



## ARTICLE OPEN

## Identifying a confused cell identity for esophageal squamous cell carcinoma

Xiangyu Pan<sup>1</sup>, Jian Wang<sup>1</sup>, Linjie Guo<sup>1</sup>, Feifei Na<sup>2</sup>, Jiajia Du<sup>1</sup>, Xuelan Chen<sup>1</sup>, Ailing Zhong<sup>1</sup>, Lei Zhao<sup>1</sup>, Lu Zhang<sup>1</sup>, Mengsha Zhang<sup>1</sup>, Xudong Wan<sup>1</sup>, Manli Wang<sup>1</sup>, Hongyu Liu<sup>1</sup>, Siqi Dai<sup>1</sup>, Ping Tan<sup>3</sup>, Jingyao Chen<sup>1</sup>, Yu Liu<sup>1</sup>✉, Bing Hu<sup>1</sup>✉ and Chong Chen<sup>1</sup>✉

The cell identity of malignant cells and how they acquire it are fundamental for our understanding of cancer. Here, we report that esophageal squamous cell carcinoma (ESCC) cells display molecular features equally similar but distinct to all three types of normal esophageal epithelial cells, which we term as confused cell identity (CCI). CCI is an independent prognostic marker associated with poor prognosis in ESCC. Further, we identify tropomyosin 4 (TPM4) as a critical CCI gene that promotes the aggressiveness of ESCC in vitro and in vivo. And TPM4 creates CCI through activating the Jak/STAT-SOX2 pathway. Thus, our study suggests an unrecognized feature of ESCC cells, which might be of value for clinic prognosis and potential interference.

Signal Transduction and Targeted Therapy (2022)7:122

; <https://doi.org/10.1038/s41392-022-00946-8>

## INTRODUCTION

The identity of malignant cells, distinguished from normal cells at various aspects, would be central for our understanding of this disease and provide insight for diagnosis and clinic interference. The hallmarks of cancer are the basic features of and fundamental principles underlying malignancies, which have been gradually revealed over the last half-century through extensive studies on the process of tumorigenesis and by comparing tumor cells with their normal counterparts.<sup>1–3</sup>

According to the traditional developmental view of tumorigenesis, it has long been proposed that tumorigenesis is an unusual event of normal development and malignant cells arise from the stem and/or progenitor cells with differentiation block.<sup>4–11</sup> One of the best examples might be acute myeloid leukemia (AML), which has been suggested to be transformed from hematopoietic stem and progenitors with differentiation-blocking mutations.<sup>6,10,12,13</sup> And thus, it is generally believed that cancer cells which are more like stem cells are more aggressive than those with differentiation features.<sup>6,10,14,15</sup> However, there are remarkable exceptions, such as multiple myeloma, which is generally thought to be developed from mast cells, the terminally differentiated B lymphocytes.<sup>16</sup> And recently, it has been argued that the so-called cancer stem cells and differentiated cells could two-way transit to each other, which, together with other arguments, challenges the developmental view of tumorigenesis.<sup>17–19</sup>

Recently, there is accumulating evidence suggesting that the lineage identity of cancer cells might be dynamic. For example, a subset of non-small cell lung cancer (SCLC) with estimated glomerular filtration rate mutations developed resistance of tyrosine kinase inhibitors and transformed into SCLC with RB1 loss.<sup>20</sup> LKB1 mutations can promote the transit of lung adenocarcinoma into squamous cell carcinoma (SCC).<sup>21</sup> And it has been

proposed that a high-plasticity cell state might underlie lung adenocarcinoma with poor outcomes and chemoresistance.<sup>22,23</sup> Similarly, p53 and Rb1 loss increase the lineage plasticity of antiandrogen resistant prostate cancer into neuroendocrine-like cancer.<sup>24–26</sup> Thus, in the latest version, the unlocked phenotypic plasticity was added as a new hallmark of cancer.<sup>2</sup>

However, the cell identity of most types of cancers, especially those derived from epithelial cells, which consist of more than 90% of human malignancies, remains unclear.<sup>27–29</sup> The molecular and cellular pathologies of SCC, one of the most common epithelial cancers with worse outcomes than others, are less understood. This type of cancer seems to be differentiated toward squamous epithelial cells to some extent and thus distinguishes it from epithelial stem cells. On the other hand, patients with low-grade differentiation have a poor prognosis.<sup>11,29</sup> To resolve these paradoxes and gain a better understanding of the cellular identity and the underlying molecular mechanisms of SCC, we analyzed the cellular identity of esophageal SCC (ESCC) at single-cell resolution in the background of normal epithelial cells with a discovery cohort of seven tumor samples and one distal normal sample from five pathologically confirmed ESCC patients. We then validated the results in a cohort of 60 patients analyzed by single-cell RNA-seq (scRNA-seq) and 9 cohorts of a total of 534 patients analyzed by bulk RNA-seq or microarray assays.

## RESULTS

Single-cell transcriptomics reveals CCI as a unique identity of ESCC cells

The ESCC samples and normal esophageal tissue in the discovery cohort were analyzed with 10× Genomics scRNA-seq (Fig. 1a and Supplementary Fig. S1a, b). To better represent the squamous

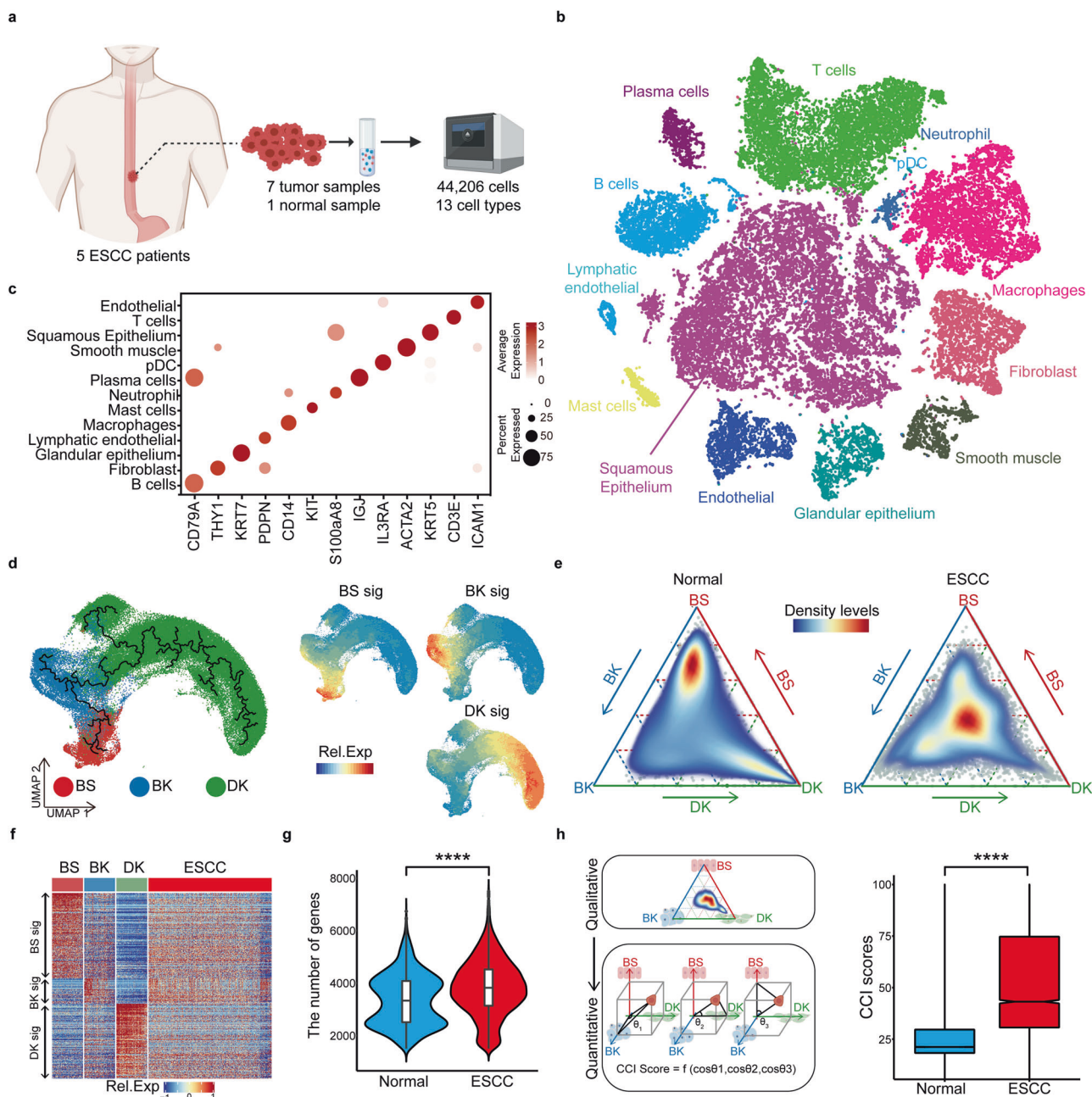
<sup>1</sup>Department of Gastroenterology, State Key Laboratory of Biotherapy and Cancer Center, West China Hospital, Sichuan University, Chengdu, Sichuan, China; <sup>2</sup>Department of Thoracic Oncology, State Key Laboratory of Biotherapy and Cancer Center, West China Hospital, Sichuan University, Chengdu 610041 Sichuan, China and <sup>3</sup>Department of Urology, Institute of Urology, West China Hospital, Sichuan University, Chengdu, China

Correspondence: Yu Liu (yuliuscu@scu.edu.cn) or Bing Hu (hubingnj@163.com) or Chong Chen (chongchen@scu.edu.cn)

These authors contributed equally: Xiangyu Pan, Jian Wang, Linjie Guo, Feifei Na

Received: 26 January 2022 Revised: 25 February 2022 Accepted: 1 March 2022

Published online: 13 April 2022



**Fig. 1** Confused cell identity (CCI) is a feature of ESCC cells at single-cell resolution. **a** The schematic graph of the experimental design. **b** The t-SNE map of single-cell ESCC landscape was colored by cell subtypes. **c** The dot plot showed the expression levels of classical signatures in each cell subtype. **d** The UMAP map of normal squamous epithelial (SE) was colored by cell subtypes. The black lines, fitted by monocle3, represented the development lineage of normal SE (left). The distribution of expression levels of BS, BK, and DK signatures on the UMAP map (right). **e** The ternary diagram showed the similarity/variation in scRNA-seq data of SE from the normal esophagus (left) and ESCC (right), with BS, BK, and DK cells. The color represented the density levels of distribution. **f** The heatmap showed the expression levels of BS, BK, and DK signatures in scRNA-seq data. **g** The boxplot showed the number of genes detected in normal and ESCC SE cells ( $n = 4052$  cells, normal;  $n = 7580$  cells, ESCC).  $p$  Values calculated by two-sided unpaired  $t$ -test. \*\*\*\* $p < 0.0001$ . **h** The schematic graph of CCI scores calculation (left). The box plot showed the CCI scores in single-cell normal samples and ESCC samples (right,  $n = 4860$  cells, normal;  $n = 6801$  cells, ESCC).  $p$  Values calculated by Wilcoxon signed-rank test. \*\*\*\* $p < 0.0001$

epithelial cells, we optimized the dissociation protocol to harvest more than a twofold number of epithelial cells than previous methods (Supplementary Fig. S1c).<sup>30</sup> Total 44,206 cells were obtained, including 16,657 normal and malignant squamous epithelial cells. The squamous epithelial cells, glandular epithelial cells, immune cells, endothelial cells, and other tumor micro-environmental cells were separated and identified with classic markers (Fig. 1b, c, Supplementary Fig. S2a and Supplementary

Table 1). On the t-SNE map, all these samples were essentially evenly distributed (Supplementary Fig. S2b–e).

