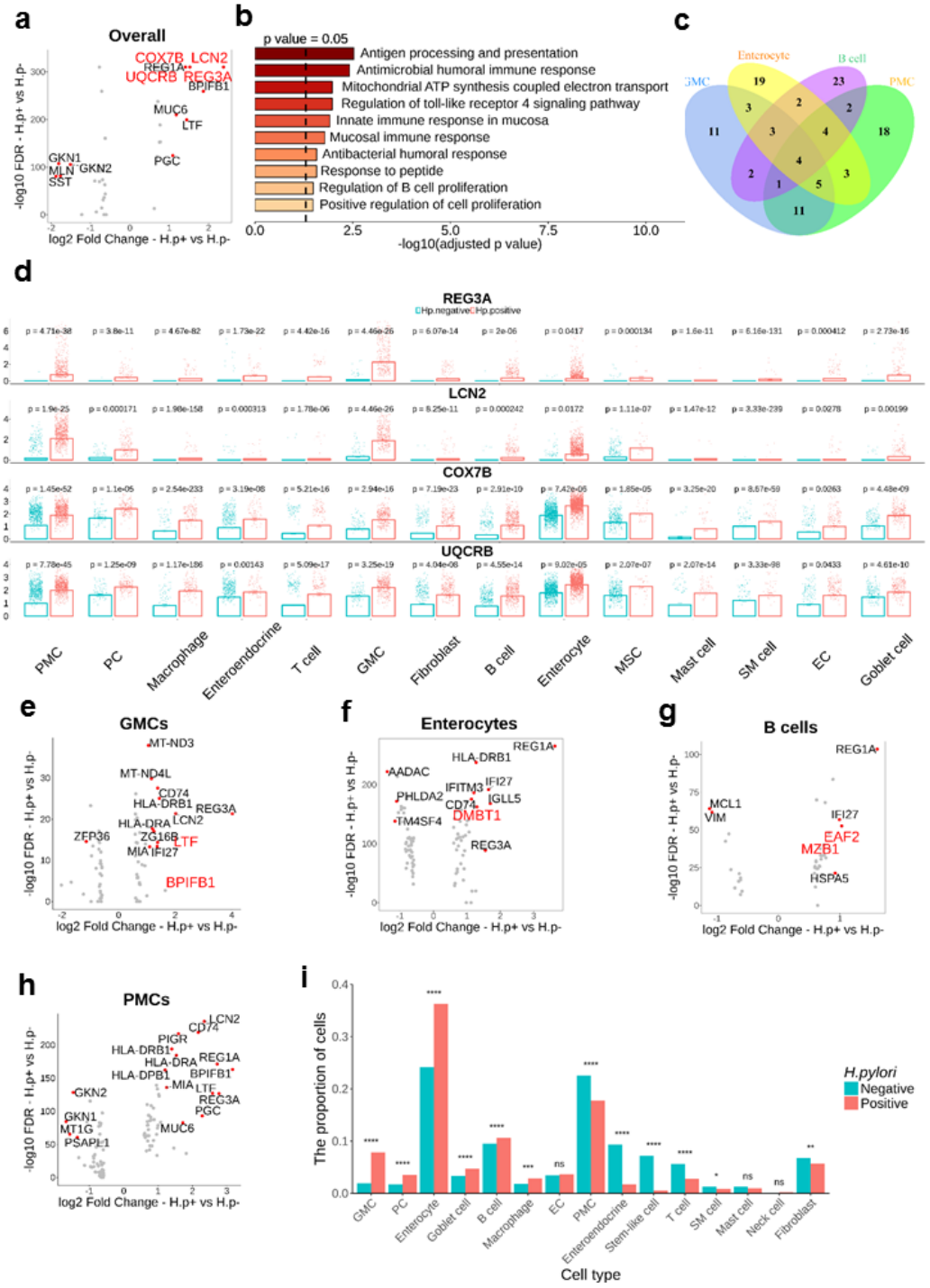

Figure S10. **The transcriptional profiles response to *H. pylori* infection across diverse cell types,** Related to **Figure 1**. a. Volcano plot of the global expression program in all cell types response to the *H. pylori* infection. b. Most enriched pathways for the up-regulated genes in all cell types. c. Veen diagram for up-regulated genes response to *H. pylori* infection for main cell types including GMC, enterocytes, B cells and PMC (cell-intrinsic response). d. box plot