Single-cell transcriptomics reveals a conserved metaplasia program in pancreatic injury

Zhibo Ma[2], Nikki K. Lytle[2], Bob Chen[3,9], Nidhi Jyotsana[1], Sammy Weiser Novak[4], Charles J. Cho[5], Leah Caplan[1], Olivia Ben-Levy[1], Abigail C. Neininger[1], Dylan T. Burnette[1,9,10], Vincent Q. Trinh[7], Marcus C.B. Tan[7,8], Emilee A. Patterson[6], Rafael Arrojo e Drigo[6,9], Rajshekhar R. Giraddi[2,@], Cynthia Ramos[2], Anna L. Means[7,8,10], Uri Manor[4], Jason C. Mills[5], James R. Goldenring[7,8,9,10,11], Ken S. Lau[1,8,9,10], Geoffrey M. Wahl[2], and Kathleen E. DelGiorno[1,8,9,10]

Affiliations:
[1]Department of Cell and Developmental Biology, Vanderbilt University, Nashville, TN, 37232
[2]Gene Expression Laboratory, Salk Institute for Biological Studies, La Jolla, CA, 92037
[3]Program in Chemical and Physical Biology, Vanderbilt University School of Medicine, Nashville, TN, 37232
[4]Waitt Advanced Biophotonics Center, Salk Insitute for Biological Studies, La Jolla, CA, 92037
[5]Division of Gastroenterology, Department of Medicine, Washington University School of Medicine, St. Louis, MO, 63110
[6]Department of Molecular Physiology and Biophysics, Vanderbilt University, Nashville, TN, 37232
[7]Department of Surgery, Vanderbilt University Medical Center, Nashville, TN, 37232
[8]Vanderbilt Digestive Disease Research Center, Vanderbilt University Medical Center, Nashville, TN, 37232
[9]Epithelial Biology Center, Vanderbilt University School of Medicine
[10]Vanderbilt Ingram Cancer Center, Nashville, TN, 37232
[11]Nashville VA Medical Center, Nashville, TN 37211

Correspondence to: Kathleen DelGiorno (kathleen.delgiorno@vanderbilt.edu)
[@]Current address: Pfizer Inc., San Diego, CA, 92121

The authors declare no potential conflicts of interest.

Author contributions: Conceptualization: K.E.D. and Z.M. Formal analysis: Z.M, N.K.L., B.C., N.J., S.W.N., C.J.C., D.T.B., V.Q.T., M.T., R.A., R.R.G., K.E.D. Funding acquisition: M.T., U.M., K.S.L, G.M.W, K.E.D. Investigation: Z.M., N.K.L., B.C., N.J., S.W.N., C.J.C., A.C.N., E.A.P, C.R., A.L.M., J.R.G. Project administration: K.E.D. Resources: J.C.M., K.S.L. Software: Z.M., B.C., S.W.N. Supervision: D.T.B., M.T., R.A., U.M., J.C.M., K.S.L, G.M.W., K.E.D Visualization: Z.M., N.K.L., B.C., S.W.N., C.J.C., A.C.N., D.T.B., R.A., K.E.D.

Data and materials availability: Sequencing data that support the findings of this study will be made available in the GEO archive.

Abstract

BACKGROUND & AIMS: Acinar to ductal metaplasia (ADM) occurs in the pancreas in response to tissue injury and is a potential precursor for adenocarcinoma. The goal of these studies was to define the populations arising from genetically wild type ADM, the associated transcriptional changes, and to identify markers of disease progression.

METHODS: Acinar cells were lineage-traced with enhanced yellow fluorescent protein (EYFP) to follow their fate upon injury. Transcripts of over 13,000 EYFP+ cells were determined using single-cell RNA sequencing (scRNA-seq). Single-cell trajectories were generated. Data were compared to gastric metaplasia and human pancreatitis. Results were confirmed by immunostaining and electron microscopy. Surgical specimens of chronic pancreatitis from 15 patients were evaluated by immunostaining.

RESULTS: scRNA-seq of ADM revealed emergence of a mucin/ductal population (*Muc6*+, *Tff2*+) resembling gastric pyloric, gland-base cells. Lineage trajectories suggest that this pyloric metaplasia is an intermediate cell identity between acinar cells and the generation of metaplastic tuft and enteroendocrine cells (EECs). 3-D electron microscopy demonstrates that all identified lineages populate ADM lesions. EECs exhibit substantial heterogeneity, including emergence of enterochromaffin (5-HT+) and delta (SST+) cells. Human pancreatitis shows a similar pyloric metaplasia phenotype and a conserved transcriptional program.

CONCLUSIONS: Under conditions of chronic injury, acinar cells undergo pyloric metaplasia to mucinous progenitor-like populations, some of which can then seed disparate tuft cell and EEC lineages. EEC subtypes are diverse with the potential to direct disease progression. This program is conserved in human pancreatitis, providing insight into early events in pancreas diseases.

Keywords: plasticity, pancreatitis, paligenosis, SPEM, Tuft cells, Enteroendocrine cells

Introduction

Damage to the exocrine compartment of the pancreas results in infiltration of immune cells, stromal deposition, loss of acinar cells, and the emergence of a metaplastic cell lineage pathologically defined by the aberrant appearance of ducts termed 'acinar to ductal metaplasia' (ADM)[1]. While many studies of ADM focus on its pre-neoplastic properties in chronic contexts, ADM is a reparative program in which digestive enzyme-producing acinar cells transdifferentiate to cells with ductal organization as part of a critical acute phase program (recently termed paligenosis) that enables tissue reconstruction following injury[2, 3]. Though the pancreas encounters damage-inducing conditions during adulthood, it lacks an adult stem cell compartment to respond with regeneration; thus, metaplastic reprogramming, which can invoke regenerative cells from mature, differentiated cells, like acinar cells, is critical to the long-term maintenance of the pancreas[4].

Under conditions of sustained injury, inherited mutation, environmental factors, or idiopathic stimuli, patients can develop chronic pancreatitis, a known risk factor for pancreatic ductal adenocarcinoma (PDAC). ADM is a pathognomonic epithelial change during chronic pancreatitis and occurs in response to oncogenic mutation in the exocrine tissue, identifying it as a pre-neoplastic event thought to be critical for eventual malignant transformation[1, 5]. The induction of injury and inflammation-induced ADM, therefore, may be one factor by which chronic pancreatitis increases the risk of PDAC.

We have recently shown that ADM results in the transdifferentiation of acinar cells to tuft cells, which are rare chemosensory cells typically absent from the normal mouse pancreas[6]. Using a combination of low-input RNA sequencing strategies, super resolution microscopy, and genetically engineered mouse models (GEMMs), we demonstrated that tuft cells inhibit pancreatic tumorigenesis by modulating tissue stroma[6-8]. These data demonstrate that rare ADM-derived cell types have the potential to significantly impact disease progression. Given the role of ADM in disease initiation and the potentially critical link between injury and tumorigenesis, we combined lineage tracing and single-cell RNA sequencing (scRNA-seq) to identify populations arising from ADM.

Here, we analyzed the transcriptional programs of lineage-traced cells resulting from injury-induced ADM. We identified residual acinar cells and emergence of tuft cells, but also enteroendocrine cells and mucin/ductal cell populations bearing gene signatures similar to spasmolytic polypeptide-expressing metaplasia (SPEM) in the stomach (*Tff2*+*Muc6*+*Gif*+). Because SPEM cells are a feature of the pyloric metaplasia that occurs after loss of parietal cells in the body of the stomach, ADM may be a manifestation of a more generalized pan-gastrointestinal injury response[9-11]. Gene expression analysis of these populations in combination with lineage trajectory and RNA velocity modeling suggests varying proclivity for seeding disparate differentiated tuft and enteroendocrine secretory cell lineages. Formation of distinct cell types was confirmed by marker immunostaining as well as by electron microscopy for cell type-specific structures. Finally, we compared our dataset to published single nucleus RNA sequencing (sNuc-seq) data from human pancreatitis and identified conserved processes[12]. These studies identify the

emergence of previously undescribed cell types in pancreas disease progression, metaplasia transcriptional programs shared between different organ systems, and a conserved response to exocrine injury in human disease.

Materials and Methods.

Mice. Mice were housed in accordance with NIH guidelines in AAALAC-accredited facilities at the Salk Institute and at Washington University. The IACUC committees at the Salk Institute and Washington University approved all mouse procedures. *Ptf1a*$^{Cre\text{-}ERTM}$, *ROSA*$^{YFP/+}$, and *ROSA*$^{Brainbow.2/+}$ mice were purchased from Jackson Laboratories (Bar Harbor, ME)[13]. All mice were backcrossed into the CD-1 strain. For stomach studies, wild type C57BL/6 mice were purchased from The Jackson Laboratory and maintained in a specified-pathogen-free barrier facility under a 12-hour light cycle.

Human samples. The Institutional Review Board of Vanderbilt University approved the distribution and use of all human samples.

Lineage tracing and pancreatitis induction. Lineage tracing was conducted using the mouse model *Ptf1a*$^{Cre\text{-}ERTM/+}$;*ROSA*$^{YFP/+}$, as previously described[7, 13]. In this model, tamoxifen treatment induces Cre activity, which then initiates expression of enhanced yellow fluorescent protein (EYFP) specifically in Ptf1a+ acinar cells. Acinar cells were labeled with daily tamoxifen treatment (Sigma, 5 mg/day, 5 days/week for two weeks) delivered in corn oil (Sigma) by oral gavage. Pancreatitis was induced with caerulein (Bachem) administered intraperitoneally (IP) at 250 μg/kg, twice a day, for five days a week for either two or four weeks and mice were then allowed to recover for two days[6]. Lineage tracing was conducted in *Ptf1a*$^{Cre\text{-}ERTM/+}$;*ROSA*$^{Brainbow.2/+}$mice with two daily treatments of 5 mg tamoxifen followed by the caerulein treatment protocol described above.

Results

Lineage tracing and scRNA-seq identifies epithelial heterogeneity in injury-induced ADM

The overall goal of these studies was to evaluate cellular heterogeneity in ADM resulting from chronic injury in genetically wild type pancreatitis. To this end, we conducted lineage tracing by labeling adult pancreatic acinar cells with EYFP in *Ptf1a*^{CreERTM};*Rosa*$^{LSL\text{-}YFP/+}$ (*CERTY*) mice with tamoxifen. We then induced pancreatic injury with repeated injection of the cholecystokinin ortholog caerulein, to induce endoplasmic reticulum (ER) stress in pancreatic acinar cells, resulting in metaplasia, cell death, and inflammation (Figure 1A)[6, 14]. Mice were sacrificed after either 2 or 4 weeks of caerulein treatment and pancreata were collected. We isolated EYFP+ cells by fluorescence activated cell sorting (FACS) and profiled the transcriptome with 10x Genomics 3' scRNA sequencing (Figure 1A, S1A). We used lineage tracing to identify metaplastic populations derived from acinar cells (i.e., ADM populations), but did not perform additional selections in order to agnostically profile the EYFP-labeled population. The combined libraries from the 2- and 4-week time points (2 mice per time point) contained a total of 21,140 cells with an average of 2,782 genes and 17,253 unique molecular identifiers (UMIs) detected per cell

(Supplemental File 1, Figure S1B). Clusters were annotated by examining both classic cell type markers and previously published single-cell gene signatures in pancreas tissue (Figure S2)[15]. Cells clustered mainly by cell type and data from the two time-points were found to overlap (Figure S1B). We identified both epithelial and stromal populations in our dataset (Figure 1B). To determine if these stromal populations truly express EYFP or if their inclusion is the product of auto-fluorescence or FACS gating, we assessed EYFP expression in these populations. At the transcriptional level, stromal populations were found to be largely negative for EYFP, whereas EYFP expression was concentrated in *Krt8*+*EpCAM*+ epithelial populations (Figure 1C-D, S3A).

We removed EYFP-neg stromal cells and further filtered with EYFP expression resulting in 13,362 EYFP+ epithelial cells. Cells from the 4 libraries (n = 2 for both time points) were integrated using the fastMNN algorithm to minimize batch effects. We identified 16 distinct clusters falling into four broad categories as well as intermediate populations (Figure 1E, S3B, Supplemental File 2). Acinar cells (*Prss1*+*Ptf1a*+*Try4*+*Cpa1*+; 4592 cells) and tuft cells (*Pou2f3*+*Siglecf*+*Trpm5*+*Gnat3*+; 1280 cells) were identified, consistent with previous studies, validating our methodology (Figure 1F, S3C)[6, 16]. In addition to these populations, we identified mucin/ductal cells (*Muc*6+*Pgc*+*Krt19*+*Car2*+; 3983 cells) and enteroendocrine cells (EECs, *Neurog3*+*Pax6*+*Chga*+*Neurod1*+; 1208 cells) (Figure 1E-F, S3C). Intermediate populations observed in the EYFP+ dataset include an acinar/mucin-ductal intermediate population (1955 cells) characterized by decreasing levels of acinar cell marker genes and increasing expression of mucin/ductal genes, as well as a *Sox4*+ EEC/tuft progenitor cell population (205 cells, Figure 1E-F, S3C).

