




OPEN ACCESS

Original research

Dissecting transcriptional heterogeneity in primary gastric adenocarcinoma by single cell RNA sequencing

Min Zhang,^{1,2} Shuofeng Hu,^{1,3} Min Min,² Yanli Ni,² Zheng Lu,² Xiaotian Sun,^{2,4} Jiaqi Wu,^{1,3} Bing Liu,^{1,2} Xiaomin Ying,^{1,3} Yan Liu ²

► Additional material is published online only. To view, please visit the journal online (<http://dx.doi.org/10.1136/gutjnl-2019-320368>).

¹Academy of Military Medical Sciences, Beijing, China

²The Fifth Medical Center of Chinese PLA General Hospital, Beijing, China

³Center for Computational Biology, Institute of Military Cognition and Brain Sciences, Academy of Military Medical Sciences, Beijing, China

⁴Department of internal medicine, Beijing South Medical District of Chinese PLA General Hospital, Beijing, China

Correspondence to

Prof Yan Liu and Prof Bing Liu, The Fifth Medical Center of Chinese PLA General Hospital, Beijing, 100071, China; liuyan1799@126.com, bingliu17@yahoo.com and Prof Xiaomin Ying, Center for Computational Biology, Institute of Military Cognition and Brain Sciences, Academy of Military Sciences, Beijing, 100850, China; yingxmbio@foxmail.com

MZ, SH and MM contributed equally.

BL, XY and YL contributed equally.

Received 27 November 2019

Revised 19 May 2020

Accepted 26 May 2020

Published Online First

12 June 2020



© Author(s) (or their employer(s)) 2021. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

To cite: Zhang M, Hu S, Min M, et al. *Gut* 2021;**70**:464–475.

ABSTRACT

Objective Tumour heterogeneity represents a major obstacle to accurate diagnosis and treatment in gastric adenocarcinoma (GA). Here, we report a systematic transcriptional atlas to delineate molecular and cellular heterogeneity in GA using single-cell RNA sequencing (scRNA-seq).

Design We performed unbiased transcriptome-wide scRNA-seq analysis on 27 677 cells from 9 tumour and 3 non-tumour samples. Analysis results were validated using large-scale histological assays and bulk transcriptomic datasets.

Results Our integrative analysis of tumour cells identified five cell subgroups with distinct expression profiles. A panel of differentiation-related genes reveals a high diversity of differentiation degrees within and between tumours. Low differentiation degrees can predict poor prognosis in GA. Among them, three subgroups exhibited different differentiation grade which corresponded well to histopathological features of Lauren's subtypes. Interestingly, the other two subgroups displayed unique transcriptome features. One subgroup expressing chief-cell markers (eg, *LIPF* and *PGC*) and *RNF43* with Wnt/ β -catenin signalling pathway activated is consistent with the previously described entity fundic gland-type GA (chief cell-predominant, GA-FG-CCP). We further confirmed the presence of GA-FG-CCP in two public bulk datasets using transcriptomic profiles and histological images. The other subgroup specifically expressed immune-related signature genes (eg, *LY6K* and major histocompatibility complex class II) with the infection of Epstein-Barr virus. In addition, we also analysed non-malignant epithelium and provided molecular evidences for potential transition from gastric chief cells into *MUC6*⁺*TFF2*⁺ spasmolytic polypeptide expressing metaplasia.

Conclusion Altogether, our study offers valuable resource for deciphering gastric tumour heterogeneity, which will provide assistance for precision diagnosis and prognosis.

INTRODUCTION

Tumour heterogeneity including morphological heterogeneity and genetic heterogeneity remains a topic of great interest in cancer research because it poses a series of challenges to both accurate diagnosis and personalised therapy. Gastric adenocarcinoma (GA) is an archetypal example of a

Significance of this study

What is already known on this subject?

- Gastric adenocarcinoma (GA) is a highly heterogeneous malignant disease that is affected by multiple genetic and environmental factors.
- The molecular and cellular heterogeneity in GA are rarely described at single-cell resolution.
- Chief cell predominant GA (GA-FG-CCP) is a rare variant of gastric cancer in the fundic gland region predominant of differentiated chief cells.

What are the new findings?

- We delineated cellular and differentiation heterogeneity within and between patients with GA at single cell resolution.
- Transcriptomically, we delineated the molecular characteristics of a rare GA type, GA-FG-CCP, at single-cell resolution and validated its presence in two public bulk transcriptomic datasets.
- Computational analysis of non-malignant epithelium uncovers a potential transition from chief cells to *MUC6*⁺*TFF2*⁺ spasmolytic polypeptide expressing metaplasia.

How might it impact on clinical practice in the foreseeable future?

- The comprehensive transcriptome atlas is a valuable resource for deciphering gastric tumour heterogeneity.
- Tumour differentiation degree is a potential good predictor of GA prognosis.
- Our results revealed that single-cell RNA sequencing can be employed in detection of rare tumour types (eg, GA-FG-CCP) and accurate diagnosis of GA.
- The trajectory analysis of potential chief cells transition presented should help to clarify the mechanism of gastric carcinogenesis.

heterogeneous malignant disease, which is featured with diverse subtypes and clinical behaviours.^{1,2}

For years, morphological heterogeneity has long been the basis of many tumour grading and prognostic classification systems. According to structural gland atypia and the differentiation degree of GA mucosa, the Lauren's classification³ generally

classifies GA into intestinal and diffuse type. Later, the mixed type has been proposed to describe intermediate histology cases. The intestinal-type GA characteristically form glandular structures and show various degrees of differentiation, which are usually developed through well-characterised sequential pathological stages, such as chronic gastritis (CG), atrophy, intestinal metaplasia and dysplasia. The diffuse-type GA generally exhibit a poorly differentiated status, characterised by poorly differentiated infiltrative growth, associated with aggressive behaviour and poor prognosis. Another histopathology-based guidelines—the WHO classification⁴ has also been developed for GA diagnosis which divides GA into papillary, tubular, mucinous (colloid) and poorly cohesive carcinomas. Although these methods are straightforward and simple to use, the pathology results may vary because of subjective discrimination and complex background factors such as *Helicobacter pylori* (*H. pylori*) bacterium⁵ and Epstein-Barr virus (EBV).⁶ Furthermore, the molecular basis of the phenotypic expression associated with gastric tumour biological behaviour remains poorly understood, limiting the dissection of tumour heterogeneity.

The advent of high-throughput sequencing technology has accelerated the pace of research on molecular profiling of many human tumours dramatically. In addition, it has been recognised that analysis of tumour behaviour at the molecular level are increasingly important for dissecting heterogeneity. To date, large-scale genomic and transcriptomic studies have characterised the genetic and epigenetic heterogeneity of GA. The molecular profiling studies such as The Cancer Genome Atlas (TCGA),⁷ Singapore⁸ and Asian Cancer Research Group⁹ genotyping have extended our knowledge for basic research of GA. The TCGA genotyping proposed a novel molecular classification which divides GA into four subtypes: EBV-positive tumour, microsatellite unstable tumour, genomically stable tumour and chromosomal instability tumour. This study systematically elucidated molecular characteristics of the four subtypes and for the first time revealed that EBV infection is an important factor in the classification of GA. However, these studies relied on the data derived from bulk tissues and obscured the molecular signatures of distinct cell subpopulations, which limits their ability to capture tumour heterogeneity and consequently impedes the precision diagnosis of GA.

Recent advances in single-cell RNA sequencing (scRNA-seq)^{10,11} and microfluidic technology^{12,13} provide methods with which we are able to characterise the transcriptional states of thousands of cells simultaneously. This method allows for an unbiased analysis of cellular characteristics within biological tissues and has been widely applied in the analysis of tumour ecosystem including cancer-immune heterogeneity.^{14–19} On the other hand, targeted biopsy²⁰ under magnifying endoscopy remains the most accurate method of GA diagnosis because it allows the clinician to observe the lesion and to capture tissues rich in tumour cells. Single-cell transcriptome experiments that incorporate the use of endoscopic biopsies will facilitate more accurate GA pathology diagnosis.

In this study, we constructed an unbiased and systemic transcriptomic landscape of GA by scRNA-seq, which covers distinct GA pathology subtypes. The aim of our single-cell study is to provide insights into GA molecular landscape for deciphering GA heterogeneity, with implications for the clinical diagnosis and prognosis of GA.

METHODS

The detailed materials and methods can be found in online supplementary materials.

