

# Integrative proteogenomic characterization of early esophageal cancer

**Lingling Li**

School of Life Sciences, Human Phenome Institute, Fudan University

**Dongxian Jiang**

Department of Pathology, Zhongshan Hospital, Fudan University

**Qiao Zhang**

Fudan University

**Hui Liu**

Henan Normal University

**Zhaoyu Qin**

**Fujiang Xu**

School of Life Sciences, Fudan University <https://orcid.org/0000-0002-7900-5319>

**Dan Shen**

Fudan University

**Haixing Wang**

Fudan University

**Chunmei Guo**

Fudan University

**Jinwen Feng**

Fudan University <https://orcid.org/0000-0003-1683-4280>

**Yang Liu**

Institute of Biomedical Sciences, Human Phenome Institute, Fudan University <https://orcid.org/0000-0003-4874-4773>

**Weijie Chen**

Fudan University

**Xue Zhang**

Fudan University

**Lin Bai**

Fudan university

**Sha Tian**

Fudan University

**Subei Tan**

Fudan University

**Chen Xu**

Neuro-Oncology Branch, Center for Cancer Research, National Cancer Institute

**Qi Song**

**Yalan Liu**

Fudan University

**Yunshi Zhong**

Fudan University

**Tianyin Chen**

Fudan University

**Pinghong Zhou**

Endoscopy Center, Zhongshan Hospital, Fudan University, Shanghai, China

**Jian-Yuan Zhao**

Fudan University <https://orcid.org/0000-0001-9861-0330>

**Yingyong Hou**

zhongshan hospital <https://orcid.org/0000-0001-9156-2305>

**Chen Ding** (✉ [chend@fudan.edu.cn](mailto:chend@fudan.edu.cn))

Fudan university

---

## Article

**Keywords:** Esophageal squamous cell carcinoma, carcinogenesis, proteogenomics, PGK1

**Posted Date:** October 12th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-227745/v2>

**License:**   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Integrative proteogenomic characterization of early esophageal cancer

Lingling Li<sup>1,4</sup>, Dongxian Jiang<sup>1,4</sup>, Qiao Zhang<sup>1,4</sup>, Hui Liu<sup>3,4</sup>, Zhaoyu Qin<sup>1,4</sup>, Fujiang Xu<sup>1,4</sup>, Dan Shen<sup>1,4</sup>, Haixing Wang<sup>1,4</sup>, Chunmei Guo<sup>1</sup>, Jinwen Feng<sup>1</sup>, Yang Liu<sup>1</sup>, Weijie Chen<sup>1</sup>, Xue Zhang<sup>1</sup>, Lin Bai<sup>1</sup>, Sha Tian<sup>1</sup>, Subei Tan<sup>1</sup>, Chen Xu<sup>1</sup>, Qi Song<sup>1</sup>, Yalan Liu<sup>1</sup>, Yunshi Zhong<sup>1,2</sup>, Tianyin Chen<sup>1,2</sup>, Pinghong Zhou<sup>1,2,\*</sup>, Jian-Yuan Zhao<sup>1,\*</sup>, Yingyong Hou<sup>1,\*</sup>, Chen Ding<sup>1,5,\*</sup>

<sup>1</sup>State Key Laboratory of Genetic Engineering and Collaborative Innovation Center for Genetics and Development, School of Life Sciences, Institute of Biomedical Sciences, Human Phenome Institute, Department of Pathology, Zhongshan Hospital, Fudan University, Shanghai 200433, China;

<sup>2</sup>Endoscopy Center and Endoscopy Research Institute, Zhongshan Hospital, Fudan University, Shanghai 200032, P.R. China;

<sup>3</sup>State Key Laboratory Cell Differentiation and Regulation, Overseas Expertise Introduction Center for Discipline Innovation of Pulmonary Fibrosis, (111 Project), College of Life Science, Henan Normal University, Xinxiang, Henan 453007, China.

<sup>4</sup>These authors contributed equally

<sup>5</sup>Lead Contact

\*Correspondence: zhou.pinghong@zs-hospital.sh.cn (P.Z.), zhaoyj@fudan.edu.cn (J.Y.Z.), hou.yingyong@zs-hospital.sh.cn (Y.H.), chend@fudan.edu.cn (C.D.)