We used co-immunofluorescence (Co-IF) for EYFP and cell type markers in normal and injured *CERTY* pancreata to confirm marker expression and to validate transcriptomically-identified populations. In uninjured, tamoxifen-treated *CERTY* pancreata, EYFP expression is restricted to acinar cells (amylase+) and is absent from ductal (pan-cytokeratin) and islet cells (CHGA+), consistent with prior reports (Figure 2A-B)[13, 16]. Under conditions of chronic injury, however, EYFP is expressed in both acinar and ductal cells (including tuft cells) as well as a large population of CHGA+ cells consistent with EEC formation (Figure 2A-B). As an orthogonal approach to confirm these cells types, we used scanning electron microscopy (SEM) on ultra-thin tissue sections from the pancreata of caerulein-treated mice. SEM readily identified acinar cells by the presence of large zymogen granules and abundant ER (Figure 2C). Tuft cells were identified by their distinct, short, blunt microvilli and deep actin rootlets (Figure 2D)[17]. We also identified mucinous cells by the presence of mucin granules, and several solitary EEC subtypes characterized by electron dense granules (Figure 3E-G). Collectively, these analyses reveal previously unrecognized epithelial heterogeneity in genetically wild type ADM.

Developmental trajectory analyses identify distinct secretory cell lineages

The cellular heterogeneity identified by scRNA-seq could arise by multiple mechanisms including, but not limited to: 1) direct transdifferentiation from acinar cells, 2) sequential generation from acinar cells to

intermediate progenitors, or 3) generation of a common intermediate for tuft and EEC populations, which can then interconvert. We used several computational biology approaches to determine the relatedness of these cell populations and to infer probable lineage trajectories (Figure 3A-D). Trajectory inference algorithms, like Monocle, learn the sequence of gene expression changes that each cell must pass through as part of a dynamic biological process, taking advantage of an individual cell's asynchronous progression[18]. Interestingly, the trajectory inferred by Monocle3 indicates that acinar cells first pass through a mucin/ductal state before forming tuft or EEC lineages (Figure 3A-B). A similar trajectory map is captured through the p-Creode algorithm, which describes transitional routes in transcriptional variation using an ensemble of graph traversals in principal component space (Figure 3C, S4A)[19].

To independently validate these inferred trajectories, we used RNA velocity, which annotates transitional cell states with inferred directional transcription dynamics by leveraging splicing kinetics[20]. We calculated RNA velocity individually in all 4 samples; all samples show a consistent flow from acinar to mucinous/ductal, then further branching into EEC and tuft cells. RNA velocity is lower magnitude in the mucin/ductal population, consistent with this population being in a relatively stable state as compared to the typical timescale of mRNA half-life (Figure 3D). This inferred trajectory is further supported by the activity of key transcription factors (TFs) associated with the cell lineages identified above (Figure S4B-C). We applied regulon analysis using the SCENIC package (Supplemental File 3)[21]. Regulon activity scores reflect the enrichment of a TF together with its correlated direct-target genes, improving the high drop-off issue for lowly expressed TFs[21]. We observed a switch from acinar-associated TFs (*Ptf1a*+*Rbpjl*+*Bhlha15/Mist1*+) to Mucin (*Spdef*+*Creb3l1*+*Creb3l4*+) and Ductal (*Sox9*+*Hnf1b*+*Onecut1*+) associated TF activation in the Acinar/MucinDuctal intermediate population (Figure S4B-C)[22-24]. The *Sox4* positive Tuft/EEC progenitor population retains expression of *Muc6* and *Krt19* (Figure S3C). Collectively, these data are consistent with a model in which injured acinar cells form a mucin/ductal progenitor population that seeds tuft cells and EECs as distinct lineages.

We tested this hypothesis by conducting lineage tracing with a Brainbow reporter ($Rosa^{Brainbow.2/+}$). Tamoxifen treatment of $Ptf1a^{CreERTM};Rosa^{Brainbow.2/+}$ mice results in the random expression of 1 of 4 reporters (YFP, GFP, RFP, CFP) in pancreatic acinar cells. Mice were first treated with a very short course of tamoxifen followed by caerulein (as shown in Figure 1A). IHC analysis of uninjured pancreata shows solitary cells expressing GFP or YFP. Injured pancreata, in contrast, are characterized by reporter expression in entire metaplastic ductal lesions, demonstrating clonal expansion from single acinar cells (Figure 3E). To examine cellular heterogeneity, we conducted co-IF and microscopy analysis on whole clones stained for markers of tuft cells (phalloidin, labels the microvilli and actin rootlets) and EECs (synaptophysin) and identified both cell types residing within the same clones (Figure 3F). To further evaluate cellular heterogeneity within individual ADM lesions, we performed SEM on 250 serial sections taken of a single metaplastic duct. Segmentation and 3-dimensional reconstruction of the processed image stack revealed an acinar cell adjacent to mucinous cells, a tuft cell, and an enteroendocrine cell with delta cell morphology (Figure 3G)[25]. Interestingly, the identified acinar cell is binuclear; how this

factors into acinar cell plasticity is currently unknown, but it has been shown that binucleate cells are resistant to cell cycle entry[26].

Injury-induced ADM as a pyloric-type metaplasia

Top differentially expressed genes (DEGs) in the progenitor-like mucin/ductal population include *Muc6*, *Tff2*, and *Gif* which are characteristic of SPEM in the stomach (Figure 1F, S3C, and 4A)[9]. SPEM is a process similar to ADM where atrophy of acid-secreting parietal cells promotes the digestive-enzyme-secreting cousins of acinar cells, chief cells, to undergo metaplasia in response to injury or oncogenic mutation[27, 28]. SPEM occurs within the overall reorganization -- known as pyloric metaplasia -- characterized by gastric units in the body (main portion) of the stomach reprogramming to resemble the distal (pyloric) region[11, 27, 29]. SPEM is believed to be a precursor for intestinal metaplasia and gastric adenocarcinoma, similar to the proposed role of ADM as a precursor for PDAC. We identified a number of canonical SPEM markers (*Cd44, Aqp5, Gkn3, Dmbt1, Wfdc2*) enriched in our mucin/ductal population as well as TF *Gata5,* a transcription factor that is widely expressed in gastric/intestinal epithelium but not in normal pancreas epithelium (Figure 4A, S5A). We also identified markers of digestive enzyme-producing gastric chief cells (*Gif* or *Cblif*, *Pga5*, *Pgc)*(Figure 4A, S5B). Co-expression of chief cell and SPEM markers is consistent with ADM adopting a pyloric-type metaplasia, which is emerging as a common, default regenerative healing program that can occur throughout the gastrointestinal tract in response to injury (e.g., Barrett's esophagus)[11]. To confirm these findings, we conducted immunohistochemistry (IHC) for CD44v9, GKN3, AQP5, WFDC2, and GIF. Consistent with our sequencing data, we identified the expression of SPEM and Chief cell markers in injured pancreata, whereas expression is absent from the exocrine tissue of the normal pancreas (Figure S5C). Consistent with the occurrence of pyloric metaplasia, we identified co-expression of all markers within single lesions (Figure 4B). Finally, we conducted SEM on ADM and identified cells with admixed mucin and digestive enzyme-containing granules, similar to the phenotype of SPEM associated with gastric injury (Figure 4C).

Based on these findings, we hypothesized that acinar cells undergo a metaplasia within the general rubric of pyloric metaplasia with specific parallels to SPEM, in response to injury. To test this hypothesis, we compared our ADM dataset to a recently published scRNA-seq study by Bockerstett et al., comparing normal gastric epithelium to that in several models of SPEM (Figure 4D and S6A-C)[30]. When we overlaid DEGs from normal gastric cell types onto our EYFP+ dataset, we found the chief cell transcriptome to be most closely related to pancreatic acinar cells (Figure 4E, S6A, Supplemental File 4). To generate SPEM signatures, we identified DEGs between SPEM populations and normal chief cells in either a mouse model of acute gastric injury (73 genes) or of chronic autoimmune injury (54 genes) and overlaid these signatures on our ADM dataset (Supplemental File 5). We found both SPEM signatures to be enriched in our mucin/ductal progenitor population (Figure 4E, S6C).

In addition to gene signatures, we identified TFs common to both disease states. *Gata5*, expressed in gastric development, and several mucin-production related TFs (*Spdef, Creb3l1* and *Creb3l4*) are enriched

in the ADM mucin/ductal cluster (Figure S4B, S6D)[22, 31, 32]. To determine if these transcriptional programs are also activated in SPEM, we examined regulon module expression in both the acute and chronic SPEM scRNA-seq datasets[21]. By this analysis, we found all mucin-associated regulons, as well as the Gata5 regulon, to be elevated in SPEM suggesting similar regulatory programs in both disease states (Figure S6E-F).

ADM results in substantial EEC heterogeneity

We next explored cellular heterogeneity within our lineage traced Tuft/EEC population. Re-clustering of the Tuft+EEC subset revealed 9 cell clusters separated into Tuft/EEC progenitors, EEC, or tuft lineages (Figure 5A, Supplemental File 6). Trajectory reconstruction using Monocle3 and RNA velocity suggests that EEC and tuft cell lineages originate from a *Sox4*+ common progenitor population (Figure 5B-C, 5F). Sox4 has previously been reported to promote tuft and EEC differentiation in the intestines[33]. Within the EEC population, we identified 5 distinct clusters including a progenitor-like cluster characterized by *Neurog3* expression, the master regulator for intestinal EEC and pancreatic endocrine lineage formation (Figure 5D)[34, 35]. Additional EEC clusters include putative enterochromaffin cells (*Ddc*+*Tac1*+*Chgb*+), delta cells (*Sst*+), epsilon cells (*Ghrl*+), and PP/gamma cells (*Ppy*+) (Figure 5E, S7A)[13]. *Gast* (gastrin) expression is detected in a subset of *Ddc*+*Tac1*+ cells; *Cck* (cholecystokinin) in a subset of *Ghrl*+ cells (Figure S7A). Insulin is not detected; minor glucagon expression is present within the *Ppy*+ cluster (Figure S7B)

To confirm expression of these hormones at the protein level, we performed IHC and could detect ghrelin, serotonin (*Ddc* and *Tph1*), somatostatin, pancreatic polypeptide, and gastrin in injury-induced ADM (Figure 5G, S7F). Ghrelin is not expressed in adult murine islets and serotonin expression is rare, however somatostatin and pancreatic polypeptide are expressed in the delta and PP/gamma cells, respectively, of normal pancreatic islets. To determine if these hormone-producing EECs are truly associated with ADM and not islets, we conducted lineage tracing with *CERTY* mice and could identify co-expression of EYFP in all 4 EEC subtypes (Figure S7G); demonstrating that all lineages inferred from the bioinformatic and histologic analyses derive from acinar cells. We then assessed EEC heterogeneity by EM. As shown in Figure 2C, we identified a number of EECs with varying granule morphology consistent with the formation of multiple EEC subtypes. Analysis of EM conducted on injured pancreata identified cells with enterochromaffin cell (serotonin-producing, pleomorphic granules) and delta cell (somatostatin-producing, round granules) morphology (Figure 5H)[36].

To identify putative regulatory processes involved in EEC subtype specification, we examined TF enrichment in the Tuft+EEC dataset. By this analysis, we found that formation of the *Neurog3*+ ECC progenitor state is associated with expression of Notch pathway inhibitors and subsequent branching into enterochromaffin and epsilon cell lineages is associated with changes in key TFs (Figure S7C-E)[37]. Enterochromaffin lineage formation correlates with *Nkx6-1* activation and a switch from *Pax4* to *Pax6* expression[38]. The epsilon lineage is associated with activation of *Arx and Isl1.* Subsequent branching into

delta and PP/gamma cells is accompanied by reactivation of *Pax6*. Delta cells are characterized by silencing of *Arx*, reactivation of *Sox9*, and expression of *Pou3f1* (Figure S7D-E)[39 40, 41]

Capture of transcriptomic changes associated with tuft cell development

A scRNA-seq study of intestinal epithelium previously demonstrated that tuft cells can be categorized as either neuronal or inflammatory-like based on marker expression[42]. We were unable to identify these sub-populations in our recently published scRNA-seq analysis of *Kras^{G12D}*-induced tuft cells, but our study contained relatively few cells[6, 8]. Further, tuft cells in our previous study were isolated by FACS for Siglecf+EpCAM+ expression and may have only represented a portion of the total tuft cell population. To identify tuft cell sub-populations in our injury-induced EYFP+ ADM dataset, we conducted DEG and GO Term analysis of tuft cell clusters (6, 5, 1, 0) including a total of 1280 cells (Figure 6A, S8A). Consistent with RNA velocity results (Figure 5F), we observed a transition from a *Sox4*+ progenitor population (cluster 6) to a *Pou2f3*+*Spib*+ population (cluster 5 and cluster1); POU2F3 is the master regulator for tuft cell formation (Figure 6A-C)[43]. Subsequent clusters 1 and 0 are enriched for known tuft cell markers and signaling molecules (i.e. *Trpm5*, *Il25*, *Chat*), suggesting that our dataset, as a whole, reflects the transcriptomic changes associated with tuft cell development, rather than separate, functional subtypes.

