ARTICLE IN PRESS

npj Precision Oncology

https://doi.org/10.1038/s41698-026-01355-8

Article in Press

Dissecting the differentiation origins of intestinal metaplasia and early intestinal-type gastric cancer in gastric antrum by single-cell RNA profiling

Received: 23 April 2025

Accepted: 23 February 2026

Cite this article as: Yin, H., Zhang, H., Zheng, S. *et al.* Dissecting the differentiation origins of intestinal metaplasia and early intestinal-type gastric cancer in gastric antrum by single-cell RNA profiling. *npj Precis. Onc.* (2026). https://doi.org/10.1038/s41698-026-01355-8

Honghao Yin, Huanyu Zhang, Shuwen Zheng, Rui Guo, Jing Wen, Mengyuan Liu, Aicun Li, Mingfang Zhao, Yuan Yuan & Yuehua Gong

We are providing an unedited version of this manuscript to give early access to its findings. Before final publication, the manuscript will undergo further editing. Please note there may be errors present which affect the content, and all legal disclaimers apply.

If this paper is publishing under a Transparent Peer Review model then Peer Review reports will publish with the final article.

ARTICLE IN PRESS

© The Author(s) 2026. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

ARTICLE IN PRESS

Dissecting the differentiation origins of intestinal metaplasia and early intestinal-type gastric cancer in gastric antrum by single-cell RNA profiling

Honghao Yin[1,2,3,5#], Huanyu Zhang[1,2,3#], Shuwen Zheng[1,2,3#], Rui Guo[1,2,3], Jing Wen[1,2,3], Mengyuan Liu[4], Aicun Li[1,2,3], Mingfang Zhao*[5], Yuan Yuan*[1,2,3] and Yuehua Gong*[1,2,3]

[1]Tumor Etiology and Screening Department of Cancer Institute and General Surgery, the First Hospital of China Medical University, Shenyang 110001, China

[2]Key Laboratory of Cancer Etiology and Prevention in Liaoning Education Department, the First Hospital of China Medical University, Shenyang 110001, China

[3]Key Laboratory of GI Cancer Etiology and Prevention in Liaoning Province, the First Hospital of China Medical University, Shenyang 110001, China

[4]Department of Gastroenterology, the First Hospital of China Medical University, Shenyang 110001, China

[5]Department of Medical Oncology, the First Hospital of China Medical University, Shenyang 110001, China

*Corresponding authors:

Yuehua Gong, yhgong@cmu.edu.cn

Yuan Yuan, yuanyuan@cmu.edu.cn

Mingfang Zhao, mfzhao@cmu.edu.cn

[#] These authors contributed equally to this work.

ARTICLE IN PRESS

ARTICLE IN PRESS

Abstract

Intestinal-type gastric cancer (IGC) is associated with a multi-step carcinogenic process, comprising non-atrophic gastritis, chronic atrophic gastritis, intestinal metaplasia, and gastric dysplasia. The risk of developing IGC gradually increases as the disease progresses. However, the origin of cell differentiation and its carcinogenic potential in different stages of gastric disease remains poorly understood. To address this issue, we analyzed the differentiation trajectory of epithelial cells in different disease stages from gastric antrum biopsies in patients with precancerous lesions and early GC using single-cell sequencing data. Our findings revealed that progenitor cells (PCs) act as the ancestors of antral gland mucous cells (GMCs) and pit mucous cells (PMCs) in the NAG/CAG stage. In the IM stage, GMCs, as well as PCs, may acquire the ability to become intestinal-like stem cell phenotypes, eventually differentiating into mature enterocyte cells. Secretory progenitor cells may differentiate into pre-secretory cells and goblet cells. In the early IGC stage, *KIAA0101*$^{+}$*PRAP1*$^{+}$ PCs may be the potential origin of early IGC. These findings provide valuable insights for further research into the molecular mechanisms underlying the development of IGC and may contribute to the development of novel prevention and treatment strategies.

Keywords: Single-cell sequencing; Intestinal metaplasia; Intestinal-type gastric cancer; Cell differentiation trajectory; Progenitor cells.

ARTICLE IN PRESS

ARTICLE IN PRESS

Introduction

Gastric cancer (GC) represents a significant global health challenge, ranking fifth in incidence and fourth in mortality according to Globocan 2022 data[1]. Despite notable advancements in diagnostic and therapeutic strategies, the incidence and mortality rates of GC are increasing in developing and middle-income countries[2,3]. According to Lauren's classification, GC is categorized into intestinal-type gastric cancer (IGC) and diffuse-type gastric cancer (DGC)[4], with IGC being more prevalent and clinically significant due to its association with distinct risk factors and biological behavior[5]. The progression of IGC adheres to *Correa's* model, which delineates a series of stages, including non-atrophic gastritis (NAG), chronic atrophic gastritis (CAG), intestinal metaplasia (IM), atypical hyperplasia, and ultimately carcinoma[6]. A comprehensive understanding of the molecular dynamics underpinning the progression to IGC is critical for identifying potential therapeutic targets.

Chronic inflammation is a known precursor of IM, where normal gastric glands transform into intestinal-like cells, such as goblet and absorptive cells[7]. Consequently, patients with IM exhibit a markedly increased risk of developing IGC. However, the cellular origins of IM and GC within the gastric mucosa remain contentious[8]. Recent studies suggest that in the gastric corpus, chief cells can re-enter the cell cycle following injury, exhibiting regenerative characteristics akin to progenitor cells in a process termed paligenosis[9-11]. This reprogramming is pivotal for tissue repair and carcinogenesis; yet, the precise origins and differentiation pathways leading to IM and IGC, particularly in the gastric antrum, remain inadequately characterized.

ARTICLE IN PRESS

Next-generation transcriptome sequencing has substantially enhanced our understanding of GC and IM at the molecular level. Nevertheless, this methodology often fails to capture the complex heterogeneity of gastric cell types and does not effectively trace their evolution during disease progression. Emerging techniques, such as organoids and lineage tracing[12], have provided valuable insights into the oncogenic transformation of gastric glands, although they possess limitations in recapitulating native tissue architecture and identifying specific stem cell markers. The advent of single-cell RNA sequencing (scRNA-seq), complemented by advanced data analysis techniques, has facilitated in-depth exploration of cellular origins and differentiation trajectories across various diseases[13]. For example, recent work by *Shuai Gao et al*. utilized scRNA-seq to investigate the spatiotemporal distribution of transcriptome profiles in the human fetal digestive tract and its divergent lineages[14]. In a landmark 2019 study, *Zhang et al*. constructed a single-cell transcriptomic atlas of gastric precancerous and early malignant mucosa, revealing distinct expression patterns of various cell types across lesions in the gastric antrum[15]. While this study sheds light on the heterogeneity of gastric mucosal cells, a comprehensive analysis of the origin and differentiation pathways of epithelial cells in gastric diseases has yet to be fully elucidated, leaving an area that remains unexplored for future investigation.

In this study, we reutilized scRNA-seq data to disclose the origins and possible regulatory mechanisms involved in IGC tumorigenesis. Additionally, we validated the expression of some key molecules in the spatial transcriptome and human tissue samples. Our findings offer new insights into the evolution of epithelial cells during gastric disease progression and suggest potential strategies for the early diagnosis, prevention, and even reversal of early GC.

ARTICLE IN PRESS

Results

1. Annotation of epithelial cell clusters

To investigate the origins of epithelial cells during the progression of gastric diseases, we initially utilized the raw scRNA-seq data uploaded by *Zhang et al*. for epithelial cell annotation. Using the Seurat package, we identified 33 distinct epithelial cell clusters, including PMC, GMC, progenitor cell (PC, with high expression of *HIST1H4C* and KIAA0101), neuroendocrine cell, transitional epithelial cell (TEC, an epithelial subpopulation with low or absent expression of lineage-specific marker genes)[16,17], intestinal-like stem cell (ILSC) with high expression of *OLFM4* and *REG1A*), goblet cell, enterocyte, and adenocarcinoma cell (Supplemental Figure S1A-S1C). Across all cell clusters, we were unable to identify a distinct population of gastric antral stem cells. Annotation based on currently established gastric stem cell markers revealed that AQP5 was only sparsely expressed within the GMC cluster and did not define an independent or clearly distinguishable cell population (Supplemental Figure S1D). Although the QC standards differed slightly from those of the original study, the observed cell types and variations were generally consistent with the reported results. Additionally, we found that TEC gradually decreased throughout the disease progression, whereas ILSC emerging as newly appearing cells during the IM stage, progressively increased over the course of disease development (Supplemental Figure S1D and S1E). These results indicate that epithelial cell annotations obtained through Seurat analysis are reliable, providing a necessary foundation for investigating the origins of cell differentiation. Building on this, we further explored the differentiation lineages and transcriptional regulation at each stage of

ARTICLE IN PRESS

gastric disease progression.

2. Differentiation lineages of epithelial cells in NAG

To explore the expression profiles and functional features of epithelial cells at the NAG stage, we conducted *t*-SNE visualization on scRNA-seq data obtained from three cases of NAG (Figure 1A, Supplemental Figure S2A). The analysis revealed a diverse population of epithelial cells, primarily comprising PC, TEC, PMC, GMC, and neuroendocrine cell (Figure 1B). To further define the molecular signatures of epithelial subtypes, the top five differentially expressed genes were identified for each cluster, followed by GO and GSVA analyses revealing distinct functional and signaling pathway enrichments (Supplemental Figures S2B-S2D).

T-SNE visualization revealed a close spatial relationship among PC, GMC, PMC, and TEC at the NAG stage (Figure 1B). Pearson correlation analysis showed strong intercellular correlations, suggesting developmental interdependence (Supplemental Figure S2E). CytoTRACE scoring identified PC as the least differentiated population, followed by TEC (Figure 1C and 1D). Pseudotime analysis with Monocle 2 delineated a bifurcating trajectory, positioning PC as the progenitor population giving rise to GMC and PMC via an intermediate undifferentiated state (Figure 1E and 1F). As PMC represent the predominant cell population, suggesting that the primary differentiation direction of PC at this stage is toward PMC, whereas differentiation into GMC occurs at a relatively lower proportion. The corresponding heatmap depicted dynamic transcriptional transitions, with cell fate 1 (red) directed toward GMC and cell fate 2 (blue) toward PMC (Supplemental Figure S2F), multi-dimensional analyses consistently place PC at the root of

ARTICLE IN PRESS

the epithelial differentiation trajectory at the NAG stage, from which GMC and PMC emerge through an intermediate transitional state.

SCENIC analysis identified *E2F4*, *E2F6*, *FOXO1*, and *TAGLN2* as candidate TFs regulating PC-to-PMC differentiation (Figure 1G; Supplemental Figure S2G). Among them, *TAGLN2* exhibited a transcriptional pattern closely mirroring *MUC5AC* (Figure 1H), and PPI network analysis confirmed extensive *TAGLN2*-associated interactions within PMC (Figure 1I), implicating it as a key TF in PMC specification. Independent validation using gastritis-related GEO datasets demonstrated a strong correlation between *TAGLN2* and *MUC5AC* expression ($R=0.79$, $P<0.005$; Supplemental Figure S2H), reinforcing its regulatory role in PMC development.