showing the differential expression pattern of four consistently up-regulated genes (REG3A, LCN2, COX7B and UQCRB) between *H. pylori* infection positive and negative cells across diver cell types. e-h. Volcano plot of the cell type-specific expression program response to *H. pylori* infection. i. The proportion of *H. pylori* positive / negative cells in each cell type.

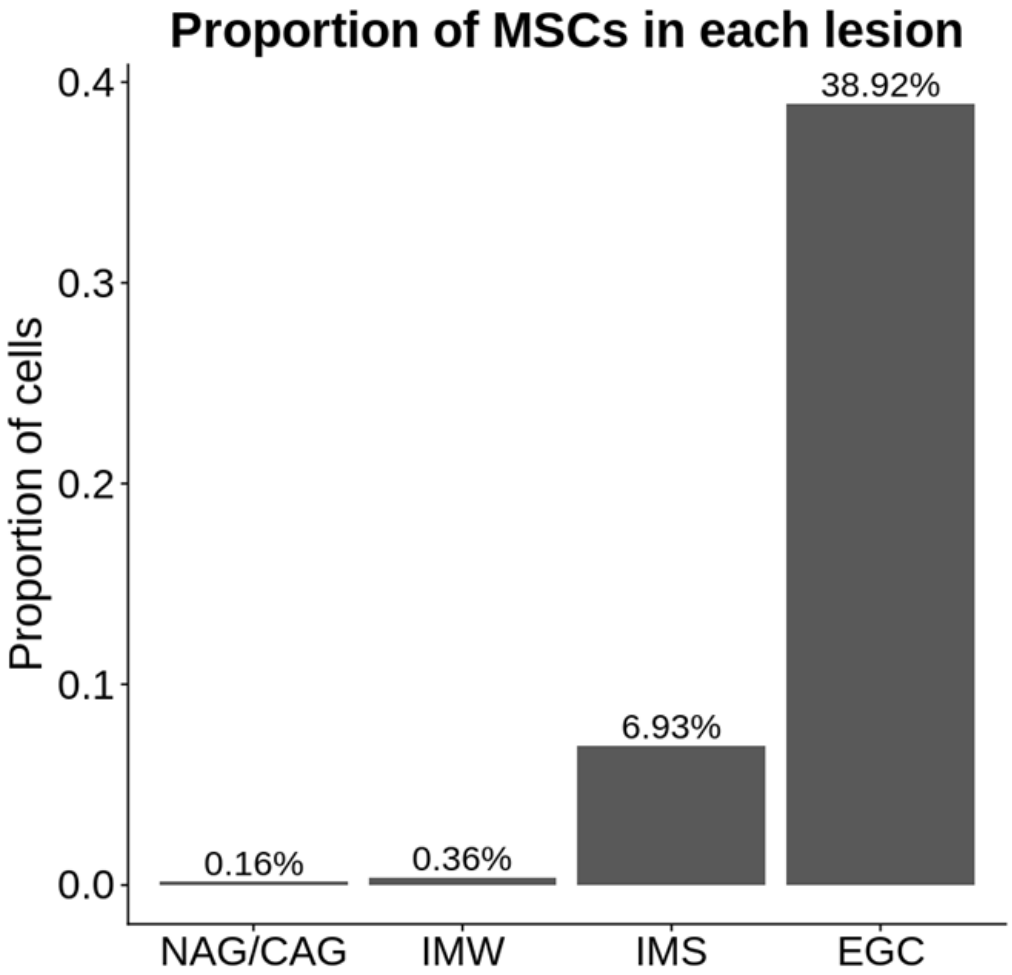

Figure S11. **The proportion of metaplastic stem-like cells (MSCs) in each lesion**, Related to **Figure 2**.