## **Abstract**

We performed a comprehensive multi-omics analysis of 756 tissue-tumor-samples from 124 esophageal squamous cell carcinoma phase (ESCC) patients, covering 9 histopathological stages in 3 phases as nontumor phase (NT phase), intraepithelial neoplasia phase (IEN phase), and ESCC phase. Proteogenomics elucidated the stage-specific molecular characterization and defined the cancer-driving waves along with the mutation accumulation in EC progression. The integrated multi-omics uncovered the chromosome 3q gain was the key event in the transition from the NT to IEN phase, disclosed the top mutation of *TP53* enhanced cell cycle and DNA replication in the IEN phase, and revealed the ESCC phase mutations of *AKAP9* and *MCAF1* elevated glycolysis and Wnt signaling, respectively. Furthermore, the trajectory analysis identified 6 major tracks related to different clinical features during ESCC progression. Growingly enhanced and hyperphosphorylated phosphoglycerate kinase 1 (PGK1, S203) was detected and considered as a drug target in ESCC progression. Collectively, this study provides insight into the understanding of ESCC molecular mechanism and a valuable resource for the development of therapeutic targets.

## **Keywords**

Esophageal squamous cell carcinoma, carcinogenesis, proteogenomics, PGK1

## Introduction

Esophageal cancer (EC) is a malignant gastrointestinal carcinoma, ranking the seventh most common cancer and the sixth leading cause of cancer-related death worldwide<sup>1,2</sup>. Esophageal adenocarcinoma (EAC) and esophageal squamous cell carcinoma (ESCC) are the two major histologic subtypes of EC, of which EAC is more prevalent in Western countries<sup>3</sup>, whereas ESCC predominantly occurs in Eastern Asia, particularly in China and Japan<sup>4</sup>, indicating the diverse lifestyle of the countries as a major etiological factor of EC. The other etiological factors of EC include gender, age<sup>5</sup>, and the habit of drinking/smoking<sup>6-9</sup>. However, the molecular signatures that are associated with the risk factors in EC progression are still unknown.

Although, several large-scale ESCC cohorts, including The Cancer Genome Atlas (TCGA)<sup>10</sup>, presented comprehensive analyses of the genomic aberration and identified high mutation frequencies of genes associated with cell cycle and cell differentiation in ESCC, including *TP53*, *FAT3*, and *FAM135B*<sup>11-13</sup>. However, the landscape of kinetic genomic aberrations and their consequences on proteomic/phosphoproteomic alterations during ESCC carcinogenesis remain largely unknown. Additionally, the molecular mechanism of the diversity and tumor heterogeneity of ESCCs is poorly understood, imposing a challenge for developing ESCC therapeutic strategies.

Histologically, the esophageal wall includes the epithelium, lamina propria, muscularis mucosa, submucosa, muscularis externa, and adventitia<sup>14</sup>. The depth of infiltration of the cancer cells, determining the stage of lesions, was measured at the deepest point of their penetration in corresponding layers<sup>15</sup>. Surgery is the predominant curative treatment strategy in advanced stages where the cells invade muscularis externa (T2 to T4 stages), with poor quality of life (QOL) and low five-year survival rate (< 30%)<sup>16</sup>. Though the recent advances of endoscopic submucosal dissection (ESD)<sup>17,18</sup> have achieved the early detection of ESCC patients (T1 stage) with higher QOL and significantly improved overall survival rate (> 90%)<sup>19-21</sup>, the complexity of the early ESCC progression and the extreme trace amount of tissue samples in different stages have limited in portraying the multi-omics molecular landscape of ESCC.

In this study, we performed a comprehensive multi-omics analysis of 756 trace-tumor-samples from 124 ESCC patients, covering 9 histopathological stages in 3 phases as nontumor phase (NT phase), intraepithelial neoplasia phase (IEN phase), and ESCC phase. The integrative multi-omics dataset elucidated the stage-specific molecular characterization and defined the cancer-driving waves along with

the mutation accumulation in EC progression. Proteogenomics uncovered the key events in the transmit of the phases and the trajectory analysis identified 6 major tracks and their molecular characteristics during the carcinogenesis of early ESCC.

