



## ARTICLE OPEN

# Single-cell transcriptomic analysis uncovers the origin and intratumoral heterogeneity of parotid pleomorphic adenoma

Xiuyun Xu<sup>1,2,3</sup>, Jiaxiang Xie<sup>1,2,3</sup>, Rongsong Ling<sup>4</sup>, Shengqi Ouyang<sup>1,2,3</sup>, Gan Xiong<sup>1,2,3</sup>, Yanwen Lu<sup>1,2,3</sup>, Bokai Yun<sup>1,2,3</sup>, Ming Zhang<sup>1,2,3</sup>, Wenjin Wang<sup>1,2,3</sup>, Xiqiang Liu<sup>5</sup>, Demeng Chen<sup>6</sup> and Cheng Wang<sup>1,2,3</sup>

Pleomorphic adenoma (PA) is the most common benign tumour in the salivary gland and has high morphological complexity. However, the origin and intratumoral heterogeneity of PA are largely unknown. Here, we constructed a comprehensive atlas of PA at single-cell resolution and showed that PA exhibited five tumour subpopulations, three recapitulating the epithelial states of the normal parotid gland, and two PA-specific epithelial cell (PASE) populations unique to tumours. Then, six subgroups of PASE cells were identified, which varied in epithelium, bone, immune, metabolism, stemness and cell cycle signatures. Moreover, we revealed that CD36<sup>+</sup> myoepithelial cells were the tumour-initiating cells (TICs) in PA, and were dominated by the PI3K-AKT pathway. Targeting the PI3K-AKT pathway significantly inhibited CD36<sup>+</sup> myoepithelial cell-derived tumour spheres and the growth of PA organoids. Our results provide new insights into the diversity and origin of PA, offering an important clinical implication for targeting the PI3K-AKT signalling pathway in PA treatment.

*International Journal of Oral Science* (2023)15:38

; <https://doi.org/10.1038/s41368-023-00243-2>

## INTRODUCTION

Pleomorphic adenoma (PA), as the most common tumour occurring in salivary glands, accounts for two-thirds of all salivary gland lesions, especially in the parotid gland.<sup>1</sup> It is a benign neoplasm but has a high recurrence rate and the potential for malignant transformation.<sup>2</sup> The term “pleomorphic adenoma” is defined due to variable cytomorphological and architectural features, which mainly contain three components: epithelial cells, myoepithelial cells and mesenchymal-like cells. Pathologically, PA presents diverse appearances with different ductal epithelial cells and myoepithelial cells growing in a variety of patterns embodied in mucoid-like tissue, myxoid-like tissue, and chondroid-like tissue, indicating that a huge intratumor heterogeneity exists. Consistent with these findings, intratumor molecular heterogeneity has been observed by evaluating the loss of heterozygosity and *PLAG1* gene rearrangements in PAs.<sup>3,4</sup> Due to diverse tissue architecture and intratumor heterogeneity, the cellular origin in PAs containing morphologically distinct components has been a controversial issue. Interestingly, a previous study showed that both the epithelial and mesenchymal elements were monoclonal using clonal analysis based on random inactivation of one of two x-chromosomes by methylation, suggesting that the original cells for both elements are identical.<sup>5</sup> Lee et al. further confirmed that both stromal and epithelial cells in PAs arose from the same origin using a human androgen receptor gene (HUMARA) assay.<sup>6</sup> Notably, recent studies have shown that epithelial cells or myoepithelial cells can transdifferentiate into mesenchymal cells in PAs, suggesting that epithelial-mesenchymal transition (EMT)

might be the basic principle of tissue heterogeneity in PAs.<sup>7–10</sup>

These findings support the notion that PA is a monoclonal tumour with a pure epithelial origin despite its diverse intratumor heterogeneity and complex tissue architecture. However, the intratumor heterogeneity of PAs is still largely unknown and there is no convincing experimental evidence to clarify the origin of PAs.

Here, we described a complete atlas of PA and uncovered its cellular complexity and intratumor heterogeneity at single-cell resolution. Our data revealed the transcriptomic profiles of the multicellular ecosystem of PA and showed that PA included acinar, ductal and basal/myoepithelial cells similar to normal parotid glands, along with two PA-specific epithelial (PASE) cell populations. Interestingly, PASE cells varied within tumours in their expression of gene signatures related to epithelial development (Cluster 0, C0), bone formation (Cluster 1, C1), immunity (Cluster 2, C2), metabolism (Cluster 3, C3), stemness (Cluster 4, C4), and the cell cycle (Cluster 5, C5). Of note, we observed that C4 cells had higher ‘differentiation’ potency and represented the starting state during PA development. Then, we revealed that CD36, a marker gene of C4 cells, was a functional cell surface marker for the enrichment of TICs in PA. In addition, activation of PI3K-AKT signalling was observed in CD36<sup>+</sup> PA cells and inhibition of the PI3K-AKT pathway suppressed the initiation and growth of PA. Taken together, these results uncover the cellular heterogeneity and the origin of PA and highlight potential intracellular signals controlling the initiation and progression of PA, which serve as a resource for further developing novel therapeutic strategies to manage PA in the future.

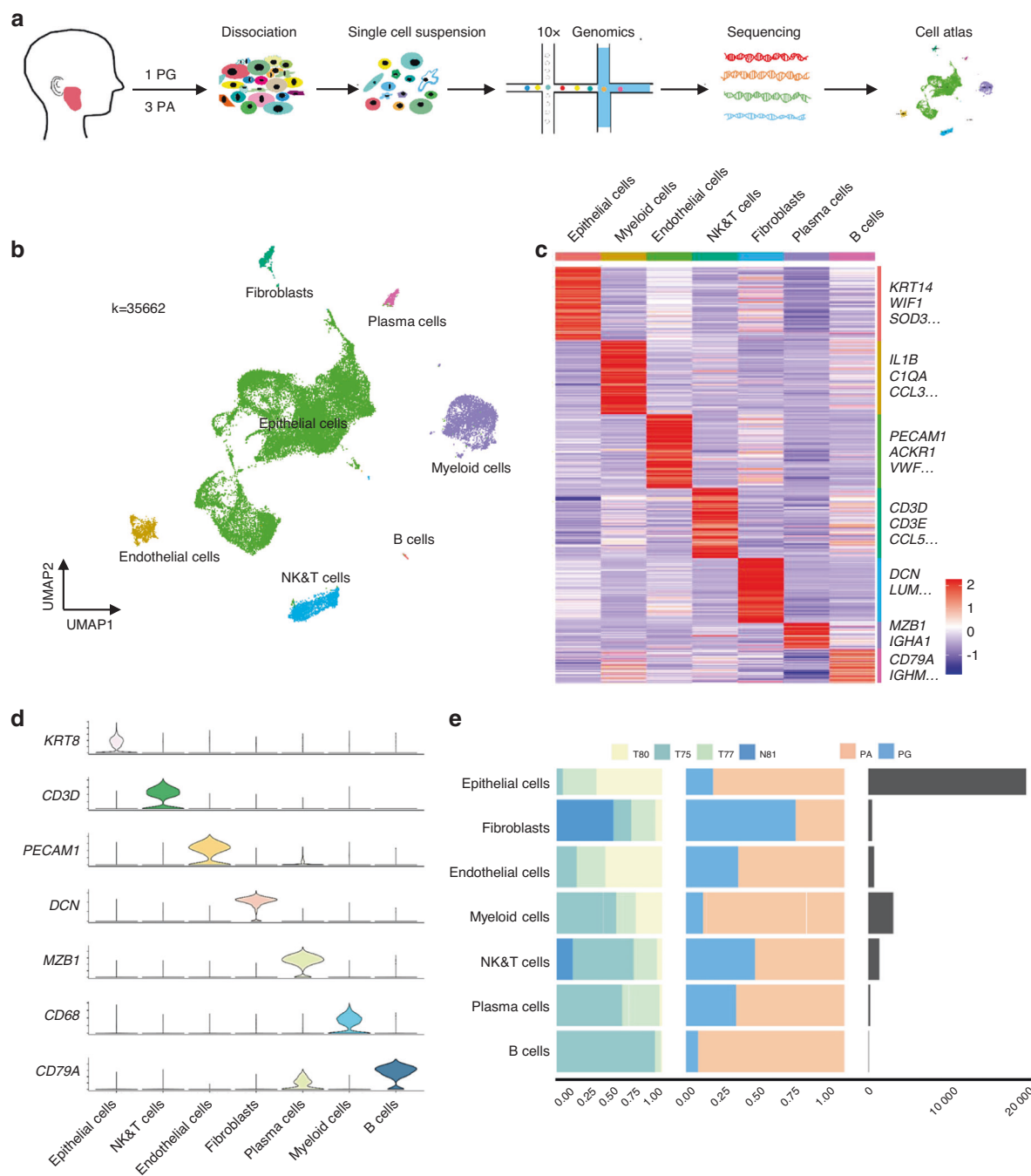
<sup>1</sup>Hospital of Stomatology, Sun Yat-sen University, Guangzhou, China; <sup>2</sup>Guangdong Provincial Key Laboratory of Stomatology, Sun Yat-sen University, Guangzhou, China; <sup>3</sup>Guanghua School of Stomatology, Sun Yat-sen University, Guangzhou, China; <sup>4</sup>Institute for Advanced Study, Shenzhen University, Shenzhen, China; <sup>5</sup>Department of Oral and Maxillofacial Surgery, Nanfang Hospital, Southern Medical University, Guangzhou, China and <sup>6</sup>Center for Translational Medicine, The First Affiliated Hospital, Sun Yat-sen University, Guangzhou, China

Correspondence: Demeng Chen (chendam29@mail.sysu.edu.cn) or Cheng Wang (wangch75@mail.sysu.edu.cn)