By contrast, no prominent TFs were activated during PC-to-GMC differentiation. *TAF7* showed a mild upregulation resembling *MUC6*, yet the magnitude was markedly lower than that of *TAGLN2* (Figure 1H). Collectively, these findings indicate that PC function as progenitors driving epithelial lineage differentiation in NAG, with *TAGLN2* serving as a central transcriptional regulator of PMC fate commitment.

3. Differentiation lineages of epithelial cells in CAG

We next characterized the epithelial landscape at the CAG stage. *t*-SNE visualization of 17,375 single cells from three CAG samples revealed distinct clustering of epithelial cell populations (Figure 2A; Supplemental Figure S3A). The overall epithelial composition in CAG remained comparable to that in NAG (Figure 2B). However, gene expression profiling and pathway

ARTICLE IN PRESS

enrichment analyses highlighted altered transcriptional programs and functional pathways across epithelial subsets (Supplemental Figure S3B-S3D), suggesting context-specific adaptations of these cell types under chronic atrophic conditions.

To investigate lineage origins, trajectory and CytoTRACE analysis demonstrated that the differentiation hierarchy among PC, GMC, PMC, and TEC during CAG largely paralleled that observed in NAG (Figure 2C and 2D; Supplemental Figure S3E and S3F). Nonetheless, the proportion of PC differentiating toward GMC increased markedly in CAG, indicating a shift in lineage bias. To further validate the biological relevance of the inferred differentiation trajectory, we performed immunohistochemical (IHC) staining to elucidate the spatial hierarchy of epithelial differentiation within gastric antral glands, thereby providing an anatomical framework for the trajectory inference (Supplemental Figure S3G and S3H）.

SCENIC analysis identified *TAGLN2* and *JUND* as major transcriptional regulators driving PC-to-PMC differentiation at this stage (Figure 2E; Supplemental Figure S3I). Conversely, during PC-to-GMC differentiation, *SPDEF* and *KLF10* were selectively activated, exhibiting expression patterns consistent with *MUC6*, a canonical GMC marker (Figure 2F). Protein-protein interaction network analysis revealed extensive interaction networks involving TAGLN2 and JUND within PMC, and SPDEF and KLF10 within GMC (Supplemental Figure S3J).

Collectively, these findings indicate that the fundamental lineage architecture established in NAG is retained during CAG, yet with a reinforced differentiation bias from PC toward GMC. This shift

ARTICLE IN PRESS

appears to be transcriptionally governed by *SPDEF* and *KLF10*, reflecting an adaptive reprogramming of epithelial differentiation as the gastric mucosa progresses from the non-atrophic to the atrophic stage.

4. PCs and GMCs as lineage sources driving enterocyte differentiation in IM

To characterize the expression landscape and functional heterogeneity of epithelial cells in IM, *t*-SNE visualization of 13,256 epithelial cells from six IM samples identified 20 distinct clusters, encompassing previously defined cell types as well as newly recognized ILSC, enterocyte, and goblet cell populations (Figure 3A and 3B; Supplemental Figure S4A). With increasing IM severity, the proportions of absorptive, goblet, and ILSC progressively expanded, accompanied by a concomitant reduction in PMC, GMC, PC, and TEC (Supplemental Figure S4B). Notably, PC exhibited elevated *OLFM4* expression at the IM stage (Figure 3C), and a subset of *MUC6*$^{+}$ GMC also expressed *OLFM4* (Figure 3D), suggesting a shared or transitional role in early intestinal differentiation. Functional enrichment analyses (KEGG and GO) further highlighted subtype-specific pathway activation patterns, reflecting the distinct phenotypic and functional states of these epithelial populations (Supplemental Figures S4C, S4D).

We next focused on the differentiation hierarchy of enterocytes at the IM stage. Six enterocyte clusters were identified, among which clusters 13 and 15 showed markedly increased expression of *FABP1* and *VIL1*, canonical markers of intestinal absorptive cells, indicating a higher degree of maturation than other subclusters (Figure 3E). Trajectory analysis further revealed that both PC and GMC served as progenitor populations, with ILSC positioned in a transitional state between

ARTICLE IN PRESS

them (Figure 3F; Supplemental Figure S4E). Consistently, Monocle 2 defined three major differentiation states (Figure 3G): state 3 captured ILSC differentiation and enterocyte maturation, state 2 represented the transition from PC to ILSC, and state 1 reflected the progression from GMC to ILSC. The convergence of these trajectories at the ILSC population highlights its pivotal role in lineage plasticity during IM.

Cellular composition analysis further confirmed this continuum. Excluding the initial PC and GMC, seven major clusters contributed to states 1 and 2, comprising enterocytes (clusters 0, 9, 11, 25) and ILSC (clusters 5, 10, 12). Specifically, clusters 0, 25, and 5 were enriched in state 2, while clusters 9, 11, 10, and 12 predominated in state 1 (Figure 3H). PC (clusters 14 and 21) differentiated into mature enterocyte (clusters 13 and 15) through intermediate ILSC (clusters 0 and 5) and immature enterocyte (cluster 25) (Figure 3I-K). Similarly, GMC (clusters 2 and 9) followed an alternative trajectory, transitioning via ILSC (clusters 10 and 12) and immature enterocyte (clusters 9 and 11) before acquiring mature enterocyte identity (Figure 3L-N).

Transcriptional regulatory analysis identified distinct TF networks governing these differentiation routes. In state 2 (PC-to-enterocyte trajectory), *CDX2, PRDM1, MAF, HNF4G, NR1H4, MXD1, FOSL1, XRCC4*, *HNF4A,* and *PPARG* were activated (Figure 3O; Supplemental Figure S4F). Their expression profiles mirrored those of enterocyte markers *FABP1* and *VIL1*, with *FOSL1*, *HNF4A*, and *PPARG* showing progressive upregulation throughout differentiation, while other TFs predominated in later maturation stages (Supplemental Figure S4G). In state 1 (GMC-to-enterocyte trajectory), SCENIC analysis identified *ESRRA*, *HNF4A*, *CDX1*, *ETS2*, *FOSL1*, *MAF*,

ARTICLE IN PRESS

MAX, *NR1H4*, *PRDM1*, *RCOR1*, and *XRCC4* as key regulators (Figure 3P; Supplemental Figure S4H). Monocle 2 further revealed marked upregulation of *ESRRA* and *CDX1* during GMC-to-enterocyte conversion, whereas the remaining TFs were associated with enterocyte maturation (Supplemental Figure S4I). CytoTRACE analysis corroborated these findings, showing higher differentiation scores for PC than GMC, consistent with their superior developmental potential (Supplemental Figure S4J and S4K). Within the ILSC subclusters, state 2 also displayed a higher differentiation status than state 1(Supplemental Figure S4J and S4K). Notably, among the identified TFs, *ESRRA* and *HNF4A* expression exhibited a persistently increasing trend throughout the process (Supplemental Figure S4G and S4I). PPI network analysis revealed extensive interactions among HNF4A-and ESRRA-associated genes within enterocyte (Figure 3Q). Functional enrichment indicated that these genes are primarily involved in lipid and fatty acid catabolism, small-molecule metabolism, and energy production (Supplemental Figure S4L). Notably, *HNF4A* was highly expressed in PC at the IM stage (Supplemental Figure S4M), suggesting that *HNF4A*-mediated metabolic reprogramming may facilitate the transition of PC into ILSC.

To validate the single-cell–derived trajectories, we integrated spatial transcriptomic data (Supplemental Figure S5A, S5B, S5C) and IHC (Supplemental Figure S5D and S5E), focusing on the differentiation of PC and GMC toward enterocytes. Spatial expression patterns showed that PC (*KIAA0101*$^+$) followed a trajectory of *KIAA0101*$^+$*OLFM4*$^+$→*OLFM4*$^+$*FABP1*$^+$→*FABP1*$^+$, while GMC (*MUC6*$^+$) underwent a similar transition through *MUC6*$^+$*OLFM4*$^+$→*OLFM4*$^+$*FABP1*$^+$→

ARTICLE IN PRESS

FABP1^{+} states. These findings indicate that both progenitor populations differentiate via a common *OLFM4*$^{+}$ ILSC intermediate, forming a continuous spatial gradient toward *FABP1*$^{+}$ enterocyte (Figure 4A, Supplemental Figure S5F). To visualize this process, stepwise differentiation scores were assigned to *MUC6* (0.5), *OLFM4* (1.0), and *FABP1* (1.5), and a weighted-expression heatmap was generated. The resulting spatial continuum captured the progressive differentiation of GMC and PC through the ILSC compartment toward mature enterocyte (Figure 4A). The IHC images clearly depict the precise distribution of KIAA0101, OLFM4, MUC6 and FABP1 along the longitudinal axis of the glands (Supplemental Figure S5D and S5E). The use of serial sections allowed us to explicitly map the transition between progenitor and differentiated states within the exact same glandular structure.

Immunofluorescence (IF) in human IM tissues corroborated these results, showing marked co-localization of HNF4A-FABP1 and ESRRA-FABP1 signals (Figure 4B). Notably, expression levels of both TFs were significantly elevated in IM compared with NAG and cancer stages (Figure 4C). Spatial transcriptional profiling further revealed strong co-expression of *ESRRA* and *HNF4A* with *FABP1* in enterocytes, confirming their regulatory involvement in terminal differentiation (Figure 4D, Supplemental Figure S5J).

Collectively, these findings confirm that PC and GMC contribute to enterocyte formation through a shared ILSC intermediate, with *HNF4A* and *ESRRA* acting as key regulators of terminal enterocyte differentiation. The spatially resolved transcriptomic landscape thus reinforces the continuous and coordinated nature of gastric-to-intestinal epithelial transformation during IM.

ARTICLE IN PRESS

5.Secretory progenitor cells as the lineage origin of goblet cells in IM

The appearance of goblet cell alongside enterocyte is a hallmark of IM. Goblet and neuroendocrine cell arise from secretory progenitor. Analysis of 1,168 cells revealed seven subclusters, including secretory progenitor (*TTR*$^{+}$), goblet cell (*SPINK4*$^{+}$), pre-goblet cell (*HES6*$^{+}$), gastric neuroendocrine cell (*GAST*$^{+}$), and enteroendocrine cell (*TPH1*$^{+}$) (Figure 5A-C). To gain insights into their functional roles, we also conducted pathway and functional enrichment analyses (Supplemental Figure S6A and S6B).

Pseudotime analysis indicated that secretory PC serve as the initial differentiation point, giving rise to goblet cell, enteroendocrine cell, and gastric neuroendocrine cell (Figure 5D-F). Goblet cell markers, such as *SPINK4*, and intestinal neuroendocrine markers, such as *TPH1*, showed a gradual increase in expression along the differentiation trajectory (Figure 5G).

SCENIC analysis identified key transcription factors significantly upregulated in goblet cell, including *HES1, CREB3L1, SPDEF, KLF4, TFF3, FOXQ1, KLF2, BCL3*, and *CEBPB* (Figure 5H; Supplemental Figure S6C). Monocle 2 analysis revealed that *CREB3L1, FOXQ1, KLF4, KLF2, TFF3*, and *SPDEF* progressively increased during goblet cell differentiation (Supplemental Figure S6D).