RESULTS

A single-cell transcriptome atlas in diverse GA types

To investigate the cell populations and the associated molecular characteristics within the gastric mucosa of various lesions, five intestinal-histology GA samples (one with EBV infection), three mixed-histology GA samples, one diffuse-histology GA sample and three non-malignant samples (control, one CG sample from adjacent tumour and two normal samples from gastric polyps) were included in our scRNA-seq survey (figure 1A,B, online supplementary figure S1). The enrolled patients were all newly diagnosed and had not been diagnosed with any other tumours. Patients did not receive chemotherapy or radiotherapy prior to surgery. The clinical characteristics of these participants, including *H. pylori* infection status, EBV infection status, pathological features and tumour classification according to Lauren's system, were recorded at the time of recruitment (table 1). After removal of low-quality cells (see 'Methods' section in online supplementary materials), 27 677 cells were retained for biological analysis, which detected a median of 1227 genes and 3809 transcripts per cell (online supplementary figure S2). After normalisation of gene expression and principal component analysis (PCA), we used graph-based clustering (see 'Methods' section in online supplementary materials) to partition the cells into 14 clusters. These clusters could be assigned to nine known cell lineages through marker genes (figure 1C–E): epithelium (10 411 cells, 37.6%, marked with *EPCAM*, *KRT18* and *KRT8*), parietal cell (215 cells, 0.8%, marked with *ATP4A*), endocrine cell (486 cells, 1.8%, marked with *CHGA*); B cells (6131 cells, 22.1%, marked with *MS4A1* and *CD79A*); T cells (6819 cells, 24.6%, marked with *CD2*, *CD3D* and *CD3E*); macrophages (1053 cells, 3.8%, marked with *CD14*); mast cells (527 cells, 1.9%, marked with *CPA3*); fibroblasts (1116 cells, 4.0%, marked with *ACTA2* and *COL1A2*); endotheliocyte (919 cells, 3.3%, marked with *ENG* and *VWF*). The proportion of each cell lineage varies greatly among different samples (figure 1F).

Classification of malignant and non-malignant epithelium

To distinguish malignant epithelium from non-malignant epithelium, we defined the initial malignant and non-malignant score for each epithelial cell based on differential expression genes between paired tumour and normal tissue samples from the TCGA stomach adenocarcinoma (STAD) dataset (online supplementary figure S3A,B; online supplementary table S1; online supplementary methods). Putative malignant and non-malignant epithelial cells were defined based on the initial scores using the k-means clustering algorithm. Since the initial recognition derived from the TCGA bulk tissues is biased due to the inclusion of non-epithelial cells, we next generated differentially expressed genes between putative malignant and non-malignant epithelial cells, re-calculated the malignant/non-malignant scores and classified epithelial cells. We repeated the process iteratively until the classification result was stable. Finally, we identified 5635 malignant and 4776 non-malignant epithelial cells (figure 2A–C). A panel of tumour-specific genes (figure 2D–F; online supplementary figure S3C, online supplementary table S2), including *CLDN4*, *CLDN7*, *TFE3* and *REG4*, was found to be upregulated in malignant epithelium compared with non-malignant epithelium ($p < 2 \times 10^{-16}$). Non-malignant epithelium exhibited high expression of marker genes associated with the secretion of gastric mucus and digestive enzymes, such as *MUC5AC*, *GKN1*, *PGC* and *LIPF* (figure 2D–F, $p < 2 \times 10^{-16}$). To decipher the molecular characteristics difference of malignant and non-malignant epithelium, we performed gene set

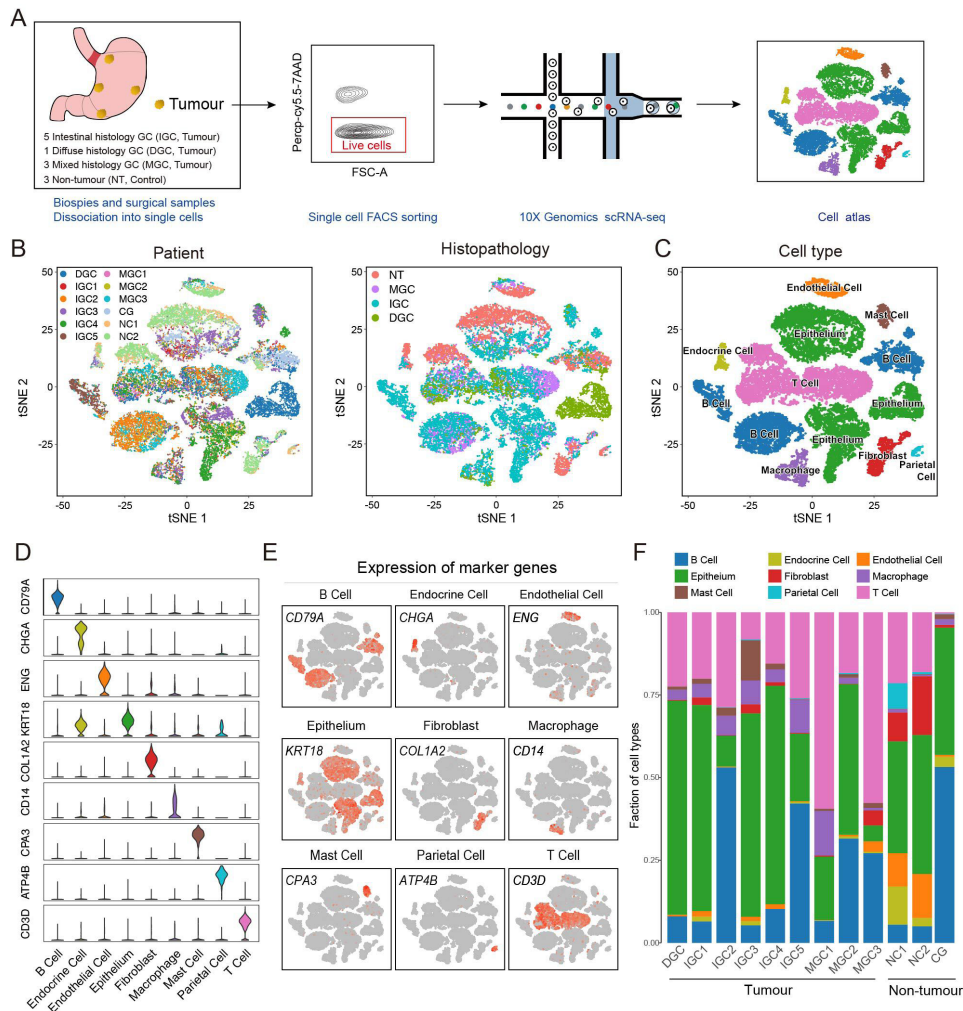


Figure 1 Cellular atlas of gastric tumours and non-tumour gastric tissues. (A) Schematic diagram for the generation of single-cell RNA sequencing (scRNA-seq) data. Nine gastric tumours (seven endoscopic samples and two surgical samples) and three non-tumour samples (one chronic gastritis sample and two normal samples) were collected. (B) t Stochastic neighbour Embedding (tSNE) plots for the 27 677 high-quality cells showing sample origin and Lauren's classification. (C) tSNE plots showing cell types for the 27 677 cells. (D) Violin plots showing the smoothed expression distribution of marker genes in nine cell types. (E) tSNE plots showing the expression levels of canonical marker genes for nine cell types. (F) The proportion of each cell type in 12 samples.

Table 1 Clinical characteristics of each sample used in scRNA-seq study and the cell number of each sample after quality control

Sample	Age	Sex	Histopathological diagnosis	Site of origin	Lauren's classification	<i>Helicobacter pylori</i>	EBER	Number of cell
Tumour samples used in scRNA-seq study								
DGC	54	F	Poorly differentiated adenocarcinoma	Cardia	Diffuse	-	-	3026
IGC1	67	F	Moderately poorly differentiated adenocarcinoma	Angular incisure	Intestinal	+	-	952
IGC2	65	M	Moderately differentiated adenocarcinoma	Cardia	Intestinal	-	-	2616
IGC3	56	M	Moderately differentiated adenocarcinoma	Cardia	Intestinal	-	-	2136
IGC4	74	M	Moderately poorly differentiated adenocarcinoma	Greater curvature	Intestinal	+	-	2141
IGC5	72	M	Moderately poorly differentiated adenocarcinoma	Angular incisure	Intestinal	-	+	2695
MGC1	56	M	Poorly differentiated adenocarcinoma	Antrum	Mixed	-	-	881
MGC2	63	M	Poorly differentiated adenocarcinoma	Angular incisure	Mixed	-	-	2405
MGC3	62	M	Poorly differentiated adenocarcinoma, partial signet ring cell carcinoma	Angular incisure	Mixed	+	-	3318
Non-tumour samples used in scRNA-seq study								
CG	53	F	Chronic gastritis	Cardia	N/A	+	-	3568
NC1	38	M	Normal	Greater curvature	N/A	-	-	1773
NC2	43	F	Normal	Greater curvature	N/A	-	-	2166

EBER, EBV-encoded RNA; N/A, not available.