These authors contributed equally: Xiuyun Xu, Jiaxiang Xie

Received: 22 February 2023 Revised: 16 August 2023 Accepted: 16 August 2023

Published online: 07 September 2023



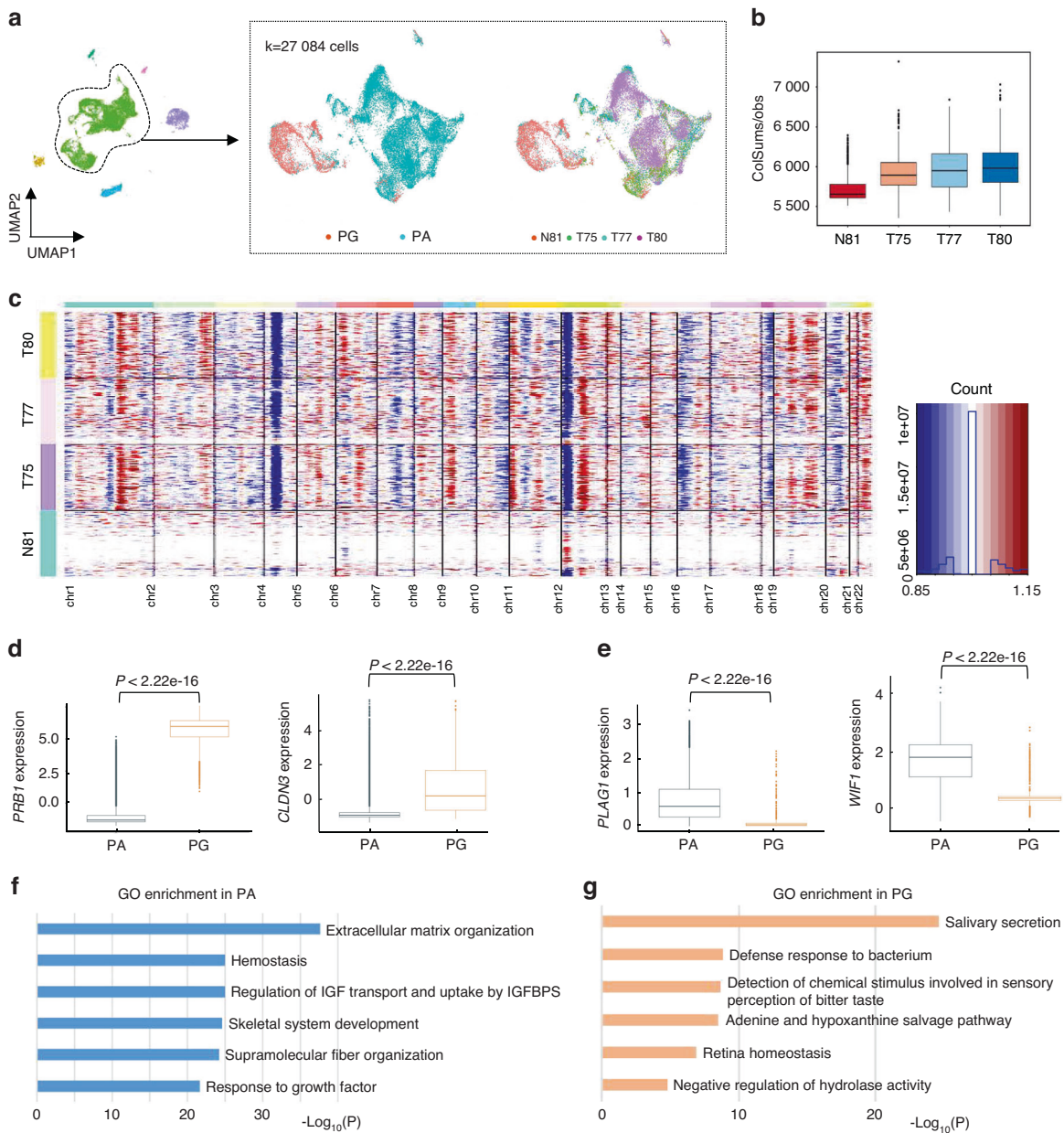
**Fig. 1** A single-cell atlas of PA and PG. **a** The workflow of single-cell RNA sequencing. **b** UMAP plot of 35 662 single cells from scRNA-seq labelled by cell type. **c** A heatmap of marker genes for each cell type. **d** StackedVlnPlot of classic genes of each cell type. **e** The cell numbers and proportions of PA cell types in different samples and tissue types

## RESULTS

### A single-cell atlas of the PA and parotid gland

To explore the cellular landscape of PA, we generated single-cell RNA-seq profiles using 1 normal parotid gland (PG) and 3 PA samples (Fig. 1a). As shown in Fig. S1a, H&E staining showed that the pathological structure of PA varied in different samples, reflecting pleomorphic architecture features. After removing low-quality cells and performing gene expression normalization, a total of 35662 single cells were processed for further analysis and 8 cell clusters were observed according to graph-based clustering and dimensional reduction with UMAP (Fig. S1b, c). As shown in Fig. S1d, e, the UMAP displayed that the cell distribution was

similar in different PA samples, but dramatically different in PA compared with the PG. This finding indicated that PA might have a unified cellular landscape despite the variable morphological architectures. Then, 7 major cell types were annotated based on the expression of canonical gene markers and assessment using the SingleR package (Fig. 1b–d, Fig. S2 and Supplementary Table 1), including epithelial cells expressing *KRT8*, *KRT14*, *KRT5*, *EPCAM*, *KRT18* and *AQP5*; endothelial cells expressing *PECAM1*, *ENG* and *CDH5*; fibroblasts expressing *COL3A1* and *DCN*; NK/T cells expressing *CD3D*, *CD3E*, *CD8A*, *NGK7* and *GNL*; B cells expressing *CD79A* and *CD79B*; plasma cells expressing *MZB1*; and myeloid cells expressing *CD68*, *CD74*, *CLEC9A* and *CD163*. Of note, we observed



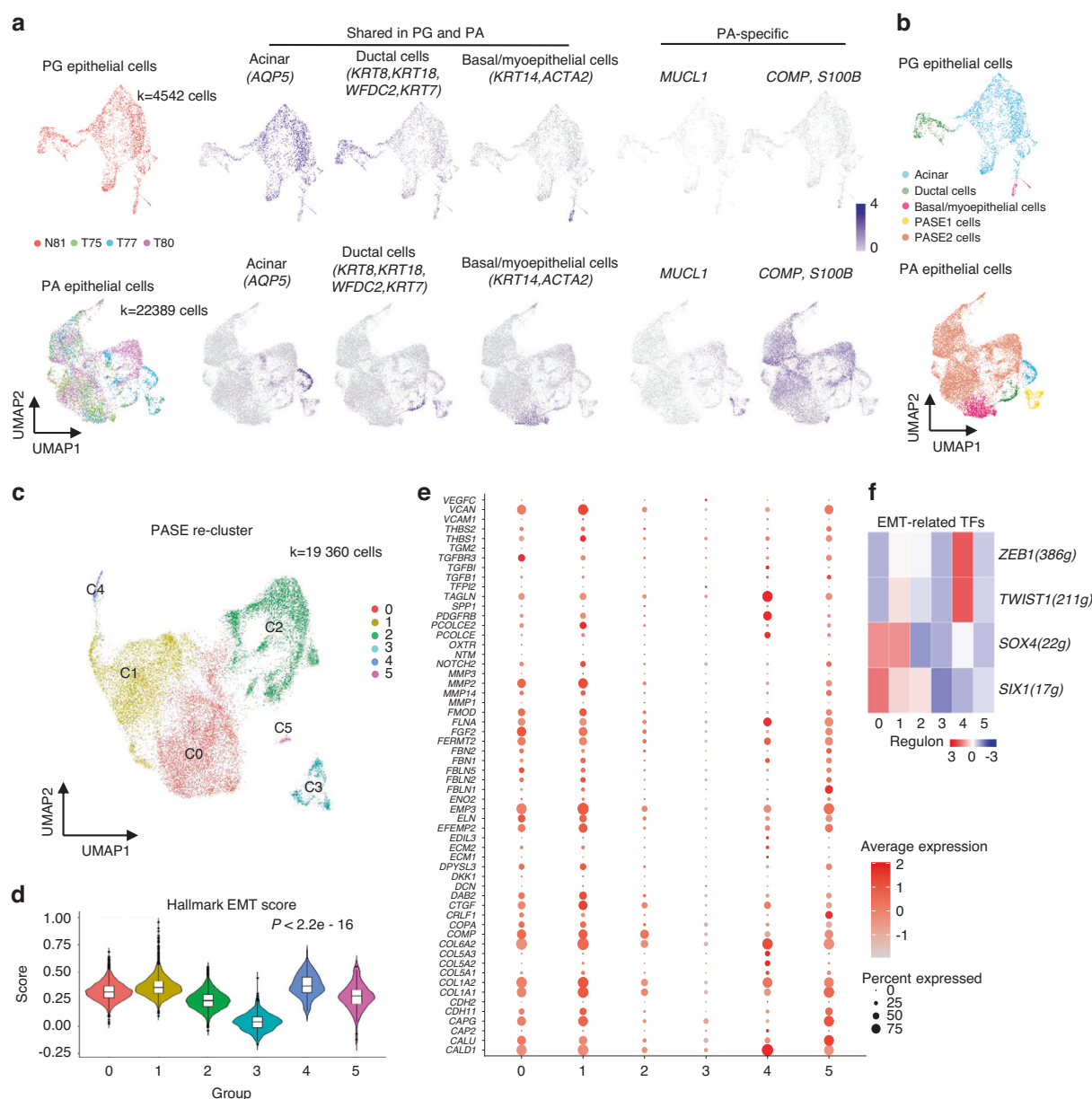
**Fig. 2** Molecular characteristics of PA and PG epithelial cells. **a** The UMAP distribution of 27 084 epithelial cells in different tissue types and samples. **b** Boxplot displaying the CNV score of epithelial cells in the four samples. **c** Chromosomal copy number variation values in epithelial cells of four samples, blue indicates deletion and red indicates amplification. **d** Boxplots showing the high differential expression of the representative genes *PRB1* and *CLDN3* in the PG sample ( $P < 2.2 \times 10^{-16}$ ). **e** Boxplots showing the high differential expression of the representative genes *PLAG1* and *WIF1* in the PA samples ( $P < 2.2 \times 10^{-16}$ ). **f** GO analysis of genes in PA epithelial cells ( $P < 0.05$ ,  $\log_2\text{FC} > 1$ ). **g** GO analysis of genes in PG epithelial cells ( $P < 0.05$ ,  $\log_2\text{FC} > 1$ ).

that each of the 7 clusters contained cells from different samples despite differing cell type proportions, indicating that the major cell types are largely consistent across the PA and PG (Fig. 1e).