These findings indicate that secretory progenitors differentiate into goblet cell via an intermediate pre-goblet state during the IM stage. This process is orchestrated by a coordinated network of TFs, underscoring the central role of secretory progenitors in driving goblet cell formation.

ARTICLE IN PRESS

6. Differentiation lineages of cancer cells in EGC

To investigate the differentiation origins of EGC cell, we performed *t*-SNE visualization on a dataset comprising 7,342 single cells, which were classified into 22 distinct clusters (Figure 6A). This clustering allowed us to define various cell types based on DEGs (Figure 6B), among which we identified 459 EGC cells. The analysis of DEGs revealed the top five DEGs across the various epithelial cell types (Supplemental Figure S7A). Furthermore, functional and pathway enrichment analyses were conducted to elucidate the biological processes and signaling pathways associated with cancer cells (Supplemental Figure S7B).

To explore the origins of EGC cell, we compared epithelial populations across disease stages. Pearson correlation analysis showed strong associations ($R > 0.85$) between cancer cell and enterocyte, goblet cell, PC, and ILSC, with PC exhibiting the highest correlation with EGC cells ($R=0.966$; Figure 6C).

ARTICLE IN PRESS

InferCNV analysis revealed elevated CNV levels in enterocyte, goblet cell, and ILSC. Notably, PC and EGC cell shared similar CNV profiles, suggesting a common genomic background (Figure 6D, E). Pseudotime analysis further indicated that PC gradually differentiate into cancer cell (Figure 6F-H), accompanied by upregulation of markers including *CEACAM6, KLK10, GPX2,* and *KRT7* (Supplemental Figure S7C).

We identified *KIAA0101* and *HMGB2* as novel PC markers, with higher expression in PC than other epithelial populations (Supplemental Figure S7D and S7E). Comparing inflammation-

ARTICLE IN PRESS

associated PC (IAPC) with cancer-associated PC (CAPC), *PRAP1* and *HSPA6* emerged as CAPC-specific markers, with *PRAP1* strongly upregulated (Supplemental Figure S7F and S7G). These results suggest that *KIAA0101*$^{+}$*PRAP1*$^{+}$PC represent key precursors of EGC. Immunofluorescence confirmed these double-positive PC were restricted to cancerous regions (Figure 6I), and spatial transcriptomics validated their co-localization within malignant areas (Supplemental Figure S7H).

Transcriptional regulatory analysis highlighted activation of *XRCC4, ELF1, ELF3, TAF7, SIRT6*, and *BHLHE40* in cancer cells (Figure 6J; Supplemental Figure S6I). *SIRT6* and *XRCC4* were upregulated during precancerous-to-cancer differentiation (Figure 6K), with immunofluorescence confirming high *SIRT6* expression in EGC tissues (Figure 6L). Notably, expression levels of *KIAA0101* and *SIRT6* were significantly elevated in cancer stages (Figure 6M). PPI analysis revealed interactions between *SIRT6* and *XRCC4*, suggesting a role in driving malignant transformation (Supplemental Figure S7J).

ARTICLE IN PRESS

Discussion

IGC progresses through multiple stages, characterized by the gradual replacement of gastric-type epithelial cells with intestinal-type cells that harbor carcinogenic potential[18]. Understanding the differentiation dynamics underlying this process is critical for uncovering the mechanisms driving IGC. By integrating single-cell and spatial transcriptomics, we mapped cellular lineage trajectories across IGC stages and traced the origins of EGC cells. We further identify *KIAA0101*$^{+}$*PRAP1*$^{+}$

ARTICLE IN PRESS

progenitors as potential precursors of EGC, with transcription factors such as *HNF4A, ESRRA*, and *SIRT6* implicated in orchestrating these lineage transitions (Figure 7). These findings provide a comprehensive view of epithelial remodeling during IGC progression and offer mechanistic insights that may inform novel strategies for IGC prevention and therapy.

Chronic inflammation of the gastric mucosa is accompanied by epithelial exfoliation, erosion, and interstitial inflammatory infiltration. In the NAG stage, we observed a notable increase in both the number and proportion of PMC compared to other stages. Trajectory analysis indicated that most PC differentiate into PMC, with a smaller fraction giving rise to GMC. Notably, *TAGLN2* was specifically activated during PC-to-PMC differentiation, but not in PC-to-GMC differentiation. PPI network analysis showed that *TAGLN2* interacts with multiple proteins in PMC, including *AGR2*, which is known to regulate gastric gland hyperplasia and lineage maturation. Previous studies, such as that by *Aparna Gupta et al*[19], demonstrated that *AGR2* knockout in mice impairs lineage maturation in the stomach. These findings suggest that *TAGLN2* may promote PMC differentiation and maturation by modulating *AGR2* expression or activity, highlighting a previously unexplored role for *TAGLN2* in gastric epithelial cell development.

In the CAG stage, a marked reduction of native glands is a defining feature. In our study, epithelial cell types in CAG without IM were similar to those in NAG but showed distinct compositional shifts[20]. Specifically, PMC decreased markedly, while GMC slightly increased, suggesting that gastric mucosal atrophy may bias PC differentiation toward GMC. Trajectory and TF analyses revealed upregulation of *SPDEF* and *KLF10* during this process. *SPDEF* is known to regulate

ARTICLE IN PRESS

goblet cell proliferation in other tissues[21] and is required for terminal maturation of mucus neck cell in the gastric antrum[22], while *KLF10* promotes *TGF-β*-related signaling critical for epithelial differentiation[23]. These results indicate that *SPDEF* and *KLF10* likely drive PC-to-GMC differentiation in CAG, representing an adaptive response to gastric mucosal atrophy.

IM plays a critical role in structural reprogramming during gastric disease progression[24], serving as a self-protective and compensatory mechanism in response to chronic inflammation[25]. Understanding the reprogramming of IM lineages may therefore provide a practical avenue for gastric disease prevention. Previous studies have mainly used cell and animal models to investigate the origins of SPEM in the gastric corpus and its role in diffuse-type GC[16,26,27], while comprehensive analyses of the antrum, the predominant site of IGC, remain limited. Using available datasets, we systematically characterized the lineage dynamics of antral IM, delineating the differentiation of PC and GMC into enterocyte and tracing the origin of goblet cell differentiation.

Trajectory analysis revealed that PC can differentiate into enterocytes via ILSC. PC were designated as the trajectory root, a selection substantiated by their peak transcriptomic potency (via CytoTRACE), primordial topological positioning in lineage inference, and functional alignment with the established ‘isthmus-centric’ model of mucosal renewal. Transcription factor analysis identified *FOSL1*, *PPARG*, and *HNF4A* as key regulators in this process. Notably, *HNF4A* is highly expressed in PCs, and its target genes are enriched in metabolic pathways. Previous studies have shown that *HNF4A* activation is critical for maintaining intestinal stem cells, including *OLFM4*$^{+}$ cells[28]. Our data suggest that *HNF4A*-mediated metabolic reprogramming may

ARTICLE IN PRESS

drive the conversion of PCs into *OLFM4*$^{+}$ ILSC. Both immunofluorescence and spatial transcriptomic analyses confirmed the co-expression of *HNF4A* with the enterocyte marker *FABP1*, indicating that *HNF4A* is predominantly active in enterocytes during the metaplastic stage. We further found that GMC can also dedifferentiate into ILSC and subsequently give rise to mature enterocyte. *ESRRA* mediates this GMC-to-enterocyte differentiation. As a nuclear receptor related to estrogen signaling[29], *ESRRA* interacts with PGC-1 family cofactors to regulate genes involved in cellular energy metabolism[30,31]. Both immunofluorescence and spatial transcriptomic analyses confirmed the co-expression of *ESRRA* with *FABP1* in intestinal metaplasia tissues, indicating that *ESRRA* is predominantly expressed in enterocyte during the metaplastic stage. These findings suggest that targeting the *PGC-1α-ESRRA* axis may effectively block both PC- and GMC-mediated pathways of absorptive cell differentiation. Collectively, our analyses indicate that isthmic PC function as the kinetic engine of gastric epithelial renewal and constitute the common trajectory root, whereas *OLFM4*$^{+}$ ILSC emerge as downstream reprogrammed intermediates during inflammation-associated remodeling.

Current evidence indicates that goblet and enteroendocrine cell belong to the secretory epithelial lineage and originate from a common progenitor[14]. Using single-cell data, we identified a population of *TTR*$^{+}$ cells with characteristics of secretory progenitors[14]. Previous studies have reported *HES6*$^{+}$ cell as primitive goblet cells[15], and we speculate that *TTR*$^{+}$ cells may serve as their progenitors. The origin of gastrointestinal enteroendocrine cell has been debated, with theories proposing derivation from either the neural crest or endodermal secretory progenitors[32]. Studies in the small intestine suggest that secretory PC can function as reserve stem cell and contribute to

ARTICLE IN PRESS

neuroendocrine tumor development[33]. In the stomach, however, the lineage of enteroendocrine cells remains unclear. Our data indicate that *TTR*$^{+}$ cell possess the potential to differentiate into gastric neuroendocrine cell, although the TFs governing this process could not be resolved due to limited cell numbers and will require further investigation.

In the EGC stage, our findings revealed that PC exhibit similar levels of CNVs as EGC cell. Further analysis of TFs *SIRT6* and *XRCC4* demonstrated significant upregulation during this process. *SIRT6*, a member of the sirtuin family of NAD-dependent enzymes, is implicated in various cellular processes, including stress resistance, genomic stability, aging, and energy homeostasis[13,34-38]. Current research suggests that *SIRT6* may enhance the expression of *TNF*[39] and *IL-8*, thereby promoting inflammation, and that *SIRT6* contributes to an increase in the intracellular levels of ADP-ribose to facilitate cancer cell migration[40]. Furthermore, we demonstrated the activation of *SIRT6* in cancer tissues using spatial transcriptomic data combined with dual-immunofluorescence, indicating that *SIRT6* exhibits high expression exclusively in the cancer cells of EGC patients compared to the precancerous conditions. Given the observed relevance of PC to cancer cell, we conducted additional analyses to compare CAPC and IAPC, revealing higher expression of *PRAP1* in the former. *PRAP1*, known as Proline-Rich Acidic Protein 1, exhibits triglyceride binding activity and has been implicated in DNA damage response, signal transduction by *P53* class mediator resulting in cell cycle arrest, deactivation of mitotic spindle assembly checkpoint, and negative regulation of apoptotic process[41]. Moreover, recent findings by *B.H. Huang et al.* suggest that *PRAP1* may also play a role in p53/TP53-dependent cancer cell survival

ARTICLE IN PRESS

and drug resistance following DNA damage[42]. Furthermore, we have utilized dual immunofluorescence techniques to validate the *PRAP1*$^{+}$PC identified in the IGC stage from EGC patients. These observations provide additional evidence supporting the involvement of *PRAP1* in GC progression and highlight its potential significance in cancer development. Clinically, the identification of these progenitor populations through endoscopic biopsy or minimally invasive molecular assays may enable early detection of high-risk individuals before overt tumor formation. Monitoring *SIRT6* and *PRAP1* expression could facilitate patient risk stratification, guide timely intervention, and potentially inform targeted therapies aimed at blocking progenitor-mediated cancer progression.