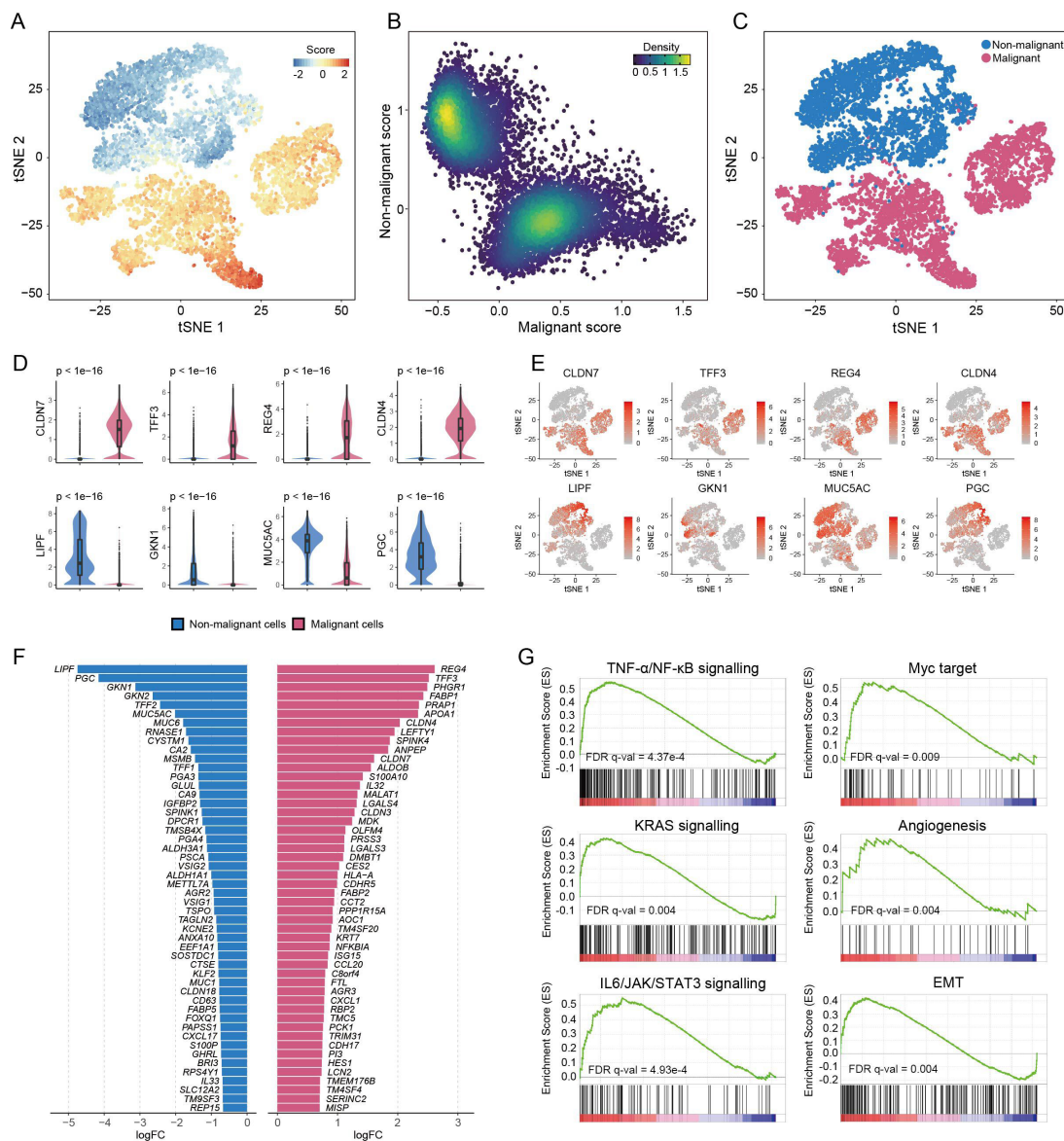


Figure 2 Classification of 10 411 epithelial cells as malignant or non-malignant. (A) tSNE of 10 411 epithelial cells, colour-coded according to malignant score minus non-malignant score. (B) Scatter plot displaying the distribution of malignant scores (x-axis) and non-malignant scores (y-axis). Each point corresponds to a cell and is colour-coded to reflect density. (C) tSNE plot of the classification of malignant and non-malignant cells. (D) Violin plots and corresponding box plots showing the expression of eight representative genes with differential expression between malignant and non-malignant cells. (E) Expression of eight representative genes with differential expression, shown using tSNE plots. (F) Bar plot showing fold changes of signature genes in malignant cells and non-malignant cells. (G) Gene set enrichment analysis (GSEA) results showing the enrichment of six gastric tumour-associated gene sets in malignant epithelial cells. EMT, epithelial-mesenchymal transition; IL6, interleukin-6; JAK, Janus kinase; NF- κ B, nuclear factor-kappa B; TNF, tumour necrosis factor- α ; STAT, signal transducer and activator of transcription.

enrichment analysis (GSEA). Compared with non-malignant epithelium, malignant epithelium was enriched for signalling pathways such as tumour necrosis factor- α /nuclear factor-kappa B, KRAS and interleukin-6/Janus kinase/signal transducer and activator of transcription (figure 2G). There are also other enriched gene sets that are crucial for cancer development and progression such as MYC target, epithelial-mesenchymal transition and angiogenesis (figure 2G).

The vast majority of epithelium (96.9%) derived from non-tumour gastric mucosa (control) was classified in the putative non-malignant group (online supplementary figure S3D), demonstrating the reliability of our method for malignant cell identification. We noticed that 79.7% (5367/6734) of epithelium from biopsy tumour samples were classified as malignant

while only 29.4% (224/762) of epithelium from surgically resected samples were classified as malignant. This difference indicates that endoscopic biopsies may be more accurate for the diagnosis of tumour malignancy. The proportion of putative non-malignant and malignant epithelium in each sample is provided in online supplementary figure S3D.

We also inferred copy-number variations (CNVs)^{21 22} in each cell based on smoothed expression profiles across chromosomal intervals. This computational method has been applied to identify malignant cells in single-cell analysis. As expected, 4/4776 of the putative non-malignant cells displayed abnormal CNV signals. However, only 25.0% of putative malignant cells exhibited high levels of CNV (online supplementary figure S4A,B). This finding supports the TCGA analysis result⁷ that most of

GA samples had low CNV signals examined by whole-exon sequencing (online supplementary figure S4C,D), indicating that malignant cell identification based on CNV signals is not suitable for GA.

Transcriptome profiling reveals the relationship among chief cell, neck cell and SPEM

The preneoplastic gastric metaplasia induced by chronic inflammation has always been a hot topic in gastric carcinogenesis. Several studies have reported that chief cells may differentiate abnormally into neck cells and then can lead to the formation of mucus cell metaplasia known as SPEM due to a loss of

parietal cells.^{23–25} According to our data, the *ATP4A* and *ATP4B* expressing parietal cell are rarely detected in cancerous and CG mucosa tissue samples compared with that of normal gastric tissues (figure 1E). We also performed subclustering of non-malignant cells, which revealed four distinct groups (figure 3A). The cells of G1 group account for 78.0% of total cells and can be distinguished as surface cells by expression of gene *TFF2*, *GKN1* and *MUC5AC* (figure 3C). The cells of group G2 exhibited a high expression level of marker genes such as *PGA3*, *PGA4* and *LIPF* and are defined as chief cells (figure 3C). These cells demonstrate a lack of expression of the bHLH transcription factor *MIST1* (data not shown), which is essential for chief