#### Molecular characteristics of PA and PG epithelial cells

As shown in Fig. 2a, the UMAP showed a substantial difference in the distribution of epithelial cells from PG and PA. To characterize the PG and PA epithelial cells, we inferred copy-number variations (CNVs) in epithelial cells of each sample based on smoothed expression profiles across chromosomal intervals.<sup>11</sup> We found that PA epithelial cells exhibited remarkably higher CNV levels than PG epithelial cells (Fig. 2b). The inferred CNV data analysis revealed significant copy number amplification on chromosomes 1, 9, 19, 22, but significant deletions along chromosomes 4, 16, 18 and 20

were observed (Fig. 2c). Of note, dramatic alterations were observed on chromosome 12. Similar genomic alterations were also observed in previous studies<sup>12–14</sup> and several other adenogenic tumours, such as breast cancer and pancreatic ductal adenocarcinoma,<sup>15,16</sup> indicating that adenogenic tumours might share some common genomic alterations. Subsequently, a panel of marker genes were identified to be upregulated in PA and PG epithelial cells. As expected, high expression of marker genes related to the secretion of salivary gland and saliva enzymes was observed in PG epithelial cells, such as *PRB1*, *CLDN3*, *PRB3*, *SMR3B*, *HTN1* and *PIGR*,<sup>17,18</sup> while several oncogenes that have been confirmed to promote PA development were increased in PA cells, including *PLAG1*, *WIF1*, *S100B*, *CDK4*, *LIFR* and *NFIB* (Fig. 2d, e and Fig. S3a–c).<sup>14,19–23</sup> Gene ontology (GO) enrichment analysis was



**Fig. 3** PA-specific epithelial cell subpopulations. **a** The UMAP distribution of epithelial cells labelled by samples (left) and the expression of acinar cells, ductal cells, basal/myoepithelial cell genes (middle) as well as the expression of tumour-specific genes (right) between PG and PA epithelial cells. **b** UMAP distribution showing the cell types shared with PG and PA samples and PA-specific cells. **c** The UMAP distribution of PA-specific epithelial cell subsets. **d** EMT score based on the HALLMARK\_EPITHELIAL\_MESENCHYMAL\_TRANSITION gene set from GSEA datasets in different PA-specific epithelial cell subsets.  $P < 2.22 \times 10^{-16}$  by pairwise Wilcoxon rank-sum tests. **e** Dotplot showing EMT-related genes. **f** Heatmap showing the EMT-related transcription factors ZEB1, TWIST1, SOX4 and SIX1

performed based on the differentially expressed genes (DEGs) and revealed that matrix formation-related pathways were involved in the tumour epithelium, such as extracellular matrix organization and skeletal system development, suggesting that abnormal matrix formation occurred in the PA epithelium (Fig. 2f). In contrast, several pathways related to the physiological function of salivary glands were mainly enriched in the PG epithelium, including salivary secretion and detection of chemical stimuli involved in sensory perception of bitter taste (Fig. 2g)

#### PA-specific epithelial cell subpopulations

After reclustering PG epithelial cells, three major cell types were identified in the PG, including acinar cells expressing *AQP5*, ductal cells expressing *KRT18*, *WFDC2*, *KRT8*, and *KRT7* and basal

cells/myoepithelial cells expressing *KRT14* and *ACTA2*, which were consistent with the known functional cells within the salivary gland.<sup>24</sup> Similarly, reclustering of PA cells produced 5 major cell types and three of them were also observed in PG, including acinar cells (*AQP5*), ductal cells (*KRT18*, *KRT8*, *KRT7*, *WFDC2*) and basal/myoepithelial cells (*KRT14*, *ACTA2*), suggesting that PA tumour cells broadly recapitulated the PG cell subpopulations with shared several common marker genes across normal PG and PA (Fig. 3a). Interestingly, two distinguished epithelial cell clusters specific to PA were discovered, which were named the PA-specific epithelial cell (PASE) subpopulations and exhibited high expression of *MUCL1* (PASE1) and *COMP* (PASE2) (Fig. 3b). Strikingly, PASE subpopulations accounted for 86.5% of tumour cells in PA.



To further investigate the heterogeneity and molecular characteristics of PASE cells, we performed analysis of 19360 PASE cells at higher resolution, which yielded 6 prominent cell subpopulations with different transcriptional profiles (C0–C5), reflecting the heterogeneity of PASE cells (Fig. 3c). It has been suggested that epithelial-mesenchymal transition (EMT) represents the basic principle of tissue heterogeneity in PA.<sup>7–9</sup> Importantly, EMT has been suggested to be responsible for generating tumor-initiating cells in many tumors.<sup>25–27</sup> Then, we first detected the EMT state in these six PASE cell populations and found that C0, C1, C2, C4 and C5 cells all exhibited high EMT scores ( $n = 200$  genes,  $P < 2.2 \times 10^{-16}$ ) (Fig. 3d). Notably, C4 cells had the highest EMT score but C3 cells had the lowest EMT score. These results indicated C4 cells were the undifferentiated cells, C3 cells were highly differentiated, and C0, C1, C2 and C5 cells showed the transition or intermediate states. Furthermore, all PASE subpopulations except C3 cells expressed high levels of EMT-related genes, such as *COL1A1*, *COL1A2*, *COMP*, *MMP2* and *EMP3* (Fig. 3e). Interestingly, several master transcription factors of EMT were also increased in C0, C1, C2 and C4 cells, including *SOX4*, *SIX1*, *ZEB1* and *TWIST1* (Fig. 3f). These results suggest that EMT occurs widely and plays a critical role in PA development, which may contribute to the generation of TICs and the formation of stromal-like components, such as myxoid-like and chondroid-like components.

Then, the transcription states of PASE subpopulations were annotated according to well-known marker genes and functional enrichment analysis. As shown in Fig. 4a, b and Supplementary Table 2, we observed that C0 cells expressed acinar cell markers (*AQP5*, *AQP1*) and basal cell markers (*KRT5*, *KRT14*)<sup>28,29</sup> with properties of epithelium development and fluid secretion, indicating that C0 cells demonstrated a hybrid state. C1 cells showed high expression of *SFRP2* and *CTGF*, which have been reported to enhance osteogenic differentiation.<sup>30,31</sup> Functional enrichment showed that C1 cells were correlated with extracellular matrix organization, bone development and cartilage development, implying that C1 cells were chondroid-like components. C2 cells expressed *IGLC2* and exhibited immune and inflammation signatures, implying that these cells might trigger immune reactions and inflammatory responses. Based on the features mentioned above, we defined C0, C1 and C2 cells as mesenchymal-like cells due to high EMT scores and the characteristics of stromal cells. In addition, C3 cells expressed *MUC1* and the canonical ductal cell markers *KRT7*, *KRT18* and *KRT19* (Fig. 4c). Functional enrichment showed that C3 cells were characterized by metabolic reprogramming. C5 cells expressed high levels of proliferation-related genes (such as *MKI67* and *TOP2A*) and had the highest cell cycle score, which were also correlated with the cell cycle gene signatures (Fig. 4d, e). Notably, we observed that C4 cells exhibited high expression levels of myoepithelial cell markers (*ACTA2*),<sup>24</sup> myogenesis-related genes (*CASQ2*, *IGFBP7*, *MYL9*),<sup>32–34</sup> angiogenesis-related genes (*SPARCL1*, *ITGA7*, *CALD1*)<sup>35–37</sup> and stemness-related genes (*CD36*, *THY1*, *ITGA7*).<sup>38–43</sup> Functional enrichment showed that C4 cells were correlated with blood vessel development and muscle structure development, implying that C4 cells might have multidifferentiation potentials. However, C4 cells lacked expression of the classic stemness-related transcription factors, including *SOX2*, *BMI1*, *GLI1*, *NANOG*, and *OCT4* (Fig. 4f). Therefore, we performed single-cell regulatory network inference and clustering (SCENIC), which nominated *MAFB*, *LEF1* and *TBX2* as master TFs potentially controlling C4 cells (Fig. 4g), which have been confirmed to be associated with stemness and EMT.<sup>44–46</sup> These findings indicated that the C4 subpopulation was a unique myoepithelial cell population with tumour-initiating potential in PA. Collectively, we then defined C0–C5 cells as hybrid, chondroid, immune modulatory, *MUC1*<sup>+</sup> ductal, progenitor and cycling PASE cells, respectively (Fig. 4h).