Compared with previous studies, this work offers several distinctive advances. First, by reconstructing cellular lineage trajectories across the *Correa*'s cascade (NAG→CAG→IM→EGC), we adopted a stage-oriented rather than a cell type-centric framework, thereby uncovering the dynamic lineage relationships driving epithelial remodeling during gastric carcinogenesis. Second, we reveal that PC serve as central hubs of epithelial reprogramming: during gastritis, PC differentiate into PMC and GMC, whereas during IM, both PC and GMC acquire ILSC properties, highlighting an unrecognized mechanism of epithelial lineage plasticity. Third, we identify *KIAA0101*$^{+}$*PRAP1*$^{+}$ PC as potential cells of origin for EGC, providing novel biological insight beyond previous analyses of the original dataset. Fourth, by integrating spatial transcriptomics with immunofluorescence, we offer direct spatial evidence that strengthens the biological interpretation of these lineage transitions. Finally, the identification of stage-specific transcription

ARTICLE IN PRESS

factors (*HNF4A, ESRRA*, and *SIRT6*) and *KIAA0101*$^{+}$*PRAP1*$^{+}$ progenitor cell provides new avenues for developing early diagnostic biomarkers and preventive therapeutic targets. Collectively, these advances establish a disease stage-based lineage framework, uncover previously unrecognized molecular drivers, and extend the translational relevance of epithelial remodeling in gastric carcinogenesis.

Although our study has yielded important insights, several limitations should be acknowledged. First, the dataset included only a limited number of IM and EGC samples, and immunostaining validation was performed in a small cohort without quantitative assessment, which may restrict the generalizability of our findings. Second, while it is widely accepted that IGC progresses through a sequence from IM to dysplasia and finally carcinoma, our analysis lacks single-cell data from the dysplastic stage. As such, the interpretation of differentiation trajectories should be viewed with caution, and future studies incorporating dysplastic samples will be essential to refine our conclusions. Third, a discrete and independent stem cell population was not identified as a standalone cluster in our current analysis. Although cells expressing the stem cell marker *AQP5* were detected, they were transcriptionally integrated within the GMC cluster rather than forming a clearly separable subpopulation. Future studies incorporating spatial proteomics or experimental models for lineage tracing will be essential to further resolve the precise spatial hierarchy and functional plasticity of these populations during disease progression. Fourth, SCENIC analysis has inherent limitations: it infers transcription factor activity based on co-expression patterns rather than direct regulatory interactions, which may not fully capture dynamic regulatory mechanisms.

ARTICLE IN PRESS

Finally, while we identified candidate cellular origins and key regulatory molecules associated with IGC progression, these findings remain largely correlative. Mechanistic validation through organoid models, lineage tracing, and gain- or loss-of-function experiments will be required to substantiate the causal roles of these factors in gastric epithelial remodeling and malignant transformation.

By single-cell analysis and spatial transcriptomics validation, we mapped the origins and differentiation trajectories of gastric epithelial cells during IGC progression. PC and GMC act as central hubs, driving enterocyte differentiation, while *KIAA0101*$^{+}$*PRAP1*$^{+}$ PC emerge as potential precursors of EGC. These insights pave the way for integrating molecular diagnostics into clinical surveillance programs and for developing targeted strategies to intercept gastric carcinogenesis at its earliest stages.

ARTICLE IN PRESS

Methods

Data acquisition

Thirteen scRNA-seq samples from nine patients were obtained from the Gene Expression Omnibus (GEO) database (accession: GSE134520), including 3 NAG, 3 CAG, 6 IM (2 mild, 4 severe), and 1 EGC samples, covering the disease continuum of IGC. Spatial transcriptomic data were retrieved from GEO (accession: GSE251950), comprising 4 IGC cases, 3 of which included adjacent IM tissues. These paired GC-IM samples were used to validate key findings from the

ARTICLE IN PRESS

scRNA-seq analyses. Additionally, gastritis-related transcriptomic datasets (GSE60662, GSE130823, GSE191275; n = 65) were incorporated to validate inter-molecular correlations.

Quality control (QC) and pre-processing

Quality control of scRNA-seq data was performed using Seurat (version 3.0.1)[18]. Cells with unique molecular identifier (UMI) counts < 400, mitochondrial gene expression > 20%, or > 7,000 detected genes (potential doublets) were removed. The remaining data were normalized using the *NormalizeData* function with the *LogNormalize* method. Raw spatial transcriptomic data were processed in R (version 4.5.1) using Seurat (version 5.0) and related packages. Count matrices were imported with the *Load10X_Spatial* function. Low-quality spots with < 200 detected genes, genes detected in < 3 spots, or mitochondrial gene content > 20% were excluded to remove low-quality or damaged cells.

ARTICLE IN PRESS

Spatial visualization and validation analysis

Spatial feature plots were generated to visualize the expression of key molecules across tissue sections. To validate spatial differentiation patterns, their weighted expression distributions were quantified and mapped. Differentiation trajectories were then constructed to assess the spatial organization of cells in distinct states. Co-expression between transcription factors and marker genes was analyzed using SpatialPlot and FeatureOverlay, enabling the identification of coordinated gene programs within histologically defined regions. All visualizations were produced in Seurat v5 and refined with ggplot2 and cowplot for publication-quality figures.

ARTICLE IN PRESS

Dimensionality reduction and batch effect elimination

Principal Component Analysis (PCA) was used to summarize the data, and the top 20 principal components were selected for dimensionality reduction using the RunT-SNE algorithm with a resolution of 2.0. Highly variable genes (HVGs) among cell subclusters were calculated using the "FindVariableFeatures" function. Batch effects within the same disease were corrected using the Harmony R package (1.0).

Cell type identification and *t*-Distributed Stochastic Neighbor Embedding (*t*-SNE) presentation

Marker genes for each cell type were identified using the "FindAllMarkers" function in Seurat. Cell types were determined based on previously identified marker genes from relevant studies. Single-cell clustering was visually analyzed using *t*-SNE to separate cells into distinct clusters in low-dimensional space.

ARTICLE IN PRESS

Function and pathway enrichment analysis of differentially expressed genes (DEGs)

DEGs in cell clusters were identified using the "FindMarkers" function in Seurat with a cutoff threshold of adjusted *P* value (adj. *P* val) < 0.05 and fold change > 0.25. DEG information was loaded into clusterProfiler for Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis. Significantly enriched pathways were determined based on adjusted *P* value < 0.05. Gene set variation analysis (GSVA) was conducted using 50 hallmark gene sets from the molecular signature database, utilizing default settings[43]. The

ARTICLE IN PRESS

limma package of R software[44] was employed to evaluate different pathways by calculating cell subcluster activity scores. Heatmaps were used to visualize the average pathway scores of the T-value data for the top 10 significantly different pathways ($P < 0.05$) in each cluster.

CytoTRACE prediction of cellular differentiation status

The CytoTRACE R package[45] was used to predict the differentiation status of individual cells. This involved calculating the top 200 genes or gene count signatures (GCS) from scRNA-seq data, determining the geometric means of these top 200 genes in our datasets, and visualizing them in a low-dimensional embedding for accurate insights into the differentiation states of individual cells.

Pseudotime trajectory analysis

Epithelial cells from various stages of gastric diseases were analyzed using Monocle 2[46] with the following parameters: num cells expressed ≥ 10, mean expression ≥ 0.125, and q-value<0.01. Trajectories were visualized using *t*-SNE plots, and dynamic expression heatmaps were generated using the "plot_pseudotime_heatmap" function.

Transcription Factor (TF) analysis

TF activity differences across cell clusters were calculated using SCENIC[47] (Version 1.1.0) with the RcisTarget motif database and GRNboost (corresponding to GENIE3 1.4.3, AUCell 1.4.1 and RcisTarget 1.2.1; with mm10__refseq-r80__10kb_up_and_down_tss.mc9nr). The limma package was used to analyze the area under the curve (AUC) of each module. Regulons with adj. $P < 0.01$ were further investigated, and results were visualized using heatmap functions.

ARTICLE IN PRESS

InferCNV analysis

Copy number variations (CNVs) of each cell were estimated using the "inferCNV" package[48] based on gene transcriptome profiles. Gland mucous cells (GMCs), pit mucous cells (PMCs), and neuroendocrine cells were used as reference cells, and a hidden Markov model was employed to predict CNV state.

Immunohistochemistry and immunofluorescent detection

Samples of cancerous tissue, adjacent IM, and distal normal gastric tissue from ten cases of patients with early GC (EGC) who underwent endoscopic submucosal dissection at the gastroenterology department of the First Hospital of China Medical University were collected and processed into paraffin slices. Paraffin-embedded tissue sections were dewaxed, rehydrated, and subjected to antigen retrieval using EDTA (pH 9.0). After blocking with 3% BSA, sections were incubated with primary antibodies overnight at 4°C, followed by fluorescent secondary antibodies for 1h at room temperature in the dark. Nuclei were counterstained with DAPI, and slides were mounted with anti-fade medium. Images were captured using a Nikon ECLIPSE CI-S fluorescence microscope (Nikon, Tokyo, Japan). IF and IHC staining was performed using the following antibodies: anti-HNF-4-alpha (ab92378, 1:300, Abcam, USA), anti-NR3B1 (NBP1-47254SS, 1:100, Novus Biologicals, USA), anti-FABP1 (RM5144, 1:1000; Biodragon, China), anti-SIRT6 (13572-1-AP, 1:200 dilution; Proteintech, China), anti-PRAP1 (11932-1-AP, 1:200, Proteintech, USA), anti-KIAA0101 (Q15004,1:100, Abcepta, China), anti-MUC6 (MAB-1011,1:100, Maxim, China), anti-MUC5AC (MAB-1079,1:100, Maxim, China), anti-OLFM4 (39141,1:100, CST, USA). This

ARTICLE IN PRESS

study was approved by the Ethics Committee of the First Hospital of China Medical University, and informed consent was obtained from all patients (2023-446-2).

Declarations statements

Ethics approval and consent to participate

This study was approved by the Ethics Committee of the First Hospital of China Medical University (protocol no. 2023-446-2) and was conducted in accordance with the ethical principles outlined in the Declaration of Helsinki. Written informed consent was obtained from all participants in this study.

Data Availability

All data included in this study are available upon request by contacting the corresponding author. Additionally, publicly available datasets were analyzed, which can be accessed at GEO: GSE134520，GSE251950，GSE60662, GSE130823, GSE191275.

Code availability

The code used for this study is not publicly available but may be made available to qualified researchers on reasonable request from the corresponding author.

Acknowledgements

This work was supported by Noncommunicable Chronic Diseases-National Science and Technology Major Project (No. 2023ZD0501400), the National Key R&D Program of China (Grant No. 2023YFB4705700), and the Education Department project of Liaoning Province (No.