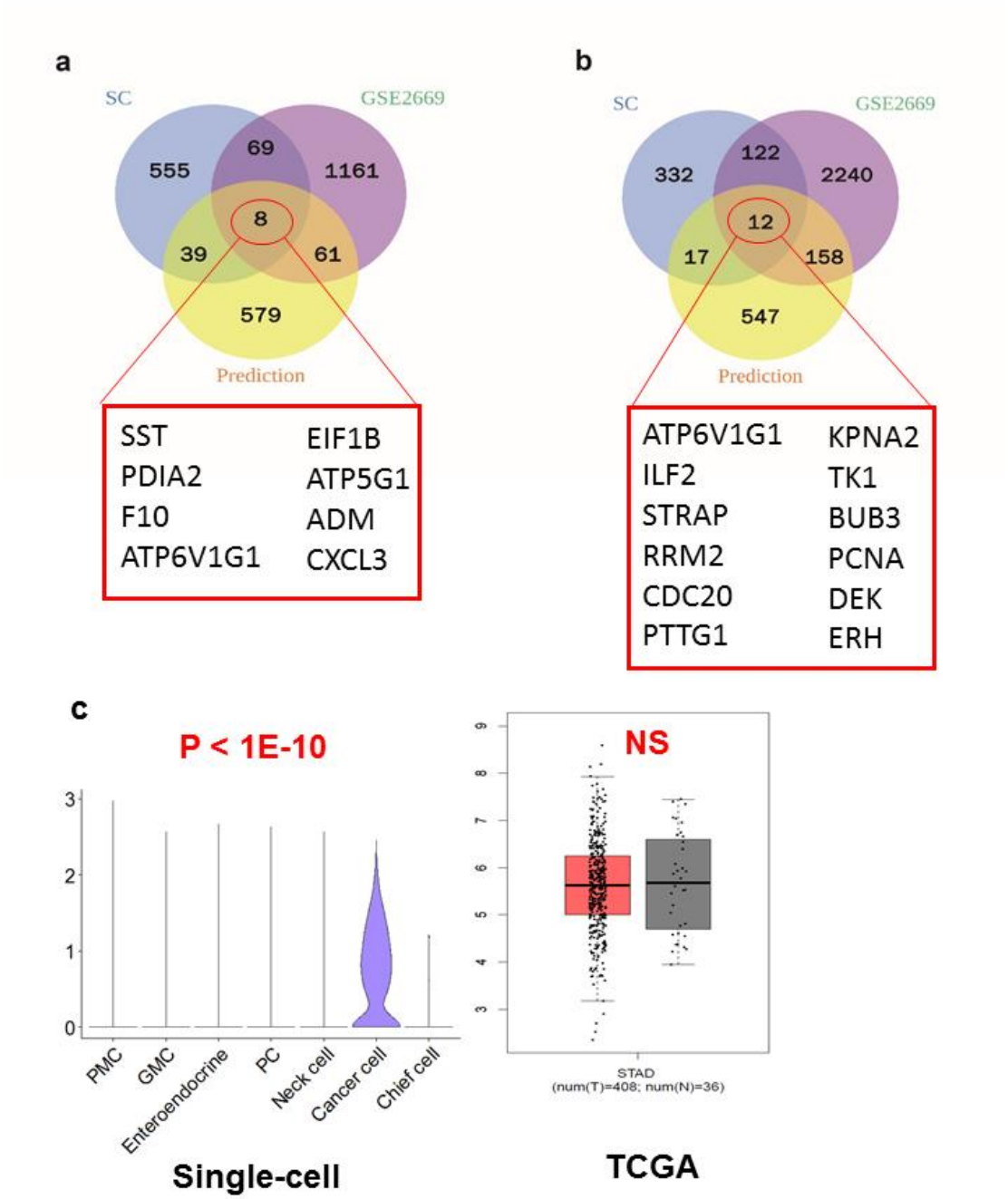

Figure S12. **Gene signatures derived from multiple evidences**, Related to **Figure 2**. a,b. Venn diagram for the gene signatures of gastritis (a) and gastric cancer (b), identified by clues including the single-cell atlas, bulk transcriptome dataset (GSE2669) and network-based prediction by the CIPHER algorithm. c. The expression distribution of inferred gastric cancer-related high-risk gene GSTM4, which ranked 5.6% in the prediction profile, within gastric epithelial cell lineages (left) and TCGA datasets (right). SC, single-cell atlas. NS, not significant.