All the esophageal squamous epithelial cells expressed high levels of marker gene KRT5 (Supplementary Fig. S2a). In patients, the squamous epithelial cells displayed obvious copy number variations, compared to other cells, calculated by inferCNV (Supplementary Fig. S2f). Thus, these cells were identified as ESCC cells. In normal esophageal tissues, the esophageal squamous

epithelium consists of three major types of cells, including basal stem cells (BS), transient proliferating basal keratinocytes (BK), and postmitotic differentiated keratinocytes (DK).<sup>31–33</sup> And these cells are distinct from each other by their morphologies, localization, and molecular features and BS can generate BK and then DK through stepwise differentiation.<sup>32,33</sup> Consistently, we found that normal squamous epithelial cells could be divided into three continuous subpopulations (Supplementary Fig. S3a). Similar three subpopulations were observed with the combinations of 4885 cells from our dataset and the human cell atlas (the HCA) esophageal dataset (Fig. 1d). One of them expressed *SOX2*, *TP63*, and *KRT15*, suggesting them as BS, the second one expressed a high level of *KRT5* and the third one expressed markers *KRT14*, *KRT13*, and *IVL*, corresponding to the definition of BK and DK, respectively (Supplementary Fig. S3b, c and Supplementary Table 2). The unsupervised fate map showed a pseudo trajectory from BS to BK and DK at the end (Fig. 1d). BK had the highest proliferating score, while DK was mostly postmitotic (Supplementary Fig. S3d). And accordingly, BS had enrichments of stemness-related pathways, BK had enrichments of MYC targets, DNA replication, and other cell cycle pathways and DK had high levels of genes related to the functions of squamous epithelium, including cornification and keratinization (Supplementary Fig. S3e).

To visualize the relationship of ESCC cells with normal esophageal squamous epithelial cells, we created a ternary map with BS, BK, and DK at each corner (Fig. 1e). This map could also distinguish all normal esophageal squamous epithelial cells from other studies (Supplementary Fig. S4a, b).<sup>34,35</sup> However, strikingly, all malignant cells fell at the center of the triangle with equal distance to each corner. This relative position of ESCC cells indicated that they might contain similar gene expressions as each of these three normal squamous epithelial cells. Indeed, the gene expression heatmap showed that ESCC cells expressed all the signature genes of BS, BK, and DK, though at slightly lower levels (Fig. 1f). And concordantly, the average numbers of genes detected in malignant cells were significantly more than those in normal cells (Fig. 1g). This result was in contrast to what the traditional stem cell view of tumorigenesis. We termed this new cell identity relative to the normal cells as the confused cell identity (CCI). To quantify the CCI of each cell on the ternary map, we designed a CCI score by calculating the standard deviation of the cosine values of the three angles with the dot of cell position as the vertex and the connecting lines of the dot to each corner and the origin as rays. The median CCI score was 21.2 for the normal cells and 43.2 for the malignant cells (Fig. 1h). With 44,730 malignant cells from 60 ESCC patients in a previous study, we also observed a similar CCI for these ESCC cells (Supplementary Fig. S4c).<sup>30</sup> And although they were from independent studies and potential batch effects, the normal esophageal squamous cells had less than 20 CCI scores while ESCC cells had a score of 35.4 (Supplementary Fig. S4d, e). Thus, we proposed that CCI was a new molecular feature of malignant cells.

CCI is a general feature of ESCC in multiple cohorts. Further, we wondered whether the CCI score was correlated with the aggressiveness of ESCC in patients. First, we found that the average CCI score was significantly higher in ESCC patients with aggressive progression (Fig. 2a). In Zhang's cohort with 60 ESCC patients, there were 16 Stage I and 44 Stage II/III patients.<sup>30</sup> Despite the inter- and intra-tumoral heterogeneity, the CCI scores of the early stage patients were significantly less than those of the advanced-stage patients (Fig. 2b). Consistently, at the single-cell level, the CCI scores were significantly positively correlated with the expressions of ESCC advanced stage-specific genes (Fig. 2c and Supplementary Table 3).

Given that most of the available gene expression data for ESCC patients were bulk transcriptomics analyzed with RNA-seq or microarray, we wanted to measure the CCI status of each patient

with bulk transcriptomes. In the TCGA ESCC cohort with 81 patients, we found all stage II/III patients located in the center of our CCI map, while the stage I patients' location was close to the corner of DK (Fig. 2d). The average CCI score of higher-grade patients was significantly higher than that of the stage I patients (22.1 vs. 38.7,  $p = 0.028$ ) (Fig. 2e). More importantly, we found that patients with high CCI scores had significantly worse prognoses than those with low CCI scores (median survival time 484 vs. 1263 days,  $p = 0.02$ ) (Fig. 2f). Further, the CCI analysis could also be applied with the microarray gene expression data. In Li's cohort<sup>36</sup> with 119 ESCC patients analyzed with microarray gene expression profiling, all the tumor samples were at the center of the CCI map while their adjacent normal tissue was close to DK (Fig. 2g). And their CCI scores were 33.1 and 14.9, respectively (Fig. 2h). Similarly, in this cohort, the survival of patients with high CCI scores was significantly shorter than that of those with low-CCI scores (21.83 vs. 45.83 months,  $p = 0.0045$ ) (Fig. 2i). To test whether other clinic factors would affect the CCI score, we analyzed the correlations of multiple clinic parameters with the CCI score in the TCGA ESCC cohort. The results showed that there were no significant differences of sex, smoking, alcohol history, treatment, BMI, and other clinic factors between the CCI low and high patients (Supplementary Table 4). These results suggested that the CCI score could be an independent prognostic marker for ESCC patients.

The molecular mechanism underlying CCI

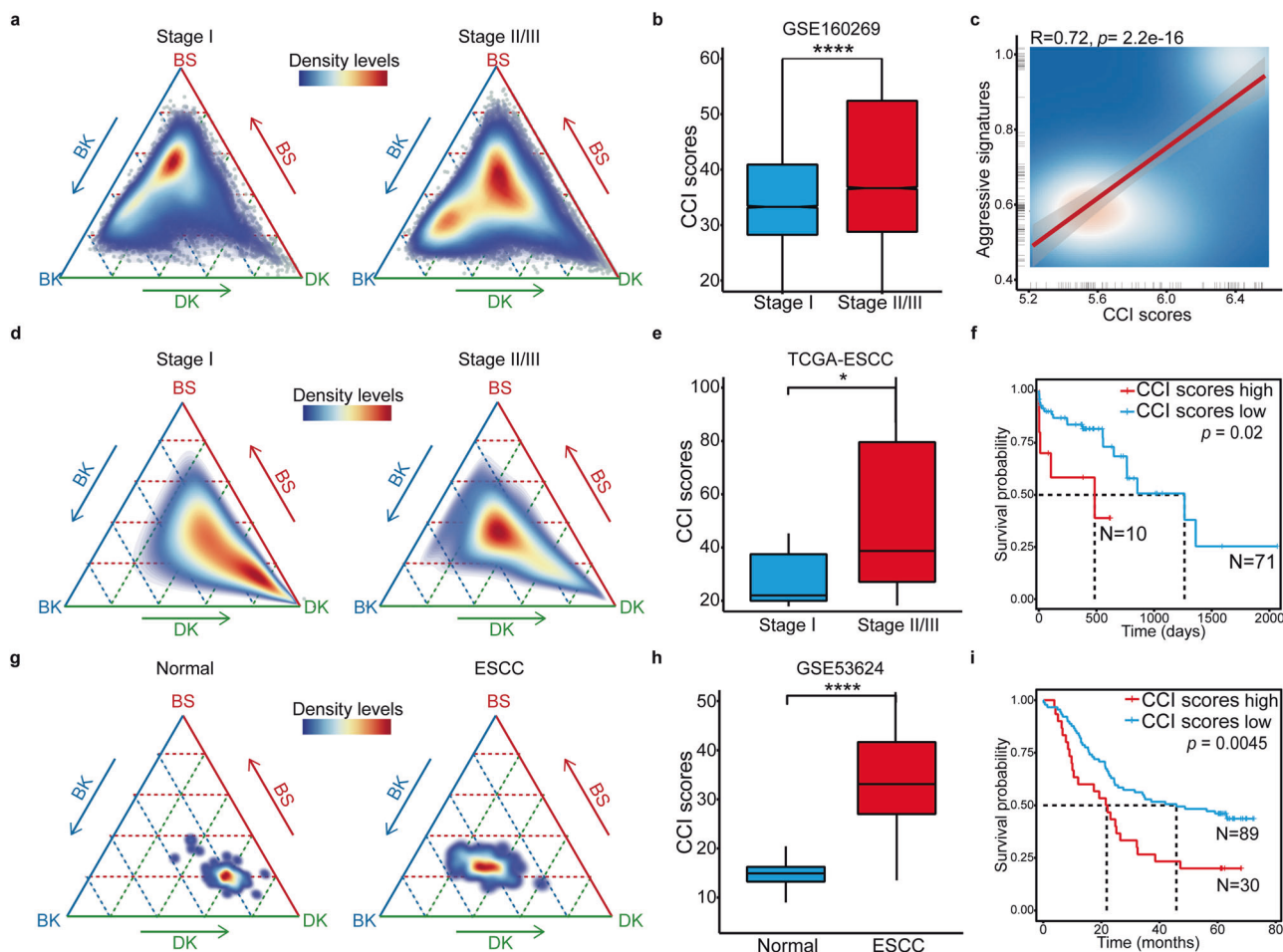
Then we explored the molecular mechanism underlying CCI in ESCC. The CCI score was positively correlated with many cancer pathways, including those related to cancer stemness, metastasis, and epithelial cell proliferation (Supplementary Fig. S5a and Supplementary Table 5). Then, we identified the CCI signature genes by comparing the CCI-positive cells with the negative epithelial cells. Total 195 genes specifically expressed in the malignant cells were scored (Supplementary Fig. S5b and Supplementary Table 6); these CCI signature genes were also upregulated in the tumors of advanced stages compared to those of early stages (Supplementary Fig. S5b). On the CCI map, these CCI signature genes were not expressed in normal cells at the three corners and were highly expressed in the malignant cells at the center (Fig. 3a). Noteworthy, the expression levels of these CCI signature genes were significantly positively correlated with the CCI score (Fig. 3b).