## Results

### Overview of proteogenomic landscape in ESCC progression

We performed multi-omics-based profiling of trace 756 samples collected from 124 ESCC patients who had not experienced prior chemotherapy or radiotherapy. Subsequently, 22 substages during ESCC progression from healthy esophageal tissue to tumor development were established for these samples following WHO and Japanese pathology diagnostic criteria<sup>22</sup>. The number of substages identified from the ESCC patients varied from 4 – 16, and the independent tissue samples from the corresponding substages were separately dissected from the formalin-fixed, paraffin-embedded (FFPE) slides (Supplementary Fig. 1a). All the early-stage ESCC samples in our cohort were dissected with 3 mm thick and stood up one by one in the embedding, and were then marked in the hematoxylin and eosin (H&E)-stained sections (Supplementary Fig. 1b). In total, 756 samples were collected, and were classified into 9 histopathological stages covering 22 substages in our cohort, including stage 1 (normal tissue stage), stage 2 (hyperplasia stage), stage 3 (Tis stage), stage 4 (lamina propria cancer stage), stage 5 (muscularis mucosa stage), stage 6 (submucosal invasion cancer stage a), stage 7 (submucosal invasion cancer stage b), stage 8 (T2 stage), and stage 9 (T3 stage) (Fig. 1a, Supplementary Fig. 1c, and Supplementary Table 1a; details shown in Methods). The tumor purity of all samples was defined as the Score 5 (>80%), indicating the high quality of all samples of our cohort (Supplementary Fig. 1d). That represented the advantages of our samples in a pathological region resolved mode, providing the chances to portray molecular profiles of ESCC in a time resolved mode. Subsequently, we performed mass spectrometry (MS) profiling of all 756 samples, phosphoproteomic profiling of 54 samples (Supplementary Table 1b), and whole-exome sequencing (WES) of 53 samples (Fig. 1a, Supplementary Fig. 2a, and Supplementary Table 1c, d).

WES profiling (131× mean coverage, 14.0G raw data) identified 5,327 non-silent mutations in 3,366 genes in the Fudan cohort (this study) (Supplementary Table 2a, b). The top mutations were *TP53* (52.2%), *TTN* (37.0%), *FAT3* (21.7%), *NR2E3* (19.6%), *AKAP9* (17.4%), *ATAD2* (17.4%), *DYTN*

(17.4%), *EPASI* (17.4%), *EPHA3* (17.4%), and *FAT4* (17.4%) (Fig. 1b). *TP53*, frequently mutated in EC<sup>11,12</sup>, was the top mutation in the Fudan cohort (our cohort), and was co-occurrence with the mutations of *HSPG2*, *ADAT2*, etc. (Supplementary Fig. 2b, c). We observed that the number of mutations gradually cumulated during ESCC carcinogenesis, ranging from 78 mutations in hyperplasia stage to 1,511 mutations in T3 stage. In our study, the novel mutations were indicated those just appearing at certain stage, such as the mutation of *FAT3*, not detected at the hyperplasia stage while observed at the Tis stage, was the novel mutation of the Tis stage. We integrated the novel mutations and observed the number of the novel mutations was peaked at the Tis stage (n = 459), and the T2 stage (n = 1,433), indicating the significant events during carcinogenesis. We then divided 9 stages of our cohort into 3 phases covering NT phase, IEN phase, and ESCC phase, which allowed us to explore the key events in ESCC progression (Fig. 1c). Observation of mutation loads of the Fudan cohort against TCGA cohort<sup>12</sup> and other ESCC cohorts (Lin 2014 cohort<sup>11</sup> and Song 2014 cohort<sup>13</sup>) showed the fewer mutation loads in the NT phase (Fig. 1d). Liu et al., revealed fewer mutations were detected in the esophageal nondysplastic epithelium (simple hyperplasia)<sup>23</sup>, which was also observed in our cohort. Specifically, *TP53* mutation was exclusively detected in the NT phase of the No.020 (Supplementary Fig. 2d). Comparing with other ESCC cohorts, we found the mutation loads of the NT phase was lower, while the mutation loads of the IEN phase were comparable (Fig. 1d), indicating the low mutation burden of our cohort was due to the low mutation loads of the early stages of ESCC.

We used non-negative matrix factorization (NMF)<sup>24</sup> to analyze the frequencies of mutated trinucleotide sequence motif and identified three mutational profiles of the IEN phase (Tis stage to submucosa stage), the ESCC phase (T2 and T3 stages), and the Fudan cohort (hyperplasia stage to T3 stage), the cosine similarity of which were then against human cancer<sup>25</sup>. As a result, the best matching mutational signatures in the Fudan cohort were (1) spontaneous deamination of 5-methylcytosine (C>T transition, Signature 1), and (2) APOBEC cytidine deaminase (C>G transition, Signature 13) (Fig. 1e and Supplementary Fig. 2e). The Signature 1, prominent in the ESCC phase, was predominant in most cancers, such as colorectum cancer (CRC)<sup>26</sup>. The Signature 13 was associated with base excision repair and DNA replication<sup>26</sup>, was prominent in the IEN phase, as well as in other ESCC cohorts including Lin 2014 cohort, Song 2014 cohort, and TCGA 2017 cohort (Fig. 1e and Supplementary Fig. 2f), whereas not detected in EAC of TCGA 2017 cohort. These results indicated Signature 13 maybe the feature of ESCC carcinogenesis which exactly happened as early as in the IEN phase.