ARTICLE IN PRESS

LJ212410159002 , LJKMZ20221142). Figure 7 was created with BioRender (https://biorender.com/; accessed on January 27, 2026).

Author contributions

Y-Y, YH-G, and HH-Y conceived the study. HH-Y, HY-Z, SW-Z, YH-G, and Y-Y drafted the manuscript and performed the analysis. HH-Y, HY-Z, MY-L, and RG conducted the literature search and data collection. HH-Y, SW-Z, and AC-L were contributed to the IHC staining and IF experiments. HH-Y, SW-Z, J-W, and MF-Z contributed to data interpretation. HH-Y, MY-L, and MF-Z were involved in the underlying computer architecture. All authors have reviewed and approved the final manuscript.

Competing interests

The authors declare no competing interests.

ARTICLE IN PRESS

References list

1 Bray, F. *et al.* Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a cancer journal for clinicians* **74**, 229-263, doi:10.3322/caac.21834 (2024).

2 Singh, D. *et al.* Global estimates of incidence and mortality of cervical cancer in 2020: a baseline analysis of the WHO Global Cervical Cancer Elimination Initiative. *The Lancet. Global health* **11**, e197-e206, doi:10.1016/s2214-109x(22)00501-0 (2023).

3 Cao, Y. Cancer-triggered systemic disease and therapeutic targets. *Holistic integrative oncology* **3**, 11, doi:10.1007/s44178-024-00077-w (2024).

4 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 pathologica et microbiologica Scandinavica* **64**, 31-49, doi:10.1111/apm.1965.64.1.31 (1965).

5 Cheung, K. S., Chan, A. O. O. & Yu Wong, B. C. in *Gastrointestinal Oncology - A Critical Multidisciplinary Team Approach 2e* 120-138 (2024).

6 Correa, P. A human model of gastric carcinogenesis. *Cancer research* **48**, 3554-3560 (1988).

7 Meyer, A. R. & Goldenring, J. R. Injury, repair, inflammation and metaplasia in the stomach. *The*

ARTICLE IN PRESS

Journal of physiology **596**, 3861-3867, doi:10.1113/jp275512 (2018).

8 Iwaya, M. *et al.* Artificial intelligence for evaluating the risk of gastric cancer: reliable detection and scoring of intestinal metaplasia with deep learning algorithms. *Gastrointestinal endoscopy* **98**, 925-933.e921, doi:10.1016/j.gie.2023.06.056 (2023).

9 Lee, J. H. *et al.* p57(Kip2) imposes the reserve stem cell state of gastric chief cells. *Cell stem cell* **29**, 826-839.e829, doi:10.1016/j.stem.2022.04.001 (2022).

10 Hata, M. *et al.* GPR30-Expressing Gastric Chief Cells Do Not Dedifferentiate But Are Eliminated via PDK-Dependent Cell Competition During Development of Metaplasia. *Gastroenterology* **158**, 1650-1666.e1615, doi:10.1053/j.gastro.2020.01.046 (2020).

11 Shiokawa, D. *et al.* Elevated stress response marks deeply quiescent reserve cells of gastric chief cells. *Communications biology* **6**, 1183, doi:10.1038/s42003-023-05550-2 (2023).

12 Lago, C. *et al.* Medulloblastoma and high-grade glioma organoids for drug screening, lineage tracing, co-culture and in vivo assay. *Nature protocols* **18**, 2143-2180, doi:10.1038/s41596-023-00839-2 (2023).

13 Jovic, D. *et al.* Single-cell RNA sequencing technologies and applications: A brief overview. *Clinical and translational medicine* **12**, e694, doi:10.1002/ctm2.694 (2022).

14 Gao, S. *et al.* Tracing the temporal-spatial transcriptome landscapes of the human fetal digestive tract using single-cell RNA-sequencing. *Nature cell biology* **20**, 721-734, doi:10.1038/s41556-018-0105-4 (2018).

15 Zhang, P. *et al.* Dissecting the Single-Cell Transcriptome Network Underlying Gastric Premalignant Lesions and Early Gastric Cancer. *Cell reports* **27**, 1934-1947.e1935, doi:10.1016/j.celrep.2019.04.052 (2019).

16 Goldenring, J. R. & Nomura, S. Differentiation of the gastric mucosa III. Animal models of oxyntic atrophy and metaplasia. *American journal of physiology. Gastrointestinal and liver physiology* **291**, G999-1004, doi:10.1152/ajpgi.00187.2006 (2006).

17 Johnson, F. R. & Young, B. A. Undifferentiated cells in gastric mucosa. *Journal of anatomy* **102**, 541-551 (1968).

18 Stuart, T. *et al.* Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888-1902.e1821, doi:10.1016/j.cell.2019.05.031 (2019).

19 Gupta, A., Wodziak, D., Tun, M., Bouley, D. M. & Lowe, A. W. Loss of anterior gradient 2 (Agr2) expression results in hyperplasia and defective lineage maturation in the murine stomach. *The Journal of biological chemistry* **288**, 4321-4333, doi:10.1074/jbc.M112.433086 (2013).

20 Shah, S. C., Piazuelo, M. B., Kuipers, E. J. & Li, D. AGA Clinical Practice Update on the Diagnosis and Management of Atrophic Gastritis: Expert Review. *Gastroenterology* **161**, 1325-1332.e1327, doi:10.1053/j.gastro.2021.06.078 (2021).

21 Liang, L. *et al.* Gut microbiota-derived butyrate regulates gut mucus barrier repair by activating the macrophage/WNT/ERK signaling pathway. *Clinical science (London, England : 1979)* **136**, 291-307, doi:10.1042/cs20210778 (2022).

22 Horst, D. *et al.* Requirement of the epithelium-specific Ets transcription factor Spdef for mucous gland cell function in the gastric antrum. *The Journal of biological chemistry* **285**, 35047-35055, doi:10.1074/jbc.M110.164541 (2010).

ARTICLE IN PRESS

ARTICLE IN PRESS

23 Subramaniam, M., Hawse, J. R., Rajamannan, N. M., Ingle, J. N. & Spelsberg, T. C. Functional role of KLF10 in multiple disease processes. *BioFactors (Oxford, England)* **36**, 8-18, doi:10.1002/biof.67 (2010).

24 Gupta, S. *et al.* AGA Clinical Practice Guidelines on Management of Gastric Intestinal Metaplasia. *Gastroenterology* **158**, 693-702, doi:10.1053/j.gastro.2019.12.003 (2020).

25 Zivny, J., Wang, T. C., Yantiss, R., Kim, K. H. & Houghton, J. Role of therapy or monitoring in preventing progression to gastric cancer. *Journal of clinical gastroenterology* **36**, S50-60; discussion S61-52, doi:10.1097/00004836-200305001-00009 (2003).

26 Hayakawa, Y. *et al.* Mist1 Expressing Gastric Stem Cells Maintain the Normal and Neoplastic Gastric Epithelium and Are Supported by a Perivascular Stem Cell Niche. *Cancer cell* **28**, 800-814, doi:10.1016/j.ccell.2015.10.003 (2015).

27 Hayakawa, Y., Fox, J. G. & Wang, T. C. The Origins of Gastric Cancer From Gastric Stem Cells: Lessons From Mouse Models. *Cellular and molecular gastroenterology and hepatology* **3**, 331-338, doi:10.1016/j.jcmgh.2017.01.013 (2017).

28 Chen, L. *et al.* HNF4 Regulates Fatty Acid Oxidation and Is Required for Renewal of Intestinal Stem Cells in Mice. *Gastroenterology* **158**, 985-999.e989, doi:10.1053/j.gastro.2019.11.031 (2020).

29 Li, F. N. *et al.* ESRRA promotes gastric cancer development by regulating the CDC25C/CDK1/CyclinB1 pathway via DSN1. *International journal of biological sciences* **17**, 1909-1924, doi:10.7150/ijbs.57623 (2021).

30 Greschik, H. *et al.* Communication between the ERRalpha homodimer interface and the PGC-1alpha binding surface via the helix 8-9 loop. *The Journal of biological chemistry* **283**, 20220-20230, doi:10.1074/jbc.M801920200 (2008).

31 Schreiber, S. N., Knutti, D., Brogli, K., Uhlmann, T. & Kralli, A. The transcriptional coactivator PGC-1 regulates the expression and activity of the orphan nuclear receptor estrogen-related receptor alpha (ERRalpha). *The Journal of biological chemistry* **278**, 9013-9018, doi:10.1074/jbc.M212923200 (2003).

32 Thompson, M. *et al.* Gastric endocrine cells share a clonal origin with other gut cell lineages. *Development (Cambridge, England)* **110**, 477-481, doi:10.1242/dev.110.2.477 (1990).

33 Sei, Y., Feng, J., Zhao, X. & Wank, S. A. Role of an active reserve stem cell subset of enteroendocrine cells in intestinal stem cell dynamics and the genesis of small intestinal neuroendocrine tumors. *American journal of physiology. Gastrointestinal and liver physiology* **319**, G494-g501, doi:10.1152/ajpgi.00278.2020 (2020).

34 Korotkov, A., Seluanov, A. & Gorbunova, V. Sirtuin 6: linking longevity with genome and epigenome stability. *Trends in cell biology* **31**, 994-1006, doi:10.1016/j.tcb.2021.06.009 (2021).

35 Chang, M. *et al.* Suppression of SIRT6 by miR-33a facilitates tumor growth of glioma through apoptosis and oxidative stress resistance. *Oncology reports* **38**, 1251-1258, doi:10.3892/or.2017.5780 (2017).

36 Tasselli, L., Zheng, W. & Chua, K. F. SIRT6: Novel Mechanisms and Links to Aging and Disease. *Trends in endocrinology and metabolism: TEM* **28**, 168-185, doi:10.1016/j.tem.2016.10.002 (2017).

37 Roichman, A. *et al.* Restoration of energy homeostasis by SIRT6 extends healthy lifespan. *Nature communications* **12**, 3208, doi:10.1038/s41467-021-23545-7 (2021).

ARTICLE IN PRESS

ARTICLE IN PRESS

38 Li, X. & Kazgan, N. Mammalian sirtuins and energy metabolism. *International journal of biological sciences* **7**, 575-587, doi:10.7150/ijbs.7.575 (2011).

39 Jiang, H. *et al.* SIRT6 regulates TNF-α secretion through hydrolysis of long-chain fatty acyl lysine. *Nature* **496**, 110-113, doi:10.1038/nature12038 (2013).

40 Bauer, I. *et al.* The NAD+-dependent histone deacetylase SIRT6 promotes cytokine production and migration in pancreatic cancer cells by regulating Ca2+ responses. *The Journal of biological chemistry* **287**, 40924-40937, doi:10.1074/jbc.M112.405837 (2012).

41 Huang, B. H. *et al.* PRAP1 is a novel executor of p53-dependent mechanisms in cell survival after DNA damage. *Cell death & disease* **3**, e442, doi:10.1038/cddis.2012.180 (2012).