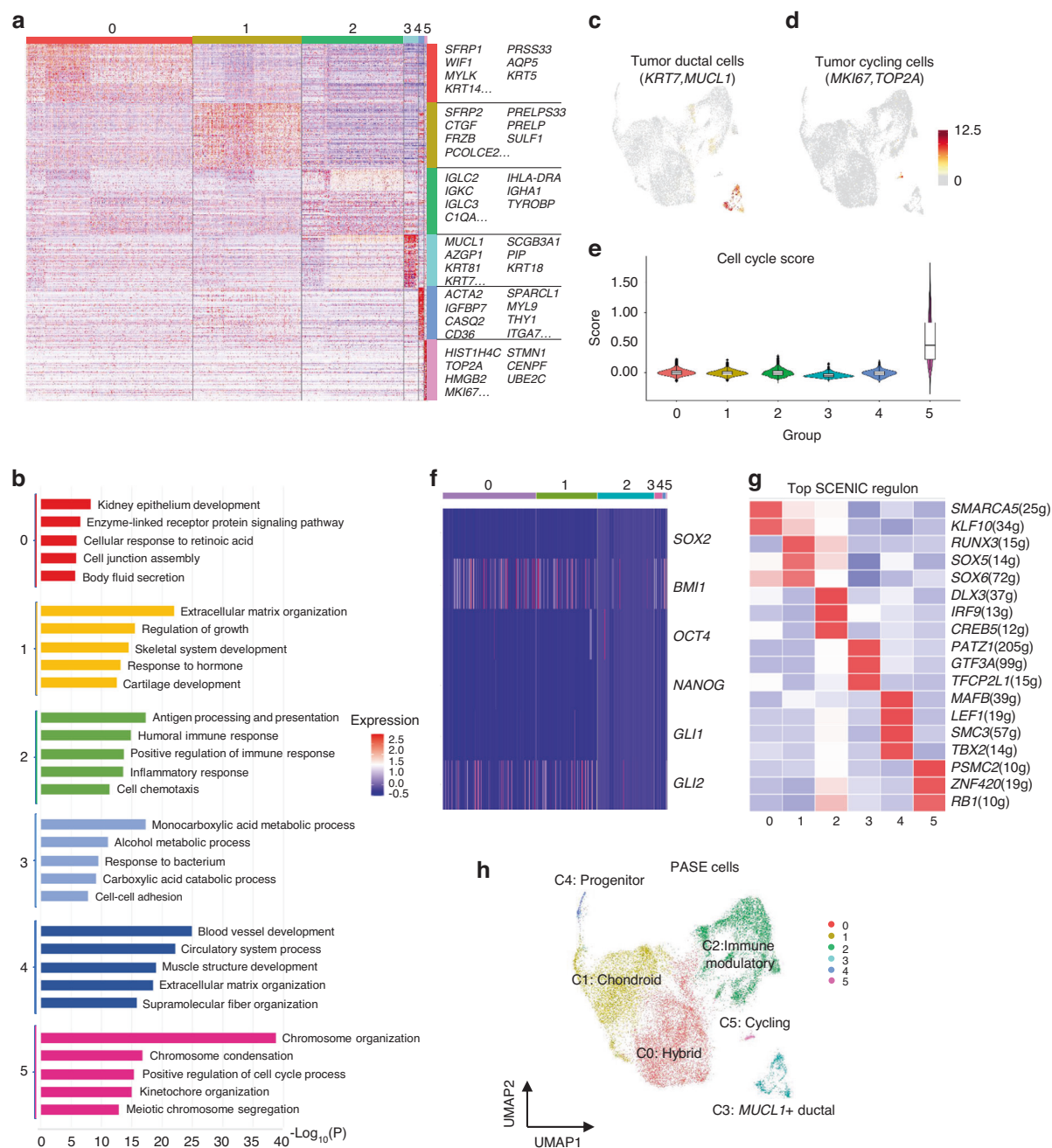
#### Tumour-initiating cells in PA

Histologically, PA is characterized by a biphasic architecture with both “epithelial” and “stromal” regions.<sup>47</sup> Currently, several studies have indicated that the epithelial and stromal elements are monoclonal and derived from the same precursor cells.<sup>5,6</sup> However, it is still not clear which cell is responsible for tumour initiation and produces the epithelial and stromal components in PA. We showed that C4 PASE cells might have tumour-initiating potential in PA, as mentioned above. To further investigate the origins and developmental trajectory of PA cells, we performed trajectory analysis using the top 100 DEGs of PASE cells. This unsupervised approach identified a continuum of cell states and showed two distinct trajectories beginning at state 2 (pre-branch) and gradually branching to state 1 and state 3 corresponding two distinct cell fates (cell-fate 1 and cell-fate 2), revealing a common origin with divergent fates (Fig. 5a, b). In the pseudotime trajectory, we investigated the distribution pattern of the six PASE subtypes. As shown in Fig. 5c and Fig. S4a, C4 cells were predominantly observed in the end of state 2 with the lowest pseudo-time position and C3 cells were mostly found in the end of state 1 with the highest pseudo-time position, supporting our previous notion that C4 cells were the progenitor cells of PA and C3 cells were the highly differentiated ductal cells. C0, C1, C2 and C5 cells were scattered along the trajectories, reflecting their transition or intermediate states. Notably, the highly expressed gene markers of C4 (*ACTA2*, *IGFBP7*, *THY1*) were located at the low pseudotime position, further indicating that C4 cells were the progenitor cells in PA. The markers that represented the differentiated epithelial cells (*KRT19*, *EPCAM*, *KRT18*) and mesenchymal-like cells (*CNMD*, *COMP*, *SOX9*) were also located at the appropriate pseudotime position (Fig. S4b–d). Moreover, C4 cells exhibited high expression of the stemness-related gene signature and tumorigenesis-related gene signature by referring to gene sets that have been reported and included in GSEA<sup>48–50</sup> (Fig. 5d–f). Taken together, these results suggested that C4 cells were the tumor-initiating cells in PA.

To study potential changes in global expression dynamics along the trajectory, we ordered the genes expression by pseudotime and conducted the enrichment analysis to investigate the precise impact of the alterations in cell fate. We found that there were 6 gene expression patterns (P1–P6) accounting for the distinctions (Fig. 5g, Supplementary Table 3). Cells undergoing cell-fate1 expressed highly homeostatic associated molecules enriched for the salivary gland function, including salivary secretion and monocarboxylic acid metabolic process (P1, P3). Cells undergoing cell-fate2 expressed high level of genes enriched for skeletal system development, extracellular matrix organization, inflammatory response and cytokine signaling in immune system (P2, P6), which were consistent with the signatures of mesenchymal-like cells mentioned above (Fig. 5g, Fig. S5). These functional enrichments reflected the cell components of PA in the two developmental routes, supporting the notion that TICs of PA could develop into the highly differentiated epithelial cells to form the duct structures and the mesenchymal-like cells with transitional states to form the stroma, including myxoid, chondroid or myxochondroid.

*CD36*<sup>+</sup> was a functional surface marker of tumour-initiating cells in PA

To further identify and isolate TICs of PA, C4 cell-specific markers were screened. Notably, we observed that *CD36*, a cell surface marker, was specifically expressed in PA epithelial cells (Fig. 6a, b) and involved in the regulation of stemness,<sup>39,40</sup> indicating that *CD36* might be a functional cell surface marker for the enrichment of TICs in PA. As shown in Fig. 6c, we further confirmed that *CD36* was mainly expressed in myoepithelial tumour cells in PA tissue by using Pan-Keratin (PCK), *ACTA2* and *CD36* multiple immunostainings. Then, primary PA epithelial



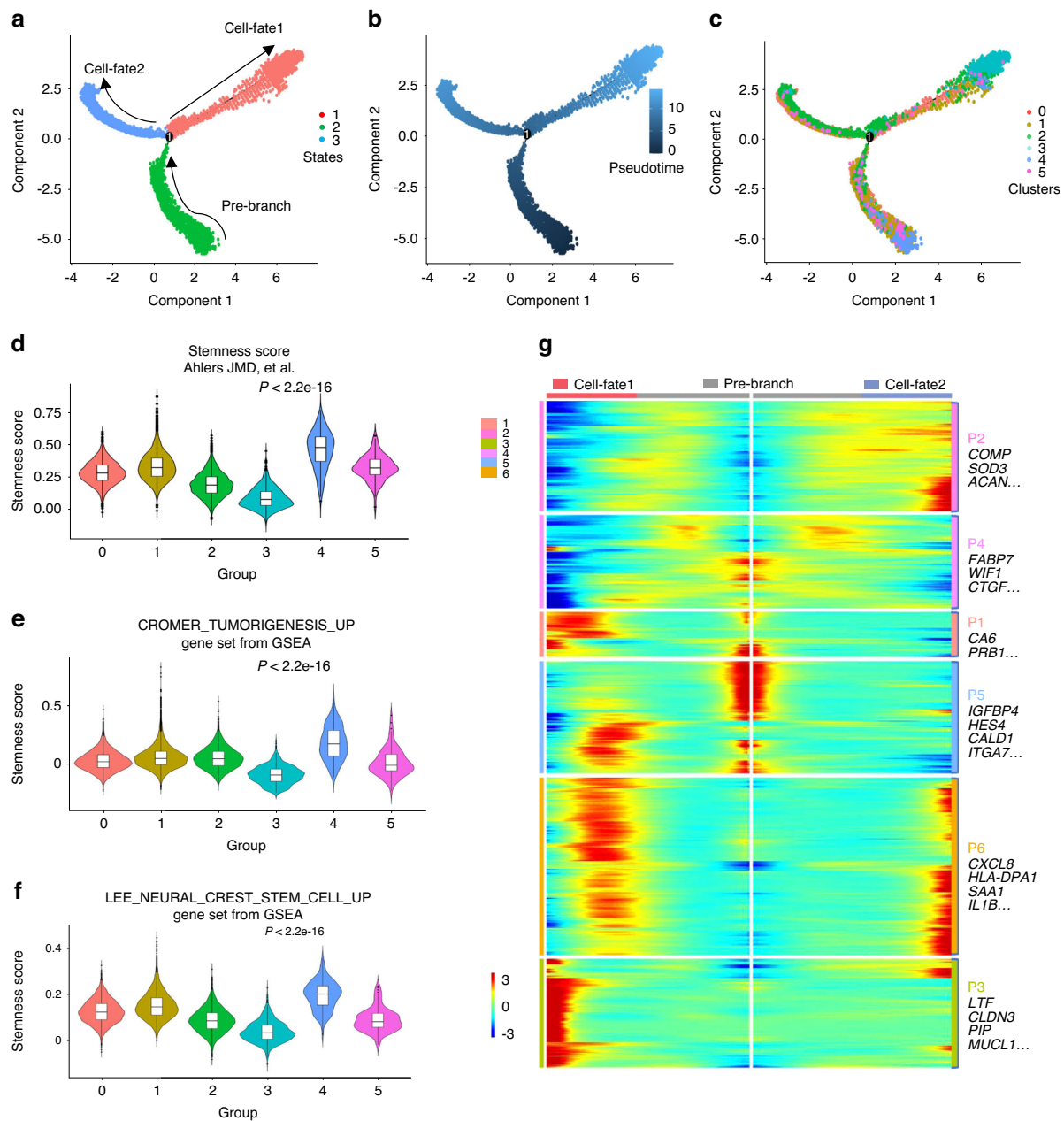
**Fig. 4** Transcriptional landscape heterogeneity of PA-specific epithelial cell subsets. **a** Heatmap of marker genes for PA-specific epithelial cell subsets. **b** GO enrichment analysis of PA specific epithelial cell subsets. **c** UMAP feature plots of expressed genes in tumour-specific ductal cells. **d** UMAP feature plots of expressed genes in cell cycle cells. **e** Cell cycle scores of PA specific epithelial cell subsets. **f** Heatmap showing stemness-related genes *SOX2*, *BMI1*, *NANOG*, *OCT4*, *GLI1* and *GLI2*. **g** Heatmap showing the top regulon of each cell cluster inferred by SCENIC. **h** The UMAP distribution of PA-specific epithelial cell subsets

cells from different patients were freshly isolated and cultured, which were further confirmed using PCK staining (Fig. 6d). Next,  $CD36^+$  and  $CD36^-$  PA epithelial cells were sorted using fluorescence-activated cell sorting (FACS) (Fig. 6e). To evaluate the tumorigenic potential of sorted cells, a tumour sphere formation assay was performed using the collected  $CD36^+$  and  $CD36^-$  PA cells. As shown in Fig. 6f,  $CD36^+$  PA epithelial cells were able to form more and larger spheres than those generated from  $CD36^-$  PA epithelial cells. More importantly, we performed RNA-seq by using  $CD36^+$  and  $CD36^-$  cells isolated from primary PA cells. GSEA results showed that the genes upregulated in  $CD36^+$  cells were enriched for the

regulation of stemness, supporting that  $CD36$  was a cell surface marker for the enrichment of TICs in PA (Fig. S6a, b).