Correlation analysis across studies using mutation frequency from cBioPortal<sup>27</sup> (<http://www.cbioportal.org/datasets>) reflected the distinct profiles of our cohort, in which the highest correlation coefficient (Spearman's correlation  $R = 0.89$ ) was detected between the IEN phase and ESCC phase, suggesting the featured mutations of the advanced stages happened in the earlier stage (Fig. 1f). The top mutation of *TP53* had much lower mutation frequency in the NT phase (Fisher's exact test,  $p = 0.04$  (Phase II),  $9.8E-5$  (Song 2014),  $0.02$  (Lin 2014)). The *FAT2* and *FAT3*, as well as *FAT1*, were more frequently mutated in the ESCC phase (50%, Fisher's exact test,  $p$  (*FAT2*) =  $0.02$  (Song 2014),  $p$  (*FAT3*) =  $5.9E-4$  (Lin 2014)) (Fig. 1f). Collectively, we built a comprehensive genomic landscape of early-stage ESCC, and these findings represented the difference between early and advanced stages of ESCC.

At the protein and phosphoprotein levels, proteomic analysis was performed using a label-free quantification strategy<sup>28,29</sup>. Protein abundance was firstly calculated by intensity-based absolute quantification (iBAQ)<sup>30,31</sup> and then normalized as the fraction of total (FOT). Whole-cell extract of HEK293T cells was used as the quality control (QC), which represented the MS was robust and was consistent based on the large Spearman's correlation coefficients (mean =  $0.91$ ) (Supplementary Fig. 2g and Supplementary Table 3a). The Spearman's correlation of 114 stage 1 (normal tissue stage) samples was  $0.83$  (mean), allowing the comparison between samples during ESCC carcinogenesis (Supplementary Fig. 2h and Supplementary Table 3b). With the advancement of the stages of ESCC, the numbers of identified proteins were slightly elevated from  $\sim 5,000$  in stage 1 (normal tissue) to  $\sim 7,000$  in stages 8 and 9 (T2 and T3 stages) with a total of 13,711 in 756 samples at 1% false discovery rate (FDR) (Methods) (Fig. 1g). Take an example, 4,613 and 6,381 proteins were detected at (sub)stage 1 (normal tissue stage) and 7\_3 (submucosal invasion cancer stage) in No.105 ESCC patient (Supplementary Fig. 2i). In addition, the reference proteome was highly dynamic based on the protein abundance, which spanned over eight orders of magnitude (Supplementary Fig. 2j, k). The top 10 proteins extensively expressed in our dataset were HBB, HBD, ALB, KRT5, HBG1, ANXA1, CSTB, HSPB1, FABP5, and S100A9 (Supplementary Fig. 2l and Supplementary Table 3b). Low abundance proteins, including NEB, DD11, AK9, LIPE, SDK1, MYH8, HELZ, TRIO, KMT2C, and DNAH5, and common transcription factors, including BAZ1A, ATF7IP, BPTF, BCAS3, CDH24, and EP300 were also covered. Additionally, the ESCC biomarkers identified in previous studies (several ESCC tissues or cell lines), including ACTA2, TAGLN, ANXA1<sup>32</sup>, HSPA9, PDIA4, PLEC, POSTN, PSAP, and THBS1<sup>33</sup> were also covered in the Fudan cohort. Comparably, the median of the identified phosphosites was 4,537, and a total of

34,395 phosphosites corresponding to 6,293 phosphoproteins were identified in 54 samples (Supplementary Fig. 2m, n, and Supplementary Table 3c). During the process of ESCC carcinogenesis, the growingly increased numbers of phosphosites and phosphoproteins were also observed at the phosphoprotein level (Fig. 1g and Supplementary Fig. 2m, n). Overall, we established a comprehensive landscape of ESCC progression at the gene, protein, and phosphoprotein levels.