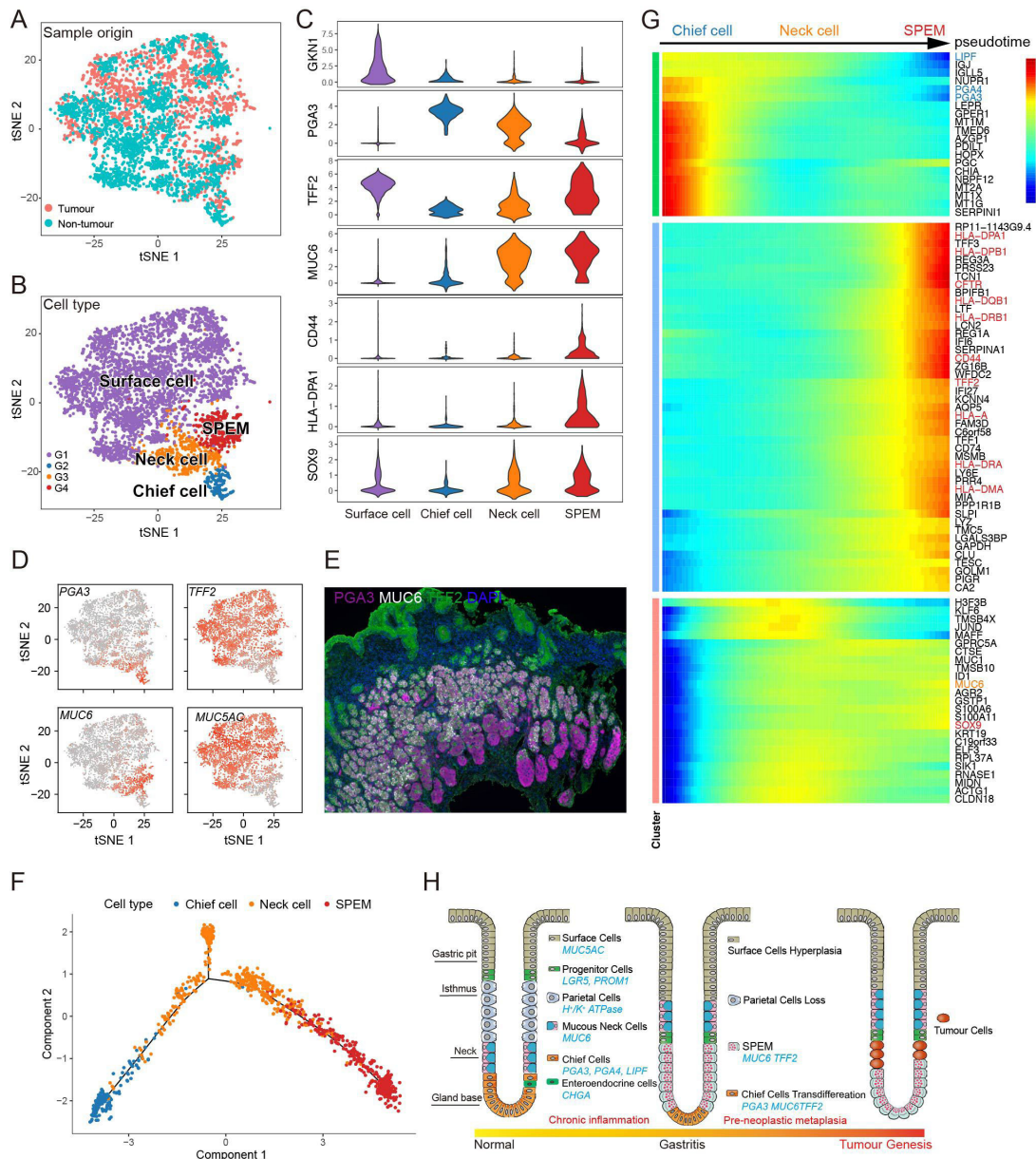


Figure 3 Cell clusters of non-malignant epithelial cells and potential transition between cell types. (A) tSNE plot of 4776 non-malignant epithelial cells colour-coded for sample origin. (B) tSNE plot of non-malignant epithelial cells colour-coded to reflect cell cluster. (C) Violin plots showing the smoothed expression distribution for marker genes for each cell type. (D) tSNE plot of non-malignant epithelial cells, colour-coded for the expression of marker genes. (E) Immunostaining showing the spatial distribution of chief cell, neck cell and spasmolytic polypeptide expressing metaplasia (SPEM). (F) Monocle 2 pseudotime analysis for chief cells, neck cells and SPEM. (G) Heatmap showing scaled expression of dynamic genes along the pseudotime. Rows of the heatmap represent genes that show dynamic changes along the pseudotime, and these genes were clustered into three groups according to their expression pattern along the pseudotime. (H) A simple model for the origin and progression of gastric metaplasia cells in human stomach.

cell maturation.²⁴ We also found that the cells of G3 expressed marker PGA3 and MUC6, which could be defined as hyperplastic/hypertrophic mucous neck cells.²⁶ The expressions were also validated by immunostaining (figure 3E). The G4 cluster is predominantly composed of cells from the mucosa of cancerous lesions. These cells are characterised by high expression levels of neck cell marker MUC6, metaplasia transcripts *TFF2*, *CD44* and *SOX9* and major histocompatibility complex class II (MHC-II) genes (eg, HLA-DPA1 and HLA-DRA, figure 3C and online supplementary figure S5A), leading us to characterise them as SPEM.^{24,26} Immunostaining also showed that chief cell, neck cell and SPEM co-exist in the same tissue, and are arranged in the order of chief cell, neck cell and SPEM from gastric gland base to gland neck gradually, supporting the inference of the transition of chief cell to neck cell to SPEM (figure 3E). We next used the Monocle²⁷ analysis toolkit to perform cell trajectory analysis to further investigate the potential transition between cell types. The pseudotime trajectory axis derived from Monocle indicates that chief cells could transdifferentiate into mucosa neck cells then into SPEM (figure 3F). Pseudotemporal expression dynamics of specific representative genes also marked the progression of chief cells to mucosa neck cells then into SPEM (figure 3G and online supplementary figure S5B). The results presented here for the first time delineate the potential differentiation paths of chief cells at a single-cell level and indicate that they may give rise to metaplastic cells (figure 3H).

Transcriptional landscape heterogeneity of GA malignant cells

We performed clustering analysis of 5635 malignant cells and revealed five prominent cell subgroups (C1–C5). The malignant cells were grouped mainly according to Lauren's histopathological type and background factor (EBV infection) (figure 4A). We also observed that the proportion of C1–C5 varies among nine patients (figure 4C). C1 consisted almost entirely (96.9%) of malignant cells from the diffuse-type sample (DGC), while C2 was predominantly derived from intestinal-type samples (IGC) (97.1%) (figure 4B). C3 contained a mixture of cells from mixed-type (MGC) and intestinal-type samples (figure 4B). C4 was mainly derived from one intestinal-type GC sample (IGC3) and C5 comprised cells from the EBV+ patient (IGC5) (figure 4B). By comparing gene expression profiles (figure 4D), we found that C1/C2/C3 were in correspondence with the three canonical subtypes of Lauren's classification and C4/C5 were two novel cell subgroups with distinct molecular characteristics compared with C1/C2/3 (figure 4A, online supplementary table S3).

Gastric tumour is well known to exhibit various differentiation degree.²⁸ Our data reveal that the expression of *KRT20*,²⁹ an epithelial differentiation-related gene, is varying among tumour cells. To quantitatively explore the differentiation heterogeneity of malignant cells within and between tumours, we defined a differentiation score based on *KRT20* and its positive-correlated genes (ie, *PHGR1*,³⁰ *MDK*,³¹ *CHDR2*, *RARRES3*,³² *GPA33*,³³ *SLC5A1*³⁴ and *MUC13*,³⁴ (online supplementary table S3)), all of which have been reported to be cellular differentiation-related makers (figure 4E). Differentiation scores varied among subgroups and patients (figure 4F), revealing a high degree of intertumour heterogeneity. Notably, we also uncovered high intratumour differentiation heterogeneity in patients such as IGC1 and IGC4, which contained two distinct cell groups with various differentiation scores (figure 4H), which was also validated by *KRT20* immunostaining (online supplementary figure S6). Furthermore, malignant cells of IGC1 and IGC4

displayed similar pseudotime trajectories from low degrees to high degrees of differentiation (figure 4H). Our results suggest that the distinct cell groups in IGC1 and IGC4 may represent two different differentiation stages. Next, we evaluated the association of tumour differentiation degree with patient outcome in bulk transcriptomic data (TCGA STAD dataset). The survival analysis in 408 patients reveals that low differentiation scores can significantly predict worse overall survival (figure 4I). This supports that gastric tumour with a low differentiation degree might be more aggressive.

Molecular features of classic Lauren's histopathology types of GA

We next investigated the molecular characteristics of C1, C2 and C3 (figure 5A). C1 highly expressed a few genes which were widely expressed in gastric epithelial cells (eg, *CLDN18* and *TFF2*) (figure 5B), but these genes were rarely detected in C2. In contrast, C2 exhibited a significant molecular feature resembling small intestine, which is supported by the widespread expression of enterocyte markers (eg, *APOA1* and *FABP2*) (figure 5C). These results were also validated by *KRT20* and *CLDN18* immunostaining (online supplementary figure S7). C3 did not show clear molecular features but partially expressed both markers from C1 and C2, suggesting it was an intermediate subgroup of C1 and C2, instead of a mixture of C1 and C2. In addition, we observed that differentiation scores vary among the three subgroups (figure 5D). The expression levels of *PHGR1*, *KRT20* and *MDK* were highest in C2 and moderate in C3, but were lowest in C1 (figure 5E). These results illustrated that C1 represented a subpopulation of undifferentiated status while C2 represented a different subpopulation of well-differentiated malignant cells with features of mature enterocyte and C3 represented an intermediate state between C1 and C2. Our result delineated the unique molecular features of different Lauren's histological types at single cell resolution.

Transcriptome profiling of GA-FG-CCP

The malignant cells of C4 (figure 6A) were mainly derived from an intestinal-type GA sample IGC3 (figure 4C). H&E staining results showed tumour sites were located in the gastric fundic gland, which were mainly composed of differentiated pellet cells (online supplementary figure S8A). Our scRNA-seq data revealed that C4 widely expressed chief cell markers (eg, *LIPF*, *PGC* and *PGA3*, figure 6B) and exhibited a moderate level of differentiation scores (figure 5D). We identified a few genes specifically expressed in C4, some of which were reported to regulate Wnt/ β -catenin signalling pathway, that is, *RNF43*³⁵ (figure 6B). Gene ontology (GO) enrichment analysis confirmed that signature genes in C4 are enriched for regulation of growth and Wnt/ β -catenin signalling pathway (figure 6C). Immunofluorescence staining showed that these cells predominate in the stomach gland base, with strong co-expression of MUC6 and pepsinogen-I (PGA3) (figure 6D). The histopathological findings and phenotypic expression pattern of the cells in subgroup C4 are highly consistent with the cell features of GA-FG-CCP.³⁶

To further validate our findings of GA-FG-CCP, we investigated whether GA-FG-CCP are presented in public transcriptomic datasets. Unsupervised non-negative matrix factorisation (NMF, online supplementary figure S8B, see 'Methods' section in online supplementary materials) was applied to cluster 86 intestinal-type patients in TCGA STAD dataset. Among the four clusters identified by NMF, we noticed that cluster 3 (n=11) highly expressed C4 subgroup-specific genes but rarely expressed