To test this hypothesis, we conducted Regulon analysis and identified candidate TFs and down-stream regulatory networks associated with tuft cell development[21]. We examined key Regulon activities as a function of Monocle3 Pseudotime (Figure 5B, 6B), moving from cluster 6 -> 5 -> 1 -> 0. Consistent with RNA velocity and DEG analysis, we identified a decrease in *Sox4* activity, an increase in *Pou2f3*, *Spib*, and *Ascl1* activity (a TF associated with taste cell function) along the Pseudotime progression (Figure 6B-D). Additional tuft cell-associated regulators include neuronal development-associated TFs (*Tead2*, *Hmgb3*, *Mef2b*), TFs associated with hematopoietic differentiation and inflammatory cell function (*Runx1*, *Myb*, *Gfi1b*, *Nfatc1*, and *Foxe1*) and TFs associated with epithelial differentiation (*Ehf*, *Klf4*) (Figure 6E)[44-46]. We next examined cell fate probability using a directed single cell-cell fate mapping approach developed in the CellRank package (see supplemental methods). Cells were decomposed into macrostates that match our seruat clusters (Figure S8B). Directed fate probabilities in the tuft cell lineage branch reveal a continuous process of tuft cell formation which is shown by a steady gain in tuft cell lineage fate probabilities moving from cluster 6 -> 0, consistent with tuft cell development and maturation (Figure 6F, S8C).

The fate probability-inferred continuum of tuft cell development is consistent with the transcriptional activity changes in many tuft cell-specific marker genes. Among many previously identified tuft cell marker genes, we observed three major groups. The first group includes genes that are enriched only in cluster 5, where expression fades throughout tuft cell development (*Gnat3*, *Mef2b*, *Nrep*)(Figure S8D, S6D, 6G, 6H-I). The second group represents a more or less steady level of gene expression throughout the pseudotemporal ordering of tuft cells from cluster 5 to cluster 0 (*Pou2f3*, *Spib*, *Avil*, *Ptgs1*, *Alox5, Gng13*, *Dclk1*, *Il13ra1 etc.*)(Figure 6H, S8F). The last group includes genes associated with the late stage of pseudotemporal ordering of tuft cell formation, many of these genes increase in expression along tuft

cell development (*Siglecf*, *Ptprc*, *Fcna*, *Folr1*, *Cdhr5*, *Bmp2*, *Wnt5a*, *Nrgn*, *Ascl1*)(Figure 6I, S8G). Many of these genes are associated with tuft cell function in immuo-regulation, cell-cell adhesion, and potential cell-cell communication, which is indicative of a mature cell state.

Finally, we explored the neuronal and inflammatory tuft cell classifications previously reported in small intestines in our EYFP+ dataset[42]. To do this, we overlaid neuronal and inflammatory tuft cell consensus gene signatures on our Tuft/EEC dataset. We found that the inflammatory signature is enriched late in tuft cell development, while the neuronal signature is enriched at a relatively early stage and remains expressed throughout tuft cell development (Figure S8D). Recently, Manco et al., used 'clump sequencing' to identify spatial expression programs of intestinal secretory cells (**doi:** https://doi.org/10.1101/2020.08.05.237917). The authors found that tuft cells associated with intestinal crypts, where cells are more stem- or progenitor-like, express *Sox4* and are enriched for the neuronal tuft gene signature. Tuft cells associated with the intestinal villus tip, where cells are most differentiated, are enriched for functional genes like *Fapb1* and *Il17rb*, and the inflammatory tuft signature. Consistent with this, we identified expression of *Fabp1*, *Il17rb*, and a number of other secretory signaling molecules and hormones (*Il25*, *Wnt5a*, *Sct*, *Chat*) enriched in differentiated tuft cell cluster 0 (Figure 6J-K, S8G). Collectively, these data identify regulatory programs and gene expression changes associated with tuft cell formation and maturation.

Human pancreatic metaplasia as a pyloric-type transition

To determine if the epithelial composition of human pancreatitis reflects the diversity identified in mouse models, we evaluated marker expression in single nucleus RNA sequencing (sNuc-seq) data collected from human pancreata and conducted immunofluorescence on tissue samples. Recently, Tosti et al., generated a comprehensive human pancreas cell atlas consisting of ~113,000 nuclei from healthy adult donors and 2,726 nuclei from two patients with chronic pancreatitis[12]. Consistent with mouse data, the pancreatitis dataset reflects an increase in a mucin/ductal population (*MUC5B*+/*KRT19*+; "MUC5B+Ductal") and the appearance of tuft cells, as compared to the normal pancreas (Figure 7A, S9A). To directly compare the human and murine mucin/ductal populations, we overlaid their respective gene signatures on their counterpart species dataset and found significant enrichment in the mucin/ductal clusters as compared to all other populations (Figure S9B).

To determine if human ADM represents a pyloric-type metaplasia transition, we humanized and overlaid the murine acute and chronic SPEM signatures on the human pancreatitis dataset. In both analyses, we found enrichment of the SPEM signature in the MUC5B+Ductal population (Figure 7B, S9C). Analysis of individual markers identifies *TFF2*, *WFDC2*, and *CD44* (Figure S9A). To validate these findings, we conducted immunostaining on pancreata from 15 patients with chronic pancreatitis, pre-screened by a pathologist for the presence of metaplasia, but varying in terms of etiology and severity. First, we evaluated the expression of SPEM markers TFF2, AQP5, and CD44v9 (generated by *Cd44*). As compared to normal pancreas, we identified up-regulation of Cd44v9 and AQP5 in acinar cells in areas of tissue

damage, in all 15 samples (Figure S9D). Occasionally, TFF2 was co-expressed in these regions (triple positive clusters present in 13 of 15 samples), but was more often found in metaplastic structures lacking acinar morphology, in PanIN-like lesions, and in duct glands associated with large ducts (Figure 7C-D).

We next evaluated the presence of tuft cells and EECs in human pancreatitis. Consistent with prior reports, we identified tuft cells by expression of marker phospho-Girdin in areas of metaplasia associated with injured acini, in normal ducts, and in incidental PanIN (10/15 patient samples, Figure 7D)[7, 47]. When we interrogated the sNuc-seq dataset for tuft cell TFs identified in our EYFP+ dataset, we found enrichment of *SOX4*, *POU2F3*, *SPIB*, and *RUNX1* in human tuft cells (Figure S9E). To assess EEC heterogeneity, we conducted IF for ghrelin (Ghrl), somatostatin (SST), pancreatic polypeptide (PPY), and serotonin (5-HT). All four hormones could be detected in islets, though the proportions of each varied by patient. All hormones were associated with aberrant ductal structures, rather than injured acini, consistent with formation later in ADM. Ghrelin and PPY expression is rare in metaplasia, but both can be found in large ducts. 5-HT and SST are expressed in metaplasia, PanIN-like lesions, and in intralobular ducts (Figure 7E-F).

Overall, the staining patterns in humans are largely reminiscent of the murine model, including the formation of distinct cell types (mucin/ductal, tuft, EECs) and expression of SPEM signatures. Differences include the lack of TFF2 in injured acini and expression of individual genes (i.e. *Muc6* vs. *MUC5B*). Further studies are required to determine if these differences reflect the heterogeneity of etiology, chronicity and severity of chronic pancreatitis in patients selected to undergo surgical resection. Altogether, these data are consistent with a model in which injured acinar cells in human or mouse undergo a pyloric-type metaplasia resulting in tuft and enteroendocrine cells.

Discussion

Using lineage tracing, scRNA-seq, multiple computational biology approaches (that rely on non-overlapping assumptions), immunostaining, and high-resolution ultramicroscopy, we have identified significantly more epithelial heterogeneity in injury-induced pancreatic ADM than previously appreciated. Brainbow lineage tracing demonstrates that acinar cells can clonally expand, forming secretory types identified in these analyses. While scRNA-seq provides information regarding cell type markers, high-resolution ultramicroscopy is critical to the identification of these populations as it provides structural information, corroborating the formation of defined cell types. Collectively, these data are consistent with a model by which acinar cells undergo a pyloric type metaplasia resulting in mucinous ductal cells; it is from this state that differentiated tuft cells or EECs emerge. Previously, we have shown that tuft cells form in response to oncogenic $Kras^{G12D}$-induced ADM in the pancreas[7]. Despite the paucity of tuft cells in this disease state, abrogation by genetic ablation of master regulator transcription factor *Pou2f3* results in acceleration of tumorigenesis, demonstrating that rare cell types can significantly impact disease progression[8]. Identifying the mechanisms of tuft cell injury mitigation and resolution in pancreatitis will provide the opportunity to co-opt or target these processes for patient benefit.

In addition to tuft cells, we have discovered that chronic injury-induced ADM results in substantial EEC heterogeneity. ADM-derived EEC formation was previously described by Farrell et al., in PanIN resulting from expression of oncogenic *Kras*^{G12D} by lineage tracing and expression of marker synaptophysin[48]. The term 'enteroendocrine cell', however represents a lineage populated by multiple subtypes characterized by the expression, or co-expression, of various hormones. While best known for regulating intestinal function, insulin secretion, nutrient assimilation and food intake in normal tissues, evidence points to poorly understood roles for several of these hormones in inflammation and tumorigenesis[36, 49]. In particular, somatostatin is known to have diabetogenic properties, which, when combined with the observation that adult-onset diabetes is a known association of pancreatic adenocarcinoma, suggests the intriguing idea that tumor-derived EEC have an important role in the endocrine/metabolic symptoms associated with this malignancy[50]. These data underscore the importance of understanding the contribution of EECs, including their regulatory processes and secretory products, in the contexts of pancreatic injury and tumorigenesis.

Altogether, we propose that this carefully orchestrated process of acinar cell de- and trans-differentiation exists to mitigate injury and may invoke a subset of a generalized pyloric metaplasia response that can occur in many organs in the GI tract. Ever more evidence is accumulating that there are shared patterns of regenerative changes that depend on cellular plasticity. Here, we observe striking transcriptional overlap between metaplastic cells in the pancreas (ADM) and those of the stomach (SPEM). Our results are in line with recent studies indicating that mature cells like gastric chief cells and pancreatic acinar cells have a shared injury-induced plasticity program, paligenosis, that governs their conversion from the differentiated to regenerative states[3, 27, 51]. Cycles of paligenosis that fuel pyloric metaplasia may allow cells to accumulate mutations and increase the risk for progression to cancer.

Our analysis of the human chronic pancreatitis sNuc-seq atlas generated by Tosti et al., and of the epithelial composition of injured pancreata by immunostaining, reflects the processes identified in mouse models[12]. However, ADM and markers of pyloric metaplasia are largely restricted to areas of low to intermediate damage and can be identified in areas pathologically defined as "normal". To us, these data suggest that ADM is an ongoing reparative process within homeostatic states, but is lost under conditions of severe injury, leading to substantial loss of the exocrine compartment. Conversely, an inability to undergo a carefully orchestrated ADM reparative process could result in chronic pancreatitis and would, therefore, increase the risk for development of PDAC. Further studies are justified to determine the pathways and cell states required to properly execute and resolve ADM.

References

1. Giroux V, Rustgi AK. Metaplasia: tissue injury adaptation and a precursor to the dysplasia-cancer sequence. Nat Rev Cancer 2017;17:594-604.
2. Storz P. Acinar cell plasticity and development of pancreatic ductal adenocarcinoma. Nat Rev Gastroenterol Hepatol 2017;14:296-304.