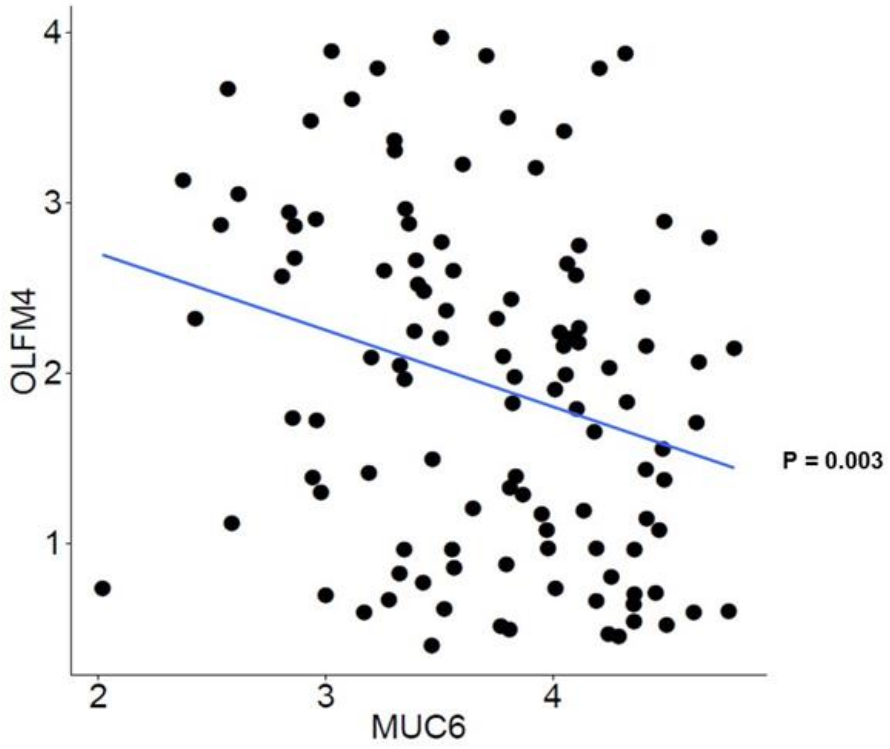

Figure S13. **The co-expression of MUC6 and OLFM4 in individual GMCs in the IM lesion**, Related to **Figure 3**.