To dissect the molecular mechanisms underlying CCI, we performed the network analysis on the CCI signature genes. The results showed that multiple gene networks were enriched for CCI, including spliceosome, ribosome, antigen processing and presentation, oxidative phosphorylation, and pathways in cancer (Supplementary Fig. S5c). Further, we looked for the CCI signature genes which were highly expressed in ESCC patients at the protein levels.<sup>37</sup> Six of them, *CBR1*, *GSTM3*, *HSPB1*, *MIF*, *PGK1*, and *TPM4* were identified (Fig. 3c). Among them was tropomyosin 4 (*TPM4*), which, by itself, had the highest diagnostic value in ESCC patients revealed by the multivariate Cox regression analysis (Supplementary Fig. S5d). *TPM4* was among the most upregulated genes in ESCC and various cancers at the protein level (Supplementary Table 7).<sup>37–40</sup> In the GSE149609 cohort, we confirmed the upregulation of *TPM4* in ESCC patients (Fig. 3d). *TPM4* high expressions were significantly associated with poor prognosis (Fig. 3e). And the expression level of *TPM4* was significantly positively correlated with the CCI score at the single-cell levels and the patient level (Fig. 3f, g). Thus, we believed that *TPM4* was a CCI gene.

*TPM4* overexpression increases the CCI score

To test the function of *TPM4*, we overexpressed it in esophageal organoids. *TPM4* overexpression did not significantly affect the size of organoids. However, once passaged, the *TPM4* organoids gave rise to substantially more secondary organoids, indicating





**Fig. 2** CCI is an independent prognostic marker in ESCC. **a** The ternary diagram showed the similarity/variation in stage I (left) and stage II/III (right) from scRNA-seq data of GSE160269, with BS, BK, and DK cells. The color represented the density levels of distribution. **b** The boxplot showed the CCI scores in stage I and stage II/III patients from GSE160269 ( $n = 13,041$  cells, Stage I;  $n = 31,506$  cells, Stage II/III).  $p$  Values calculated by Wilcoxon signed-rank test. \*\*\*\* $p < 0.0001$ . **c** The scatter plot showed the correlation between the CCI scores and the expression levels of ESCC aggressive stage-specific signatures in scRNA-seq. **d** The ternary diagram showed the similarity/variation in stage I (left) and stage II/III (right) from bulk RNA-seq data of TCGA-ESCC, with BS, BK, and DK cells. The color represented the density levels of distribution. **e** The boxplot showed the CCI scores in stage I and stage II/III patients from TCGA-ESCC ( $n = 7$  patients, Stage I;  $n = 72$  patients, Stage II/III).  $p$  Values calculated by two-sided unpaired  $t$ -test. \* $p < 0.05$ . **f** The Kaplan–Meier survival curves of TCGA-ESCC patients with low and high CCI scores. **g** The ternary diagram showed the similarity/variation in adjacent normal (left) and ESCC (right) from array data of GSE53624, with BS, BK, and DK cells. The color represented the density levels of distribution. **h** The boxplot showed the CCI scores in adjacent normal and ESCC samples from GSE53624 ( $n = 119$  patients, normal;  $n = 119$  patients, ESCC).  $p$  Values calculated by Wilcoxon signed-rank test. \*\*\*\* $p < 0.0001$ . **i** The Kaplan–Meier survival curves of GSE53624 patients with low and high CCI scores

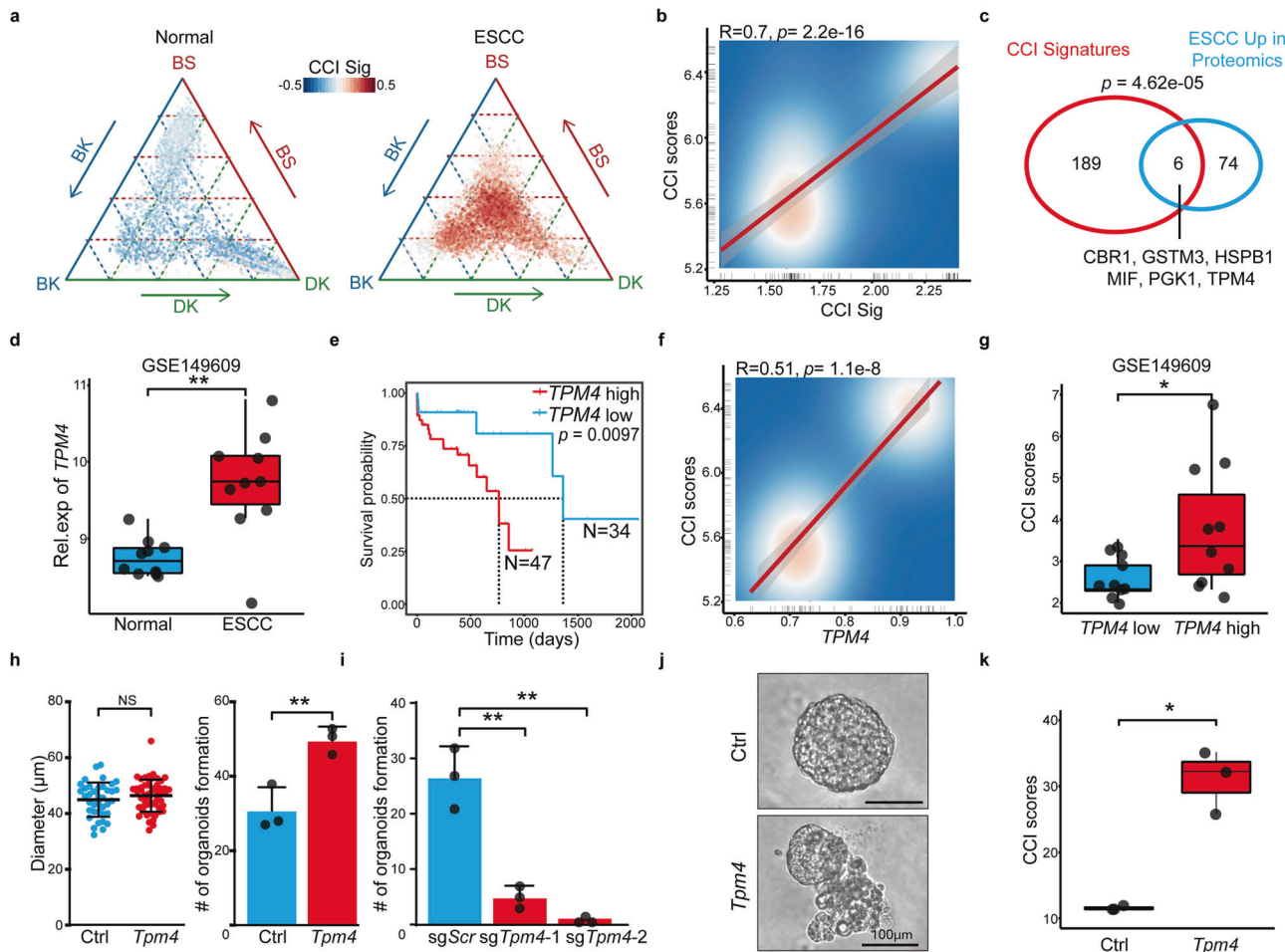
that *TPM4* promoted the self-renewal of progenitor cells (Fig. 3h). And reversely, disruption of *Tpm4* significantly reduced the formation of esophageal organoids (Fig. 3i). Thus, *TPM4* might not be required for cell proliferation or growth but is essential for the self-renew of esophageal epithelial organoids. And interestingly, more than half of the *TPM4* overexpressed organoids displayed a dramatic morphology change those cells migrated out of the organoids (Fig. 3j and Supplementary Fig. S5e). *TPM4* can affect cell migration by regulating the formation of F-actin.<sup>41,42</sup> And consistent, the migrating cells in the *Tpm4* overexpressed organoids expressed high levels of F-actin (Supplementary Fig. S5f).

In agreement with its function in the self-renewal and migration of esophageal epithelial cells, the *TPM4* expression was positively correlated with multiple related pathways, including EMT, stem cell maintenance, tumor invasiveness (Supplementary Fig. S6a). And the enriched pathways in the *TPM4* high ESCC samples were primarily overlapped with those in the CCI gene signature (Supplementary Fig. S5a). Then, we measured the CCI status of control and *Tpm4*-overexpressed organoids in the CCI map. While

the control organoids were close to the corner of BK, those with *Tpm4* overexpression moved toward the center (Supplementary Fig. S6b and Supplementary Table 8). And the CCI scores of *Tpm4* overexpressed organoids were significantly higher than those of the control ones (Fig. 3k). Taken together, *TPM4* promoted CCI in esophageal epithelial cells.