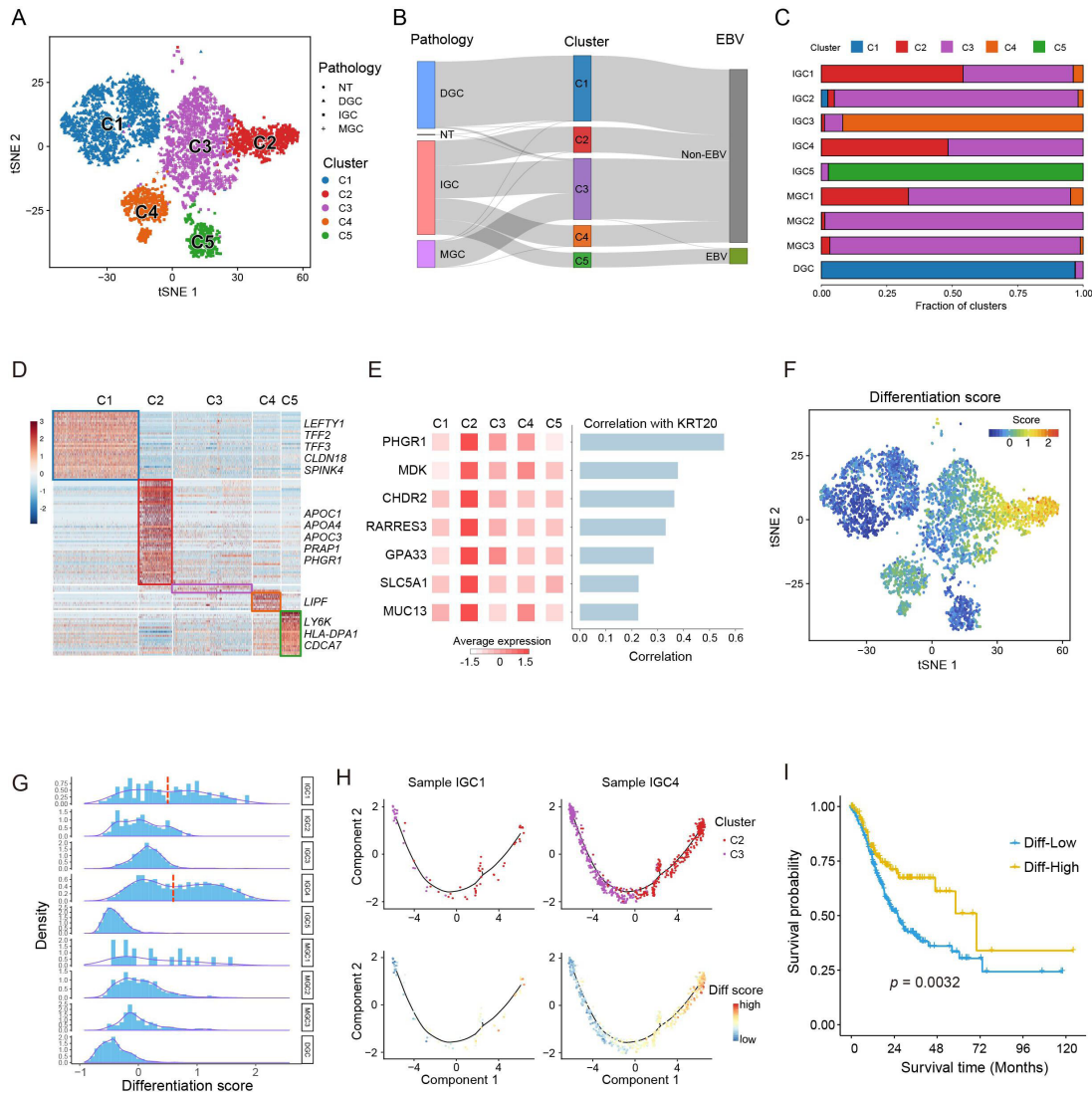


Figure 4 The landscape of intratumour and intertumour heterogeneity in gastric adenocarcinoma (GA). (A) tSNE plot of 5635 tumour cells, colour-coded for pathology and five molecular clusters. (B) Sankey diagram showing the distribution of Lauren's histological characteristics and Epstein-Barr virus (EBV) infection in five clusters. (C) Fraction of five cell subgroups in nine patients. (D) Relative expression of signature genes with adjusted p value < 0.01 and fold change > 1 . (E) Average expression of seven epithelial differentiation-related genes in five subgroups. Right, Pearson's correlation coefficient between seven genes and *KRT20*. (F) tSNE plot showing differentiation scores of 5635 tumour cells. (G) Histogram of differentiation scores in nine patients. The red dash lines partition two peaks of differentiation score distribution. (H) Trajectory of malignant cells in IGC1 and IGC4 constructed by Monocle 2. Each point corresponds to a single cell and is colour-coded by cell subgroup (top) and differentiation score (bottom). (I) The overall survival of Diff-high and Diff-low groups of patients. $P = 0.0032$, calculated using log rank test.

enterocyte markers (figure 6E). The average expression of C4-specific genes used for NMF was also highest in cluster 3 (figure 6F). Intriguingly, *CTNNB1*,³⁷ which acts as an important signal transducer in Wnt/ β -catenin signalling pathway, exhibited significantly higher expression in cluster 3 than others (Student's t -test, $p = 0.003$) (figure 6G). A similar cluster with molecular features of GA-FG-CCP was also identified in the Singapore dataset⁸ (online supplementary figure S8C). We further examined the histopathological images of the inferred GA-FG-CCP cases in GDC Data Portal website (<https://gdc.cancer.gov/>). Interestingly, most of tumour tissue samples ($n = 10$) are characterised by differentiated fundic gland columnar cells (online supplementary figure S8D and online supplementary figure S9), which is highly consistent with the histopathological findings of IGC3. Altogether, our study demonstrated the molecular features of GA-FG-CCP at single-cell level and validated its presence in bulk

transcriptomic datasets, which may enhance our understanding of gastric tumour molecular and cellular diversity.

EBV+ malignant cells highly express immune signature gene

The C5 subgroup comprised almost entirely of cells from intestinal-type patient (IGC5) with EBV infection (figures 7A and 4C). The vast majority of cells in C5 displayed unique molecular features: high levels of lymphocyte antigen-6 (Ly6) family members (eg, *LY6K*) and MHC-II genes (*HLA-DPA1* and *HLA-DPB1*) (figure 7B), which are tightly linked with immune responses. GO enrichment analysis indicated that this subgroup showed high levels of immune-related signalling pathways such as interferon- α (IFN- α) response, antigen processing and presentation (figure 7D), suggesting possible cell-cell communication between immune cells and tumour cells. The immunostaining

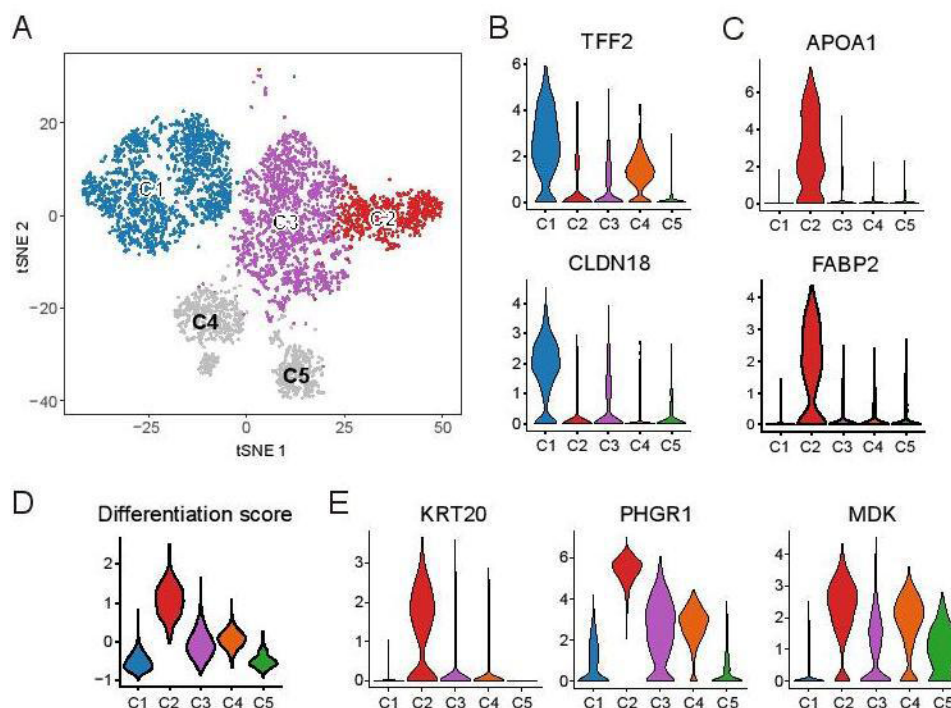


Figure 5 Molecular characteristics of subgroup C1, C2 and C3. (A) tSNE plot of tumour cells, colour-coded for C1, C2 and C3. (B and C) Violin plots showing the expression of *CLDN18*, *TFF2*, *APOA1* and *FABP2*. (D) Violin plots showing differentiation scores of five clusters. (E) Violin plots showing the expression of *KRT20*, *PHGR1* and *MDK* in five cell subgroups.

assay verified our findings that HLA class II molecules and LY6K exhibited higher expression levels in EBV+ cases than EBV- cases (figure 7E). The TCGA bulk sequencing result also revealed that HLA class II molecules highly expressed in EBV+ samples (figure 7G). Besides, as EBV infects B cells, we compared the cell cycle phases of B cells between EBV+ and EBV- patients and found more cycling cells (S and G2M) in IGC5 than EBV- patients (figure 7H). Altogether, our analysis uncovers unique molecular features of malignant cells from EBV-infected patients with GA, which highlights the necessity of EBV diagnosis for histological classification of GA.