3. Willet SG, Lewis MA, Miao ZF, et al. Regenerative proliferation of differentiated cells by mTORC1-dependent paligenosis. EMBO J 2018;37.
4. Mills JC, Sansom OJ. Reserve stem cells: Differentiated cells reprogram to fuel repair, metaplasia, and neoplasia in the adult gastrointestinal tract. Sci Signal 2015;8:re8.
5. Murtaugh LC, Keefe MD. Regeneration and repair of the exocrine pancreas. Annu Rev Physiol 2015;77:229-49.
6. DelGiorno KE, Naeem RF, Fang L, et al. Tuft Cell Formation Reflects Epithelial Plasticity in Pancreatic Injury: Implications for Modeling Human Pancreatitis. Front Physiol 2020;11:88.
7. Delgiorno KE, Hall JC, Takeuchi KK, et al. Identification and manipulation of biliary metaplasia in pancreatic tumors. Gastroenterology 2014;146:233-44 e5.
8. DelGiorno KE, Chung CY, Vavinskaya V, et al. Tuft Cells Inhibit Pancreatic Tumorigenesis in Mice by Producing Prostaglandin D2. Gastroenterology 2020.
9. Schmidt PH, Lee JR, Joshi V, et al. Identification of a metaplastic cell lineage associated with human gastric adenocarcinoma. Lab Invest 1999;79:639-46.
10. Weis VG, Sousa JF, LaFleur BJ, et al. Heterogeneity in mouse spasmolytic polypeptide-expressing metaplasia lineages identifies markers of metaplastic progression. Gut 2013;62:1270-9.
11. Goldenring JR. Pyloric metaplasia, pseudopyloric metaplasia, ulcer-associated cell lineage and spasmolytic polypeptide-expressing metaplasia: reparative lineages in the gastrointestinal mucosa. J Pathol 2018;245:132-137.
12. Tosti L, Hang Y, Debnath O, et al. Single nucleus and in situ RNA sequencing reveals cell topographies in the human pancreas. Gastroenterology 2020.
13. Pan FC, Bankaitis ED, Boyer D, et al. Spatiotemporal patterns of multipotentiality in Ptf1a-expressing cells during pancreas organogenesis and injury-induced facultative restoration. Development 2013;140:751-64.
14. Sah RP, Garg SK, Dixit AK, et al. Endoplasmic reticulum stress is chronically activated in chronic pancreatitis. J Biol Chem 2014;289:27551-61.
15. Elyada E, Bolisetty M, Laise P, et al. Cross-Species Single-Cell Analysis of Pancreatic Ductal Adenocarcinoma Reveals Antigen-Presenting Cancer-Associated Fibroblasts. Cancer Discov 2019;9:1102-1123.
16. Schlesinger Y, Yosefov-Levi O, Kolodkin-Gal D, et al. Single-cell transcriptomes of pancreatic preinvasive lesions and cancer reveal acinar metaplastic cells' heterogeneity. Nat Commun 2020;11:4516.
17. Sato A. Tuft cells. Anat Sci Int 2007;82:187-99.
18. 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;32:381-386.
19. Herring CA, Banerjee A, McKinley ET, et al. Unsupervised Trajectory Analysis of Single-Cell RNA-Seq and Imaging Data Reveals Alternative Tuft Cell Origins in the Gut. Cell Syst 2018;6:37-51 e9.
20. La Manno G, Soldatov R, Zeisel A, et al. RNA velocity of single cells. Nature 2018;560:494-498.
21. Aibar S, Gonzalez-Blas CB, Moerman T, et al. SCENIC: single-cell regulatory network inference and clustering. Nat Methods 2017;14:1083-1086.
22. Gregorieff A, Stange DE, Kujala P, et al. The ets-domain transcription factor Spdef promotes maturation of goblet and paneth cells in the intestinal epithelium. Gastroenterology 2009;137:1333-45 e1-3.
23. Seymour PA. Sox9: a master regulator of the pancreatic program. Rev Diabet Stud 2014;11:51-83.
24. Pin CL, Rukstalis JM, Johnson C, et al. The bHLH transcription factor Mist1 is required to maintain exocrine pancreas cell organization and acinar cell identity. J Cell Biol 2001;155:519-30.
25. Arrojo EDR, Jacob S, Garcia-Prieto CF, et al. Structural basis for delta cell paracrine regulation in pancreatic islets. Nat Commun 2019;10:3700.
26. Wollny D, Zhao S, Everlien I, et al. Single-Cell Analysis Uncovers Clonal Acinar Cell Heterogeneity in the Adult Pancreas. Dev Cell 2016;39:289-301.

27. Nam KT, Lee HJ, Sousa JF, et al. Mature chief cells are cryptic progenitors for metaplasia in the stomach. Gastroenterology 2010;139:2028-2037 e9.
28. Choi E, Hendley AM, Bailey JM, et al. Expression of Activated Ras in Gastric Chief Cells of Mice Leads to the Full Spectrum of Metaplastic Lineage Transitions. Gastroenterology 2016;150:918-30 e13.
29. Saenz JB, Mills JC. Acid and the basis for cellular plasticity and reprogramming in gastric repair and cancer. Nat Rev Gastroenterol Hepatol 2018;15:257-273.
30. Bockerstett KA, Lewis SA, Wolf KJ, et al. Single-cell transcriptional analyses of spasmolytic polypeptide-expressing metaplasia arising from acute drug injury and chronic inflammation in the stomach. Gut 2020;69:1027-1038.
31. Sakamoto N, Fukuda K, Watanuki K, et al. Role for cGATA-5 in transcriptional regulation of the embryonic chicken pepsinogen gene by epithelial-mesenchymal interactions in the developing chicken stomach. Dev Biol 2000;223:103-13.
32. Asada R, Saito A, Kawasaki N, et al. The endoplasmic reticulum stress transducer OASIS is involved in the terminal differentiation of goblet cells in the large intestine. J Biol Chem 2012;287:8144-53.
33. Gracz AD, Samsa LA, Fordham MJ, et al. Sox4 Promotes Atoh1-Independent Intestinal Secretory Differentiation Toward Tuft and Enteroendocrine Fates. Gastroenterology 2018;155:1508-1523 e10.
34. Gradwohl G, Dierich A, LeMeur M, et al. neurogenin3 is required for the development of the four endocrine cell lineages of the pancreas. Proc Natl Acad Sci U S A 2000;97:1607-11.
35. Lopez-Diaz L, Jain RN, Keeley TM, et al. Intestinal Neurogenin 3 directs differentiation of a bipotential secretory progenitor to endocrine cell rather than goblet cell fate. Dev Biol 2007;309:298-305.
36. Gunawardene AR, Corfe BM, Staton CA. Classification and functions of enteroendocrine cells of the lower gastrointestinal tract. Int J Exp Pathol 2011;92:219-31.
37. Bastidas-Ponce A, Tritschler S, Dony L, et al. Comprehensive single cell mRNA profiling reveals a detailed roadmap for pancreatic endocrinogenesis. Development 2019;146.
38. Oster A, Jensen J, Edlund H, et al. Homeobox gene product Nkx 6.1 immunoreactivity in nuclei of endocrine cells of rat and mouse stomach. J Histochem Cytochem 1998;46:717-21.
39. Baron M, Veres A, Wolock SL, et al. A Single-Cell Transcriptomic Map of the Human and Mouse Pancreas Reveals Inter- and Intra-cell Population Structure. Cell Syst 2016;3:346-360 e4.
40. Heller RS, Jenny M, Collombat P, et al. Genetic determinants of pancreatic epsilon-cell development. Dev Biol 2005;286:217-24.
41. Andralojc KM, Mercalli A, Nowak KW, et al. Ghrelin-producing epsilon cells in the developing and adult human pancreas. Diabetologia 2009;52:486-93.
42. Haber AL, Biton M, Rogel N, et al. A single-cell survey of the small intestinal epithelium. Nature 2017;551:333-339.
43. Yamashita J, Ohmoto M, Yamaguchi T, et al. Skn-1a/Pou2f3 functions as a master regulator to generate Trpm5-expressing chemosensory cells in mice. PLoS One 2017;12:e0189340.
44. Ramsay RG, Gonda TJ. MYB function in normal and cancer cells. Nat Rev Cancer 2008;8:523-34.
45. Kas K, Finger E, Grall F, et al. ESE-3, a novel member of an epithelium-specific ets transcription factor subfamily, demonstrates different target gene specificity from ESE-1. J Biol Chem 2000;275:2986-98.
46. Ghaleb AM, Yang VW. Kruppel-like factor 4 (KLF4): What we currently know. Gene 2017;611:27-37.
47. Kuga D, Ushida K, Mii S, et al. Tyrosine Phosphorylation of an Actin-Binding Protein Girdin Specifically Marks Tuft Cells in Human and Mouse Gut. J Histochem Cytochem 2017;65:347-366.
48. Farrell AS, Joly MM, Allen-Petersen BL, et al. MYC regulates ductal-neuroendocrine lineage plasticity in pancreatic ductal adenocarcinoma associated with poor outcome and chemoresistance. Nat Commun 2017;8:1728.

49. Gribble FM, Reimann F. Function and mechanisms of enteroendocrine cells and gut hormones in metabolism. Nat Rev Endocrinol 2019;15:226-237.
50. Molina-Montes E, Coscia C, Gomez-Rubio P, et al. Deciphering the complex interplay between pancreatic cancer, diabetes mellitus subtypes and obesity/BMI through causal inference and mediation analyses. Gut 2021;70:319-329.
51. Miao ZF, Lewis MA, Cho CJ, et al. A Dedicated Evolutionarily Conserved Molecular Network Licenses Differentiated Cells to Return to the Cell Cycle. Dev Cell 2020.
52. Stuart T, Butler A, Hoffman P, et al. Comprehensive Integration of Single-Cell Data. Cell 2019;177:1888-1902 e21.
53. Haghverdi L, Lun ATL, Morgan MD, et al. Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors. Nat Biotechnol 2018;36:421-427.
54. Cao J, Spielmann M, Qiu X, et al. The single-cell transcriptional landscape of mammalian organogenesis. Nature 2019;566:496-502.
55. Wolf FA, Angerer P, Theis FJ. SCANPY: large-scale single-cell gene expression data analysis. Genome Biol 2018;19:15.
56. Horstmann H, Korber C, Satzler K, et al. Serial section scanning electron microscopy (S3EM) on silicon wafers for ultra-structural volume imaging of cells and tissues. PLoS One 2012;7:e35172.
57. Seligman AM, Wasserkrug HL, Hanker JS. A new staining method (OTO) for enhancing contrast of lipid--containing membranes and droplets in osmium tetroxide--fixed tissue with osmiophilic thiocarbohydrazide(TCH). J Cell Biol 1966;30:424-32.
58. Walton J. Lead aspartate, an en bloc contrast stain particularly useful for ultrastructural enzymology. J Histochem Cytochem 1979;27:1337-42.
59. Harris KM, Perry E, Bourne J, et al. Uniform serial sectioning for transmission electron microscopy. J Neurosci 2006;26:12101-3.
60. Burel A, Lavault MT, Chevalier C, et al. A targeted 3D EM and correlative microscopy method using SEM array tomography. Development 2018;145.
61. Berger DR, Seung HS, Lichtman JW. VAST (Volume Annotation and Segmentation Tool): Efficient Manual and Semi-Automatic Labeling of Large 3D Image Stacks. Front Neural Circuits 2018;12:88.
62. Lee CT, Laughlin JG, Angliviel de La Beaumelle N, et al. 3D mesh processing using GAMer 2 to enable reaction-diffusion simulations in realistic cellular geometries. PLoS Comput Biol 2020;16:e1007756.
63. Jorstad A, Blanc J, Knott G. NeuroMorph: A Software Toolset for 3D Analysis of Neurite Morphology and Connectivity. Front Neuroanat 2018;12:59.
64. Schneider CA, Rasband WS, Eliceiri KW. NIH Image to ImageJ: 25 years of image analysis. Nat Methods 2012;9:671-5.
65. Saenz JB, Burclaff J, Mills JC. Modeling Murine Gastric Metaplasia Through Tamoxifen-Induced Acute Parietal Cell Loss. Methods Mol Biol 2016;1422:329-39.
66. Ramsey VG, Doherty JM, Chen CC, et al. The maturation of mucus-secreting gastric epithelial progenitors into digestive-enzyme secreting zymogenic cells requires Mist1. Development 2007;134:211-22.

Acknowledgments: We thank Richard DiPaolo, Christopher Wright, Eunyoung Choi, Ela Contreras Panta, and members of the Goldenring, Choi, and DelGiorno laboratories for helpful discussions. We thank Dr. Tom Bartol for his help with 3DEM stack alignment using SWiFT-IR (funded by NIH P41GM103712 and NSF DBI-1707356). Finally, we thank the scientists whose work this study builds on, but cannot be discussed here due to space constraints.