#### *TPM4* promotes ESCC metastasis in mice

We further confirmed the function of *TPM4* in ESCC in vivo. *Tpm4* was overexpressed in mouse ESCC organoids and then subcutaneously transplanted into recipient mice with control ESCC organoids on the other side. Tumors with *Tpm4* overexpression grew significantly faster than control tumors, measured by fluorescence living imaging (Supplementary Fig. S7a, b). Histology analyses suggested that *Tpm4* tumors were a more advanced stage than control ESCC, indicating a higher nuclear-to-cytoplasmic ratio, denser nuclei, and less differentiation (Supplementary Fig. S7c). And there were also significantly more Ki67+ tumor cells in *Tpm4* overexpression ESCC than control tumors



**Fig. 3** TPM4 is a CCI gene. **a** The ternary diagram showed the expression levels and distribution of CCI signatures in scRNA-seq from normal SE cells (left) and ESCC SE cells (right). **b** The scatter plot showed the correlation between the expression levels of CCI signatures and CCI scores in scRNA. **c** The Venn plot showed the intersection between CCI signatures and upregulated proteins in ESCC patients. The statistic power was calculated by a hypergeometric test. **d** The box plot showed the expression levels of *TPM4* in adjacent normal and ESCC samples from GSE149609 ( $n = 10$  patients, normal;  $n = 10$  patients, ESCC).  $p$  Values calculated by Wilcoxon signed-rank test.  $**p < 0.01$ . **e** The Kaplan–Meier survival curves of TCGA-ESCC patients with low and high expression of *TPM4*.  $p = 0.0097$ . **f** The scatter plot showed the correlation between the expression levels of *TPM4* and CCI scores in scRNA. **g** The box plot showed the log<sub>2</sub>-CCI scores in ESCC's samples from GSE149609, which were grouped by the median of the expression levels of *TPM4* ( $n = 10$  patients, *TPM4*-low;  $n = 10$  patients, *TPM4*-high).  $p$  Values calculated by two-sided unpaired  $t$ -test.  $*p < 0.05$ . **h** The dot plot showed the diameter of organoids with control and *Tpm4* overexpressed (left). The number of control and *Tpm4* overexpressed organoids in each well was counted (right). Data are shown as mean  $\pm$  SD ( $n = 3$  biological replicates, Ctrl;  $n = 3$  biological replicates, *Tpm4*).  $p$  Values calculated by two-sided unpaired  $t$ -test. NS nonsignificant;  $**p < 0.01$ . **i** The number of *sgScr* and *sgTpm4* organoids in each well was counted. Data are shown as mean  $\pm$  SD ( $n = 3$  biological replicates, *sgScr*;  $n = 3$  biological replicates, *sgTpm4-1*;  $n = 3$  biological replicates, *sgTpm4-2*). Three independent biological replicates were performed for each group.  $p$  Values calculated by two-sided unpaired  $t$ -test.  $**p < 0.01$ . **j** The morphology of control and *Tpm4* overexpressed esophageal organoid. Scale bar, 100  $\mu$ m. **k** The boxplot showed the CCI scores in control and *Tpm4* overexpressed organoids ( $n = 3$  organoids, Ctrl;  $n = 3$  organoids, *Tpm4*).  $p$  Values calculated by two-sided unpaired  $t$ -test.  $*p < 0.05$

(Supplementary Fig. S7d). Then, we orthotopically implanted the control and *Tpm4*-overexpressed ESCC organoids under the epithelial layer of the recipient's esophagus. Fluorescence living imaging showed the *Tpm4* overexpressed tumors had a dramatically more vigorous fluorescence intensity than the control tumor (Fig. 4a, b). Consistently, the biopsy revealed that the massive tumors with *Tpm4* overexpression spread to the whole stomach while the control tumors were limited to the transplanted site (Fig. 4c). H/E and IHC staining confirmed that both tumors with and without *Tpm4* overexpression were typical ESCC expressing diagnostic markers p63 and CK14, highly similar to the human disease (Fig. 4d, e and Supplementary Fig. S7e). These pathologic analyses and p63 staining suggested that *Tpm4* overexpressed tumors were more aggressive than the control.

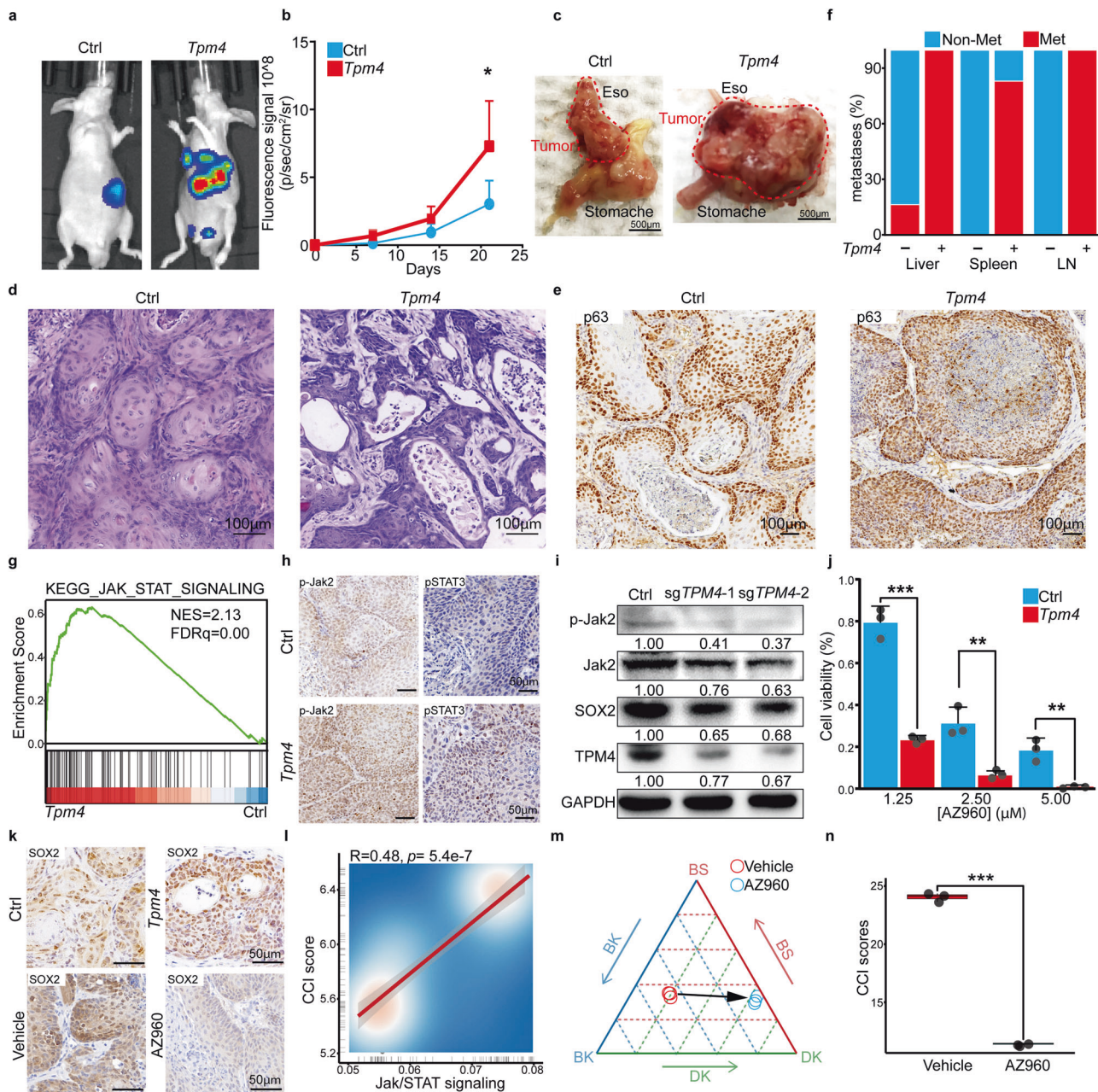
And intriguingly, we noticed strong fluorescence signaling at multiple sites on the recipient mice of *Tpm4* tumors (Fig. 4a).

Biopsy showed that the *Tpm4* mice displayed specific signaling on numerous organs, including liver, spleen, and lymph nodes, hotspots of ESCC metastases in patients (Supplementary Fig. S7f). The histology analyses of metastases in lymph nodes and liver confirmed these tumor cells as ESCC metastases (Supplementary Fig. S7g). All six *Tpm4* recipients had metastases in the liver and lymph nodes and five of them had metastases in the spleen. In comparison, only one of the six recipients with control ESCC had metastases in the liver and none of them had metastases in other organs (Fig. 4f).

#### The TPM4-Jak/STAT-SOX2 axis for CCI in ESCC

To gain insight into the TPM4-promoting CCI, we performed transcriptome analyses of control and *Tpm4* overexpressed esophageal organoids. The results showed that the Jak/STAT signaling pathway was among the most upregulated pathways by





**Fig. 4** TPM4 promotes CCI and ESCC aggressiveness through the Jak/STAT pathway. **a** Representative living images after orthotopic transplantation of control (left) and *Tpm4* (right) overexpressed tumor organoids, each group consisting of six mice. **b** The luciferase fluorescence signal intensity of the control and *Tpm4* overexpressed mice after orthotopic transplantation. Data shows the means  $\pm$  SEM. ( $n = 6$  independent biological replicates were performed for each group).  $p$  Values calculated by two-sided unpaired  $t$ -test. \*,  $p < 0.05$ . **c** Representative bright-field images of the orthotopic esophageal tumor tissue derived from the control (left) and *Tpm4* overexpressed (right). Scale bar, 500  $\mu$ m. **d** Representative H&E staining images of primary orthotopic tumors derived from the control (left) and *Tpm4* overexpressed (right). Scale bar, 100  $\mu$ m. **e** Representative immunohistochemistry images of p63 in orthotopic control and *Tpm4* overexpressed tumors. Scale bar, 100  $\mu$ m. **f** The percentages of the control and *Tpm4* overexpressed mice with indicated numbers of metastases in liver, spleen, and lymph nodes (LN) were summarized. (Each group consists of  $n = 6$  mice). **g** The GSEA plot showed the Jak/STAT signaling pathways enriched in *Tpm4* overexpressed samples. **h** Representative immunohistochemistry images of p-Jak2 and pSTAT3 in subcutaneous control and *Tpm4* tumors. Scale bar, 50  $\mu$ m. **i** The representative western blotting pictures showed the p-Jak2, Jak2, SOX2, TPM4, and GAPDH levels in the control or TPM4 disrupted organoids. **j** The viability of control and *Tpm4* overexpressed tumor organoids in different concentrations of AZ960. Data are shown as mean  $\pm$  SD. ( $n = 3$  independent biological replicates were performed for each group).  $p$  Values calculated by two-sided unpaired  $t$ -test. \*\* $p < 0.01$ ; \*\*\* $p < 0.001$ . **k** Representative immunohistochemistry images of SOX2 in subcutaneous control and *Tpm4* overexpressed tumors (top) and in subcutaneous *Tpm4* overexpressed tumors treated with vehicle and AZ960 (bottom). Scale bar, 50  $\mu$ m. **l** The scatter plot showed the correlation between the expression levels of Jak/STAT and CCI scores in scRNA. **m** The ternary diagram showed the similarity/variation in RNA-seq data of *Tpm4* overexpressed organoids treated with vehicle and AZ960, with BS, BK, and DK cells. Three independent biological replicates were performed for each group. **n** The boxplot showed the CCI scores in *Tpm4* overexpressed organoids treated with vehicle and AZ960 ( $n = 3$  organoids, Vehicle;  $n = 3$  organoids, AZ960). Three independent biological replicates were performed for each group.  $p$  Values calculated by two-sided unpaired  $t$ -test. \*\*\* $p < 0.001$