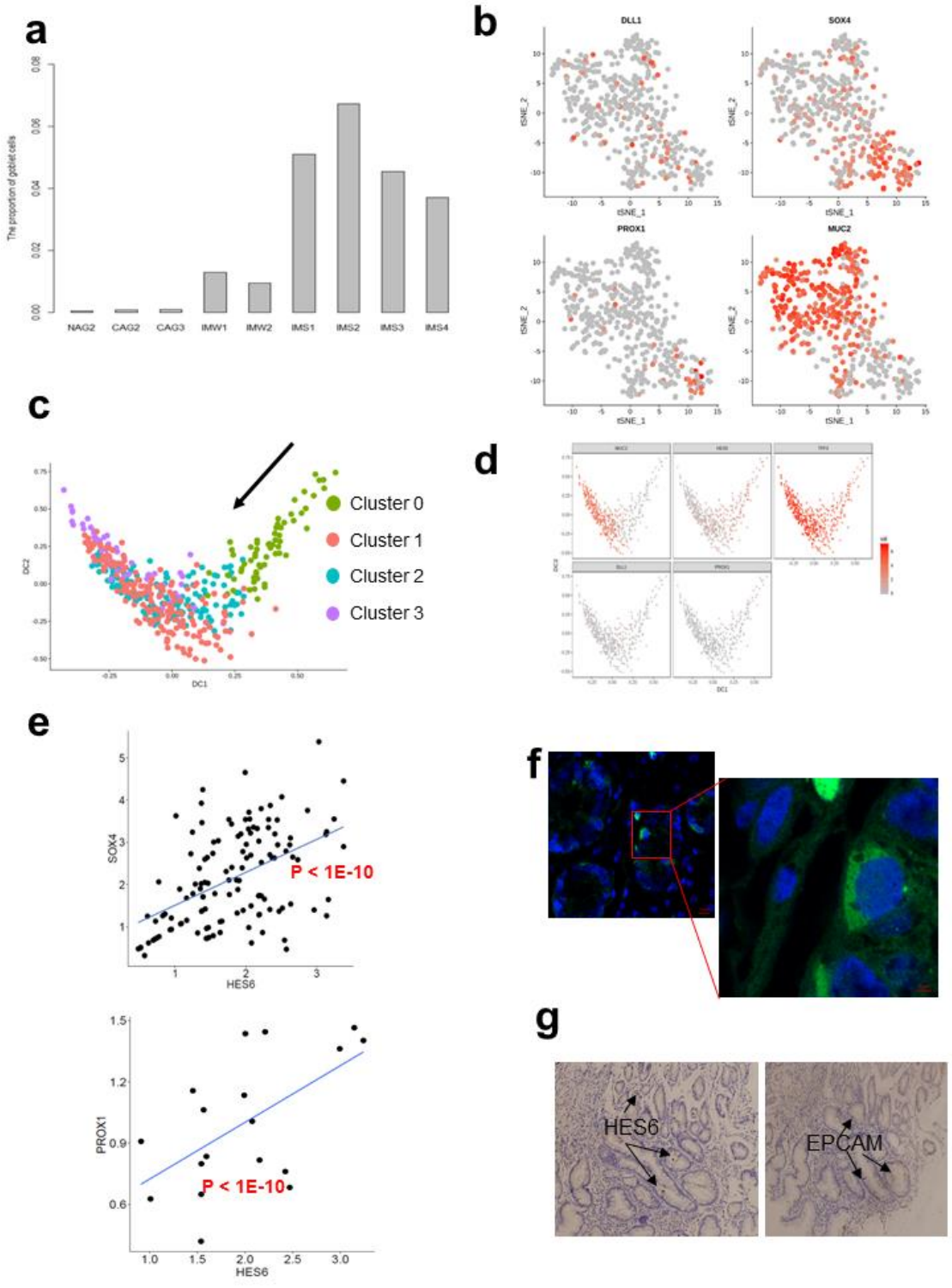

Figure S14. **The heterogeneity within goblet cells**, Related to **Figure 5**. a. The proportion of goblet cells in each premalignant lesion. b. t-SNE plot shows the expression levels of previously implicated secretory progenitor markers (DLL1, SOX4 and PROX1) and differentiated goblet cell markers (MUC2), in the 'Goblet cell' cluster. c. the putative trajectory that potentially indicates the differentiation process of goblet cells from the putative secretory progenitors (Cluster 0) toward differentiated goblet cells (Cluster 1, 2 and 3). It was inferred by using diffusion mapping. d. The expression profiles of some secretory differentiation-committing transcription factors along the putative

trajectory in c. e. The positive correlation between the expression profiles of SOX4 and HES6 (upper), as well as PROX1 and HES6 (bottom), in individual goblet cells. f. The high-resolution morphological view (10um) of HES6+ cells. It was observed that HES6+ cells were round, immature cells with a small amount of cytoplasm and a dark nucleus. g. Immunohistochemistry (IH) staining HES6 (left) and EPCAM (right) in the IM samples.