42 Wolfarth, A. A. *et al.* Proline-Rich Acidic Protein 1 (PRAP1) Protects the Gastrointestinal Epithelium From Irradiation-Induced Apoptosis. *Cellular and molecular gastroenterology and hepatology* **10**, 713-727, doi:10.1016/j.jcmgh.2020.06.011 (2020).

43 Hänzelmann, S., Castelo, R. & Guinney, J. GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC bioinformatics* **14**, 7, doi:10.1186/1471-2105-14-7 (2013).

44 Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic acids research* **43**, e47, doi:10.1093/nar/gkv007 (2015).

45 Gulati, G. S. *et al.* Single-cell transcriptional diversity is a hallmark of developmental potential. *Science (New York, N.Y.)* **367**, 405-411, doi:10.1126/science.aax0249 (2020).

46 Qiu, X. *et al.* Reversed graph embedding resolves complex single-cell trajectories. *Nat Methods* **14**, 979-982, doi:10.1038/nmeth.4402 (2017).

47 Aibar, S. *et al.* SCENIC: single-cell regulatory network inference and clustering. *Nat Methods* **14**, 1083-1086, doi:10.1038/nmeth.4463 (2017).

48 Patel, A. P. *et al.* Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science (New York, N.Y.)* **344**, 1396-1401, doi:10.1126/science.1254257 (2014).

ARTICLE IN PRESS

Figures and associated legends

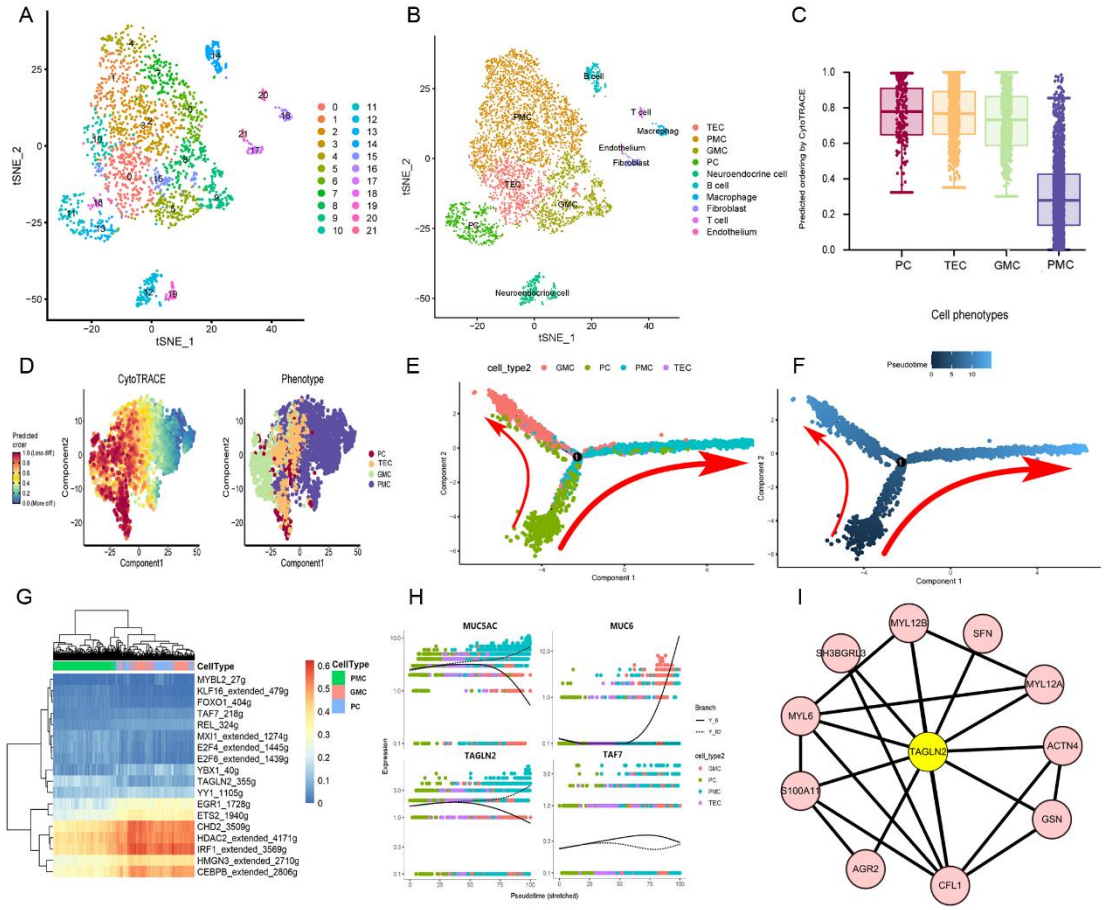

Figure 1. Expression characteristics and differentiation trajectories of epithelial cells at the NAG stage.

(A) *t*-SNE plot showing 5,272 high-quality epithelial cells grouped into 22 distinct clusters.

(B) Cell-type annotation of epithelial cells based on the expression of known marker genes. Each cluster represents a specific epithelial lineage.

(C) Boxplots showing the predicted differentiation potential of epithelial cells estimated by CytoTRACE.

(D) Prediction of epithelial cell differentiation states by CytoTRACE. The heatmap (left) displays the inferred differentiation status, while the right panel visualizes the spatial distribution of cells with distinct differentiation potentials.

(E, F) Monocle 2 trajectories depicting the differentiation dynamics of epithelial cells during the NAG stage, with pseudotime reflecting the progression of differentiation.

ARTICLE IN PRESS

(G) Heatmap illustrating the activity (AUC scores) of TFs regulating epithelial cell differentiation.

(H) Expression patterns and dynamic changes of key marker genes and TFs along distinct differentiation trajectories. Y_6 denotes the trajectory of PCs differentiating into GMC, while Y_82 denotes the trajectory of PC differentiating into PMC.

(I) STRING protein-protein interaction network showing the regulatory role of TAGLN2 in PMC cells, visualized using Cytoscape.

ARTICLE IN PRESS

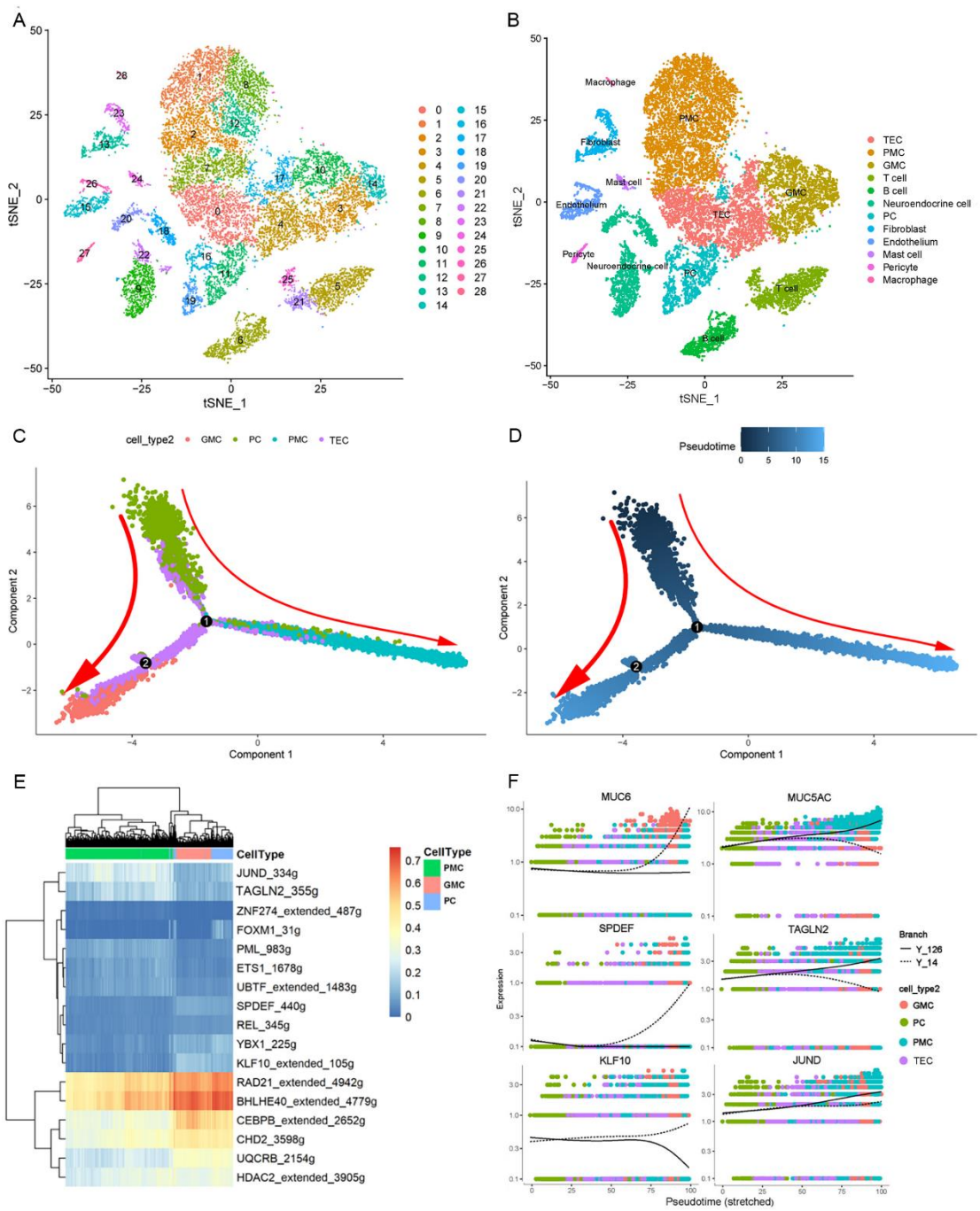

Figure 2. Expression characteristics and differentiation trajectories of epithelial cells at the CAG stage.

(A) *t*-SNE plot showing 17,375 high-quality epithelial cells grouped into 29 distinct clusters.

(B) Cell-type annotation based on the expression of known marker genes. Each cluster corresponds to a specific epithelial cell type.

ARTICLE IN PRESS

(C, D) Monocle 2 pseudotime trajectories illustrating the differentiation dynamics of epithelial cells at the CAG stage. Pseudotime reflects the progression of cellular differentiation.

(E) Heatmap showing the regulatory activity (AUC scores) of TFs inferred by the SCENIC analysis.

(F) Expression profiles and dynamic changes of key marker genes and TFs along distinct differentiation branches, highlighting potential lineage-specific regulatory patterns. Y_14 denotes the trajectory of PC differentiating into GMC, while Y_126 denotes the trajectory of PC differentiating into PMC.