*Tpm4* (Fig. 4g, Supplementary Fig. S8a and Supplementary Table 9). And the advanced ESCC also expressed higher levels of these genes than early ESCC (Supplementary Fig. S8b). Of note, the IHC staining confirmed that *Tpm4* tumors had increased p-Jak2 and p-STAT3 levels compared to the control tumors (Fig. 4h). And similarly, western blotting showed that *Tpm4* overexpression increased the p-Jak2 level and reversely, *Tpm4* loss decreased the p-Jak2 level (Fig. 4i and Supplementary Fig. S8c). And importantly, ESCC organoids with *Tpm4* overexpression were significantly more sensitive to the Jak2 specific inhibitor AZ960 than the control ESCC organoids (Fig. 4j). AZ960 treatment also reduced the growth of *Tpm4* overexpressed ESCC in vivo, indicated by reduced Ki67 staining (Supplementary Fig. S8d).

It has been reported that the Jak/STAT pathway can upregulate SOX2, a key transcription factor for cell identity and ESCC progress.<sup>24,25,43,44</sup> Indeed, ESCC tumors with *Tpm4* overexpression had increased SOX2 staining compared to control tumors (Fig. 4k). And consistently, western blotting showed that *TPM4* overexpression upregulated SOX2 expression while *TPM4* disruption down-regulated it (Fig. 4i). Further, AZ960 treatment reduced the SOX2 level in vitro and in vivo (Fig. 4k and Supplementary Fig. S8e).

In the end, we found that the expression levels of the Jak/STAT signaling pathway genes were significantly positively correlated with the CCI score in ESCC (Fig. 4l). And importantly, inhibition of the Jak/STAT pathway by AZ960 significantly reduced the CCI scores of ESCC cells and drove the *Tpm4* overexpressed ESCC cells from the center to the DK side on the CCI map (Fig. 4m, n and Supplementary Table 10).

## DISCUSSION

By comparing the single-cell transcriptomes of ESCC cells and their counterparts, we propose that CCI, a status with a mixture of molecular features of multiple types of normal squamous cells, is a hallmark of ESCC cells at the single-cell and patient level. The identification of CCI suggests that ESCC carcinogenesis is not a simple differentiation block of normal development, and instead, the tumor cells acquire a new identity with partial similarities to multiple normal cell types. CCI is a cell identity simultaneously possessing features of multiple cell types and thus distinguished from lineage plasticity that cells switch their identities from one lineage to other lineages.<sup>26</sup> Given the similarity of ESCC cells to BS, BK, and DK, it might be possible that multiple types of squamous epithelial cells could be the cell of origin of this disease. Thus, CCI describes a previously unrecognized feature of malignant cells, which is distinguished from the developmental view of tumorigenesis or the lineage plasticity theory.<sup>4-11,24-26</sup>

We show that CCI is not only associated with the aggressiveness of ESCC cells but also a potential independent prognostic marker for patients. Since CCI is measured by a set of signature genes, which might be more stable than single genes. It is robust and can also apply to various transcriptome profiling methods, including microarray profiling, bulk RNA-seq, and scRNA-seq. It would be interesting to test it in further clinic practice for ESCC patients. Further, once the lineage identification of normal cell types has been established for other tissues, similar strategies might be used to test whether CCI also plays role in other human cancers and in other contexts, including drug resistance.

Multiple molecular pathways might be underlying CCI, such as spliceosome, ribosome, antigen processing and presentation, and oxidative phosphorylation. We found that *TPM4*, one of the top CCI signature genes which have not been functionally studied in ESCC, promotes the CCI and ESCC aggressiveness. Mechanistically, *TPM4* can activate the Jak/STAT signaling pathway, which further activates SOX2. SOX2 is frequently amplified and/or upregulated in ESCC and recently it has been shown to be able to induce ESCC in mice.<sup>45-49</sup> The Jak/STAT signaling pathway has been shown to be activated in various human cancers and serve as a clinic

therapeutic target at least in some hematopoietic malignancies.<sup>50</sup> It can coordinate stem cell proliferation and lineage differentiation.<sup>51</sup> SOX2 is also reported to be critical for lineage identification in both normal and tumor cells.<sup>24-26,52</sup> Interestingly, the Jak/STAT signaling pathway itself is on the CCI gene network. Thus, we believe that these CCI signature genes might be functionally linked in ESCC cells, and their interactions together contribute to the unique identity of malignant cells.

Identifying CCI and understanding its underlying molecular mechanisms reveal the therapeutical vulnerability of ESCC. We show that inhibiting the Jak/STAT signaling pathway can decrease the CCI score of ESCC and reduce the growth of ESCC cells. Further studies on the potential benefits of the Jak/STAT inhibitors, some of which have been used for patients with other syndromes, for ESCC patients.

## MATERIALS AND METHODS

This study complied with all relevant ethical regulations and was approved by the Ethics Committee on Biomedical Research, West China Hospital of Sichuan University.

### Study subjects and biospecimens

Individuals with ESCC ( $n = 5$ ) were recruited from West China School of Medicine/West China Hospital of Sichuan University (WCSM/WCH; Sichuan, China) between 2018 and 2019. None of the subjects underwent radiotherapy, chemotherapy, or immunotherapy prior to surgery. ESCC tumor tissue and normal adjacent tissue ( $\geq 5$  cm from the tumor margin) were obtained after Lugol's iodine solution staining. The pathological diagnosis was made by pathologists in the Department of Pathology, West China Hospital, Sichuan University, who confirmed that there were no tumor cells in normal tissues and no necrosis in tumor tissues, and the proportion of tumor cells in tumor tissues was more than 60%. Written informed consent was obtained from every subject and this study was approved by the Institutional Review Board of WCSM/WCH.

### Human esophageal squamous carcinoma cell lines

TE10 (#CL-0453) were purchased from Procell Corp and cultured in RPMI-1640 medium with 10% FBS supplemented, 10% heat-inactivated FBS (Gibco, #10099-141) and 1% penicillin-streptomycin (Gibco, #10378016). All cell lines were tested negative for mycoplasma. TE10 cell lines were infected with the SpCas9 lentiviral vector with sg*TPM4* or lentiviral containing *TPM4* cDNA. To identify the phosphorylated JAK2 and phosphorylated STAT3, all cell lines were stimulated with human IFN- $\gamma$  (50 ng/ml, novoprotein, #CO14) for at least 48 h.

### Mice

All mice used in this study were maintained in the SPF level mouse facility at State Key Laboratory of Biotherapy of Sichuan University. All the animal experiment protocols have been approved by the institutional Animal Care and Use Committees of Sichuan University. *Rosa26-CAG-Cas9-iRES-GFP* mice hereafter referred to as Cas9 (Cat# JAX:026179, RRID: IMSR\_JAX:026179), *Trp53*<sup>-/-</sup> mice were obtained from The Jackson Laboratory (Cat# JAX:000664, RRID: IMSR\_JAX:000664). Recipient nude mice were purchased from Hfkbio (Cat# BALB/cA-nu Mice). All animal experiments were performed in accordance with relevant guidelines and regulations and were approved by the West China Hospital, Sichuan University. Mice were deeply anesthetized by gas inhalation isoflurane before surgery or euthanasia.

### Murine esophageal organoids culture

Six to eight weeks old male mouse was sacrificed for culture. Esophagus tissue was separated and rinsed with cold phosphate-buffered saline buffer (Gibco, Cat# 10010002) supplemented with antibiotics (Gibco, Cat# 15240096) with vigorous shaking. After



several times washing transferred the tissue to a new 15 ml centrifuge tube and minced with fine scissors. Then 5 ml 0.25% trypsin-EDTA (Gibco, Cat# 25200072) was added to digest the tissues in a 37 °C water bath for 1 h. When most of the tissue dissolved, the digestion reaction was stopped by a serum containing DMEM medium. Filtered the mixture through a 70- $\mu$ m cell strainer (JET BIOFIL, #CSS-013-070) to obtain a single-cell suspension, then centrifuged at 400g, 5 min and washed once with phosphate-buffered saline (PBS) subsequently. After the liquid in the centrifuge tube was removed, the Matrigel (BD, Cat# 354230) was used to resuspend the cell pellet. After being mixed on ice evenly, 30  $\mu$ l of the mixture was added to the bottom of the 48-well plates to form a hemispherical shape. After solidification, organoid medium was applied (DMEM/F12 (Gibco, Cat# A4192001) medium supplemented with GlutaMax (Gibco, Cat# 35050079), N2 Supplement (Gibco, Cat# 17502001), B27 Supplement (Gibco, Cat# 17502001), 1 mM N-acetylcysteine (Sigma, Cat# A0737), 10 mM Nicotinamide (Sigma, Cat# N0636), 50 ng/mL human epidermal growth factor (PEPROTECH, Cat# 900-M05), and 100 ng /mL fibroblast growth factor-10 (PEPROTECH, Cat# 400-29A), 10  $\mu$ m Y27632 (Selleck, Cat# S6390), 500 nM A83-01 (Selleck, Cat# S7692), 10% conditioned medium containing Wnt3a (homemade), Noggin and R-spondin (homemade). After 3–4 days, the esophageal organoids would be seen. For passage, organoids were dissociated by TrypLE™ Express (Gibco, Cat# 12605010) at 37 °C for 10 min. The process disrupted the spherical organoids into cell aggregates or single cell that were then embedded in fresh Matrigel.