DISCUSSION

GA is a highly heterogeneous disease that is affected by multiple genetic and environmental factors, which poses a series of challenges to both accurate diagnosis and personalised therapy. Emerging scRNA-seq has been widely used for the study of tumour heterogeneity, including analysis of tumour developmental hierarchies, drug resistance programmes, cell-cell communication and immune infiltration patterns.^{21 38 39} In this study, we employed the technology to generate a comprehensive landscape of distinct GA types at single cell resolution.

Focused on malignant cells, we identified five subgroups (C1–C5) with different transcriptomic characteristics. C1 and C2 exhibited molecular features of diffuse-type and intestinal-type GA, respectively. C3 which were mainly derived from mixed-histology patients showed no unique features but partially expressed signature genes of C1 and C2. Furthermore, we found that the differentiation degree varies among the five subgroups and within/between tumours, revealing a high degree of tumour heterogeneity. In addition, the differentiation degree is low in diffuse-type tumours and varies in intestinal-type cases. These results support that cancer cells of diffuse-histology GA are poorly differentiated, while the intestinal-type

cases can exhibit various differentiation degree.⁴⁰ Additionally, our results show that differentiation degree is associated with patients' outcome. Low tumour differentiation degrees can predict poor prognosis in GA, suggesting GA with low differentiation degree may be more aggressive. Together, this is the first report bridging the Lauren's pathological types to molecular evidences at single cell resolution, which may facilitate more accurate GA diagnosis and prognosis assessment in GA clinical management.

A type of well-differentiated GA, GA-FG-CCP was reported by limited published studies mainly from Japan.^{36 41 42} Most of the tumours are at early stages when diagnosed, which are stained positive for PGA3 and MUC6,³⁶ and the Wnt/ β -catenin signalling pathway is activated. However, whether such tumours are also present in the patients with advanced stage gastric cancer and more detailed features of such tumours remain poorly understood. In this study, we found that an additional panel of markers such as *DPEP1* and WNT pathway associated gene *RNF43*³⁵ also expressed specifically in GA-FG-CCP. We then validated the presence of GA-FG-CCP in two public bulk transcriptomic datasets (TCGA and Singapore dataset) and found that both datasets harbour subclusters with transcriptome features of GA-FG-CCP. The histopathological images of the inferred GA-FG-CCP cases were examined by two experienced pathologists. The results show that the histopathological features are highly consistent with the histopathological features of GA-FG-CCP. Our findings suggest that GA-FG-CCP may be a GA type at early and advanced stages occurred both in Asian and western population. Since the inferred GA-FG-CCP cases account for only 2%–3% (10/440) of total TCGA STAD patients, it may be a rare GA type. To our knowledge, this is the first study demonstrating the histopathological and phenotypic features of GA-FG-CCP at single cell resolution, which may enhance our understanding of gastric tumour cellular and molecular diversity.

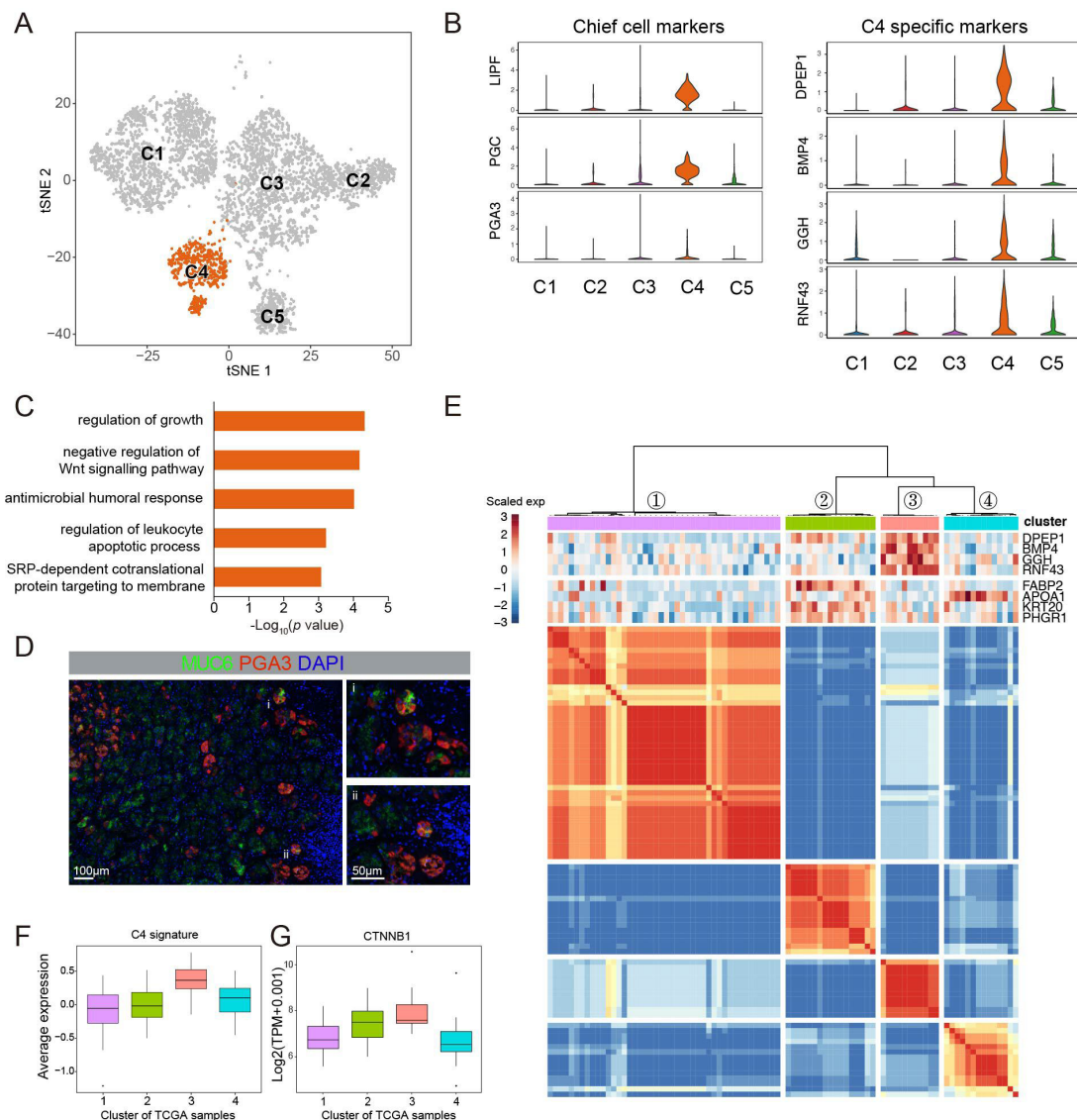


Figure 6 Molecular characteristics of subgroup C4 and definition of GA-FG-CCP. (A) tSNE plot of tumour cells, coloured for C4. (B) Violin plots showing the expression of chief cell markers, differentiation-related marker and C4-specific genes in five subgroups. (C) The enriched gene ontology terms for top 50 signature genes in C4 using Metascape (www.metascape.org). (D) Immunofluorescence staining indicates the co-expression of MUC6, PGA3 and DAPI (nuclei) on malignant cells of GA-FG-CCP. Scale bars, 100 μ m. (E) Heatmap showing the clustering result of 86 intestinal-type patients in TCGA STAD dataset. The colours in heatmap correspond to the concordance between samples (see 'Methods' section in online supplementary materials). Upper panel showing the relative expression of four specific genes in C4 and two differentiation-related markers. (F) Box plot showing the average expression of C4-specific genes selected for clustering of TCGA samples in four clusters. (G) Box plot showing the expression *CTNNB1* in four clusters of TCGA samples.

The EBV-positive GA was first proposed in 2014⁷ based on genomic and epigenomic data at population level, which displays recurrent *PIK3CA* mutations, extreme DNA hypermethylation and amplification of *JAK2* and immune-related genes such as *CD274* (also known as *PD-L1*) and *PDCD1LG2* (also known as *PD-L2*). In this study, C5 was derived principally from an EBV-positive sample (IGC5), characterised by high expression of MHC-II genes (*HLA-DPA1* and *HLA-DPB1*) and enriched for immune active GO terms. Recent works have shown that higher expression of MHC genes in tumours was associated with response to checkpoint blockade immunotherapy.^{43–44} Additionally, it has been reported that virus infection can lead to cellular phenotypic alteration.^{45–46} Our results demonstrate that EBV infection is a pathogenic factor of GA that may play an important role in immune system activation and tumour cellular

phenotypic alteration and further proved that EBV-positive GA is immune active. Therefore, EBV infection detection should be a requisite component of the clinical management of GA.