ARTICLE IN PRESS

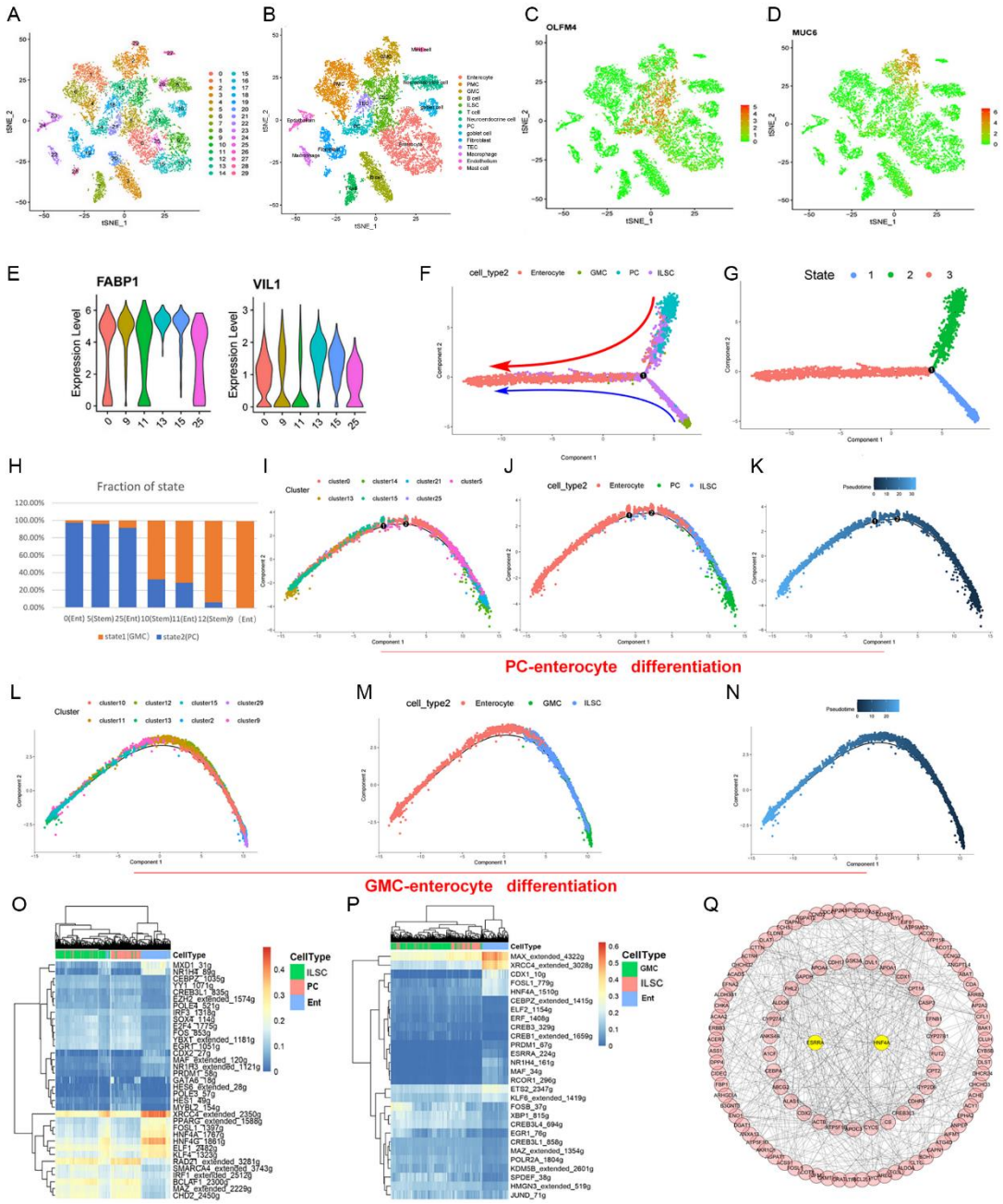

Figure 3. Expression characteristics and differentiation trajectories of epithelial cells at the IM stage.

(A) *t*-SNE plot showing 13,256 high-quality epithelial cells grouped into 30 distinct clusters.

(B) Cell-type annotation based on the expression of known marker genes. Each cluster corresponds to a specific epithelial cell lineage.

(C, D) *t*-SNE visualization of ILSC marker *OLFM4* (C) and gland mucous cell (GMC) marker

ARTICLE IN PRESS

MUC6 (D) across all epithelial cell types.

(E) Violin plots showing the expression levels of enterocyte marker genes (*FABP1*, *VIL1*) across six enterocyte subclusters.

(F, G) Monocle 2 pseudotime trajectories illustrating the overall differentiation dynamics of epithelial cells during the IM stage. Colors indicate cell types(F) and states(G).

(H) Stacked histograms showing the proportions of cells in two distinct differentiation states among different stem cell and enterocyte clusters.

(I-K) Monocle 2 trajectories depicting the differentiation from PC to ILSC and enterocyte during the IM stage.

(L-N) Monocle 2 trajectories depicting the differentiation from GMC to ILSC and enterocyte during the IM stage.

(O-P) Heatmaps showing transcription factor regulatory activity (AUC scores) inferred by SCENIC during the differentiation of PC (O) and GMC (P) into enterocyte.

(Q) STRING protein-protein interaction network illustrating the regulatory roles of *ESRRA* and *HNF4A* in enterocyte, visualized using Cytoscape.

ARTICLE IN PRESS

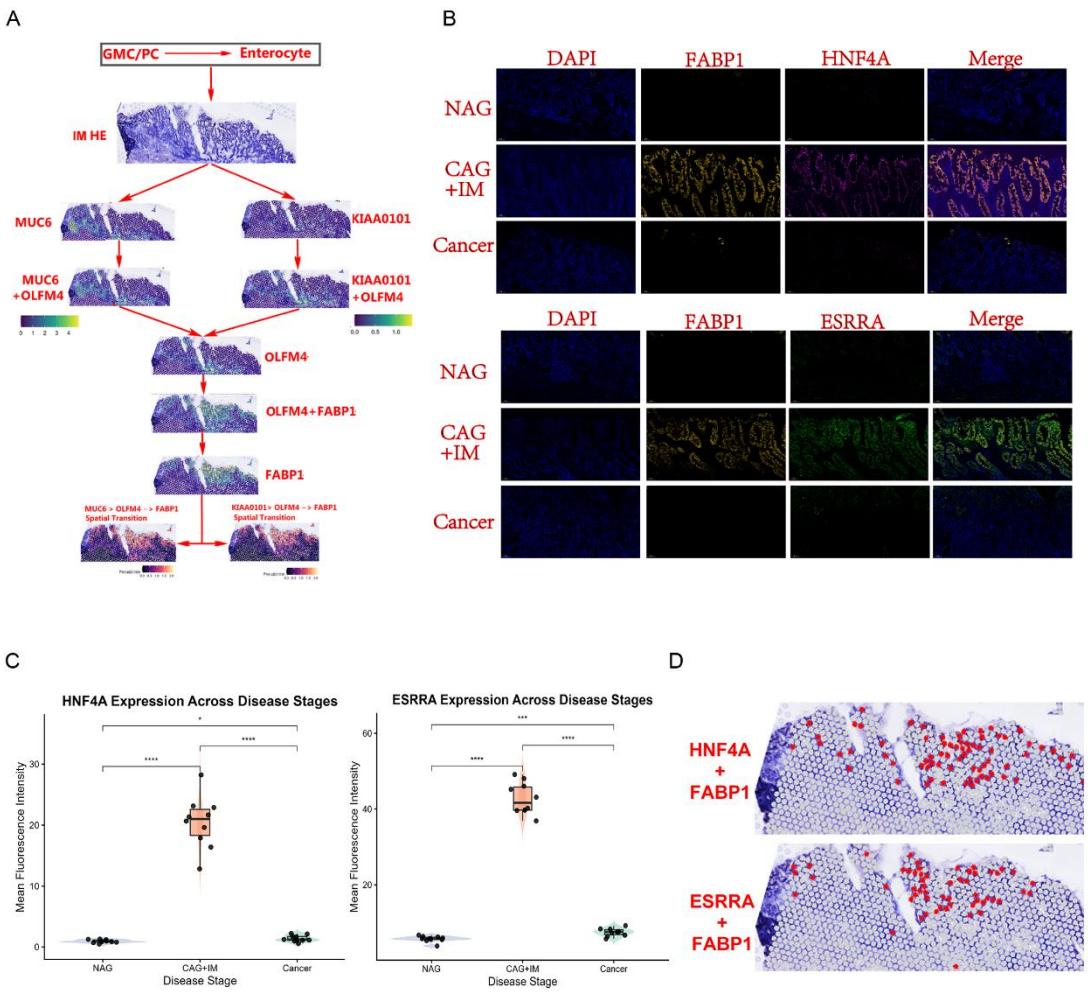

Figure 4. Spatial Transcriptomic Analysis of Intestinal Metaplasia Reveals Differentiation Trajectories and Transcriptional Regulation of Enterocyte.

(A) Spatial trajectory of gastric epithelial differentiation during IM. Schematic diagram illustrates the stepwise differentiation of *MUC6*$^{+}$ GMC and *KIAA0101*$^{+}$ PC into *FABP1*$^{+}$ enterocyte via *OLFM4*$^{+}$ ILSC. HE staining shows the histomorphology of gastric mucosa. Spatial transcriptomic heatmaps display the expression gradients of *MUC6*, *KIAA0101*, *OLFM4*, and *FABP1* (color scales represent normalized expression levels).

(B) Immunofluorescence validation of enterocyte markers and TFs in human gastric tissues. Representative images of DAPI (nuclear stain, blue), *FABP1* (enterocyte marker, yellow), *HNF4A* (red), and *ESRRA* (green) staining in NAG, IM, and GC tissues (scale bar: 50 μm).

(C) Quantification of *HNF4A* and *ESRRA* expression across disease stages. Violin plots show mean

fluorescence intensity of *HNF4A* and *ESRRA* (n=10 biological replicates per group). *HNF4A* and *ESRRA* expression is significantly upregulated in IM compared to NAG and cancer tissues (one-way ANOVA with Tukey's post-hoc test, $P < 0.001$).

(D) Spatial co-localization of TFs and *FABP1* in IM tissues. High-resolution spatial transcriptomic images confirm co-expression of *HNF4A* and *ESRRA* with *FABP1*. Red dots indicate spatial overlap of *HNF4A*/*ESRRA* and *FABP1* transcripts (scale bar: 50 μm).