Funding: N.K.L. was supported by the Salk Institute Cancer Training Grant T32CA009370. C.J.C. is funded by National Institutes of Health/National Cancer Institute (T32-CA009547). Salk Core facilities are supported, in part, by NIH-NCI CCSG: P30 014195. The Salk Next Generation Sequencing Core is additionally supported by the Chapman Foundation and the Helmsley Charitable Trust. The Salk Waitt

Advanced Biophotonics Core Facility and the Manor Laboratory is additionally supported by the Waitt Foundation, the Chan-Zuckerberg Initiative Imaging Scientist Award, and by NSF NeuroNex Award No. 2014862. The Flow Cytometry Core Facility of the Salk Institute is additionally funded by Shared Instrumentation Grant S10-OD023 (Aria Fusion cell sorter). M.C.B.T is supported by a Vanderbilt Digestive Disease Center Pilot and Feasibility Grant (P30 058404) and a Nikki Mitchell Foundation Pancreas Club Seed Grant. R.AeD. is supported by start-up funds from the Department of Molecular Physiology and Biophysics (Vanderbilt University) and NIH grants 1U01 DK120447-01 and 1R03 DK127484-01. J.C.M. is supported by the National Institute of Diabetes and Digestive and Kidney Diseases (R21 DK111369, R01DK094989, R01DK105129, R01DK110406) and The National Institutes of Health National Cancer Institute (R01 CA239645, R01 CA246208). The Goldenring lab is supported by a Department of Veterans Affairs Merit Review Award IBX000930, NIH RO1 DK101332, DOD CA160479, and a Cancer UK Grand Challenge Award 29075. The Lau laboratory is supported by NIH R01 DK103831, P50 CA236733, U01 CA215798, and T32 LM012412. Work in the Wahl laboratory is supported, in part, by the Cancer Center Core Grant (CA014195), National Institutes of Health/National Cancer Institute (R35 CA197687), National Institutes of Health Cancer Training Grant (T32 CA009370), the Isacoff Gastrointestinal Research Foundation, the Freeberg Foundation, and the Leona M. and the Harry B. Helmsley Charitable Trust (2012-PG-MED002). The DelGiorno laboratory is supported by the Vanderbilt-Ingram Cancer Center Support Grant (P30 CA068485) and start-up funds from the Department of Cell and Developmental Biology (Vanderbilt University).

SUPPLEMENTAL METHODS

Isolation of pancreatic single cell suspensions. EYFP+ cells were isolated from the pancreata of caerulein-treated mice. The pancreas was quickly dissected and minced in HBSS. Supernatant and fat were removed and pancreatic tissue was then incubated in 10 ml DMEM supplemented with 1 mg/ml collagenase I (Sigma), 1 mg/ml soybean trypsin inhibitor (Gibco), 0.5-1 mg/ml hyaluronidase (Sigma), and 250 ml of DNAse I, shaking gently at 37°C for a maximum of 30 min. Digestion was monitored and tissue was further digested mechanically by pipetting. Digested tissue was passed through a 100 mm filter, washed with PBS containing 1 mM EDTA and 0.5% BSA, and incubated with ACK lysing buffer (Gibco) to remove red blood cells. Single cell suspensions were incubated on ice with DAPI (molecular probes, 1:1000), used to exclude dead cells. EYFP+ cells were FACS purified at the Salk Institute's Flow Cytometry core facility on a BD Biosciences Influx and an Aria Fusion cell sorter (100-µm size nozzle, 1 x PBS sheath buffer with sheath pressure set to 20 PSI). Cells were sorted in 1-drop Single Cell sort mode for counting accuracy.

Single cell preparation and 10x 3' library construction. FACS-isolated EYFP+ cells were loaded onto a microfluidic chip with barcoded beads according to 10X Chromium Next GEM Single Cell 3' (v3.1, Catalog # 1000128, 10x Genomics Inc.). Subsequent cell lysis, first strand cDNA synthesis and amplification were carried out according to the 10X v3.1 protocol, with cDNA amplification set for 11 cycles. Following sample indexing and bead-based library purification, the final library size distribution and concentration were measured using TapeStation (Agilent Biosystems) and Qubit (ThermoFisher). The libraries were pooled in equal molar ratio and sequenced on Illumina NovaSeq 6000.

10X single cell data processing. The raw sequencing data from each mouse was first processed separately in Cell Ranger 4.0.0 using a customized reference based on refdata-gex-mm10-2020-A-R26 to allow quantification of EYFP expression. The EYFP reporter transgene was added to the refdata-gex-mm10-2020-A-R26 reference and rebuilt by running cellranger mkref with default parameters (10x Genomics). Each dataset captured a similar number of cells (~9k cells/mouse) with comparable sequencing depth and quality (see library QC summary in Supplemental File 1). The filtered gene-by-cell count matrices from 10x Cellranger were further QCed and analyzed in R package Seurat (v3.2.3)[52]. Each library was processed separately in Seurat with NormalizeData(normalization.method = "LogNormalize") function. The normalized data were further linear transformed by ScaleData() function prior to dimension reduction. Principal components analysis (PCA) was performed on the scaled data by only using the most variable 2000 genes (identified using the default "vst" method). Cells were examined across all clusters to determine the low quality cell threshold that accommodates the variation between celltypes. Low quality cells were removed with the same filtering parameters across libraries (percent.mt <=6 & nCount_RNA >= 1000). The QCed dataset was reprocessed as described above and clusters were labeled with cell types based on marker gene expression and previously published scRNAseq derived gene signatures in pancreas. Doublets were marked using the doubetFinder R package (largely consistent with the results from Scrublet package) and removed. In addition to the doublets labeled by douletFinder. Two small clusters of cells with a high percentage of inferred doublets label, high number of UMI counts, and shared expression of marker genes were also removed as doublets. QCed datasets were integrated by using the fastMNN algorism included in the SeuratWrapper package to minimize the technical batch variations[53]. A union of the most variable 1500 genes from each sample was used as integration features for fastMNN. The first 45 MNN principal components were used in UMAP embedding and cluster identification. New clusters were re-labeled by checking both known individual marker gene expression and the module scores of previously generated cell-type-specific gene signatures from a pancreas scRNA-seq study[15]. EYFP negative clusters of stromal contamination were removed and the epithelial population was further filtered for EYFP transcripts resulting in 13362 EYFP+ cells (detectable EYFP transcripts) for downstream analysis. Geneset enrichment was performed in Seurat using the addModuleScore() function.

Tuft and enteroendocrine cells were subsequently subset, re-scaled, and re-clustered to better reveal cellular heterogeneity. A small cluster of 107 cells become evident as doublets after re-clustering and was removed from subsequent analysis. This cluster expresses tuft, acinar, and mucin/Ductal cells and has a higher number of genes and UMI counts detected per "cell". Differentially expressed genes were identified by using FindAllMarkers() function with logfc.threshold=0.5 and min.pct=0.2.

GO enrichment analysis. GO analysis was done with R package topGO v2.42.0 (algorithm = "elim", statistic = "fisher"))(https://bioconductor.org/packages/release/bioc/html/topGO.html). Genes meeting the filtering threshold in DEGs analysis (pct.min=0.2, min.diff.pct=-inf, min.cell.feature =3, logfc.threshold > 0) were used as allGenes to create the topGOdata object. The top 20 enriched GO terms were plotted in ggplot2. Enrichment score was calculated by taking -log10(p-value of Fisher's exact test).

In generating heatmaps of selected genes, cells were grouped by clusters and cells within a cluster group were re-ordered by increasing pseudotime within each cluster to better reveal expression pattern changes.

RNA velocity estimation. RNA velocity estimation was performed separately in each sample using the Velocyto.R package (v0.6)[20]. The spliced/unspliced/ambiguous reads classification was done by running the `velocyto run10x` pipeline on cellranger output. The resulting loom file was converted to Seurat object and subset to keep the 13362 EYFP+ cells. RNA velocity estimation was performed by using `RunVelocity()` function in Velocyto.R package with the following parameters: `deltaT=1, ambiguous="ambiguous", kCells=25, fit.quantile=0.02,reduction = "mnn", ncores=30`. Cell-cell distance used in the `Runvelocity()` step was calculated from the fastMNN corrected principle components space. Calculated velocity was visualized on the UMAP embedding through the `show.velocity.on.embedding.cor()` function with the parameters setting: min.grid.cell.mass=0.5, show.grid.flow=TRUE, grid.n=40, do.par=T, n.cores=30,cell.border.alpha=0.1, arrow.scale=1.

CellRank analysis

The aggregate fate probabilities of tuft+eec lineage were calculated using the CellRank package (v.1.2.0) (**doi:** https://doi.org/10.1101/2020.10.19.345983). Briefly, the count matrix from the RNA assay in the Seurat object of Tuft+EEC data subset was converted into h5ad format using the SeuratDisk package and subsequently loaded using scranpy for CellRank analysis. New neighborhood graph was recomputed in scranpy using the first 30 fastMNN principal components. To get cell-cell transition probabilities, we used a PalantirKenel on monocle3 psudotime. 9 macrostates corresponding to the 9 seurat clusters were identified using the `compute_macrostate()` function in the GPCCA estimator of the CellRank package. Macrostates corresponding to the Seurat clusters 0, 4, 7, and 8 were labeled as terminal states and the fate probabilities towards the tuft terminal state (seurat cluster 0) was calculated using the `pl.cluster_fates()`function. Similarly, the regulon activity score matrix was imported into scranpy for plotting regulon activity score over monole3 pseudotime. The plot was generated using the `pl.gene.trends(n_test_point=500)` function by fitting a GAM model between regulon activity score and monocle3 pseudotime in the cellrank package.

Regulon Analysis

Gene regulatory network (Regulon) analysis in this study was performed using the SCENIC R package (v1.2.0)[21] The mouse mm10 500bp_up_100bp_down and 10kb_up_and_down databases were downloaded from the cistarget database; the TSS database was downloaded from the Aerts lab website (https://resources.aertslab.org/cistarget/) and used in this analysis. The filtered count matrix of the 13362 EYFP+ cells were further filtered to exclude non-expressed genes using the function `geneFiltering(exprMat, scenicOptions=scenicOptions, minCountsPerGene=100, minSamples=25)` in the SCENIC package. Spearman correlation between each TF and other genes was run by

runCorrelation() function with the default settings. TF-target gene co-expression module was identified in pySCENIC by running the Run_arboreto_with_multiprocessing.py pipeline with default setting and random seed 915. The adjacency file was imported into SCENIC R package and the normalized weight was used as weight subsequent steps. Regulon and regulon activity was calculated by running the `runSCENIC_1_coexNetwork2modules()`, `runSCENIC_2_createRegulons()`,`runSCENIC_3_scoreCells()` functions with default settings. Filtered Count matrix was used in step 3 for regulon activity scoring using the AUCell algorithm. The resulting regulon activity score matrix was subsequently imported into Seurat through `CreateAssayObject()` function. Regulon activity score was plotted in Seurat without further normalization or scaling.

Monocole3 Trajectory analysis

Seurat object of the 13362 EYFP+ cells was converted into Monocle3 CellDataSet object by using the `as.cell_data_set()`function in the SeuratWrapper Package. Subsequent trajectory analysis is done in the monocle3 package[54]. Cells were re-clustered in monocle by running the `cluster_cells()` function with setting `resolution= 2.5e-4`. Then cells were ordered in pseudotime by running the `order_cells()` function with default setting. Root state was manually set to acinar cells. Trajectory plot was generated using the `plot_cells()` function. Tuft+EEC monocle3 trajectories was done similarly as above. 2586 Tuft+EEC cells subset was used and the `cluster_cells()` function was run with `resolution = 1e-3`.The root state was manually set to Cluster 6 based on the RNA velocity information.

p-Creode trajectory analysis. Count matrices were first filtered using the dropkick algorithm to identify cells versus ambient barcodes (https://doi.org/10.1101/2020.10.08.332288). Counts data were then normalized to the median total read count across all retained barcodes, transformed using an inverse hyperbolic sine function, and Z-score standardized through functions available in Scanpy (v1.6.1)[55]. p-Creode (v2.2.0) was performed on normalized data in supervised mode with a max of 10 principal components across 1000 scored bootstrapped runs. The highest scoring run, producing the most representative graph from the ensemble, was chosen to visualize overlays and pseudotime gene expression dynamics. These gene expression dynamics were smoothened and plotted using a generalized additive model, implemented through pyGAM (v0.8.0, zenodo.org). p-Creode density parameters were chosen according to automatic thresholding, or the best_guess function. Noise and target density parameters varied between the different datasets analyzed, resulting in the use of approximately 20% to 25% of cells in each bootstrapped replicate.