There is an argument regarding gastric chief cells being the point of origin of metaplasia and cancer cells.^{25–47–50} Chief cells are generally considered to originate from mucous neck cell progenitors in the mid-portion of the oxyntic mucosa. This transition involves the expression of the bHLH transcription factor *MIST1*.^{23–24} However, parietal cell death caused by chronic inflammation has been reported to affect the expression of *MIST1*, with subsequent effects on chief cell maturation and differentiation. Investigators have also reported that loss of parietal cells⁵¹ is a prerequisite for the induction of metaplastic cell lineages and that this loss can induce chief cell to transdifferentiate back into mucosa neck cells.⁵² In addition, lineage tracing⁵²

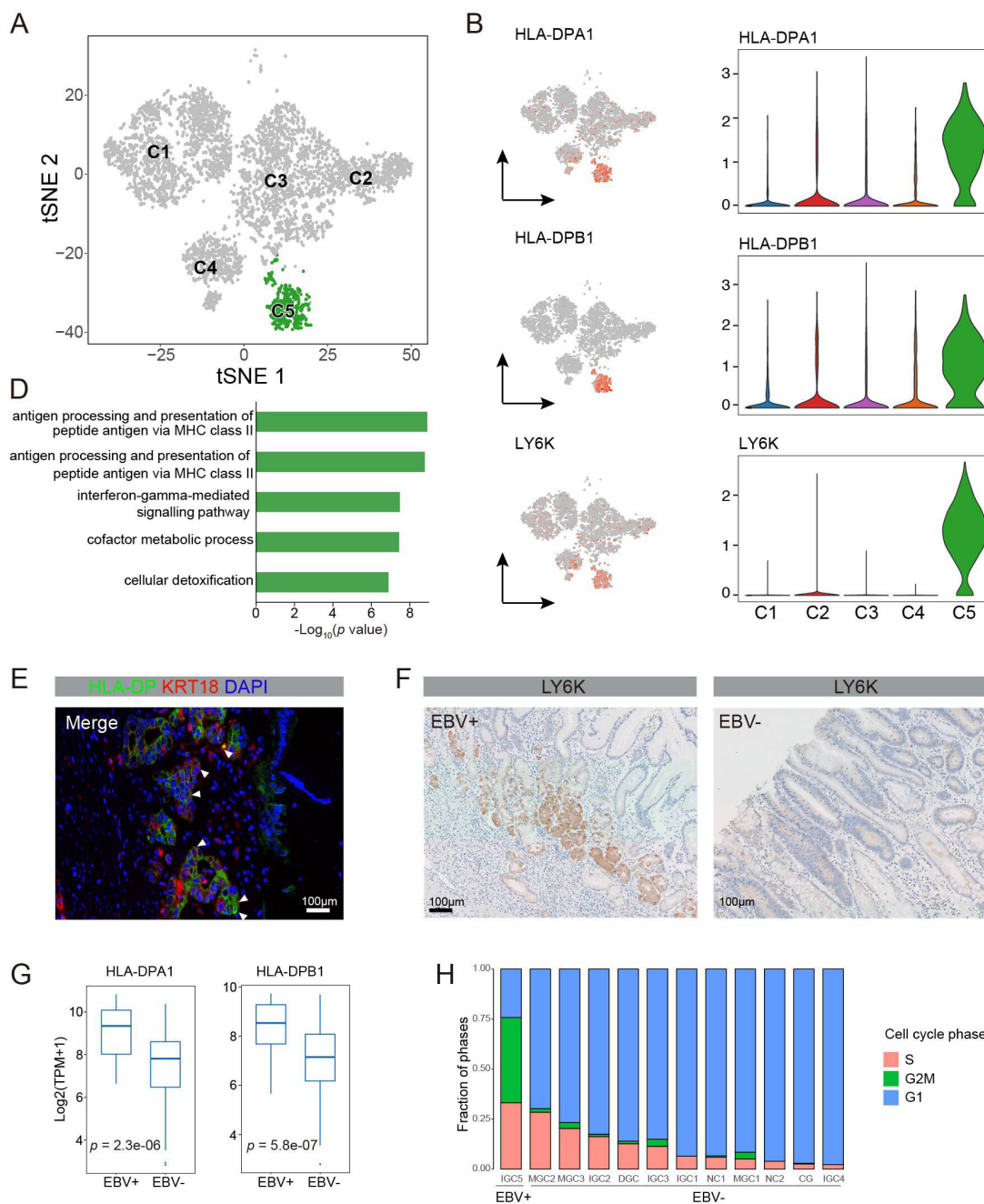


Figure 7 Molecular characteristics of subgroup C5. (A) tSNE plot of tumour cells, coloured for C5. (B) tSNE plots showing the expression of *HLA-DPA1*, *HLA-DPB1* and *LY6K* in tumour cells. (C) Violin plots showing the expression of *HLA-DPA1*, *HLA-DPB1* and *LY6K* in five subgroups. (D) The enriched gene ontology terms for top 50 signature genes in C5 using Metascape (www.metascape.org). (E) Immunofluorescence staining indicates the co-expression of HLA-DP, KRT18 and DAPI (nuclei) on malignant cells of Epstein-Barr virus (EBV)+ gastric adenocarcinoma (GA). Scale bars, 100 μ m. (F) Immunohistochemistry staining indicates the expression of *LY6K* on EBV+ and EBV- tumour samples, Scale bars, 100 μ m. (G) Box plots showing the expression *HLA-DPA1* and *HLA-DPB1* in EBV+ and EBV- patients of TCGA STAD dataset, p value calculated by Student's t-test. (H) Fraction of G1, S and G2M B cells in 12 samples.

or BrDU tracking⁴⁹ experiments in mouse gastric corpus and immunohistochemical examinations of preneoplastic human stomach lesion biopsies⁴⁹ show that metaplastic cells may arise from the transdifferentiation of mature chief cells. In our study, Monocle pseudotime trajectory analysis and immunostaining also reveals that metaplastic cells may originate from chief cells transdifferentiation. The above results indicate that chief cells might be the origin of metaplastic cell lineages as well as gastric neoplastic cells, which further enhances our understanding of gastric carcinogenesis.

In summary, we constructed a comprehensive single-cell transcriptome atlas of 27 677 cells from 9 samples of GA and 3 sample of non-malignant gastric mucosa. Focusing on epithelial cells, we identified a panel of biomarkers for discriminating between benign and malignant epithelium. The results of potential chief cells transition presented should help to clarify the mechanism of gastric carcinogenesis. We also depicted the diverse differentiation degrees within and between tumours and predicted the poor outcomes of patients with low differentiation degrees. Furthermore, we identified GA-FG-CCP and EBV+

GA using both scRNA-seq and bulk transcriptomic dataset. As this is a preliminary study, the limitation of our analysis is the small number of patient cases enrolled in this study and our findings require verification in a larger patient cohort. On the other hand, we employed scRNA-seq method of which the sequencing depth is generally lower than RNA-seq of bulk tissues. The low depth brings several issues. First, the scRNA-seq data are relatively sparse owing to the low capture efficiency and high drop-outs. Second, transcript levels are subject to temporal fluctuation which further contributes to the high frequency of zero observations in scRNA-seq data. Finally, the low depth of scRNA-seq also make the detection of low-expressed genes (eg, non-coding RNA) and alternative splicing difficult. Therefore, the numbers of expressed genes detected by scRNA-seq are typically smaller compared with bulk RNA-seq. Nonetheless, despite the above limitations, scRNA-seq provides us a technique to decipher the transcriptome of single cells, which is a big step forward. The molecular profiling of GA characterised by single-cell transcriptome profiles may pave the way for dissecting tumour heterogeneity, with implications for clinical practice.

Acknowledgements The authors would like to thank the following agencies and foundations for their financial support: National Key R&D Program of China, Natural Science Foundation of Beijing Municipality and Outstanding Youth Training Fund of the PLA General Hospital. The authors would also like to thank Yun Shao and Yanhong Tai of the Department of Pathology for assistance with the assessment of gastric tumour pathology in both samples used in our scRNA-seq and TCGA dataset.

Contributors MZ designed the experiments, collected the biopsies, analysed the data and wrote the manuscript. SH was responsible for data acquisition, analysis and interpretation, and writing the manuscript. MM collected the biopsies. YN was responsible for FACS sorting. YL, XY and BL were responsible for the study concept, design, interpretation and revising the manuscript.

Funding This work was supported by National Key R&D Program of China (Grant No. 2017YFC0908300), Natural Science Foundation of Beijing Municipality (Grant No. 7192201), Outstanding Youth Training Fund of the Chinese PLA General Hospital (Grant No. 2019-JQPY-001).

Competing interests None declared.

Patient and public involvement Patients and/or the public were involved in the design, conduct, reporting or dissemination plans of this research. Refer to the 'Methods' section in online supplementary materials for further details.

Patient consent for publication Obtained.