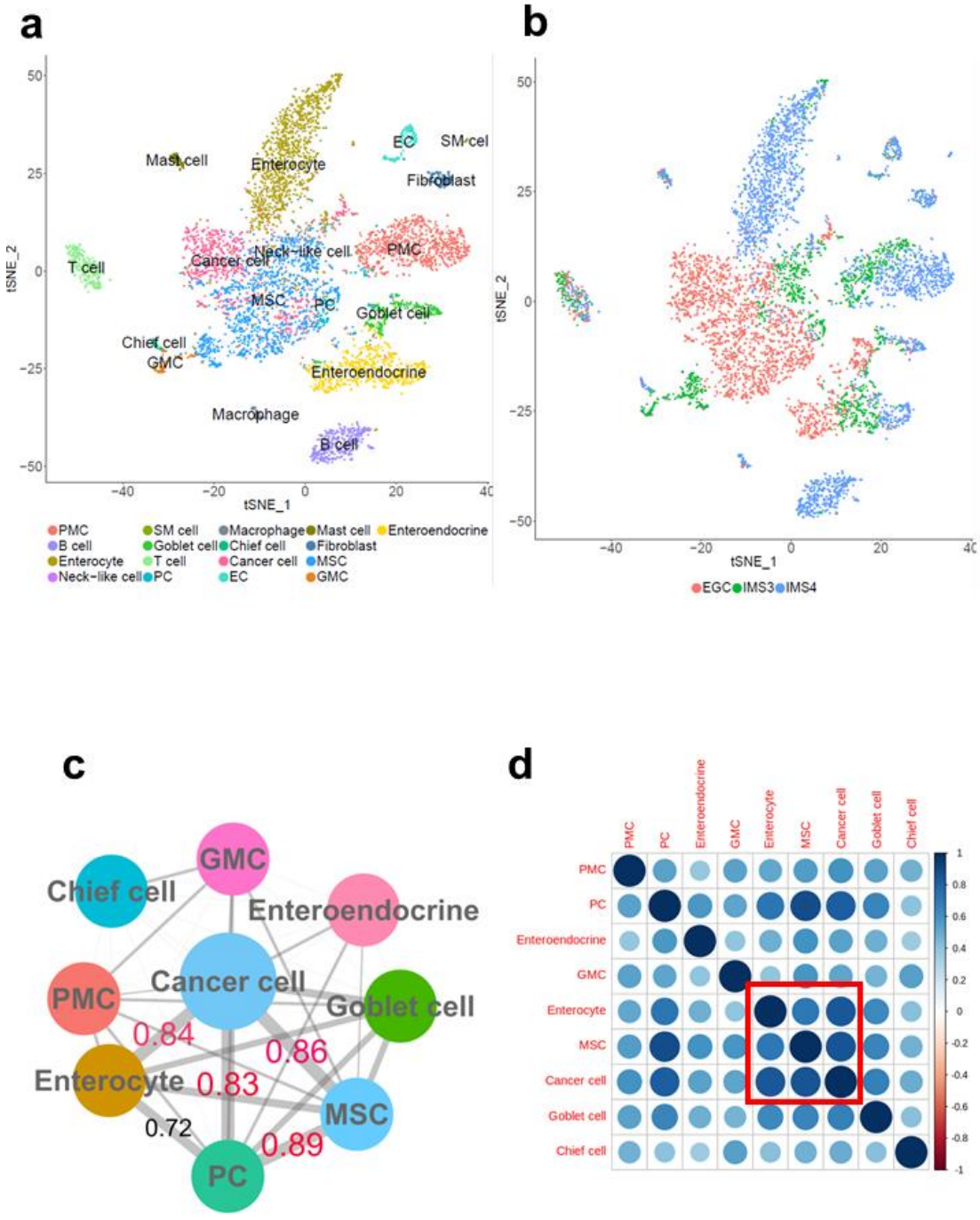

Figure S15. **The t-SNE plot showing the distribution of diverse cell populations (a) and biopsies (b) in the same EGC patient (P8)**, Related to **Figure 6**. The similarities between these cell populations were showed in the form of network (c) and correlation matrix (d). The thickness of the edges in the network was denoted as the Pearson correlation coefficient between the centroids of any

pair of cell types.