Histological staining. Tissues were fixed overnight in zinc-containing, neutral-buffered formalin (Fisher Scientific), embedded in paraffin, cut in 5 μm sections, mounted, and stained. Sections were deparaffinized in xylene, rehydrated in ethanol, and then washed in PBST and PBS. Endogenous peroxidase activity was blocked with a 1:50 solution of 30% H_2O_2:PBS followed by microwave antigen retrieval in 100 mM sodium citrate, pH 6.0. Sections were blocked with 1% bovine serum albumin (BSA)

and 5% normal goat or rabbit serum in 10 mM Tris (pH 7.4), 100 mM $MgCl_2$, and 0.5% Tween-20 for 1hr at room temperature. Primary antibodies were diluted in blocking solution and incubated overnight. Anti-Gkn3 was generated by the Mills Laboratory (Washington University, St. Louis, MO), anti-Tff2 was a generous gift from Sir Nicholas Wright (Barts Centre, London), and anti-Gif was a generous gift from the Alberts Laboratory (Washington University, St. Louis, MO). Information on additional primary antibodies is provided in Table S1. Slides were then washed, incubated in streptavidin-conjugated secondaries (for rabbit or mouse antibodies, Abcam, for rat or goat antibodies, Vector) and developed with DAB substrate (Vector). Hematoxylin and eosin (H&E) staining was performed to assess tissue morphology. All slides were scanned and imaged on an Olympus VS-120 or Olympus VS-200 Virtual Slide Scanning microscope.

Fluorescence microscopy. Tissues were fixed overnight in 4% paraformaldehyde, washed 3x with PBS and floated overnight in 30% sucrose. Tissues were then incubated in a 1:1 mixture of 30% sucrose and Tissue-Tek optimal cutting temperature compound (OCT, VWR) for 30 min, embedded in OCT and frozen at -80°C. 7 μm tissue sections were cut, permeabilized with 0.1% Triton X-100 in 10 mM PBS, and blocked with 5% normal donkey serum and 1% BSA in 10 mM PBS for 1 hour at room temperature. Tissue sections were stained with primary antibodies in 10 mM PBS supplemented with 1% BSA and 0.1% Triton X-100 overnight (Table S1). Sections were then washed 3 x 15 min in PBS with 1% Triton X-100, incubated in Alexa Fluor secondary antibodies and/or phalloidin (Invitrogen), washed again for 3 x 5 min, rinsed with distilled water, and mounted with Prolong Gold containing Dapi (Invitrogen). Immunofluorescence on paraffin-embedded tissues followed the immunohistochemistry protocol until the blocking step. Instead, tissues were blocked in the donkey serum block described above and then followed the protocol for fluorescence microscopy described here. All slides were imaged on an Olympus VS-120, an Olympus VS-200 Virtual Slide Scanning microscope, or Zeiss 880 Airyscan Super-Resolution microscope.

For multiplex staining featured in Figure 2B, primary antibodies (Table S1) were used at the listed concentrations and incubated in tris buffered saline (TBS) overnight. The next day slides were washed 4 x 5 min in TBS and incubated for 1 hour at room temperature with DAPI (H-1000, Invitrogen) and secondary antibodies raised in donkey or goat and conjugated to AlexaFluor 488, 555, 594 or 647 (Invitrogen or JacksonImmunoresearch), used at a 1:400 dilution. After incubation, samples were washed 4 x 5 min in TBS, covered with Vectashield (Invitrogen) and cover-slipped. Multi-plant confocal images were acquired using an up-right Leica Stellaris X5 microscope with a 20x/0.75 N.A. multi-immersion objective. After acquisition, images were processed using ImageJ/Fiji (NIH) and denoised using a median filter of 1px. Images shown are maximum projections of confocal stacks.

instant Structured Illumination Microscopy (iSIM). Image-stacks of Brainbow lineage tracing were captured with a system assembled by BioVision Technologies (Exton, PA) consisting of a VisiTech International (Sunderland, United Kingdom) iSIM module attached to a Nikon (Melville, NY) Ti2 stand

equipped with a Nikon 100x 1.49 Plan Apo objective. Raw image stacks were then deconvolved using a Microvolution (Cupertino, CA) plugin in Fiji (ImageJ).

3DEM Methods. Samples of injured pancreas tissue were processed and imaged as described previously with some modifications[56]. Materials were sourced from Electron Microscopy Sciences (Hatfield, PA) unless noted otherwise. Briefly, fresh biopsies of injured tissues were dissected into small pieces (<1mm) and quickly immersed in 37°C buffered fixative (2.5% glutaraldehyde, 2% paraformaldehyde, 3mM $CaCl_2$, 0.1M sodium cacodylate) before overnight storage at 4°C in the same solution[57]. Samples were serially rinsed with ice-cold cacodylate/$CaCl_2$ buffer and stained with buffered reduced osmium (1.5% Osmium, 1.5% potassium ferrocyanide, 3mM $CaCl_2$, 0.1M sodium cacodylate) for 45 minutes in the dark at room temperature. Samples were serially rinsed with ice cold distilled water before 30 minutes of treatment with filtered 1% aqueous thiocarbohydrazide (Ted Pella) that was prepared at 60°C for an hour before use. Samples were serially rinsed with ice cold distilled water before staining with 1.5% aqueous osmium tetroxide for 45 minutes in the dark at room temperature. Samples were rinsed serially with ice cold water before staining overnight with 1% uranyl acetate at 4°C. Samples were rinsed with distilled water and stained with Walton's lead aspartate for 30 minutes at 60°C[58]. Samples were rinsed again with distilled water and serially dehydrated in ice cold ethanol. Samples were then infiltrated with Durcupan resin (hard formulation) and polymerized at 60°C for 3 days.

A sample block was prepared for serial sectioning as previously described and a ribbon of 250 serial sections, each of dimension ~150µmx400µmx70nm (x-y-z), were collected onto a silicon wafer diced to a dimension of 35x6mm using Diatome diamond knives mounted in a Leica UC8 ultramicrotome[59, 60]. The chip was imaged using a Zeiss Sigma VP scanning electron microscope equipped with a Gatan backscattered electron detector and Atlas5 (Fibics) control software. Multi-resolution maps were generated for a subset of sections throughout the ribbon, of which images were used for Figure 3G and Movie S1, and a region of interest (ROI) for 3DEM was identified. Images were acquired with an accelerating voltage of 1.5kV, a 20µm aperture, at a working distance of 9mm, and with pixel size of 8nm. Contiguous images of the ROI were assembled into a series of 8-bit .tiff and aligned.

A delta cell, a tuft cell, two mucinous cells, an acinar cell, and their associated basement membrane were identified by their salient ultrastructural features. Their boundaries, along with their respective nuclei, were traced using VAST Lite annotation software[61]. Cells and nuclei were exported using the Matlab tools bundled with VAST Lite as .obj files, which were imported into Blender 2.79 (Blender Foundation) with add-ons BlendGAMer and Neuromorph both installed[62, 63]. Cells and nuclei were processed with GAMer to optimize geometry for visualization, before superimposition with 3DEM data using Neuromorph and animation and rendering using Blender. For video figures, rendered scenes were assembled into .avi files using ImageJ, post-processed using After Effects (Adobe), and compressed using Handbrake (Handbrake.fr)[64].

Transmission electron microscopy of murine pancreas

Tissues were fixed in a solution of 2% paraformaldehyde, 2.5% glutaraldehyde, and 2 mM $CaCl_2$ in 0.15 M sodium cacodylate buffer (pH 7.4) for 2 hours at room temperature. They were then post-fixed in 1% osmium tetroxide for 40 min and 1.5% potassium ferricyanide in sodium cacodylate buffer for 1 hr at 4°C in the dark. Tissues were stained en bloc in 1% aqueous uranyl acetate (4C in the dark), dehydrated in a series of graded ethanols, and embedded in Eponate12 resin (Ted Pella). Ultra-thin sections (70 nm) were obtained using a diamond knife (Diatome) in an ultramicrotome (Leica EM UC7) and placed on copper grids (300 mesh). Sections were imaged on a Zeiss Libra 120 TEM operated at 120 kV.

Generation of SPEM and transmission electron microscopy of murine stomach. Gastric SPEM was induced with tamoxifen (5mg/20g mouse body weight; Toronto Research Chemicals Inc., Toronto, Canada) administered by intraperitoneal injection for three consecutive days[65]. Tamoxifen was prepared by sonication in a 90% sunflower seed oil (Sigma, St Louis, MO) and 10% ethanol solution. For transmission electron microscopy, stomachs were prepared and imaged as previously described[66]. Briefly, stomach corpus was collected as described, fixed overnight at 4°C in fixative (modified Karnovsky's), and sectioned. Tissue was processed by the Washington University in St. Louis Department of Pathology and Immunology Electron Microscopy Facility.

All images (IHC, IF, EM) were digitally enhanced to edit the color, brightness and contrast levels using Zen (Carl Zeiss), ImageJ/Fiji (NIH), and/or Photoshop (Adobe) software.

Figures. Models shown in Figures 1A and in the graphical abstract were generated using BioRender.com.

Antibody	Company	Catalog #	Dilution, IF	Dilution, IHC
Amylase	Abcam	ab21156	N/A	1:5000
Aqp5	Sigma	HPA-065008	1:500	1:500
Cd44v9 (anti-mouse)	Cosmo Bio Co.	LKG-M002	1:10,000	1:10,000
Cd44v9 (anti-human)	Cosmo Bio Co.	LKG0M003	1:10,000	N/A
Cd45	Biolegend	103101	1:100	N/A
ChgA	Santa Cruz	sc-1488	1:100	N/A
Gamma actin	Santa Cruz	sc-65638	1:100	N/A
Gastrin	Biogenex	PU019-UP	N/A	1:100
GFP	Abcam	ab6673	1:1000	1:1000
Ghrelin	Cell Signaling	31865	1:1000	1:2000
GIF	N/A	N/A	1:1000	1:1000
Gkn3	N/A	N/A	N/A	1:150
Insulin	Agilent	IR00261-2	1:4	N/A
Krt19	Epitomics	AC-0073	1:100	N/A
Mmp7	Vanderbilt MBCF	N/A	1:10,000	1:10,000
Muc5ac	Thermofisher	MA5-12178	N/A	1:500
Pan-CK	Abcam	ab9377	1:100	N/A
Phospho-Girdin (pY1798)	IBL America	28143	1:100	N/A

PPY	Abcam	ab255827	1:80,000	1:80,000
Serotonin	Santa Cruz	sc-58031	1:200	1:500
Somatostatin	Millipore	MAB354	1:1000	1:2000
Synaptophysin	Cell Marque	336R-94	1:500	
TFF2	N/A	N/A	1:500	N/A
Wfdc2	Invitrogen	PA5-80227	N/A	1:2000

Table S1. Primary antibodies used in immunofluorescence (IF) and immunohistochemistry (IHC) studies.

Movie S1. Animated visualization of the 3-D reconstruction of a metaplastic duct showing a heterogeneous consortium of cell types encapsulated by basement membrane with closely apposed vasculature. Volumetric reconstructions of cell plasma membranes and nuclei, as well as a portion of the basement membrane rendered in gray, demonstrate the morphological variability between cell types. Binucleate acinar cell, orange; mucinous cells, green; tuft cell, magenta; enteroendocrine cell, blue.

Supplemental File 1: Library QC Matrices
Supplemental File 2: EYFP-positive cell type markers
Supplemental File 3: EYFP-positive Regulons
Supplemental File 4: Gastric cell type signatures
Supplemental File 5: SPEM signatures
Supplemental File 6: Tuft and EEC markers

What you need to know:

Background and context: Acinar to ductal metaplasia (ADM) is a reparative process in the pancreas in response to injury and is a precursor lesion for pancreatic ductal adenocarcinoma.

New Findings: In response to chronic injury, acinar cells undergo pyloric metaplasia, transitioning into progenitor-like populations, tuft cells, and enteroendocrine cells generating a gastric pyloris-like phenotype.

Limitations: Functional studies are required to identify the role of ADM populations in pancreatic injury and regeneration.

Impact: This study identifies significant acinar cell plasticity and the formation of metaplastic epithelial populations with the potential to fuel disease progression.

Short summary: Single cell RNA sequencing of lineage-traced acinar cells in pancreatic metaplasia identifies a transition to a gastric pyloric metaplasia with the formation of tuft cells and enteroendocrine cells with unknown roles in disease progression.