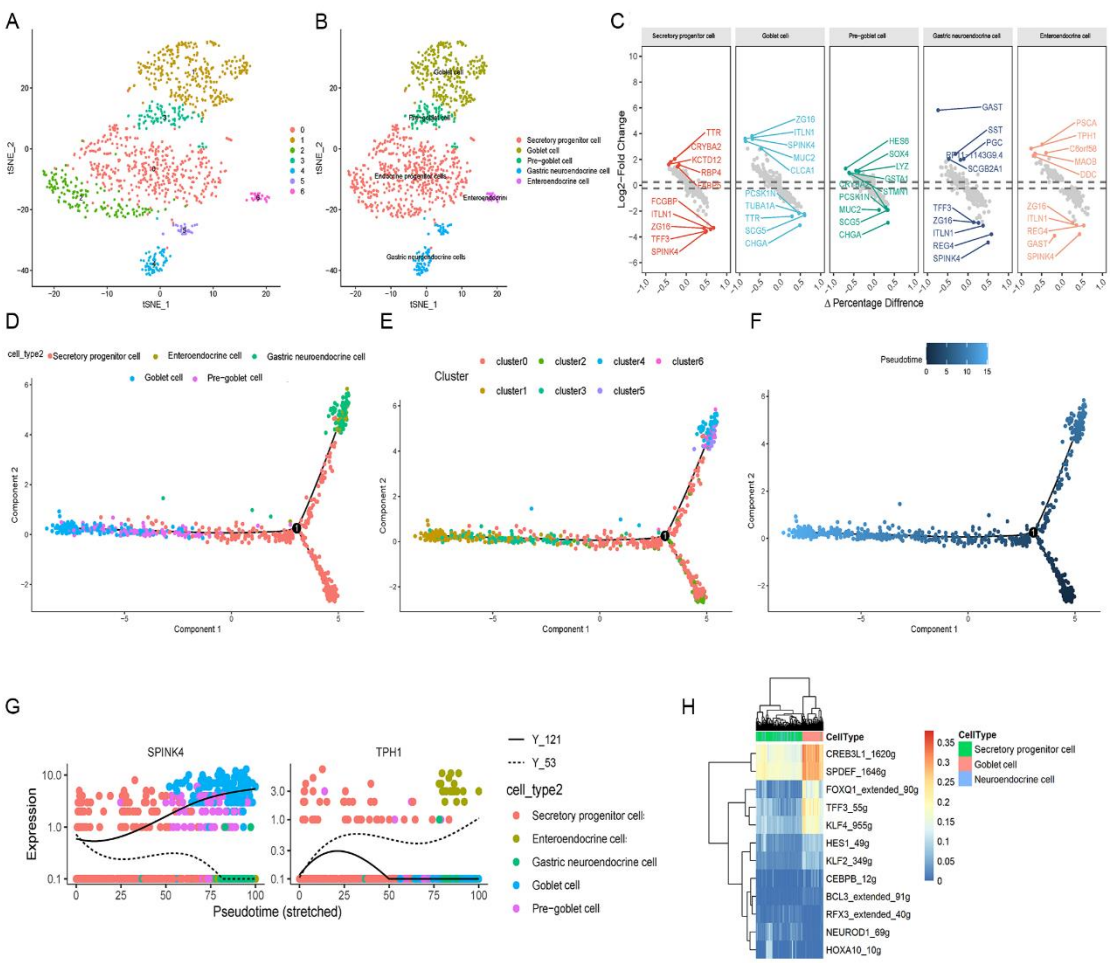

Figure 5. Origin and Differentiation Trajectory of Goblet Cells in IM stage.

(A) *t*-SNE plot showing 1,168 high-quality secretory lineage cells from IM tissues, unsupervised clustered into 7 distinct clusters

(B) *t*-SNE plot as in (A), with cells colored by identified cell types based on canonical marker

ARTICLE IN PRESS

genes.

(C) Volcano plots showing the top 5 upregulated (red labels) and downregulated (blue labels) genes in 5 secretory subgroups. Differential expression was determined by |log2(fold change) | > 1, ∆Percentage Difference (proportion of expressing cells between groups) > 20%, and adjusted p-value < 0.05 (Wilcoxon rank-sum test with Benjamini-Hochberg correction).

(D-F) Pseudotime trajectory reconstruction of secretory lineage cells, colored by (D) cell type, (E) cluster identity (from A), and (F) pseudotime (blue gradient, 0-15 arbitrary units). The trajectory reveals a branched differentiation path originating from secretory progenitors.

(G) Scatter plots showing expression levels of *SPINK4* (Goblet cell marker, left) and *TPH1* (enteroendocrine cell marker, right) along stretched pseudotime. Curves represent loess-smoothed expression trends for each cell type. Y_121 denotes the trajectory of Secretory progenitor cell differentiating into goblet cell, while Y_ 53 denotes the trajectory of Secretory progenitor cell differentiating into gastric neuroendocrine cell.

(H) Heatmap displaying AUC scores (calculated by SCENIC) for TFs regulons, reflecting TF-mediated gene expression regulation. Rows represent TFs with significant activity (AUC>0.1), columns represent cell types; color intensity indicates relative regulatory strength.

ARTICLE IN PRESS

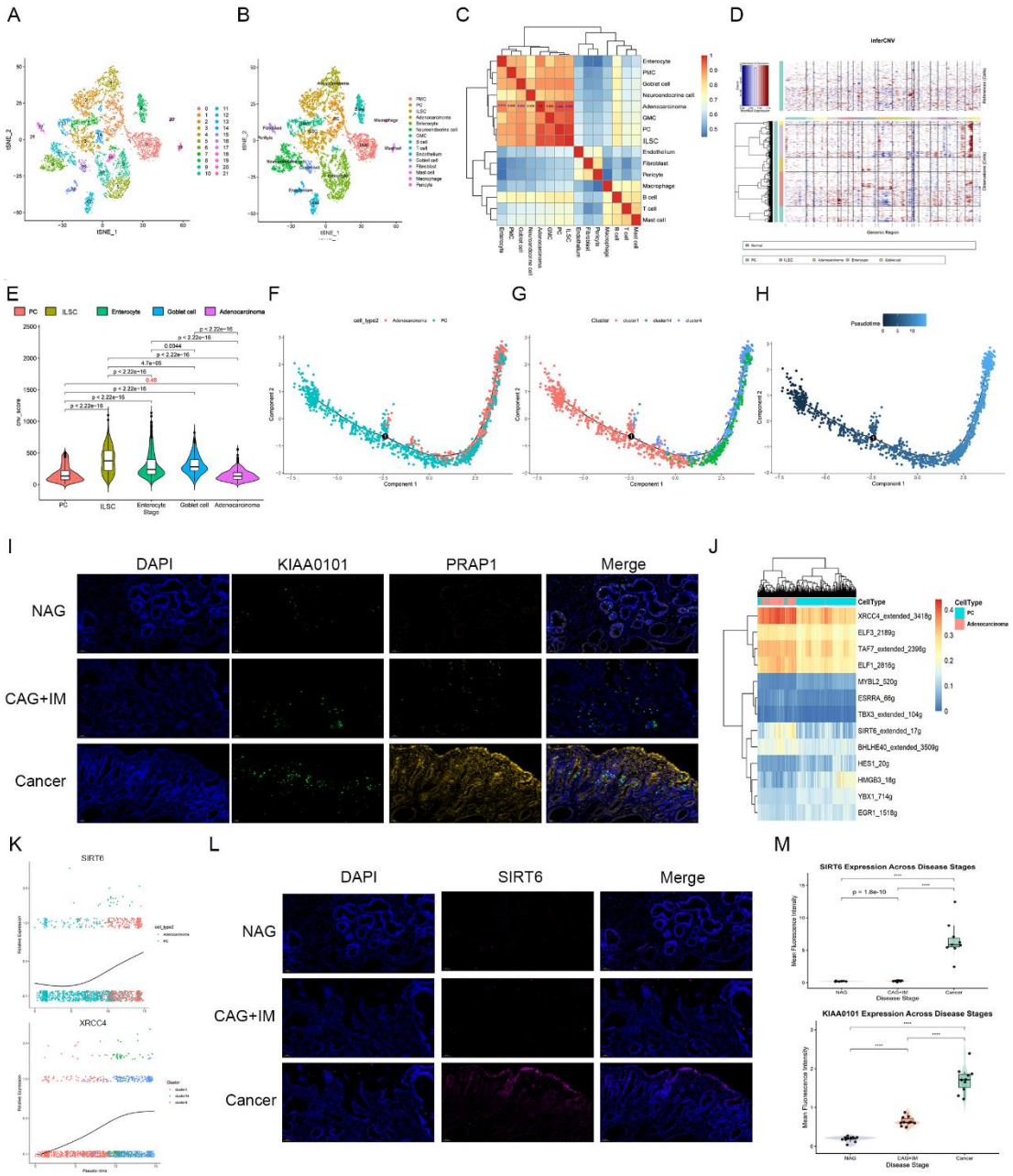

Figure 6. The origin and differentiation trajectory of cancer cells in EGC stage.

(A) *t*-SNE plot of 7,342 high-quality cells from EGC tissues, revealing 22 distinct clusters.

(B) *t*-SNE plot as in (A), with cells colored by identified cell types based on canonical marker genes (see color legend for full annotations).

(C) Heatmap showing Pearson correlation coefficients (R) between cell types. R>0.85 (red) indicates highest similarity. Blue numbers denote correlation coefficients between each cell type

and Cancer cell, highlighting closest association with PC.

(D, E) Copy number variation (CNV) profiling of EGC cells. (D) inferCNV heatmap showing CNV changes across chromosomes (x-axis) in each cell type, using PMC, GMC, and neuroendocrine cells as normal controls. (E) Violin plot quantifying CNV scores by cell type.

(F-H) Pseudotime trajectory reconstruction of PC-Cancer lineage, colored by (F) cell type, (G) subcluster, and (H) pseudotime (blue gradient, 0-50 arbitrary units). Trajectory reveals a branched differentiation path originating from PC.

(I) Co-staining of *KIAA0101* (green, enterocyte/proliferative marker) and *PRAP1* (yellow) in disease stages (20x magnification; DAPI: blue nuclear stain).

(J) Heatmap of transcription factor regulatory activity (AUC scores, calculated by SCENIC) showing cell type-specific transcription factor modules.

(K) Expression dynamics of key transcription factors (e.g., *SIRT6*, *XRCC4*) along PC-Cancer pseudotime, with color-coded cell types.

(L) Immunofluorescence validation of *SIRT6* in clinical tissues. *SIRT6* (red) staining in NAG, CAG+IM, and EGC tissues (20x magnification; DAPI: blue nuclear stain).

(M) Violin plots quantifying *SIRT6* and *KIAA0101* fluorescence intensity, with highest expression in EGC tissues (n=10 patients per group; one-way ANOVA, $P<0.05$, $P<0.01$).

ARTICLE IN PRESS

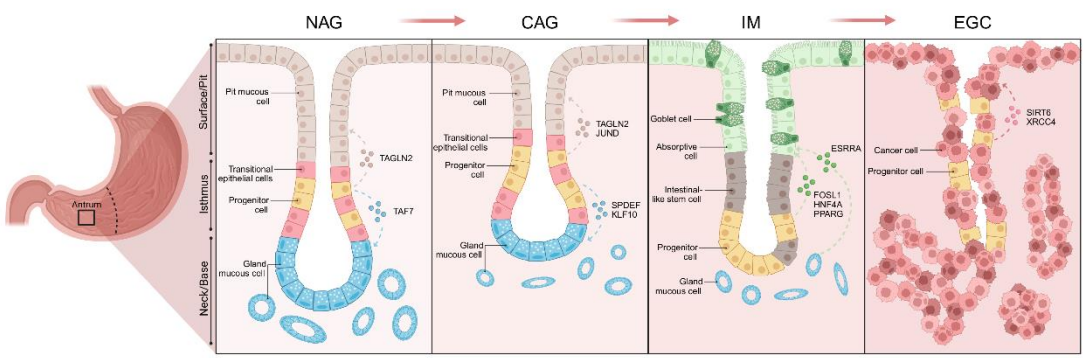

Figure 7. Schematic of the differentiation trajectory of epithelial cells at different diseases states and potential key molecules involved in the EGC evolution process.