Targeting the PI3K-AKT pathway eliminated tumour-initiating cells and suppressed tumour growth in PA

Next, to clarify the key signalling pathway controlling TICs in PA, we performed Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis based on our scRNA-seq data, which indicated that PI3K-AKT signalling pathway was activated in C4 cells (Fig. 7a). Similar results were also observed in our RNA-seq data generated from  $CD36^+$  and  $CD36^-$  PA cells (Fig. 7b). Then, two small molecular inhibitors MK2206 and GDC0068 were used to validate the functional role of



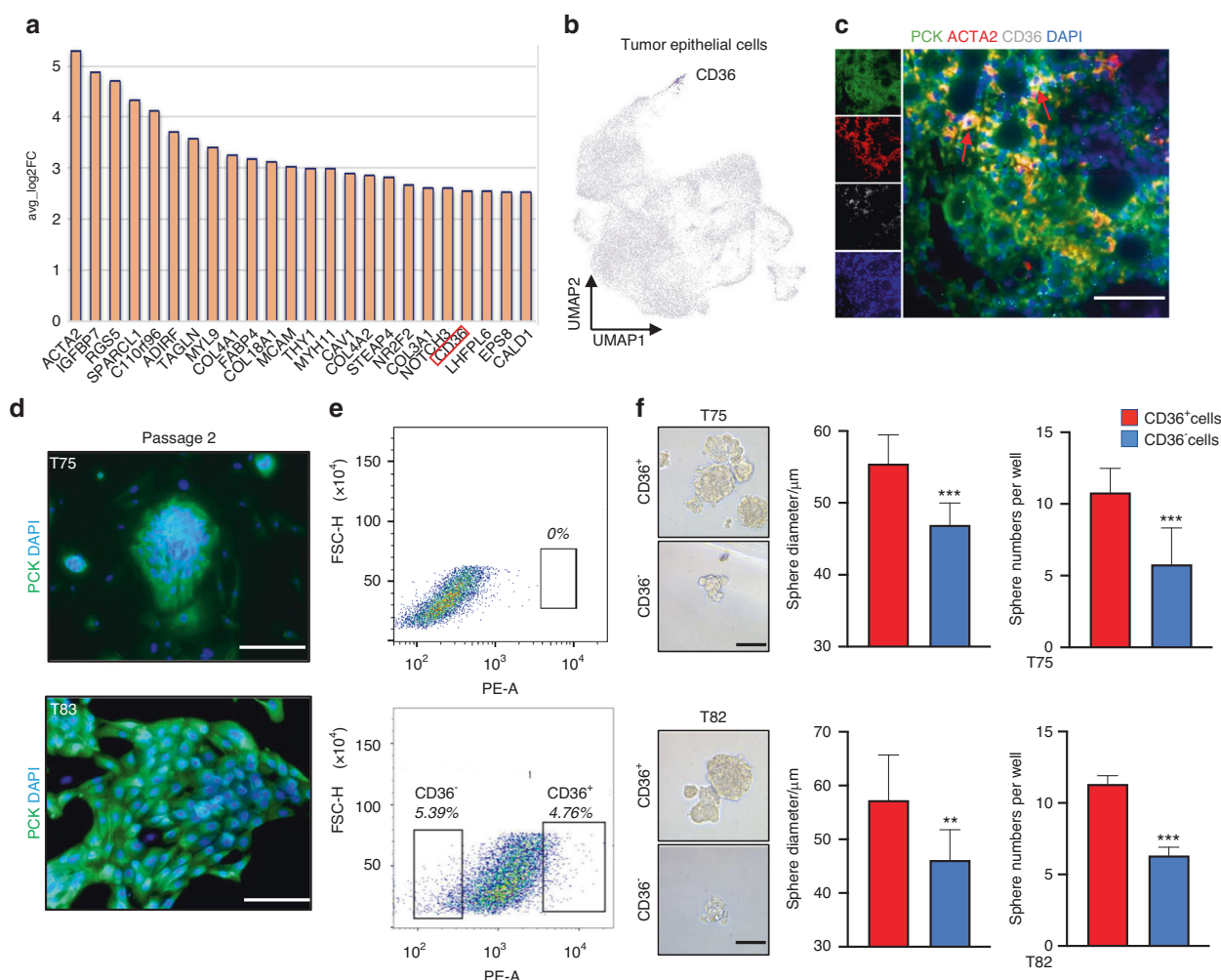
**Fig. 5** Tumour-initiating cells in PA. **a** The pseudotime trajectory is shown coloured by cell state designation. **b** Pseudotime is shown coloured in a gradient from dark to light blue and the start of pseudotime is indicated. **c** The pseudotime trajectory is shown coloured by cell subset designation. **d** The violin plot of stem cell score based on the published article<sup>50</sup> (Supplementary Table 4) in different PA specific epithelial cell subsets (log2FC > 0.5).  $P < 2.2e-16$  by pairwise Wilcoxon rank-sum tests. **e** Violin plot of stem cell score based on the CROMER\_TUMORIGENESIS\_UP gene set from GSEA datasets in different PA-specific epithelial cell subsets.<sup>48</sup>  $P < 2.2e-16$  by pairwise Wilcoxon rank-sum tests. **f** Violin plot of stem cell scores based on the LEE\_NEURAL\_CRESCENT\_STEM\_CELL\_UP gene set from GSEA datasets in different PA specific epithelial cell subsets.<sup>49</sup>  $P < 2.2e-16$  by pairwise Wilcoxon rank-sum tests. **g** Heatmap showing the ordering gene expression dynamics during the cellular-fate transition process and expression dynamics

the PI3K-AKT pathway in PA tumorigenesis. As shown in Fig. S7a, inactivation of the PI3K-AKT pathway with MK2206 and GDC0068 dramatically inhibited sphere formation generated from CD36<sup>+</sup> PA cells. These findings suggested that targeting the PI3K-AKT pathway may be an effective therapeutic strategy for the treatment of PA. To further evaluate the therapeutic value of targeting the PI3K-AKT pathway in PA, PA patient-derived organoids (PPDOs) were cultured and confirmed by H&E staining and double staining of PCK and CD36 (Fig. 7c, d). As expected, we observed that pharmacological inhibition of the PI3K-AKT pathway significantly impaired the formation and growth of PPDOs (Fig. 7e).

## DISCUSSION

The complex cytomorphological and architectural features of PA have been well recognized in past decades. However, the landscapes of cell constituents and their genetic heterogeneities are still mysterious and have never been documented. In the present study, we employed scRNA-seq technology to generate a comprehensive gene expression atlas of PA and identify seven major cell types with UMAP clustering that contribute to PA lesions. Notably, we showed that PA epithelial cells were dramatically different from their normal counterparts. High copy number variations were observed in PA epithelial cells compared





**Fig. 6** CD36<sup>+</sup> is a functional surface marker of tumour-initiating cells in PA. **a** High expression genes of the C4 cluster ( $P < 0.05$ , avg\_log2FC > 2.5). **b** The UMAP distribution of CD36 gene expression in PA epithelial cell subsets. **c** Representative images of multicolor immunostaining showing the expression of PCK, ACTA2, CD36 in PA sections; scale bar = 50 μm. **d** Immunofluorescence of PCK in PA epithelial cells after two passages culture; scale bar = 50 μm. **e** Flow cytometry analysis of CD36<sup>+</sup> PA cells and CD36<sup>-</sup> PA cells. **f** Spheroids cultured from CD36<sup>+</sup> PA cells and CD36<sup>-</sup> PA cells, scale bar = 50 μm (left). Quantification results are shown. \*\* $P < 0.01$  and \*\*\* $P < 0.001$  by Student's  $t$  test (right)

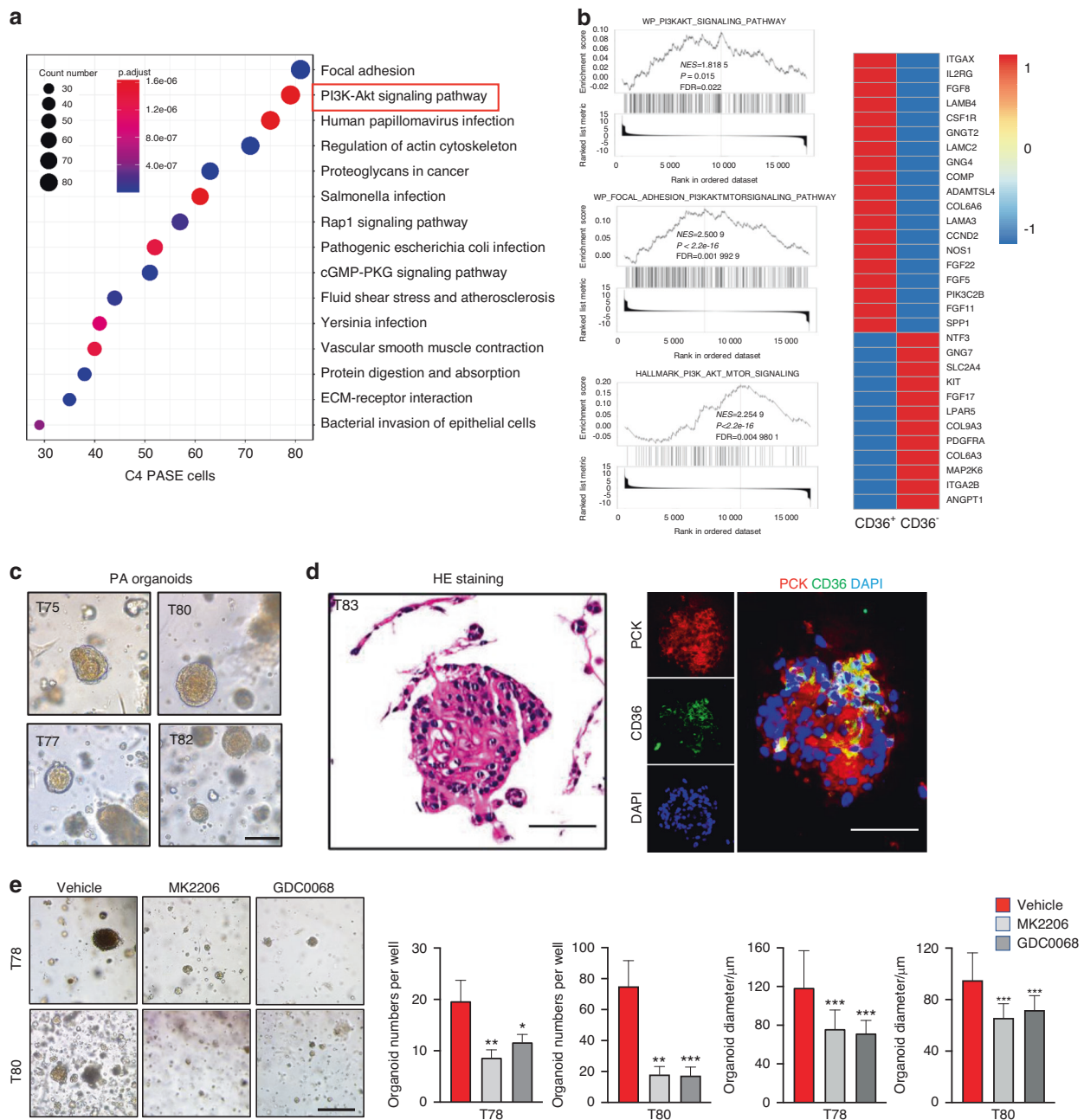
to normal salivary epithelial cells despite the benign nature of PA. Obvious CNVs were detected on chromosomes 1, 4, 9, 12, 19 and 22, implying that the occurrence of PA results from multiple genomic alterations. It has been confirmed that PA is characterized by recurrent chromosome rearrangements, particularly translocations with breakpoints at 8q12 and 12q13-15, which activate the overexpression of *PLAG1* and *HMG2*.<sup>14,51</sup> In our study, the predominant copy number alterations were observed in chromosome 12, in which some key genes driving tumorigenesis of PA were located, including *HMG2*, *WIF1*, *MGP* and *MUCL1*. These findings suggest that the alteration of chromosome 12 may be the genomic hallmark of PA.