Ethics approval The study was approved by the the ethical review community of the Fifth Medical Center of Chinese PLA General Hospital (No. ky-2017-8-11). All study participants provided written informed consent.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available in a public, open access repository. Data are available on reasonable request. The accession number for the raw data reported in this paper have been deposited in the Genome Sequence Archive (GSA) under accession number HRA000051 that can be accessed at: <http://bigd.big.ac.cn/gsa-human/>

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iD

Yan Liu <http://orcid.org/0000-0002-1019-6567>

REFERENCES

- Tan P, Yeoh K-G. Genetics and molecular pathogenesis of gastric adenocarcinoma. *Gastroenterology* 2015;149:1153–62.
- Hudler P. Challenges of deciphering gastric cancer heterogeneity. *World J Gastroenterol* 2015;21:10510–27.
- Lauren P. 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* 1965;64:31–49.
- Bosman FT, Carneiro F, Hruban RH, et al. *WHO classification of tumors of the digestive system*, 2009.
- Eid R, Moss SF. Helicobacter pylori infection and the development of gastric cancer. *N Engl J Med* 2002;346:65–7.
- Liang Q, Yao X, Tang S, et al. Integrative identification of Epstein-Barr virus-associated mutations and epigenetic alterations in gastric cancer. *Gastroenterology* 2014;147:1350–62.
- Cancer Genome Atlas Research Network. Comprehensive molecular characterization of gastric adenocarcinoma. *Nature* 2014;513:202–9.
- Lei Z, Tan IB, Das K, et al. Identification of molecular subtypes of gastric cancer with different responses to PI3-kinase inhibitors and 5-fluorouracil. *Gastroenterology* 2013;145:554–65.
- Cristescu R, Lee J, Nebozhyn M, et al. Molecular analysis of gastric cancer identifies subtypes associated with distinct clinical outcomes. *Nat Med* 2015;21:449–56.
- Baslan T, Hicks J. Unravelling biology and shifting paradigms in cancer with single-cell sequencing. *Nat Rev Cancer* 2017;17:557–69.
- Wen L, Tang F. Single-Cell sequencing in stem cell biology. *Genome Biol* 2016;17:71.
- Macosko EZ, Basu A, Satija R, et al. Highly parallel genome-wide expression profiling of individual cells using Nanoliter droplets. *Cell* 2015;161:1202–14.
- Prakadan SM, Shalek AK, Weitz DA. Scaling by shrinking: empowering single-cell 'omics' with microfluidic devices. *Nat Rev Genet* 2017;18:345–61.
- Pastushenko I, Brisebarre A, Sifrim A, et al. Identification of the tumour transition states occurring during EMT. *Nature* 2018;556:463–8.
- Tirosh I, Venteicher AS, Hebert C, et al. Single-Cell RNA-seq supports a developmental hierarchy in human oligodendrogloma. *Nature* 2016;539:309–13.
- Venteicher AS, Tirosh I, Hebert C, et al. Decoupling genetics, lineages, and microenvironment in IDH-mutant gliomas by single-cell RNA-seq. *Science* 2017;355:science.aai8478.
- Chevrier S, Levine JH, Zanotelli VRT, et al. An immune atlas of clear cell renal cell carcinoma. *Cell* 2017;169:e18:736–49.
- Savas P, Virassamy B, Ye C, et al. Single-Cell profiling of breast cancer T cells reveals a tissue-resident memory subset associated with improved prognosis. *Nature Medicine* 2018.
- Zheng C, Zheng L, Yoo J-K, et al. Landscape of infiltrating T cells in liver cancer revealed by single-cell sequencing. *Cell* 2017;169:e16:1342–56.
- Purnell SD, Bloom JB, Valera V, et al. Targeted biopsy: benefits and limitations. *Curr Opin Urol* 2018;28:219–26.
- Patel AP, Tirosh I, Trombetta JJ, et al. Single-Cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science* 2014;344:1396–401.
- Tirosh I, Izar B, Prakadan SM, et al. Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* 2016;352:189–96.
- Lennerz JKM, Kim S-H, Oates EL, et al. The transcription factor Mist1 is a novel human gastric chief cell marker whose expression is lost in metaplasia, dysplasia, and carcinoma. *Am J Pathol* 2010;177:1514–33.
- Goldenring JR, Nam KT, Mills JC. The origin of pre-neoplastic metaplasia in the stomach: chief cells emerge from the mist. *Exp Cell Res* 2011;317:2759–64.
- Giroux V, Rustgi AK. Metaplasia: tissue injury adaptation and a precursor to the dysplasia–cancer sequence. *Nat Rev Cancer* 2017;17:594–604.
- Bockerstett KA, Lewis SA, Wolf KJ, et al. Single-Cell transcriptional analyses of spasmodic polypeptide-expressing metaplasia arising from acute drug injury and chronic inflammation in the stomach. *Gut* 2020;69:1027–38.
- Trapnell C, Cacchiarelli D, Grimsby J, et al. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat Biotechnol* 2014;32:381–6.
- Machado JC, Nogueira AM, Carneiro F, et al. Gastric carcinoma exhibits distinct types of cell differentiation: an immunohistochemical study of trefoil peptides (TFF1 and TFF2) and mucins (MUC1, MUC2, MUC5AC, and MUC6). *J Pathol* 2000;190:437–43.
- Chan CWM, Wong NA, Liu Y, et al. Gastrointestinal differentiation marker cytokeratin 20 is regulated by homeobox gene Cdx1. *Proc Natl Acad Sci U S A* 2009;106:1936–41.
- Oltedal S, Skaland I, Maple-Grødem J, et al. Expression profiling and intracellular localization studies of the novel proline-, histidine-, and glycine-rich protein 1 suggest an essential role in gastro-intestinal epithelium and a potential clinical application in colorectal cancer diagnostics. *BMC Gastroenterol* 2018;18:26.
- Aridome K, Tsutsui J, Takao S, et al. Increased midline gene expression in human gastrointestinal cancers. *Jpn J Cancer Res* 1995;86:655–61.
- Shyu R-Y, Jiang S-Y, Chou J-M, et al. RARRES3 expression positively correlated to tumour differentiation in tissues of colorectal adenocarcinoma. *Br J Cancer* 2003;89:146–51.
- Rageul J, Mottier S, Jarry A, et al. Klf4-Dependent, PPARgamma-induced expression of GPA33 in colon cancer cell lines. *Int J Cancer* 2009;125:2802–9.
- Lee H-J, Nam KT, Park HS, et al. Gene expression profiling of metaplastic lineages identifies CDH17 as a prognostic marker in early stage gastric cancer. *Gastroenterology* 2010;139:213–25.
- Gao Y, Cai A, Xi H, et al. Ring finger protein 43 associates with gastric cancer progression and attenuates the stemness of gastric cancer stem-like cells via the Wnt- β /catenin signaling pathway. *Stem Cell Res Ther* 2017;8:98.

- 36 Miyazawa M, Matsuda M, Yano M, *et al.* Gastric adenocarcinoma of fundic gland type: five cases treated with endoscopic resection. *World J Gastroenterol* 2015;21:8208–14.
- 37 Jiang X, Hao H-X, Growney JD, *et al.* Inactivating mutations of RNF43 confer Wnt dependency in pancreatic ductal adenocarcinoma. *Proc Natl Acad Sci U S A* 2013;110:12649–54.
- 38 Ren X, Kang B, Zhang Z. Understanding tumor ecosystems by single-cell sequencing: promises and limitations. *Genome Biol* 2018;19:211.
- 39 Puram SV, Tirosh I, Parkh AS, *et al.* Single-Cell transcriptomic analysis of primary and metastatic tumor ecosystems in head and neck cancer. *Cell* 2017;171:1611–24.
- 40 Baniak N, Senger J-L, Ahmed S, *et al.* Gastric biomarkers: a global review. *World J Surg Oncol* 2016;14:212.
- 41 Fukatsu H, Miyoshi H, Ishiki K, *et al.* Gastric adenocarcinoma of fundic gland type (chief cell predominant type) treated with endoscopic aspiration mucosectomy. *Dig Endosc* 2011;23:244–6.
- 42 Miyazawa M, Matsuda M, Yano M, *et al.* Gastric adenocarcinoma of the fundic gland (chief cell-predominant type): a review of endoscopic and clinicopathological features. *World J Gastroenterol* 2016;22:10523–31.
- 43 Rodig SJ, Gusenleitner D, Jackson DG, *et al.* Mhc proteins confer differential sensitivity to CTLA-4 and PD-1 blockade in untreated metastatic melanoma. *Sci Transl Med* 2018;10:scitranslmed.aar3342.
- 44 Chowell D, Morris LGT, Grigg CM, *et al.* Patient HLA class I genotype influences cancer response to checkpoint blockade immunotherapy. *Science* 2018;359:582–7.
- 45 Kefalakes H, Rehermann B. Inflammation drives an altered phenotype of mucosal-associated invariant T cells in chronic hepatitis D virus infection. *J Hepatol* 2019;71:237–9.
- 46 Hufbauer M, Maltseva M, Meinrath J, *et al.* Hpv16 increases the number of migratory cancer stem cells and modulates their miRNA expression profile in oropharyngeal cancer. *Int J Cancer* 2018;143:1426–39.
- 47 Goldenring JR, Nam KT, Wang TC, *et al.* SPEM and Im: time for re-evaluation of metaplasias and the origins of gastric cancer 2010;138:2207.
- 48 Kinoshita H, Hayakawa Y, Koike K. Metaplasia in the Stomach-Precursor of gastric cancer? *Int J Mol Sci* 2017;18:ijms18102063.
- 49 Radyk MD, Burclaff J, Willet SG, *et al.* Metaplastic cells in the stomach arise, independently of stem cells, via dedifferentiation or transdifferentiation of chief cells. *Gastroenterology* 2018;154:839–43.
- 50 Weis VG, Goldenring JR. Current understanding of SPEM and its standing in the preneoplastic process. *Gastric Cancer* 2009;12:189–97.
- 51 Bredemeyer AJ, Geahlen JH, Weis VG, *et al.* The gastric epithelial progenitor cell niche and differentiation of the zymogenic (chief) cell lineage. *Dev Biol* 2009;325:211–24.
- 52 Nam KT, Lee H-J, Sousa JF, *et al.* Mature chief cells are cryptic progenitors for metaplasia in the stomach. *Gastroenterology* 2010;139:2028–37.