#### Genomic DNA extraction, PCR amplification

Five-hundred microlitre DNA lysis buffer (homemade, 10 mM Tris, 100 mM NaCl, 10 mM EDTA, 0.5% SDS 0.4 mg/ml proteinase K) was used to lysis cells at 55 °C for at least 2 h. The supernatant was combined with 500  $\mu$ l isopropanol, mixed, and centrifuged at 13,000 rpm for 10 min. Removed the supernatant and then washed with 1 mL 75% ethanol once. Volatilized at 55 °C for about 5 min, the DNA pellet was then dissolved in double-distilled water. One-hundred nanogram DNA was used for each genotyping PCR reaction. PCR reactions were performed on a ProFlex™ PCR system (Applied Biosystems).

#### Plasmid construction and virus production

The full-length *TPM4* and *Tpm4* cDNA were cloned into the LentiCRISPR-v2 puro (RRID: Addgene\_52961) lentiviral vector, the full-length *Kras*<sup>G12D</sup> and *cMyc* cDNA was cloned into the MSCV-cDNA -IRES-luciferase retroviral vector, and the guide RNAs were cloned into the vectors as previously described. All plasmids' sequences were confirmed by Sanger sequencing. Subsequently, these resulting plasmids were transfected with the helper plasmids psPAX2 (RRID: Addgene\_12260), pMD2.G (RRID: Addgene\_12259) into HEK293T (CCLV Cat# CCLV-RIE 1018, RRID: CVCL\_0063) cells by the calcium phosphate transfection method. The virus in the supernatant at 36 and 48 h were respectively collected. And all the sequences of sgRNA used in this study were listed in Supplementary Table 11.

#### CRISPR-Cas9 knockout organoids

Organoids were disassociated with TrypLE™ for 15–30 min at 37 °C in a 15 ml conical tube. Pipetted several times every 5 min until most of them became single cells and then centrifuged at 400g for 5 min. The cells were resuspended with lentiviral supernatant and centrifuged for 1 h at 800g followed by another 2 h incubating at 37 °C. After that, cells were harvested and seeded with Matrigel in a 48-well plate. Two days later, 1  $\mu$ g ml<sup>-1</sup> puromycin (Gibco, #A1113803) was added to select the positive cells for several passages. To validate the efficiency of CRISPR in purified clones, their genome was extracted for PCR reaction and then T7E1 enzymes were performed. Sanger sequencing was used to validate the mutation status finally.

#### Subcutaneous tumor growth assay

After one week of cells/organoids transplantation, the length and width of the tumor were measured using calipers every 3 days. Tumor size was calculated with the formula like this: (length  $\times$  width<sup>2</sup>)  $\times$  0.5. For AZ960 treatment, after transplantation 1 week, mice were injected intraperitoneally with AZ960 (Sichuan Shuyan Ltd., #905586-69-8) at a dose of 40 mg/kg or vehicle daily.

#### Living image assay

After transplanted subcutaneously or orthotopically with organoids, mice would be periodically imaged to detect the luciferase fluorescence signal intensity with IVIS Spectrum (PerkinElmer) system. The mice were anesthetized with isoflurane after 250  $\mu$ l 15 mg/ml D-luciferin, Sodium salt (Bio Vision, #7903-1G) in PBS intraperitoneally injected.

#### Subcutaneous and orthotopic implantation

For the organoid mouse model, the tumor-forming background we selected was the mutation of *Pten*, *Smad4*, and overexpression of *Myc* and *Kras*<sup>G12D</sup> in the esophageal organs of *Trp53*<sup>-/-</sup>; Cas9 mice (TPSCK). The organoid model of this genotype could form ESCC at four weeks in immunodeficient mice. The H&E and IHC staining had confirmed the pathological characteristic of ESCC generated from TPSCK organoids. Of note, organoids should be dissociated to approximately single cells using TrypLE™ mentioned above before subcutaneous and orthotopic transplantation. For the AZ960 treatment experiment, one well of TPSCK/ TPSCK *Tpm4* tumor organoids were injected subcutaneously into the flank of nude mice (6–8 weeks old, male, six mice per group). For tumor formation of orthotopic implantation from primary TPSCK/TPSCK *Tpm4* overexpressed organoids, 2 wells of organoids were collected with 30  $\mu$ l of the mixture (1:1 Matrigel/medium ratio) for orthotopically injecting into the esophagus of nude mice (6–8 weeks old, male, six mice per group). The nude mice were anesthetized with isoflurane, and an incision was made through the left upper abdominal pararectal line and peritoneum. The stomach and esophagus were carefully exposed. With fixing the stomach by forceps, the cells were slowly injected into the junction of the esophagus and stomach with an insulin needle. The abdominal wall and skin were closed with 5-0 Ethicon sutures (ET-661G). The entire process was performed on a clean bench, and the wound was disinfected with iodophor after the operation.

#### Immunoblotting

Cells were lysed in lysis buffer (50 mM Tris-HCl pH 7.4, 150 mM NaCl, 1% Triton-X100 supplemented with 1 mM phenylmethylsulfonyl fluoride before use). Cell lysate in SDS loading buffer (50 mM Tris-HCl (pH 6.8), 10% glycerol, 2% SDS, 1% 2-mercaptoethanol, 0.1% bromophenol blue) was boiled and analyzed by SDS-PAGE (Bio-Rad) and transferred to 0.45  $\mu$ m polyvinylidene fluoride (PVDF) membranes with transfer buffer (25 mM Tris base, 190 mM glycine, 20% methanol, and 80% ddH<sub>2</sub>O, pH 8.3). Then the PDVF membranes were blocked with 2% milk (2% bovine serum albumin for phosphorylated antibody) with TBST buffer (150 mM NaCl, 20 mM Tris-HCl pH 7.6, and 0.1% Tween-20) for 1 h before incubation with primary antibodies for 2 h at room temperature or 4 °C overnight. After washing with TBST, membranes were incubated with secondary antibody for 1 h and then washed three times with TBST. Cells need to be treated with 50 ng/mL IFN- $\gamma$  48 h in advance for phosphorylated protein detection. The antibodies, including anti-Jak2 (Cell Signaling Technology Cat# 3230, RRID: AB\_2128522), anti-pJak2 (Cell Signaling Technology Cat# 3771, RRID: AB\_330403), anti-STAT3 (Cell Signaling Technology Cat# 12640, RRID: AB\_2629499), anti-pSTAT3 (Fluidigm Cat# 3158005, RRID: AB\_2661827), anti-TPM4 (Abcam Cat# 181085), anti-SOX2 (Abcam Cat# ab92494, RRID: AB\_10585428), anti-GAPDH (BioXcell, #BX-008) were used in this study.



H&E, immunohistochemistry, and immunofluorescence staining Fresh tumor tissues were fixed in 4% PFA, embedded in paraffin, and cut into 5- $\mu$ m -thick sections. The sections were stained with hematoxylin for 1 min to identify the cell nucleus and eosin for 30 s to identify the cytoplasm. For immunohistochemistry, the sections were incubated with antibodies of pJAK2 (Abcam Cat# ab32101, RRID: AB\_775808), pSTAT3 (Abcam Cat# ab267373), Sox2 (Abcam Cat# ab92494, RRID: AB\_10585428) and Ki67 (Abcam Cat# ab16667, RRID: AB\_302459) overnight at 4 °C after process of antigen retrieval and block with 2% goat serum. After being washed with PBST, the sections were incubated with secondary antibodies at room temperature for 1 hour. Then DAB was applied for signal detection (ZSGB-BIO, ZLI-9018), followed by staining with hematoxylin for seconds and rinsed with water for 1 min. Then the sections were blocked with the Permount mounting medium. For immunofluorescence staining, deparaffinization and rehydration procedures were as described above, antibodies of P63 (Abcam Cat# ab735, RRID: AB\_305870), CK13 (Leica Biosystems Cat# NCL-CK13, RRID: AB\_563792), CK14 (ARP American Research Products Cat# 03-GPCK14, RRID: AB\_1541011) and CK5 (Thermo Fisher Scientific Cat# MA5-17057, RRID: AB\_2538529) were used. After secondary antibodies were added, the slides were mounted by mounting solution with DAPI (Solarbio, Cat#C0060). And the images were taken with fluorescence microscopy (Zeiss, 880).

#### Organoid growth assay

For organoid growth assay, 2000 single cells were cultured in a 96-well plate embedded in 10  $\mu$ l Matrigel (3 independent biological replicates were contained in each group). And the number of organoids was counted after 5 days of culture.

#### JAK2 inhibitor treatment assay

For drug response, 1000 single cells were cultured in a 96-well plate embedded in 10  $\mu$ l Matrigel and then subjected to different concentrations of AZ960 (Selleck, #S2214) or DMSO. After treatment for 48 h, the number of organoids was counted in each well. The number of organoids in the experimental group normalized by the DMSO group was the cell viability.

#### Statistical analysis

The unpaired Student's *t*-test (two-tailed) was used to evaluate the gene expression, tumor volume, organoid number. Expression differences among three groups were analyzed by ANOVA. The statistical analysis was performed using GraphPad Prism (RRID: SCR\_002798) version 9. Data are represented as mean  $\pm$  SEM. All results are considered statistically significant at  $P < 0.05$ .

#### Bulk RNA-seq data analysis

TRIzol reagent (Invitrogen, Cat# 15596018) was used to extract the total RNA in each sample. The library was constructed by Illumina Stranded mRNA Sample Preparation Kit (NEB, #E7770) followed by the manufacturer's protocol. And the Illumina NovaSeq 6000 was used to sequence with paired-end 150 bp.

The RNA-seq data were aligned by STAR (v2.6) (RRID: SCR\_004463) with genome reference mm10 for mice and hg19 for humans. The quantification of transcripts in each sample was implemented by R packages, GenomicAlignments, DESeq2 (RRID: SCR\_000154) was used to identify the differentiated expressed genes with  $|\log_2\text{-FC}| > 1$  and  $p$ -value  $< 0.05$ . The GO and KEGG (RRID: SCR\_012773) enrichment was performed by clusterProfiler. The normalized data, generated by DESeq, was processed by log-transformation for subsequent visualization and quantification of specific signatures.