Then, we characterized the features of gene expression profiles in PA epithelial cells and revealed intratumour heterogeneity and diversity. Consistent with a previous study in cutaneous squamous cell carcinoma,<sup>52</sup> we observed that several PA tumour subpopulations (acinar, ductal and basal/myoepithelial cells) were also similar to their normal counterparts, which shared several marker genes across normal PG and PA, suggesting that PA might recapitulate stages of PG development. Strikingly, two PA subpopulations unique to tumour were identified, which accounted for most of the tumour cells in PA. Then, we identified six subgroups (C0–C5) with different transcriptomic characteristics

in tumour specific epithelial cells. C0, C1, C2, C4 and C5 cells were all characterized with properties of EMT but exhibited different differentiation degrees and proliferative activities, indicating that EMT was widely activated during the tumorigenesis of PA. C0 cells were associated with epithelial development, C1 cells were correlated with chondrogenesis and osteogenesis, and C2 cells displayed dual characteristics of epithelial cells and immune cells, revealing a high degree of tumor heterogeneity. In addition, C3 was a ductal cell population with a low EMT score and high expression of *MUCL1* gene, which demonstrated obvious epithelial features and might form the luminal structures of PA. Increasing studies have confirmed that *MUCL1* functions as a unique oncogene to promote proliferation and metastasis in several adenogenous malignancies, including breast cancer,<sup>53</sup> gastric adenocarcinoma,<sup>54</sup> and colorectal adenocarcinoma,<sup>55</sup> indicating that the *MUCL1* gene may be a potential marker to evaluate the aggressiveness of PA. However, further studies are required to examine the expression of *MUCL1* and *MUCL1*<sup>+</sup> ductal cell subpopulations in PA.

It has been suggested that TICs or tumour stem cells are responsible for tumour recurrence, metastatic spread and resistance to treatment.<sup>56–58</sup> Notably, we revealed that the C4 cluster housed the TICs based on single-cell algorithms rather than





**Fig. 7** Targeting the PI3K-AKT pathway eliminated tumour-initiating cells and suppressed tumour growth in PA. **a** KEGG enrichment scatter plot of C4. The y-axis represents the name of the pathway, and the x-axis represents the rich factor, and the degree of KEGG pathway enrichment. The dot size represents the number of genes and the colour indicates the  $q$ -value. **b** GSEA showing that the PI3K-AKT pathway signatures were significantly enriched in CD36<sup>+</sup> cells compared with CD36<sup>-</sup> cells. The top DEGs (log2FC > 1) of these GSEA terms are presented in the heatmap. **c** Representative images showing patient organoids derived from different patient pleomorphic adenomas; scale bar = 100  $\mu$ m. **d** Haematoxylin and eosin (H&E) staining of organoid sections; scale bar = 100  $\mu$ m (left). Representative immunofluorescence images showing the expression of PCK and CD36 in PA organoids; scale bar = 100  $\mu$ m (right). **e** The suppressive effect of AKT inhibitors (MK2206, 10  $\mu$ mol·L<sup>-1</sup> and GDC0068, 2.5  $\mu$ mol·L<sup>-1</sup>) in the organoid assay, scale bar = 100  $\mu$ m (left). Quantification results are shown. \* $P$  < 0.05, \*\* $P$  < 0.01 and \*\*\* $P$  < 0.001 by one-way ANOVA (right)

relying on predefined surface markers, which often target unrelated cells.<sup>59,60</sup> Then, we confirmed that the cells within C4 showing high expression levels of CD36 were the tumour stem cells resulting in the formation of PA. Similar results were also observed in glioblastoma,<sup>39</sup> oral carcinomas<sup>40</sup> and leukaemia.<sup>61</sup> Next, we further revealed that the PI3K-AKT pathway was activated in C4 cells and CD36<sup>+</sup> tumour cells, and inhibition of PI3K-AKT signalling suppressed PA tumorsphere formation and organoid growth. Currently, accumulating evidence has confirmed

that the PI3K-AKT pathway plays an important role in cell mitosis, proliferation and apoptosis, especially in maintaining the pluripotency of mesenchymal stem cells and embryonic stem cells.<sup>62,63</sup> Importantly, previous studies have shown that CD36 promotes tumour progression by activating PI3K/AKT in several cancer types.<sup>64–67</sup> These findings support the notion that TICs of PA are controlled by PI3K-AKT signalling.

Taken together, our findings provide the first resource for deciphering comprehensive gene expression landscapes of PA

and reveal the intratumoral heterogeneity and origin of PA at a single-cell resolution despite the small number of patients enrolled. Furthermore, we identified a unique myoepithelial cell population with high expression of CD36 and activation of PI3K-AKT signalling that contributes to tumorigenesis of PA, which provides a novel therapeutic target and lays a new foundation for the development of precision therapies in PA.

## MATERIALS AND METHODS

### Salivary tumour specimens

In compliance with all relevant ethical regulations, we collected fresh pleomorphic adenoma and adjacent nontumour tissue from patients at the Hospital of Stomatology, Sun Yat-sen University. Briefly, freshly harvested tissues were mechanically and enzymatically digested into single cell suspensions with gentleMACS (Miltenyi) according to the instructions of a tumour dissociation kit (Cat #130-095-929, Miltenyi Biotec, Bergisch Gladbach, Germany). Then, the dissociated cells were filtered with a 70- $\mu$ m SmartStrainer and centrifuged at  $400 \times g$  for 5 min to remove large pieces of debris. After the supernatant was removed, the pelleted cells were incubated with red blood cell lysis buffer (Thermo Fisher Scientific, Carlsbad, CA) for 1–2 min to lyse red blood cells. Finally, the cells were washed twice with 1 $\times$ PBS (GIBCO), and the cell pellets were counted and tested for cell viability using trypan blue staining.

### Tissue processing

The remaining tissues were fixed in 4% paraformaldehyde (Biosharp, BL539A) for 48 h at 4°C overnight. Dehydration and embedding in paraffin were performed using routine methods. Paraffin sections (5  $\mu$ m) were cut and adhered to glass slides. Then, the paraffin sections were placed at the room temperature.

### Haematoxylin and eosin (H&E) staining

The paraffin sections were placed in a 70°C oven for 15 min before deparaffinization in xylene and successively hydrated in 100%, 90%, 80% and 70% alcohol. H&E staining was performed according to a standard protocol (Solarbio, G1120-100), and stained slides were dehydrated in ethanol and sealed in neutral resin.

### Immunofluorescence staining

Formalin-fixed paraffin-embedded sections of PA specimens were collected at the Hospital of Stomatology, Sun Yat-sen University. To detect the protein expression levels of marker genes, the PA paraffin sections were stained four-color-multilabeled immunofluorescence staining kit (Absin, Cat#abs50012) according to the manufacturer's protocols. Briefly, the sections were incubated in 3% H<sub>2</sub>O<sub>2</sub> for 10 mins for the first time. Every time the sections were incubated with antibodies against CD36 (ab17044, Abcam), Pan-keratin (PCK) (26411-AP, Proteintech), ACTA2 (ab220179, Abcam). After each incubation of the primary antibody, heat-induced epitope recovery and 5% BSA blocking were performed. The HRP conjugate and three wavelengths (520, 570 and 650 nm) were utilized to attach the different primary antibodies. Then, the slides were counterstained with DAPI for nuclear visualization, and subsequently coverslipped with fluorescence quenching mounting medium. Images were acquired using a fluorescence microscope. The organoid immunofluorescence staining was performed as described previously with antibodies against CD36 and PCK<sup>68,69</sup>. All primary antibodies were used at a dilution of 1:200.

### 10x library preparation and sequencing

According to the standard manufacturer's protocol, single cells were resuspended in DMEM buffer at 1 000 cells per  $\mu$ L and loaded onto the Chromium chips. All the remaining procedures including capturing, barcoding and cDNA library preparation were

carried out using the Chromium Single Cell 3'Library v3 chemistry (10x Genomics).

### Single cell RNA-seq data processing

Sequenced reads were aligned and quantified using the Cell Ranger 4.0.0 pipeline against the GRCh38 human reference genome. Next, by counting unique molecular identifiers (UMIs) and removing low quality barcodes, a gene barcode matrix containing the barcoded cells and gene expression counts was generated. The expression matrix passed the quality control based on three metrics step by step, including the total cell count, number of detected genes and the proportion of mitochondrial gene count per cell, by using the Seurat R package (v4.0)<sup>70,71</sup> to generate Seurat Objects for downstream analyses. High quality cells ( $200 < \text{nFeature\_RNA} < 9000$ ,  $\text{nCount\_RNA} > 1000$  and percent.mt < 10) were included. Then, the following Seurat functions were performed on remaining 35662 cells using the scan and scater packages in R.<sup>72</sup>

### Integration of multiple scRNA-seq datasets

We ran the CCA<sup>+</sup> anchors, an algorithm originating from Seurat v3.0 that can identify the shared cell state among different scRNA-seq datasets to remove the batch effects of 4 samples. The results from CCA integration and batch correction were used as input data for highly-variable gene identification and dimension reduction.

### Dimension reduction and unsupervised clustering

Single-cell data were processed for dimension reduction and unsupervised clustering by following the workflow in Seurat v3.0. In brief, 2 000 highly variable genes were summarized by principal component analysis (PCA) and a total of 20 principal components were selected for dimensional reduction using the default settings of the RunPCA function. Finally, the dimensionality of each dataset was further reduced using uniform manifold approximation and projection (UMAP) for visualization.

### Major cell types determination and marker genes identification

Doublets with mix features were deleted from further analysis using the DoubletFinder (v2.0.3) function. The remaining cells were used for the downstream analysis. We determined cell types by using Seurat package in combination with the SingleR (v1.4.1) package. The specific gene markers were identified using the FindAllMarkers function. Cell clusters were annotated as biological cell types according to canonical marker genes. To refine the classification of various cell types, we further compared our cell-type annotation to the reference by SingleR (v1.4.1).