### **The gain of chromosome 3q was the key event in the transmit from the NT phase to the IEN phase**

To explore the impacts of genomic aberrations, we performed whole exome-based somatic copy number alterations (SCNAs) analyses based on WES data and examined the regulatory effects of 23,107 SCNAs on protein expressions of 53 samples in ESCC progression (Supplementary Fig. 3a, b, and Supplementary Table 2b). The integrated genomic, proteomic, and phosphoproteomic data of all phases illustrated the gain of chromosomes 3q (chr3q) was the key event in the transmit from the NT phase to the IEN phase (Fig. 2a and Supplementary Fig. 3c). To explore the biological functions of the gain of chr3q, we performed *cis* effects of the genes with CNA regions at the protein level, and found 28 genes at chr3q gain perturbation profiles had significantly positive *cis* effects on their associated proteins (Fig. 2b). Interestingly, nearly 32.14% genes at chr3q gain were positive associated with Ca<sup>2+</sup> signal.

The Ca<sup>2+</sup> signal is identified as a key regulator of processes in other excitable cells (e.g., neurons) and non-excitable cells (including those of the epithelia)<sup>34</sup>, where it can control a diverse array of processes such as secretion, proliferation<sup>35</sup>, and promote cancer cell survival<sup>36</sup>. In our cohort, all the amplification of genes in the Ca<sup>2+</sup> signal (e.g., *PRKCI*, *GMPS*, *ATP1B3*, *ATP2C1*, and etc.) were detected in the IEN phase (Fig. 2c), suggesting the gain of chr3q event was prominent in the ESCC early-stage. In addition, we found the amplifications of those genes also had significantly positive effects on other parallel proteins in the Ca<sup>2+</sup> signal at the protein level, such as the amplification of *GMPS* was significantly associated with the upregulation of *PRKCI* ( $p = 0.037$ , Spearman's  $R = 0.32$ ), *PLCH1* ( $p = 9.6E-3$ , Spearman's  $R = 0.40$ ), *GFM1* ( $p = 1.5E-4$ , Spearman's  $R = 0.55$ ), *ATP1B3* ( $p = 0.04$ , Spearman's  $R = 0.32$ ), and so on (Fig. 2c). To assess the effects of these amplified genes, we annotated outliers for the degree of which CRISPR- or short hairpin RNA (RNAi)-mediated depletion reduced ESCC cell lines<sup>37</sup>. As the result, we found the deletion of three genes (*GMPS*, *GFM1*, and *ATP2C1*) had negative effects on the proliferation of ESCC cell lines, in which *GMPS*, *GFM1*, and *ATP2C1*, exhibited positive correlation with PCNA (Pearson's correlation,  $R = 0.39$  (*GMPS*) and  $0.72$  (*GFM1*),  $p = 6.1E-3$  (*GMPS*)

and 3.5E-8 (GFM1)) and MKI67 (Pearson's correlation,  $R = 0.37382$  (GMPS), 0.41 (GFM1), and 0.34 (ATP2C1),  $p = 0.01$  (GMPS), 5.1E-3 (GFM1), and 0.02 (ATP2C1)), makers of tumor cells proliferation<sup>38</sup> (Fig. 2d and Supplementary Table 4a, b). Taken together, the chr3q gain was the driven event from the NT phase to the IEN phase, leading to the enhancement of  $Ca^{2+}$  signal and cell proliferation in the IEN phase (Fig. 2e).

### **The impacts of genomic aberrations in ESCC progression**

Mutually exclusive (ME) analysis is generally used to identify driver genes, pathways, and investigate their functional relationships<sup>39</sup>. Therefore, we performed ME analysis of the identified mutations in the Fudan cohort to elucidate the mutation dynamics in ESCC progression. Specifically, the mutations in genes related to DNA repair (e.g., *YY1* (12.5% in stage 2), *POLE3* (8.3% in stage 3), and *ATR* (50.0% in stage 9)) (Bartlett's test,  $p < 1.0E-4$ ), and cell cycle (e.g., *ARID1B* (12.5% in stage 2), *E2F4* (10.0% in stage 7), and *SMAD1* (50.0% in stage 8)) (Bartlett's test,  $p < 1.0E-4$ ) were ME from each other (Supplementary Fig. 3d), indicating the impacts of genomic aberrations of DNA repair and cell cycle in ESCC progression.