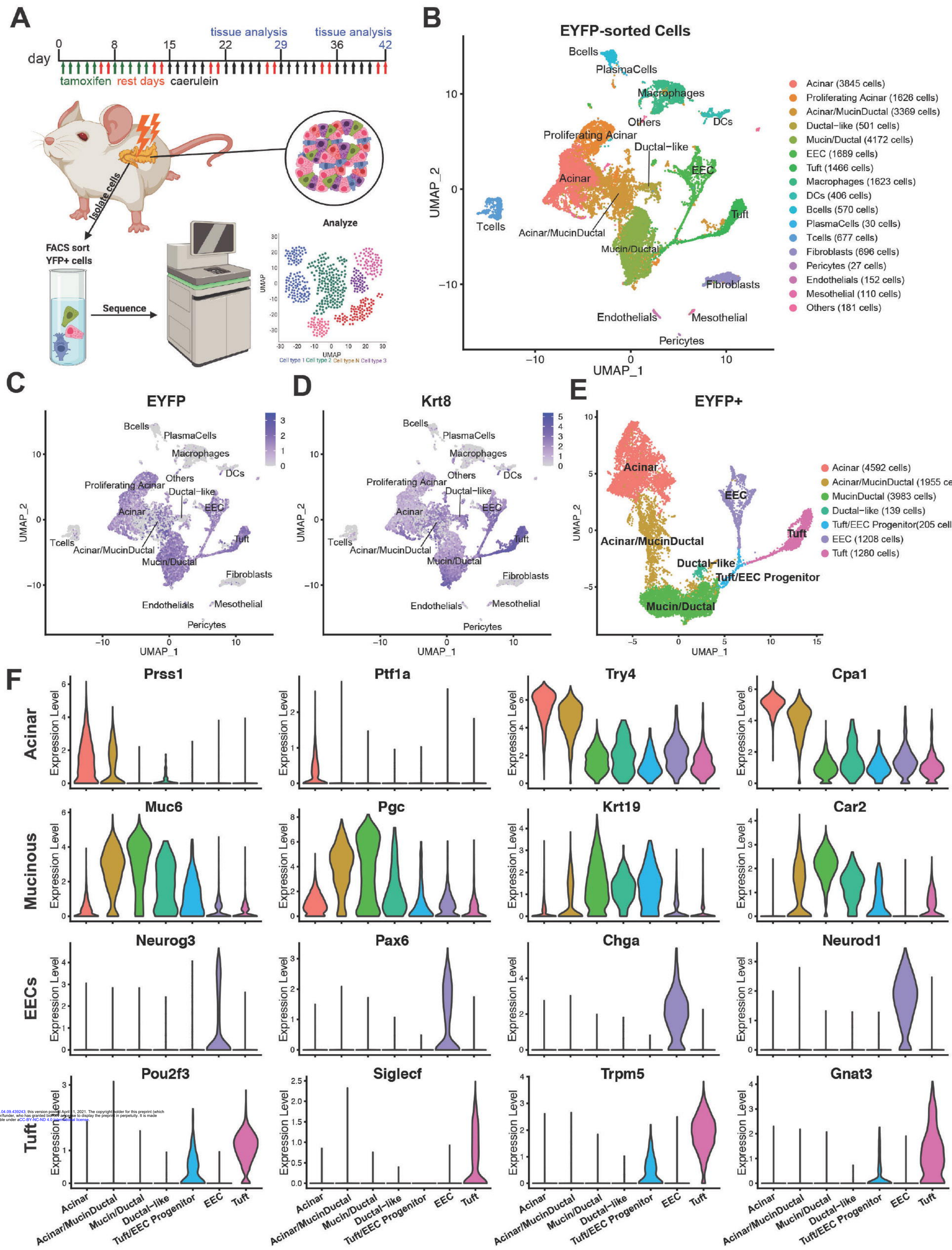

Figure 1. Identification of acinar to ductal metaplasia cell lineages using scRNA-seq and unbiased clustering. (**A**) Schematic of lineage tracing and scRNA-seq strategy to identify ADM populations. (**B**) Uniform manifold approximation and projection (UMAP) showing annotated clusters from FACS collected EYFP+ cells. EECs, enteroendocrine cells. (**C**) UMAP of *EYFP* and (**D**) *Krt18* expression in (B). (**E**) UMAP showing annotated clusters in the *EYFP*+ population. (**F**) Violin plots of marker genes used to identify acinar cells, mucin/ductal cells, EECs, tuft cells, and intermediate populations.

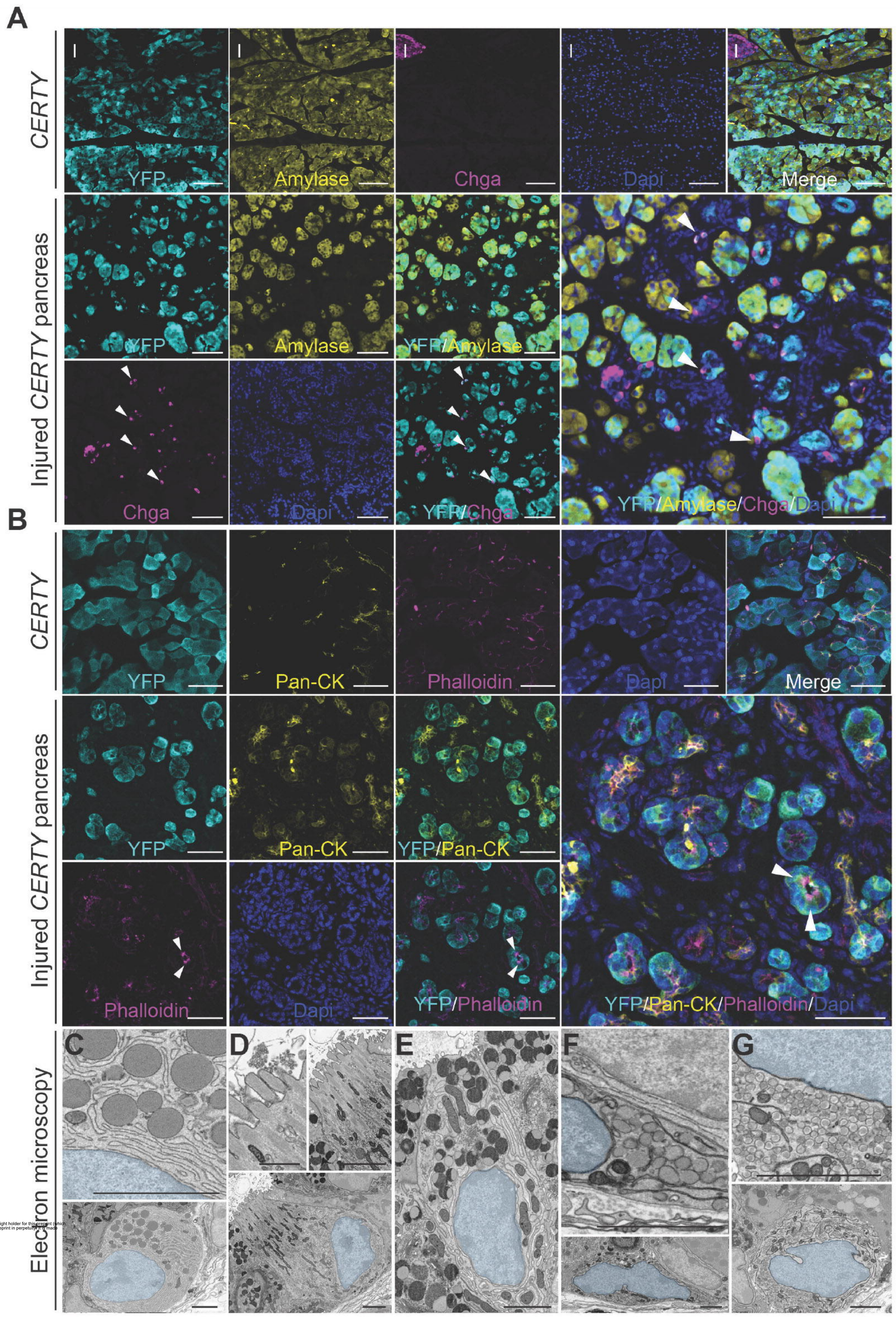

Figure 2. Validation of ADM epithelial populations identified by scRNA-seq. (**A**) Co-immunofluorescence (Co-IF) for EYFP (cyan), acinar marker amylase (yellow), islet marker chromogranin A (Chga, magenta), and Dapi, blue, in *CERTY* pancreata. White arrowheads, EYFP+ EECs. Scale bars, 100 μm. (**B**) Co-IF for EYFP (cyan), ductal marker pan-cytokeratin (Pan-CK, yellow), tuft cell marker phalloidin (microvilli and actin rootlets, magenta), and Dapi, blue, in *CERTY* pancreata. White arrowheads, EYFP+ tuft cells. Scale bars, 50 μm. (**C**) Scanning electron microscopy (SEM) of thin sections from the injured pancreas highlighting an acinar cell, (**D**) a tuft cell, (**E**) a mucinous cell, and (**F-G**) two EECs. Scale bars, 500 nm; tuft cell inset, 166.7 nm.

Figure 3. Lineage trajectory analysis of ADM populations. (**A**) Monocle3 trajectory analysis overlaid on the UMAP of *EYFP*+ cells from injured *CERTY* pancreata. (**B**) Pseudotime projection analysis and (**C**) p-Creode approximation of lineage progression of *EYFP*+ cells. (**D**) RNA velocity analysis of all four datasets from two time points. Arrow direction and length indicate the probable lineage trajectory and velocity, respectively. (**E**) IHC for GFP and YFP in untreated and injured pancreata from *Ptf1a*^{CreERTM};*Rosa*$^{Brainbow.2/+}$ mice. Scale bar, 25μm. (**F**) iSIM analysis of immunofluorescence of a whole YFP+ clone containing a tuft cell (phalloidin) and an EEC (synaptophysin) within the same lesion. Scale bars, 10 μm. (**G**) 3-D EM reconstruction of ADM showing a heterogeneous consortium of cell types encapsulated by basement membrane with closely apposed vasculature. From left to right, two partially annotated cross-sections from the 3-D EM volume (volume: 58x29x17.5μm; voxel size: 8-8-70nm, x-y-z). Volumetric reconstructions of cell plasma membranes and nuclei, as well as a portion of the basement membrane rendered in gray, demonstrating the morphological variability between cell types. See Movie S1 for an animated visualization. Scale bars, 10μm.