Table S1. The pathological grade of each biopsy used in the study, Related to **Figure 1** and **STAR Methods.**

Sample ID	Histological diagnosis*	Patient
NAG1	NAG	P1
NAG2	NAG	P2
NAG3	NAG	P9
CAG1	CAG	P3
CAG2	CAG	P4
CAG3	CAG	P4
IMW1	IM-W (incomplete IM)	P5
IMW2	IM-W (incomplete IM)	P6
IMS1	IM-S (incomplete IM)	P7
IMS2	IM-S (incomplete IM)	P7
IMS3	IM-S (incomplete IM)	P8
IMS4	IM-S (incomplete IM)	P8
EGC	EGC	P8

* NAG, Non Atrophic Gastritis CAG, Chronic Atrophic Gastritis, IM-W, Intestinal metaplasia with wild level; IM-S, Intestinal metaplasia with severe level; EGC, Early Gastric Cancer

Table S2. The clinical information of patients enrolled in the study, Related to **Figure 1** and **STAR Methods.**

Patient ID	*General conditions*						TCM-related manifestations				
	Age	Gender	Smoking (cigs/day)	Alcohol (units/week)	Family history of GC	*H. pylori* positive (P vs N)*	Dry mouth	Scorching stomach pain	Yellow tongue fur	Thick tongue fur	Stick stool
P1	58	male	0	0	N	N	N	N	N	Y	Y
P2	56	Female	0	0	N	N	Y	Y	N	N	Y
P3	51	male	2	4	N	N	N	N	Y	Y	N
P4	62	Female	0	0	N	N	Y	N	Y	Y	N
P5	63	male	1	3	N	P	N	N	N	N	N
P6	48	Female	0	0	N	N	Y	Y	N	N	N
P7	68	male	0	0	Y	P	Y	Y	N	N	N
P8	67	male	0	1	N	N	Y	Y	Y	Y	N
P9	62	male	1	0	N	N	Y	N	N	N	N

* P: *H. pylori positive*, N: *H. pylori* negative.

Table S3. The number of high-quality cells from each sample. Related to **Figure 1** and **STAR Methods.**

Sample ID	NAG1	NAG2	NAG3	CAG1	CAG2	CAG3	IMW1	IMW2	IMS1	IMS2	IMS3	IMS4	EGC
# of cells	2030	1634	1146	2991	6570	4831	1327	1258	2201	1776	1566	2782	2220

Table S4. The known markers for cell lineages in stomach. Related to **Figure 1** and **STAR Methods.**

Categories	Cell lineages	Marker genes
Endocrine	G cell	GAST
	X cell	GHRL
	D cell	SST
Mucous and secretory lineages	pit mucous cell (PMC)	MUC5AC
	gland mucous cell (GMC)	MUC6
	parietal cell	ATP4A,ATP4B,GIF
	chief cell	PGA4,PGA3,LIPF
Immune cells	T cell	CD2, CD3D,CD3E,CD3G
	B cell	CD79A,CD19
	mast cells	TPSAB1,TPSB2
	Macrophage	CD14, CD163, CD68, CSF1R
Stromal cells	Fibroblasts	FAP, PDPN,COL1A2,DCN, COL3A1, COL6A1
	Endothelial cells	PECAM1,VWF,ENG,MCAM
Stem cell	stem cell	OLFM4,SOX2,LGR5,CCKBR
Myocytes	Smooth muscule cell (SMC)	ACTA2,ACTN2,MYL2,MYH2
Proliferative cell	proliferative cell (PC)	MKI67,BIRC5,CDK1
Intestinal cells	goblet cell	TFF3,SPINK4,MUC2
	enteroendocrine cell	CHGA,CHGB,TAC1,TPH1,NEUROG3
	enterocytes	FABP1,CA1,VIL1