Visualization of genomic aberrations ( $> 5\%$  mutation frequency) in ESCC progression disclosed that there were more mutations in the IEN phase ( $n = 333$ ) and the ESCC phase ( $n = 223$ ), 17 of which were also recorded in cancer associated genes (CAGs,  $n = 593$ ) (Fig. 2f, Supplementary Fig. 3e, and Supplementary Table 4c). Among these 17 mutations, we found that *TP53* mutation (Chi-square test,  $p = 0.039$ ) was prevalent in the IEN phase, and the mutations of *MACF1* (Chi-square test,  $p = 0.048$ ), *FAT1* (Chi-square test,  $p < 1.0E-4$ ), *SPTA1* (Chi-square test,  $p = 7.8E-3$ ), *KMT2D* (Chi-square test,  $p = 7.8E-3$ ), *CHD3* (Chi-square test,  $p = 1.1E-3$ ), and *MGA* (Chi-square test,  $p = 1.1E-3$ ), were prominent in the ESCC phase (Fig. 2f). In addition, *TP53*, a prominent mutated gene in ESCC<sup>4,23</sup>, was the only CAG that was mutated in all stages of ESCC, which enhanced its protein level (Kruskal-Wallis test,  $p = 1.7E-3$ ) (Fig. 2h). To explore the impacts of *TP53* mutation in ESCC progression, we incorporated the alterations of significantly up-regulated proteins (SUPs, t-test,  $p < 0.05$ , fold change (FC) ( $TP53$  Mut/WT)  $\geq 2$ ) in all phases and the up-regulated proteins (UPs, FC ( $TP53$  Mut/WT)  $\geq 2$ ) both detected in the IEN phase and the ESCC phase (Fig. 2g and Supplementary Fig. 3f). The Gene Oncology (GO) enrichment disclosed that those overlapped SUPs were associated with DNA replication and cell cycle (e.g., MCM3/4/6/7, SSBP1, PPP3CA, etc.), and ECM signaling (e.g., FN1, VCAN, ITGB1/4, etc.) (Fig. 2h).

Corresponding to the ESCC progression stages, DNA replication and cell cycle were predominant in the IEN phase, especially in the mucosa stages (stages 4 and 5), while ECM signaling was predominant in the ESCC phase (stages 8 and 9) (Fig. 2i).

To delineate the consequences of the genomic aberrations of the ESCC phase prominent mutations (*MACF1*, *FAT1*, *SPTA1*, *KMT2D*, *CHD3*, and *MGA*) in ESCC progression, we performed the mutation-protein correlation analysis, and found only the mutations of *MACF1* and *FAT1* had significantly positive impacts on their counterpart proteins levels (t-test,  $p = 9.8E-3$  (*MACF1*),  $6.7E-4$  (*FAT1*)), while only the depletion of *MACF1* reduced the proliferation of ESCC cell lines according to the depmap<sup>37</sup> (<https://depmap.org/portal>) (Fig. 2j).

*MACF1*, is a multidomain protein that can associate with microfilaments and microtubules, and associates with a complex (e.g., *CTNNB1*, *GSK3B*, etc.) in Wnt signaling<sup>40</sup>. In our cohort, we found significantly correlation between *MACF1* and *GSK3A* (Pearson's  $R = 0.36$ ,  $p = 0.01$ ), and *MACF1* and *GSK3B* (Pearson's  $R = 0.36$ ,  $p = 0.01$ ) at the protein level, thus we inferred the mutation of *MACF1* would elevate Wnt signaling (Fig. 2k). Of note, the incorporation of the proteins of Wnt signaling affected by the mutation of *MACF1*, disclosed the enhanced proteome-levels of *GSK3A*, *GSK3B*, *CSNK1A1*, *CSNK2B*, *CTNNB1*, *CTNND1*, and so on (Fig. 2l). Furthermore, the CRISPR-mediated depletion of *CSNK2B* and *CTNND1* reduced the proliferation of ESCC cell lines (Fig. 2l). At the phosphoprotein level, a similar trend was also observed that the phosphorylation of *CTNNB1* (S45, S552, and S191), *CTNNB1* (S545), *GSK3B* (Y216), and *CTNND1* (S346) was prominent in the ESCC phase, in which the expression of *CTNNB1* S191 was highly correlated with phosphoprotein level of *GSK3B* Y216 (Pearson's correlation,  $R = 0.78$ ,  $p = 4.0E-12$ ) (Fig. 2m and Supplementary Fig. 3g). Therefore, the mutation of *MACF1* enhanced the protein level of *MACF1*, which bound to the Wnt complex and activated Wnt signaling in the ESCC phase (Fig. 2n). Collectively, we found *TP53* mutation observed in the whole ESCC stages exhibited diverse impacts during ESCC carcinogenesis, while *MACF1* mutation, had roles in promoting ESCC progression in the ESCC phase.