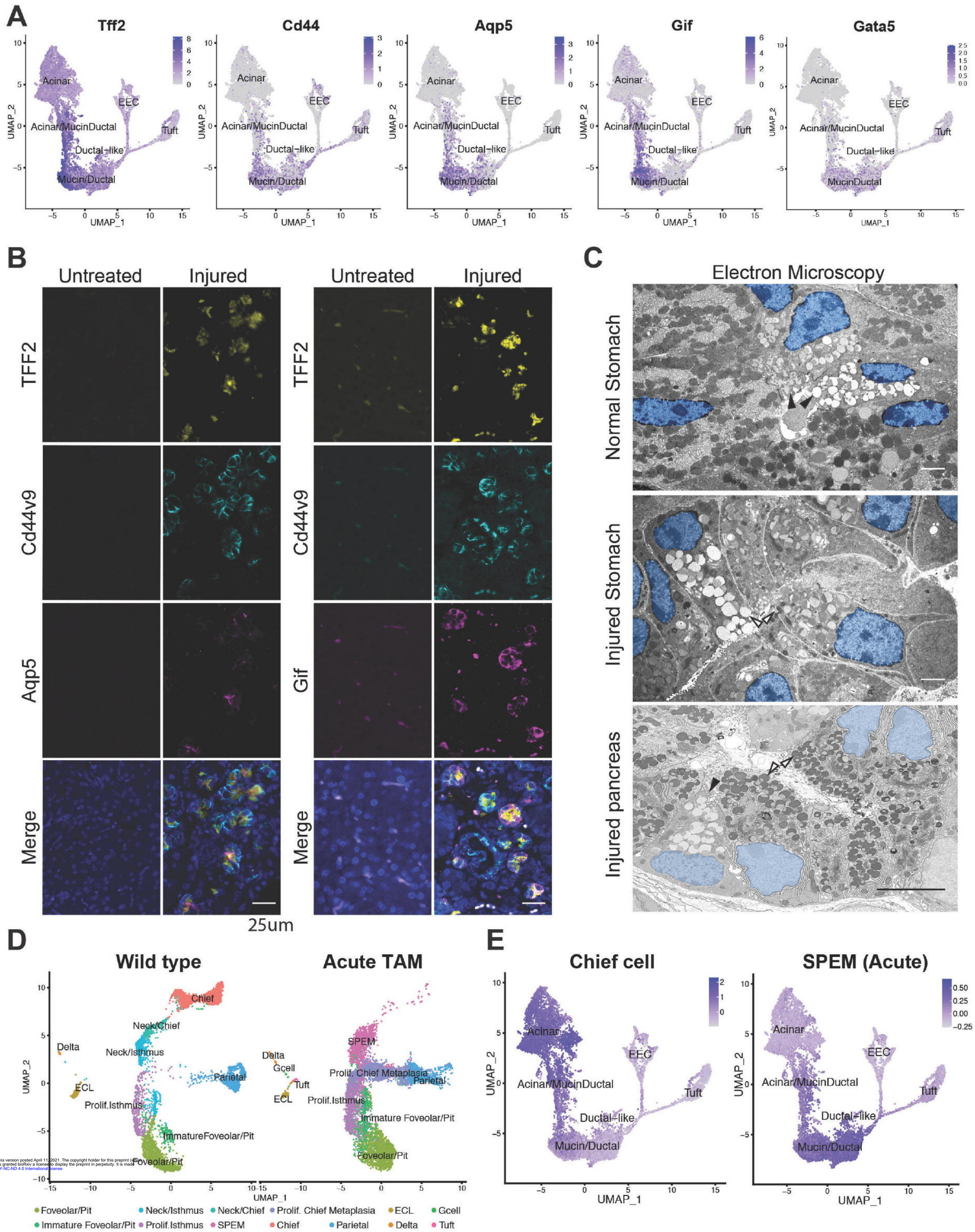

Figure 4. ADM as a pyloric metaplasia event. (**A**) Expression of SPEM marker genes *Tff2*, *Cd44*, and *Aqp5*, chief cell marker gene *Gif*, and TF *Gata5* overlaid on the UMAP of *EYFP*+ cells. (**B**) Co-IF for SPEM markers TFF2 (yellow), CD44v9 (cyan), and AQP5 or GIF (magenta) and dapi (blue). Scale bars, 25 μm. (**C**) EM of mature mucus neck cells (black arrowheads) in the normal stomach, SPEM in an acute model of gastric injury (white arrowheads), and mucin/ductal cells from pancreatic ADM. ADM cells with similar granules to gastric cell types are marked by arrowheads. Scale bars, 2 μm. (**D**) UMAP of scRNA-seq data collected from normal murine stomach and a SPEM model of acute gastric injury, from Bockerstett et al. (**E**) Expression of chief cell and acute injury-induced gastric SPEM signatures overlaid on the UMAP of *EYFP*+ cells. Gene signatures generated from (D). EC, enterochromaffin; Prolif, proliferating.

Figure 5. ADM results in substantial enteroendocrine cell heterogeneity (**A**) UMAP of cell clusters identified in the combined tuft and EEC dataset identified in Figure 1E and labeled by cell type. (**B**) Monocle3 trajectory analysis overlaid on the UMAP in (A). (**C**) Expression of TFs *Sox4* and (**D**) *Neurog3* overlaid on the UMAP from (A). (**E**) Expression of epsilon cell marker *Ghrl*, enterochromaffin cell marker *Ddc*, delta cell marker, *Sst*, and PP/gamma cell marker *Ppy* overlaid on the UMAP from (A). (**F**) RNA velocity analysis of all four datasets from two time points. Arrow direction and length indicate the probable lineage trajectory and velocity, respectively. (**G**) IHC for ghrelin (*Ghrl*), serotonin (generated by *Ddc*), somatostatin (*Sst*), and pancreatic polypeptide (*Ppy*) in untreated and injured pancreata. Black arrows, positive cells. I, islets. Scale bar, 25 μm. (**H**) TEM of an enterochromaffin cell or a delta cell in the injured pancreas. Scale bars, 5 and 2 μm respectively.

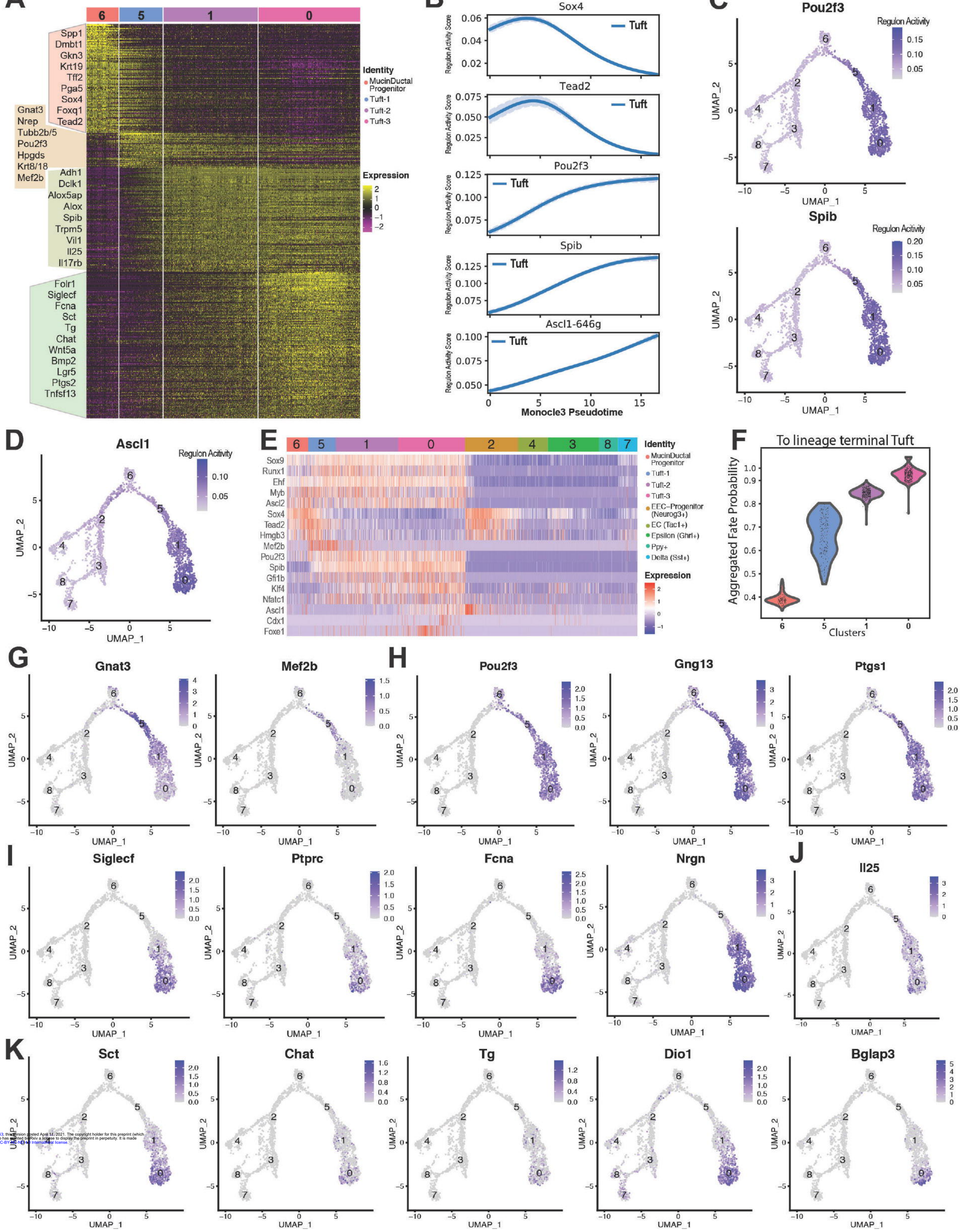

Figure 6. Transcriptomic analysis of pancreatic tuft cell formation and maturation. (**A**) Heatmap of scaled DEGs for the 4 clusters of the tuft cell lineage. A selection of known marker genes are labeled on the left. Cells were ordered by increasing pseudotime within each cluster group. (**B**) Smoothed regulon activity score trend of the tuft cell lineage in Monocle3 pseudotime. (**C-D**) Regulon activity scores of *Pou2f3*, *Spib* and *Ascl1* overlaid on the Tuft+EEC UMAP. (**E**) Heatmap of scaled RNA expression of tuft-related TFs. Cells are ordered by increasing pseudotime within each cluster group. **(F)** The aggregated fate probability towards terminal tuft lineage formation (shown in Figure 8B). (**G-K**) Expression of select genes overlaid on the Tuft+EEC UMAP, labeled by Seurat cluster.

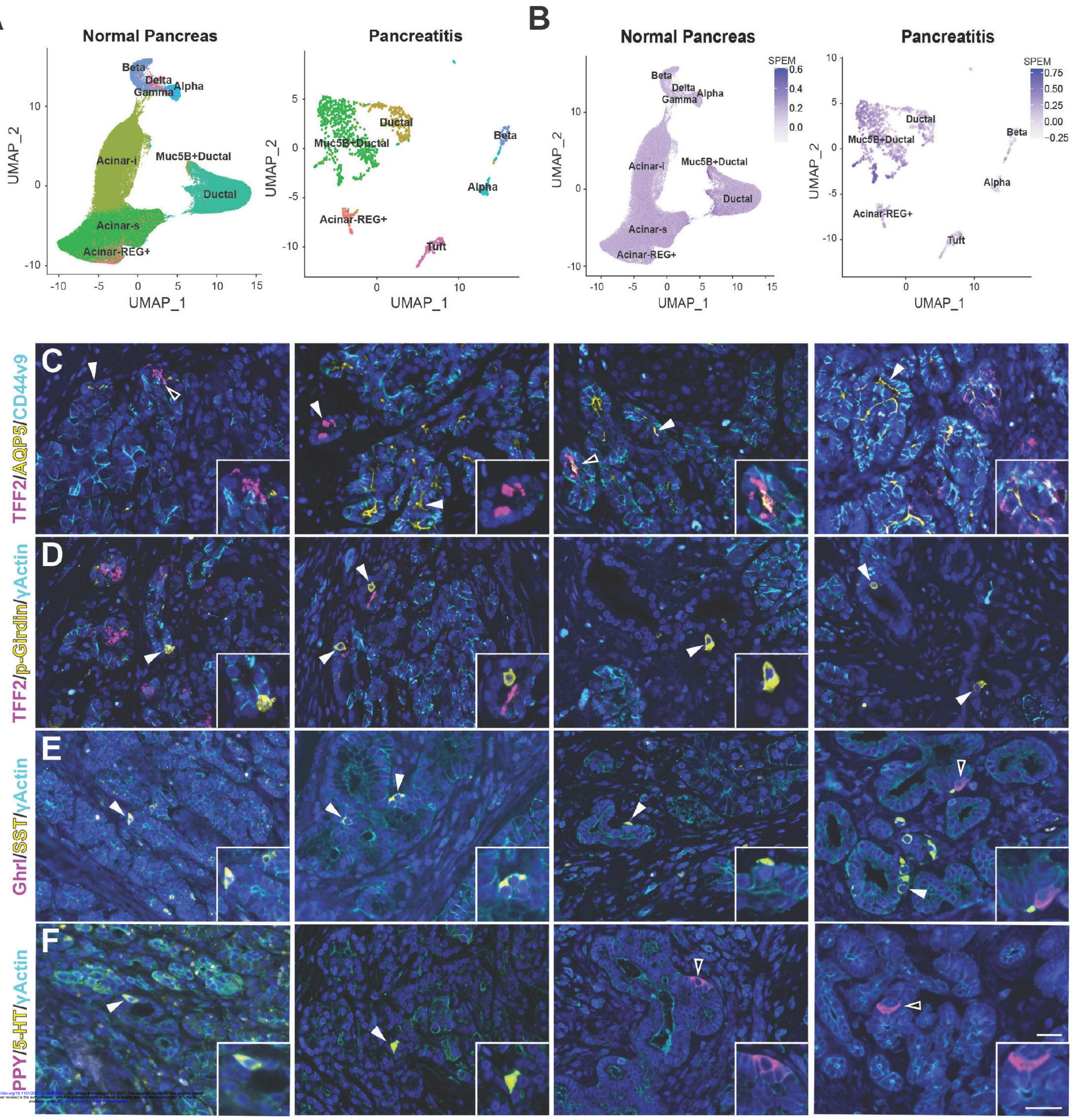

Figure 7. Human chronic pancreatitis as a pyloric-type metaplasia (**A**) UMAPs showing annotated clusters from sNuc-seq data collected from normal pancreata (~113k nuclei) or the injured pancreata of two patients with chronic pancreatitis (2,726 nuclei) derived from Tosti et al. (**B**). Expression of the humanized acute SPEM signature (Figure 4D) overlaid on the UMAPs from (A). (**C**) Co-IF for SPEM markers (TFF2, magenta; AQP5, yellow; CD44v9, cyan), (**D**) TFF2 (magenta), γActin (cyan), and tuft cell marker phospho-Girdin (yellow), (**E**) epsilon and delta cell markers (Ghrl, magenta; SST, yellow; γActin, cyan), and (**F**) gamma and enterochromaffin cell markers (PPY, magenta; 5-HT, yellow; γActin, cyan). Dapi, blue. Examples from 4 different patients are shown. Scale bars, 25 μm.