### Subclustering of some major cell types

To identify subclusters with some major cell types, we reclustered cells belonging to epithelial cells, myeloid cells and NK/T cells separately using the workflow in Seurat v3.0. Applying the graph-based clustering approach, with a unique resolution and other default parameters, they were reclustered by its principal components. For visualization purposes, these informative principal components were converted into UMAP plots. To further investigate the characteristics of various clusters, cluster marker genes were detected using the FindMarkers function.

### CNV estimation based on scRNA-seq data

Initial CNVs for tumour epithelial cells of each sample were calculated by the inferCNV R package.<sup>73</sup> For each sample, the CNV from the single-cell RNA-seq dataset was explored by the expression intensity of genes across positions of the genome compared with a set of reference cells. We set 1 000 cells from normal samples for reference, and 1 000 cells each from other tumour samples for observations and ran inferCNV with cut-off = 0.1. The expected output is a heatmap of observed

expression relative to the reference, shedding light on regions of chromosomal gain and loss.

#### EMT score, cell cycle score and stemness score

According to the reference gene sets, subsets were scored using the Seurat function `AddModuleScore`. We calculated the mean abundance levels of the subset function-related mRNA genes against the abundance of the control gene sets as the score of each epithelial subset.

#### Delineation of cell differentiation trajectories

To uncover the cell state transition, the Monocle package<sup>74</sup> was applied to depict single cell trajectories. Briefly, the Seurat Object was converted into the Monocle cell dataset using the 'importCDs' function. The top 100 differentially expressed genes in the subclusters of epithelial cells were selected using the differential-GeneTest function in pseudotime order. We then applied 'DDRtree' to reduce dimensions and the visualization functions 'plot\_cell\_trajectory' to plot the minimum spanning tree on cells.

#### Analysis of transcription factor expression

To assess the transcriptional activity of tumour-specific epithelial cells, single-cell regulatory network inference and clustering (SCENIC) analysis was used to map the gene regulatory network and identify stable cellular states by evaluating the activity of GRNs (Gene Regulatory Networks) in each cell. The analysis used the 20-thousand motifs database for RisTarget and GRNboost (corresponding to GENIE3 1.4.3, AUCell 1.4.1 and RisTarget 1.2.1; with hg19.motifDatabases.20k).<sup>75</sup> The input matrix was normalized expression matrix, output from Seurat.

#### Metascape analysis GO enrichment

We performed the analysis via website (<https://metascape.org/gp>). Gene sets were input into the Metascape database for GO enrichment analysis.

#### Flow cytometry

Fresh pleomorphic adenoma samples obtained from patients were digested with collagenase IV (Thermo Fisher Scientific, 17104019) and a single-cell suspension was acquired. Then, we cultured the cells in 10 cm petri dishes with common medium. After two passages, tumour epithelial cells were purified. Next, we obtained a single-cell suspension. Cells were sorted by PE anti-human CD36 antibodies (Biolegend, Cat#336205) with FACS Aria I (BD Biosciences, Germany).

#### Sphere culture

The sorted cells were seeded at 5 000 cells per  $\mu\text{L}$  in low-attachment 96-well plate which added into DMEM/F12 (Thermo Fisher Scientific, Cat#C11330500BT), supplemented with 20 ng·mL<sup>-1</sup> human EGF (PeproTech, Cat#AF-100-15), 20 ng· $\mu\text{L}^{-1}$  FGF2 (Sino Biological Inc, Cat#10014-HNAE) and 1×B27 supplement (Thermo Fisher Scientific, Cat#12587010). Cells were cultured for 7 days before harvesting. Images were acquired using an inverted microscope.

#### Organoid culture

For organoid culture, the tumour suspensions were seeded at  $1 \times 10^4$  cells per well in 200  $\mu\text{L}$  Matrigel (Corning, 354234) in 24-well plates. After polymerization, organoids were cultured in self-configured medium containing DMEM/F12 (Thermo Fisher Scientific, Cat#C11330500BT), 1×B27 supplement (Thermo Fisher Scientific, Cat#12587010), 1.25 mmol·L<sup>-1</sup> N-acetyl-L-cysteine (Sigma, Cat#A7250), 10 mmol·L<sup>-1</sup> nicotinamide (Sigma, Cat#N0636), 50 ng·mL<sup>-1</sup> human EGF (PeproTech, Cat#AF-100-15), 500 nmol·L<sup>-1</sup> A83-01 (PeproTech, Cat#9094360), 10 ng·mL<sup>-1</sup> human FGF10 (PeproTech, Cat#100-26-5), 5 ng·mL<sup>-1</sup> human FGF2 (Sino Biological

Inc, Cat#10014-HNAE), 1  $\mu\text{mol} \cdot \text{L}^{-1}$  prostaglandin E2 (MCE, Cat#HY-101952), 0.3  $\mu\text{mol} \cdot \text{L}^{-1}$  CHIR 99021 (Sigma, Cat#SML1046), 1  $\mu\text{mol} \cdot \text{L}^{-1}$  forskolin (Abcam, Cat#ab120058), 50 ng·mL<sup>-1</sup> R-spondin (R&D Systems Cat#3266-RS) and 25 ng·mL<sup>-1</sup> Noggin (PeproTech, Cat#120-10 C). The medium was changed twice a week.

#### Statistical analysis

Data are shown as the mean  $\pm$  SD. Statistical analyses were conducted by using Prism 8.0 (Graphpad Software). Statistical comparisons were performed using Student's test or one-way ANOVA.  $P < 0.05$  was considered to be statistically significant.

#### DATA AVAILABILITY

The raw sequencing data of human PA tissue and cells have been deposited in GSA-Human under accession number HRA003970, including scRNA-seq data and RNA-seq data. All the other data associated with this study are available in the article and its supplementary information files and from the corresponding author upon reasonable request.

#### ACKNOWLEDGEMENTS

This project was financially supported by the National Natural Science Foundation of China (82141112, 82073265). Guangdong Financial Fund for High-Caliber Hospital Construction (174-2018-XMZC-0001-03-0125/D-14) to C.W.

#### AUTHOR CONTRIBUTIONS

C.W. and D.C. conceived and supervised the study. X.X. and J.X. performed most experiments. Y.L., S.O., X.L., B.Y., G.X., M.Z. and W.W. were engaged in sample collection and preparation. X.X. and R.L. contributed to bioinformatics and statistical analysis. X.X. drafted the manuscript. All authors approved the final manuscript.

#### ADDITIONAL INFORMATION

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41368-023-00243-2>.

**Competing interests:** The authors declare no competing interests.

#### REFERENCES

- Califano, J. & Eisele, D. W. Benign salivary gland neoplasms. *Otolaryngol Clin North Am* **32**, 861–873 (1999).
- Xu, W. et al. Recurrent pleomorphic adenoma of the parotid gland: a population-based study with emphasis on re-recurrence and malignant transformation. *Head Neck*, <https://doi.org/10.1002/hed.27286> (2022).
- Gomes, C. C., Galvao, C. F., do Carmo, A. C., Pereira, N. B. & Gomez, R. S. Intratumor molecular heterogeneity in pleomorphic adenoma of the salivary glands. *Oral Surg. Oral Med. Oral Pathol. Oral Radiol.* **121**, 158–163 (2016).
- Kandasamy, J., Smith, A., Diaz, S., Rose, B. & O'Brien, C. Heterogeneity of PLAG1 gene rearrangements in pleomorphic adenoma. *Cancer Genet Cytogenet* **177**, 1–5 (2007).
- Noguchi, S. et al. Demonstration of monoclonal origin of human parotid gland pleomorphic adenoma. *Cancer* **77**, 431–435 (1996).
- Lee, P. S., Sabbath-Solitare, M., Redondo, T. C. & Ongcapin, E. H. Molecular evidence that the stromal and epithelial cells in pleomorphic adenomas of salivary gland arise from the same origin: Clonal analysis using human androgen receptor gene (HUMARA) assay. *Human Pathol.* **31**, 498–503 (2000).
- Kim, H. et al. SLUG is a key regulator of epithelial-mesenchymal transition in pleomorphic adenoma. *Lab Invest* **102**, 631–640 (2022).
- Matsumiya-Matsumoto, Y., Morita, Y. & Uzawa, N. Pleomorphic adenoma of the salivary glands and epithelial-mesenchymal transition. *J Clin Med.* **11**, <https://doi.org/10.3390/jcm11144210> (2022).
- Koutlas, I. G., Dolan, M., Lingen, M. W. & Argyris, P. P. Plasmacytoid cells in salivary pleomorphic adenoma: an alternative interpretation of their immunohistochemical characteristics highlights function and capability for epithelial-mesenchymal transition. *Oral Surg. Oral Med. Oral Pathol. Oral Radiol.* **128**, 515–529 (2019).
- Devi, A. et al. Potential immunohistochemical markers to characterize epithelial-mesenchymal transition in pleomorphic adenoma. *J. Exp. Ther. Oncol.* **13**, 1–7 (2019).