### **A carcinogenesis path with eight dynamic waves in ESCC progression**

The proteogenomics on 3 phases indicated a temporal correlation between genomic aberrations/proteomic alterations and ESCC progression. To further portray molecular profiles of ESCC

in a time resolved mode, we split the whole ESCC progression into 22 substages. Visualization of the abundance of the most variable proteins ( $n = 6,687$ , at least 1/6 samples in certain stages; Methods) by principal component analysis (PCA) (Methods) differentiated the proteome profiles for the 22 ESCC substages, which clearly discriminated the proteomes of the early and the advanced ESCC stages. Additionally, PCA of early ESCCs displayed obvious diversity among the substages along with EC progression (Fig. 3a). Recent studies have provided considerable evidence that tumor stroma plays a crucial role in the induction of immune tolerance to human cancers<sup>41</sup>. In this study, the assessment of the immune-score, stromal-score, and microenvironment-score of 756 ESCC samples confirmed the significant improvement of immune score (Kruskal-Wallis test,  $p < 2.2E-16$ ), stroma score (Kruskal-Wallis test,  $p = 4.8E-3$ ), and microenvironment score (Kruskal-Wallis test,  $p < 2.2E-16$ ) in ESCC progression, which was consistent with the pathological features (Supplementary Fig. 4a and Supplementary Table 5a). These results further showed the distinct profiles of the substages in ESCC progression.

To investigate specific characteristics of the 22 substages, we performed the substage-based supervised clustering analysis (Methods), and the substages were subsequently assigned to 8 panels on the basis of the histopathological stages and temporal expression patterns (Fig. 3b and Supplementary Fig. 4b). We then represented a panel-based analysis to elucidate the molecular dynamic models that drove carcinogenesis from early to progressive ESCC. In (sub)stage 1 (normal tissue), the primary metabolic machinery was predominant, such as lipid metabolism (e.g., HMGCS1 (Kruskal-Wallis test,  $p < 2.2E-16$ ) and ALOX12 (Kruskal-Wallis test,  $p < 2.2E-16$ )), pyruvate metabolism (e.g., ACOX1 (Kruskal-Wallis test,  $p < 2.2E-16$ ) and ACOX3 (Kruskal-Wallis test,  $p < 2.2E-16$ )), and amino acid metabolism (e.g., ANXA1 (Kruskal-Wallis test,  $p < 2.2E-16$ ) and CTSB (Kruskal-Wallis test,  $p < 2.2E-16$ )) (Supplementary Fig. 4c). Particularly, ANXA1, a biomarker of ESCC<sup>32</sup>, was gradually decreased during ESCC progression (FC (advanced stages (AS)/stage 1 (S1)) = 0.029) (Supplementary Fig. 4d). In stage 2 (hyperplasia stage), the expression of proteins, including PPP2R5A (Kruskal-Wallis test,  $p < 2.2E-16$ ), SERPINB3 (Kruskal-Wallis test,  $p < 2.2E-16$ ), and CSTA (Kruskal-Wallis test,  $p < 2.2E-16$ ) involved in the immune response to external damage, were elevated. Enhancement of ERBB2 signaling and NOTCH signaling, and the related proteins (e.g., AGK (Kruskal-Wallis test,  $p < 2.2E-16$ ), MOV10 (Kruskal-Wallis test,  $p = 9.3E-8$ ), and GRB7 (Kruskal-Wallis test,  $p = 1.0E-4$ )) were detected in stage 3 (Tis stage). In stage 4 (lamina propria stage), insulin signaling (e.g., INSR (Kruskal-Wallis test,  $p = 4.1E-$