#### Single-cell RNA-seq data analysis

The library was prepared by using the 10 $\times$  genomics platform, Chromium Single Cell 3' Reagent Kits v3, following the manufacturer's protocol. The average 7000 total cells (from 7000 to 9000) of

each sample were added to the individual channel with an average recovery of 5525 cells. The Gel Beads were used to partition cells in the Chromium instrument. After library generating, Illumina NovaSeq 6000 was used to sequence with paired-end 150 bp for each sample.

The cellranger (v2) was used to align the clean reads with genome reference hg19. The Seurat (v3) pipeline was used to quantify and visualize single-cell RNA data. The poor-quality cells whose detectable gene number is lower than 200 or higher than 7,500, and whose mitochondrial genes expression proportion is higher than 20%, would be removed. The poor-quality genes that detected lower than 3 cells also would be removed.

The vst model was used to identify 4000 high variable genes for subsequent processing. To reduce the systematic and random batch effects between samples, the embedding of harmony was used to cluster the population and reduce dimension. And the 30 statistically significant PCA components were used for harmony embedding calculation. The t-SNE and UMAP maps were generated from the Seurat formula, RunTSNE, and RunUMAP, respectively. The phase maps were generated from phateR. Each subpopulation was identified by the classical signatures, and top markers could be available. The inferCNV was used to identify the malignant squamous epitheliums cells compared with immune cells and other epitheliums in an individual sample.

The monocle (v3) was used to inference the normal squamous epithelium development lineage. The principal graph of phate and UMAP map was learned by reversed graph embedding firstly. Then, the formula  $\text{order\_cells}$ , assigned the pseudo-time value in each cell based on their projection. Besides, slingshot also was used to confirm the predicted tumor lineage with 300 approximate points.

#### Gene-sets enrichment and gene signatures identified

All the enrichment gene-sets were downloaded from GSEA (v7 database). Gene-sets variation analysis (GSVA) in omics data was performed on the pseudo-bulk transformation data, which normalized single-cell data into 100 bins to accelerate the speed and save the memory of calculation. The proliferation scores of the epithelium subpopulation were calculated and quantified based on the GSVA scores from proliferation-related pathways. Besides, the cell cycle status in the normal single-cell esophageal squamous epitheliums cells was calculated and quantified by CellCycleScoring with g2m genes and s specific genes.<sup>53,54</sup> The percentages of cell cycle status had been summarized and visualized by ggplot2.

The signatures of subpopulations were detected by FindAllMarkers with 0.25 thresholds and 0.25 minimum fractions. And the BS, BK, and DK signatures were identified with  $\text{pct.2} < 0.4$ . The progression signatures of ESCC were identified from commonly upregulated in stage II/III patients compared with stage I patients in TCGA-ESCA<sup>55</sup> and GSE160269.<sup>30</sup> And the CCI signatures were identified by CCI SE exclusively upregulated genes with fold-change  $> 0.2$ , base-mean  $> 0.8$ , and  $\text{pct.1} > 0.8$  compared with normal SE.

#### CCI scores quantification

To quantify the expression levels of specific pathways or gene-sets on transcriptome data, the normalized matrix was used to calculate, as following formula processing

$$SSi = \frac{\sum_{i=1}^N Mi}{\text{Number of genes}}$$

$Mi$  means gene value in sample/single-cell  $i$ ,  $N$  means the number of pathways/gene-sets, and  $SSi$  means signature score in sample/single-cell  $i$ .

After BS, BK, and DK signatures were quantified in each sample/single cell, ggtern was used to quantify the distribution of them on BS, BK, and DK dimensions and project the distribution of dots/samples on the ternary maps. The two-dimension kernel density

of distribution on the ternary map was estimated by `stat_densify_tern` with `geom = 'polygon'` and 500 bins. To better compare the distribution on the ternary map in different conditions, the same measurement scale of three axes should be reserved in each sample/single cell. Hence, the limits were set to the same number in each sample/single cell. And all visualization was complemented by package `ggtern` and `ggplot2`.

To make the qualitative ternary map quantitative, the CCI scores were created to define the distribution of each sample/single cell among BS, BK, and DK dimensions. Firstly, we defined three angles  $\theta_1$ – $\theta_3$  to describe the variation among three axes with sample  $S$  ( $a$ ,  $b$ ,  $c$ ), respectively. Each angle exists between the one axis and the segments included the origin point and sample  $S$  ( $a$ ,  $b$ ,  $c$ ). Based on the regular cosine law, we could generate the cosine value of each angle, as the following formula:

$$\cos \theta_1 = \frac{a^2 + \sqrt{[2]a^2 + b^2 + c^2} - \sqrt{[2]b^2 + c^2}}{2 \times a \times \sqrt{[2]a^2 + b^2 + c^2}} = \frac{a}{\sqrt{[2]a^2 + b^2 + c^2}}$$

$$\cos \theta_2 = \frac{b^2 + \sqrt{[2]a^2 + b^2 + c^2} - \sqrt{[2]a^2 + c^2}}{2 \times b \times \sqrt{[2]a^2 + b^2 + c^2}} = \frac{b}{\sqrt{[2]a^2 + b^2 + c^2}}$$

$$\cos \theta_3 = \frac{c^2 + \sqrt{[2]a^2 + b^2 + c^2} - \sqrt{[2]a^2 + b^2}}{2 \times c \times \sqrt{[2]a^2 + b^2 + c^2}} = \frac{c}{\sqrt{[2]a^2 + b^2 + c^2}}$$

The  $a$ ,  $b$ ,  $c$  represented the BS, BK, and DK signatures value of sample  $S$ . And the  $\theta_1$ ,  $\theta_2$ ,  $\theta_3$  represented the angles generated from sample  $S$ .

Next, we calculate the inverse value of standard deviation among three cosine values, and after multiplying by a constant  $f$ , CCI scores would be generated in each sample/single cell.

$$\text{CCI Score} = \frac{f}{sd(\cos \theta_1, \cos \theta_2, \cos \theta_3)}$$

#### CCI signature identification

To better identify the CCI signature genes, we designed a workflow for analyze. Firstly, we analyzed the distribution of the CCI scores of ESCC cells. The area under the curve (AUC) of CCI scores distribution had been calculated. Of note, the AUC curve was significantly fitted to the GLM binomial distribution. And the first intersection point between the AUC curve and predicted curve would be the cut-point of CCI scores for ESCC cells on the single-cell ternary diagram for calculating the CCI signature genes. The ESCC cells with their CCI scores higher than the cut point were used for identifying the CCI signature genes. Then we calculated the CCI signature genes by comparing these CCI positive cells with other squamous epithelium cells with the standards of fold-change > 0.25, base-mean > 0.8, and pct.1 > 0.8. And 195 genes were identified as the CCI signature genes for ESCC (Supplementary Table 6).

#### Omics data analysis of ESCC database

The array data of transcriptome were downloaded by GEOquery, including GSE20347,<sup>56</sup> 17 paired normal and ESCC samples; GSE23400\_1,<sup>57</sup> 53 paired normal and ESCC samples; GSE23400\_2,<sup>58</sup> 51 paired normal and ESCC samples; GSE70409,<sup>59</sup> 17 paired normal and ESCC samples; GSE53624,<sup>36</sup> 119 paired normal and ESCC samples. The probe id in GSE53624 had been converted to gene name by GPL18109 platform, and others had done by feature data of getGEO results.

The TPM data of bulk RNA-seq and the data of proteomics from GSE149609<sup>37</sup> had been used in this study. In TCGA-ESCA cohorts,

the 81 ESCC patients were extracted to analyze. Besides, the transcriptome data, survival information, and clinical classification of ESCC patients, from TCGA-ESCC and GSE53624 cohorts, were used to validate the clinical correlation between CCI and ESCC diagnosis. The Cox regression and the Kaplan-Meier survival curves were calculated and visualized by R package, `survminer` and `survival`. The determination of the cut-points for numerical variables in survival plots was based on the `maxstat` package, which also was integrated into the `survminer` package.

The single-cell data of normal esophageal tissue was downloaded from HCA projects and the human cell landscape. And squamous epitheliums were extracted following the origin annotation. After the same scRNA data pipeline analysis, the HCA and our normal sample were combined to reduce the dimension, and the UMAP map of normal esophageal squamous epitheliums was generated, and the expression levels of classical genes and gene-sets were projected and summarized.

The single-cell data of ESCC were downloaded from GSE160269. The squamous epitheliums were identified following the origin annotation. The clinical information of ESCC patients was used to divide patients into two groups including stage I and stage II/III.

#### DATA AVAILABILITY

The raw data of RNA-seq and scRNA-seq data in this study are deposited in NCBI GEO (GSE188955).

#### MATERIALS AVAILABILITY

The processed files and codes can be found in [GitHub](#).

#### ACKNOWLEDGEMENTS

We thank the members of the CC-LY laboratory for their technical support and suggestions, and the Core Facilities of West China Hospital. This work was supported by the National Key R&D Program of China (2017YFA0505600), the National Natural Science Foundation of China (81522003, 82170171, 81770157, and 82003156), the Sichuan Science and Technology Program (2020YFQ0059, 2018JZ0077, 2017TJPT0005, 2022YFS0171 and 2022YFS0205), 1.3.5. Project for Disciplines of Excellence, West China Hospital, Sichuan University, and the innovation initiative of Sichuan University, Grant no. 2018SCUH0060. We thank the Chengdu OrganoidMed Medical Laboratory for their technical support. We also thank BioRender (biorender.com) for providing items for drawing scheme graphs.

#### AUTHOR CONTRIBUTIONS

X.P., J.W., L.G., and F.N. designed and performed experiments, analyzed the data, and wrote the paper. X.P., J.W., L.G., and F.N., J.D., X.C., A.Z., L. Zhao, L. Zhang, M.Z., X.W., M.W., H.L., S.D., P.T., and J.C. performed experiments and analyzed data. Y.L., B.H., and C.C. conceived the project, designed experiments, analyzed data, and wrote the paper. All authors have read and approved the article.

#### ADDITIONAL INFORMATION

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41392-022-00946-8>.