11. Zhang, M. et al. Dissecting transcriptional heterogeneity in primary gastric adenocarcinoma by single cell RNA sequencing. *Gut* **70**, 464–475 (2021).
12. Mariano, F. V. et al. Somatic copy number alterations in pleomorphic adenoma and recurrent pleomorphic adenoma. *Oral Surg. Oral Med. Oral Pathol. Oral Radiol.* **129**, 59–64 (2020).
13. Mariano, F. V. et al. Genomic copy number alterations of primary and secondary metastasizing pleomorphic adenomas. *Histopathology* **67**, 410–415 (2015).
14. Stenman, G. et al. Chromosome translocations, gene fusions, and their molecular consequences in pleomorphic salivary gland adenomas. *Biomedicines* **10**, <https://doi.org/10.3390/biomedicines10081970> (2022).
15. Xu, Q., Chen, S., Hu, Y. & Huang, W. Single-cell RNA transcriptome reveals the intra-tumoral heterogeneity and regulators underlying tumor progression in metastatic pancreatic ductal adenocarcinoma. *Cell Death Discov* **7**, 331 (2021).
16. Xu, K. et al. Single-cell RNA sequencing reveals cell heterogeneity and transcriptome profile of breast cancer lymph node metastasis. *Oncogenesis* **10**, 66 (2021).
17. Chen, M. et al. Transcriptomic mapping of human parotid gland at single-cell resolution. *J. Dent. Res.* **101**, 972–982 (2022).
18. Padiglia, A. et al. Extensive characterization of the human salivary basic proline-rich protein family by top-down mass spectrometry. *J. Proteome Res.* **17**, 3292–3307 (2018).
19. Rajkumar, T. et al. Identification and validation of plasma biomarkers for diagnosis of breast cancer in South Asian women. *Sci. Rep.* **12**, 100 (2022).
20. Yen, M. C. et al. S100B expression in breast cancer as a predictive marker for cancer metastasis. *Int. J. Oncol.* **52**, 433–440 (2018).
21. Lelliott, E. J. et al. CDK4/6 inhibition promotes antitumor immunity through the induction of T-cell memory. *Cancer Discov* **11**, 2582–2601 (2021).
22. Tang, W. et al. LIF/LIFR oncogenic signaling is a novel therapeutic target in endometrial cancer. *Cell Death Discov.* **7**, 216 (2021).
23. Semenova, E. A. et al. Transcription factor NFIB is a driver of small cell lung cancer progression in mice and marks metastatic disease in patients. *Cell Rep.* **16**, 631–643 (2016).
24. Song, E. C. et al. Genetic and scRNA-seq analysis reveals distinct cell populations that contribute to salivary gland development and maintenance. *Sci. Rep.* **8**, 14043 (2018).
25. Lambert, A. W. & Weinberg, R. A. Linking EMT programmes to normal and neoplastic epithelial stem cells. *Nat. Rev. Cancer* **21**, 325–338 (2021).
26. Guen, V. J. et al. EMT programs promote basal mammary stem cell and tumor-initiating cell stemness by inducing primary ciliogenesis and Hedgehog signaling. *Proc. Natl Acad. Sci. USA* **114**, E10532–E10539 (2017).
27. Mani, S. A. et al. The epithelial-mesenchymal transition generates cells with properties of stem cells. *Cell* **133**, 704–715 (2008).
28. Sekiguchi, R., Martin, D., Genomics, Computational Biology, C & Yamada, K. M. Single-Cell RNA-seq identifies cell diversity in embryonic salivary glands. *J. Dent Res.* **99**, 69–78 (2020).
29. Oyelakin, A. et al. Transcriptomic and single-cell analysis of the murine parotid gland. *J. Dent Res.* **98**, 1539–1547 (2019).
30. Jin, L. et al. SFRP2 enhances the osteogenic differentiation of apical papilla stem cells by antagonizing the canonical WNT pathway. *Cell Mol. Biol. Lett.* **22**, 14 (2017).
31. Safadi, F. F. et al. Expression of connective tissue growth factor in bone: its role in osteoblast proliferation and differentiation in vitro and bone formation in vivo. *J. Cell Physiol.* **196**, 51–62 (2003).
32. Liberzon, A. et al. The molecular signatures database (MSigDB) hallmark gene set collection. *Cell Syst.* **1**, 417–425 (2015).
33. Rossi, D. et al. Calsequestrin, a key protein in striated muscle health and disease. *J. Muscle Res. Cell Motil.* **42**, 267–279 (2021).
34. Huang, C. H., Schuring, J., Skinner, J. P., Mok, L. & Chong, M. M. W. MYL9 deficiency is neonatal lethal in mice due to abnormalities in the lung and the muscularis propria of the bladder and intestine. *PLoS One* **17**, e0270820 (2022).
35. Gu, C. et al. Comprehensive analysis of angiogenesis-related genes and pathways in early diabetic retinopathy. *BMC Med. Genom.* **13**, 142 (2020).
36. Girard, J. P. & Springer, T. A. Modulation of endothelial cell adhesion by hevin, an acidic protein associated with high endothelial venules. *J. Biol. Chem.* **271**, 4511–4517 (1996).
37. Naschberger, E. et al. Matricellular protein SPARCL1 regulates tumor microenvironment-dependent endothelial cell heterogeneity in colorectal carcinoma. *J. Clin. Invest.* **126**, 4187–4204 (2016).
38. Shaikh, M. V., Kala, M. & Nivsarkar, M. CD90 a potential cancer stem cell marker and a therapeutic target. *Cancer Biomark* **16**, 301–307 (2016).
39. Hale, J. S. et al. Cancer stem cell-specific scavenger receptor CD36 drives glioblastoma progression. *Stem Cells* **32**, 1746–1758 (2014).
40. Pascual, G. et al. Targeting metastasis-initiating cells through the fatty acid receptor CD36. *Nature* **541**, 41–45 (2017).
41. Koren, E. et al. Thy1 marks a distinct population of slow-cycling stem cells in the mouse epidermis. *Nat. Commun.* **13**, 4628 (2022).
42. Haas, T. L. et al. Integrin alpha7 is a functional marker and potential therapeutic target in glioblastoma. *Cell Stem Cell* **21**, 35–50.e39 (2017).
43. Ming, X. Y. et al. Integrin alpha7 is a functional cancer stem cell surface marker in oesophageal squamous cell carcinoma. *Nat. Commun.* **7**, 13568 (2016).
44. Chen, Y. et al. MAFB promotes cancer stemness and tumorigenesis in osteosarcoma through a Sox9-mediated positive feedback loop. *Cancer Res.* **80**, 2472–2483 (2020).
45. Wojahn, I., Ludtke, T. H., Christoffels, V. M., Trowe, M. O. & Kispert, A. TBX2-positive cells represent a multi-potent mesenchymal progenitor pool in the developing lung. *Respir. Res.* **20**, 292 (2019).
46. Rosmaninho, P. et al. Zeb1 potentiates genome-wide gene transcription with Lef1 to promote glioblastoma cell invasion. *EMBO J* **37**, <https://doi.org/10.15252/embj.201797115> (2018).
47. Mendenhall, W. M., Mendenhall, C. M., Werning, J. W., Malyapa, R. S. & Mendenhall, N. P. Salivary gland pleomorphic adenoma. *Am. J. Clin. Oncol.* **31**, 95–99 (2008).
48. Cromer, A. et al. Identification of genes associated with tumorigenesis and metastatic potential of hypopharyngeal cancer by microarray analysis. *Oncogene* **23**, 2484–2498 (2004).
49. Lee, G. et al. Isolation and directed differentiation of neural crest stem cells derived from human embryonic stem cells. *Nat. Biotechnol.* **25**, 1468–1475 (2007).
50. Ahlers, J. M. D. et al. Single-cell RNA profiling of human skin reveals age-related loss of dermal sheath cells and their contribution to a juvenile phenotype. *Front. Genet.* **12**, 797747 (2021).
51. Kas, K. et al. Promoter swapping between the genes for a novel zinc finger protein and beta-catenin in pleiomorphic adenomas with t(3;8)(p21;q12) translocations. *Nat. Genet.* **15**, 170–174 (1997).
52. Ji, A. L. et al. Multimodal analysis of composition and spatial architecture in human squamous cell carcinoma. *Cell* **182**, 497–514.e422 (2020).
53. Li, Q. H. et al. Small breast epithelial mucin promotes the invasion and metastasis of breast cancer cells via promoting epithelial to mesenchymal transition. *Oncol. Rep.* **44**, 509–518 (2020).
54. King, R. J., Yu, F. & Singh, P. K. Genomic alterations in mucins across cancers. *Oncotarget* **8**, 67152–67168 (2017).
55. Abdulla, M. et al. Targeting MUC1 protein inhibits cell proliferation and EMT by deregulating beta catenin and increases irinotecan sensitivity in colorectal cancer. *Int. J. Oncol.* **60**, <https://doi.org/10.3892/ijo.2022.5312> (2022).
56. Bussolati, B., Bruno, S., Grange, C., Ferrando, U. & Camussi, G. Identification of a tumor-initiating stem cell population in human renal carcinomas. *Faseb J.* **22**, 3696–3705 (2008).
57. Gassenmaier, M. et al. CXCR4 chemokine receptor 4 is essential for maintenance of renal cell carcinoma-initiating cells and predicts metastasis. *Stem Cells* **31**, 1467–1476 (2013).
58. Atala, A. Re: microvesicles released from human renal cancer stem cells stimulate angiogenesis and formation of lung premetastatic niche editorial comment. *J. Urol.* **187**, 1506–1507 (2012).
59. Pattabiraman, D. R. & Weinberg, R. A. Tackling the cancer stem cells - what challenges do they pose? *Nat. Rev. Drug Discov.* **13**, 497–512 (2014).
60. Jordan, C. T. Cancer stem cells: controversial or just misunderstood? *Cell Stem Cell* **4**, 203–205 (2009).
61. Ye, H. et al. Leukemic stem cells evade chemotherapy by metabolic adaptation to an adipose tissue niche. *Cell Stem Cell* **19**, 23–37 (2016).
62. Takahashi, K., Murakami, M. & Yamanaka, S. Role of the phosphoinositide 3-kinase pathway in mouse embryonic stem (ES) cells. *Biochem. Soc. Trans.* **33**, 1522–1525 (2005).
63. Chen, J. Multiple signal pathways in obesity-associated cancer. *Obes. Rev.* **12**, 1063–1070 (2011).
64. Luo, X. et al. The fatty acid receptor CD36 promotes HCC progression through activating Src/PI3K/AKT axis-dependent aerobic glycolysis. *Cell Death Dis* **12**, 328 (2021).
65. Li, Q. et al. HSCs-derived COMP drives hepatocellular carcinoma progression by activating MEK/ERK and PI3K/AKT signaling pathways. *J. Exp. Clin. Cancer Res.* **37**, 231 (2018).
66. Guerrero-Rodriguez, S. L., Mata-Cruz, C., Perez-Tapia, S. M. & Velasco-Velazquez, M. A. Role of CD36 in cancer progression, stemness, and targeting. *Front. Cell Dev. Biol.* **10**, 1079076 (2022).
67. Wang, C. et al. Apolipoprotein C-II induces EMT to promote gastric cancer peritoneal metastasis via PI3K/AKT/mTOR pathway. *Clin. Transl. Med.* **11**, e522 (2021).
68. Wang, X. et al. Human primary epidermal organoids enable modeling of dermatophyte infections. *Cell Death Dis* **12**, 35 (2021).
69. Qin, Y. et al. A multi-scale map of cell structure fusing protein images and interactions. *Nature* **600**, 536–542 (2021).

70. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36**, 411 (2018).
71. Stuart, T. et al. Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888 (2019).
72. McCarthy, D. J., Campbell, K. R., Lun, A. T. L. & Wills, Q. F. Scater: pre-processing, quality control, normalization and visualization of single-cell RNA-seq data in R. *Bioinformatics* **33**, 1179–1186 (2017).
73. Patel, A. P. et al. Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science* **344**, 1396–1401 (2014).
74. Trapnell, C. et al. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* **32**, 381–386 (2014).
75. Yao, J. et al. Single-cell transcriptomic analysis in a mouse model deciphers cell transition states in the multistep development of esophageal cancer. *Nat. Commun.* **11**, 3715 (2020).



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, 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 license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license 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 license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023