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

#### REFERENCES

- Hanahan, D. & Weinberg, R. A. The hallmarks of cancer. *Cell* **100**, 57–70 (2000).
- Hanahan, D. Hallmarks of cancer: new dimensions. *Cancer Discov.* **12**, 31–46 (2022).
- Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: the next generation. *Cell* **144**, 646–674 (2011).
- Battle, E. & Clevers, H. Cancer stem cells revisited. *Nat. Med.* **23**, 1124–1134 (2017).
- Blanpain, C., Horsley, V. & Fuchs, E. Epithelial stem cells: turning over new leaves. *Cell* **128**, 445–458 (2007).
- Kelly, L. M. & Gilliland, D. G. Genetics of myeloid leukemias. *Annu. Rev. Genomics Hum. Genet.* **3**, 179–198 (2002).

7. Magee, J. A., Piskounova, E. & Morrison, S. J. Cancer stem cells: impact, heterogeneity, and uncertainty. *Cancer Cell*. **21**, 283–296 (2012).
8. Pardal, R., Clarke, M. F. & Morrison, S. J. Applying the principles of stem-cell biology to cancer. *Nat. Rev. Cancer* **3**, 895–902 (2003).
9. Perez-Losada, J. & Balmain, A. Stem-cell hierarchy in skin cancer. *Nat. Rev. Cancer* **3**, 434–443 (2003).
10. Reya, T., Morrison, S. J., Clarke, M. F. & Weissman, I. L. Stem cells, cancer, and cancer stem cells. *Nature* **414**, 105–111 (2001).
11. Sanchez-Danes, A. & Blanpain, C. Deciphering the cells of origin of squamous cell carcinomas. *Nat. Rev. Cancer* **18**, 549–561 (2018).
12. Chen, C. et al. Cancer-associated IDH2 mutants drive an acute myeloid leukemia that is susceptible to Brd4 inhibition. *Genes Dev.* **27**, 1974–1985 (2013).
13. Chen, C. et al. MLL3 is a haploinsufficient 7q tumor suppressor in acute myeloid leukemia. *Cancer Cell*. **25**, 652–665 (2014).
14. Sell, S. Stem cell origin of cancer and differentiation therapy. *Crit. Rev. Oncol. Hematol.* **51**, 1–28 (2004).
15. Visvader, J. E. Cells of origin in cancer. *Nature* **469**, 314–322 (2011).
16. Berenson, J. R., Vescio, R. A. & Said, J. Multiple myeloma: the cells of origin—a two-way street. *Leukemia* **12**, 121–127 (1998).
17. Quintana, E. et al. Efficient tumour formation by single human melanoma cells. *Nature* **456**, 593–598 (2008).
18. Blanpain, C. & Simons, B. D. Unravelling stem cell dynamics by lineage tracing. *Nat. Rev. Mol. Cell Biol.* **14**, 489–502 (2013).
19. Thankamony, A. P. et al. Cancer stem cell plasticity—a deadly deal. *Front. Mol. Biosci.* **7**, 79 (2020).
20. Niederst, M. J. et al. RB loss in resistant EGFR mutant lung adenocarcinomas that transform to small-cell lung cancer. *Nat. Commun.* **6**, 6377 (2015).
21. Han, X. et al. Transdifferentiation of lung adenocarcinoma in mice with Lkb1 deficiency to squamous cell carcinoma. *Nat. Commun.* **5**, 3261 (2014).
22. LaFave, L. M. et al. Epigenomic state transitions characterize tumor progression in mouse lung adenocarcinoma. *Cancer Cell*. **38**, 212–228 e213 (2020).
23. Marjanovic, N. D. et al. Emergence of a high-plasticity cell state during lung cancer evolution. *Cancer Cell*. **38**, 229–246 e213 (2020).
24. Ku, S. Y. et al. Rb1 and Trp53 cooperate to suppress prostate cancer lineage plasticity, metastasis, and antiandrogen resistance. *Science* **355**, 78–83 (2017).
25. Mu, P. et al. SOX2 promotes lineage plasticity and antiandrogen resistance in TP53- and RB1-deficient prostate cancer. *Science* **355**, 84–88 (2017).
26. Quintanal-Villalonga, A. et al. Lineage plasticity in cancer: a shared pathway of therapeutic resistance. *Nat. Rev. Clin. Oncol.* **17**, 360–371 (2020).
27. Cancer Genome Atlas, N. Comprehensive molecular portraits of human breast tumours. *Nature* **490**, 61–70 (2012).
28. Perou, C. M. et al. Molecular portraits of human breast tumours. *Nature* **406**, 747–752 (2000).
29. Dotto, G. P. & Rustgi, A. K. Squamous cell cancers: a unified perspective on biology and genetics. *Cancer Cell*. **29**, 622–637 (2016).
30. Zhang, X. et al. Dissecting esophageal squamous-cell carcinoma ecosystem by single-cell transcriptomic analysis. *Nat. Commun.* **12**, 5291 (2021).
31. Kalabis, J. et al. A subpopulation of mouse esophageal basal cells has properties of stem cells with the capacity for self-renewal and lineage specification. *J. Clin. Invest.* **118**, 3860–3869 (2008).
32. DeWard, A. D., Cramer, J. & Lagasse, E. Cellular heterogeneity in the mouse esophagus implicates the presence of a nonquiescent epithelial stem cell population. *Cell Rep.* **9**, 701–711 (2014).
33. Morrisey, E. E. & Rustgi, A. K. The lung and esophagus: developmental and regenerative overlap. *Trends Cell Biol.* **28**, 738–748 (2018).
34. Han, X. et al. Construction of a human cell landscape at single-cell level. *Nature* **581**, 303–309 (2020).
35. Madisson, E. et al. scRNA-seq assessment of the human lung, spleen, and esophagus tissue stability after cold preservation. *Genome Biol.* **21**, 1 (2019).
36. Li, J. et al. LncRNA profile study reveals a three-lncRNA signature associated with the survival of patients with oesophageal squamous cell carcinoma. *Gut* **63**, 1700–1710 (2014).
37. Cao, W. et al. Multi-faceted epigenetic dysregulation of gene expression promotes esophageal squamous cell carcinoma. *Nat. Commun.* **11**, 3675 (2020).
38. Lomnyska, M. I. et al. Differential expression of ANXA6, HSP27, PRDX2, NCF2, and TPM4 during uterine cervix carcinogenesis: diagnostic and prognostic value. *Br. J. Cancer* **104**, 110–119 (2011).
39. Qi, Y. et al. Comparative proteomic analysis of esophageal squamous cell carcinoma. *Proteomics* **5**, 2960–2971 (2005).
40. Li, Y. et al. Phosphoproteomics reveals therapeutic targets of esophageal squamous cell carcinoma. *Signal Transduct. Target Ther.* **6**, 381 (2021).
41. Lin, J. J., Eppinga, R. D., Warren, K. S. & McCrae, K. R. Human tropomyosin isoforms in the regulation of cytoskeleton functions. *Adv. Exp. Med Biol.* **644**, 201–222 (2008).
42. Zhao, X., Jiang, M. & Wang, Z. TPM4 promotes cell migration by modulating F-actin formation in lung cancer. *Onco Targets Ther.* **12**, 4055–4063 (2019).
43. Niwa, H., Ogawa, K., Shimosato, D. & Adachi, K. A parallel circuit of LIF signalling pathways maintains pluripotency of mouse ES cells. *Nature* **460**, 118–122 (2009).
44. Gen, Y., Yasui, K., Nishikawa, T. & Yoshikawa, T. SOX2 promotes tumor growth of esophageal squamous cell carcinoma through the AKT/mammalian target of rapamycin complex 1 signaling pathway. *Cancer Sci.* **104**, 810–816 (2013).
45. Bass, A. J. et al. SOX2 is an amplified lineage-survival oncogene in lung and esophageal squamous cell carcinomas. *Nat. Genet.* **41**, 1238–1242 (2009).
46. Wu, Z. et al. Reprogramming of the esophageal squamous carcinoma epigenome by SOX2 promotes ADAR1 dependence. *Nat. Genet.* **53**, 881–894 (2021).
47. Cancer Genome Atlas Research, N. et al. Integrated genomic characterization of oesophageal carcinoma. *Nature* **541**, 169–175 (2017).
48. Song, Y. et al. Identification of genomic alterations in oesophageal squamous cell cancer. *Nature* **509**, 91–95 (2014).
49. Wu, C. et al. Joint analysis of three genome-wide association studies of esophageal squamous cell carcinoma in Chinese populations. *Nat. Genet.* **46**, 1001–1006 (2014).
50. Thomas, S. J., Snowden, J. A., Zeidler, M. P. & Danson, S. J. The role of JAK/STAT signalling in the pathogenesis, prognosis and treatment of solid tumours. *Br. J. Cancer* **113**, 365–371 (2015).
51. Beebe, K., Lee, W. C. & Micchelli, C. A. JAK/STAT signaling coordinates stem cell proliferation and multilineage differentiation in the *Drosophila* intestinal stem cell lineage. *Dev. Biol.* **338**, 28–37 (2010).
52. Garraway, L. A. & Sellers, W. R. Lineage dependency and lineage-survival oncogenes in human cancer. *Nat. Rev. Cancer* **6**, 593–602 (2006).
53. Tirosh, I. et al. Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* **352**, 189–196 (2016).
54. Nestorowa, S. et al. A single-cell resolution map of mouse hematopoietic stem and progenitor cell differentiation. *Blood* **128**, e20–31 (2016).
55. Kim, J. et al. Integrated genomic characterization of oesophageal carcinoma. *Nature* **541**, 169–175 (2017).
56. Hu, N. et al. Genome wide analysis of DNA copy number neutral loss of heterozygosity (CNNLOH) and its relation to gene expression in esophageal squamous cell carcinoma. *BMC Genomics*. **11**, 576 (2010).
57. Hyland, P. L. et al. Pathway, in silico and tissue-specific expression quantitative analyses of oesophageal squamous cell carcinoma genome-wide association studies data. *Int. J. Epidemiol.* **45**, 206–220 (2016).
58. Su, H. et al. Global gene expression profiling and validation in esophageal squamous cell carcinoma and its association with clinical phenotypes. *Clin. Cancer Res.* **17**, 2955–2966 (2011).
59. Chen, Y. K. et al. Plasma matrix metalloproteinase 1 improves the detection and survival prediction of esophageal squamous cell carcinoma. *Sci. Rep.* **6**, 30057 (2016).



**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) 2022