5) and IRAK1 (Kruskal-Wallis test,  $p < 2.2E-16$ ) and IGF signaling (e.g., PPP1CC (Kruskal-Wallis test,  $p = 3.1E-5$ ) and PPP3CC (Kruskal-Wallis test,  $p = 1.2E-7$ )) were prevalent. Cell cycle (e.g., ORC3 (Kruskal-Wallis test,  $p = 3.1E-11$ ) and RAD21 (Kruskal-Wallis test,  $p < 2.2E-16$ )) and TLR signaling (e.g., TLR3 (Kruskal-Wallis test,  $p = 6.2E-3$ ) and FGFR1OP (Kruskal-Wallis test,  $p = 3.1E-7$ )) were predominant in stage 5 (muscularis mucosa stage), whereas DNA repair (e.g., BRCC3 (Kruskal-Wallis test,  $p = 2.8E-3$ ) and MSH6 (Kruskal-Wallis test,  $p < 2.2E-16$ )) and mTOR signaling (e.g., MTOR (Kruskal-Wallis test,  $p < 2.2E-16$ ) and LAMTOR1 (Kruskal-Wallis test,  $p < 2.2E-16$ )) were enhanced in stage 6 (sm a stage). In stage 7 (sm b stage), PI3K-AKT signaling (e.g., AKT2 (Kruskal-Wallis test,  $p = 5.0E-3$ ) and PIK3R4 (Kruskal-Wallis test,  $p < 2.2E-16$ )) and EMT signaling (e.g., VWF (Kruskal-Wallis test,  $p = 6.7E-13$ ), TK1 (Kruskal-Wallis test,  $p < 2.2E-16$ ), and MMP8 (Kruskal-Wallis test,  $p < 2.2E-16$ )) were dominant, of which EMT signaling is involved in tumor-initiation and motility<sup>42,43</sup>. In addition, the associated protein markers, such as MMP1 and MMP8, were identified as specific indicators of high-grade malignancy<sup>44,45</sup>. Overrepresentation of glycolysis (e.g., PGK1 (Kruskal-Wallis test,  $p < 2.2E-16$ ), ENO2 (Kruskal-Wallis test,  $p = 9.8E-13$ ), and PGM1 (Kruskal-Wallis test,  $p = 1.2E-4$ )) and Wnt signaling (e.g., CSNK2B (Kruskal-Wallis test,  $p < 2.2E-16$ ), CTNNB1 (Kruskal-Wallis test,  $p < 2.2E-16$ ), and GSK3B (Kruskal-Wallis test,  $p = 7.3E-13$ )) was observed in stages 8 (T2 stage) and 9 (T3 stage) (Supplementary Fig. 4c). PGK1, an important glycolytic enzyme, though not mutated at the genome level, was highly elevated at the proteome level, especially in the T2 and T3 stages of ESCC, which provided further evidence for the crucial role of glycolysis, and suggested the potential function of PGK1 in ESCC progression.

Furthermore, the cellular characteristics of ESCC progression were elucidated by xCell<sup>46</sup> (<https://xcell.ucsf.edu>) (Methods) of 756 samples in the Fudan cohort. The analysis revealed that the specific characteristics of the esophageal tissue, such as epithelial cells (Kruskal-Wallis test,  $p < 2.2E-16$ ) and keratinocytes (Kruskal-Wallis test,  $p < 2.2E-16$ ), the signature of hematopoietic stem cells (HSC) (Kruskal-Wallis test,  $p < 2.2E-16$ ), and Tregs (Kruskal-Wallis test,  $p < 2.2E-16$ ), were gradually diminished with ESCC progression (Supplementary Fig. 4a and Supplementary Table 5b). In contrast, the innate/acquired immune response signatures, such as CD8<sup>+</sup> T cells (Kruskal-Wallis test,  $p < 2.2E-16$ ), endothelial cells (Kruskal-Wallis test,  $p = 4.3E-3$ ), Th1 cells (Kruskal-Wallis test,  $p < 2.2E-16$ ), macrophage (Kruskal-Wallis test,  $p < 2.2E-16$ ), and granulocyte macrophage progenitor (GMP) (Kruskal-Wallis test,  $p < 2.2E-16$ ), were significantly elevated with ESCC progression, indicating the

active immune response in ESCC progression.

At the gene level, we observed the mutations of *FAT3* (Fisher's exact test,  $p = 0.043$ ), *NR2E3* (Fisher's exact test,  $p = 0.034$ ), *FAT4* (Fisher's exact test,  $p = 0.027$ ), *SYNE1* (Fisher's exact test,  $p = 0.027$ ), *AKAP9* (Fisher's exact test,  $p = 0.027$ ), *EPAS1* (Fisher's exact test,  $p = 0.027$ ), *PCDHB16* (Fisher's exact test,  $p = 0.020$ ), *KIAA1109* (Fisher's exact test,  $p = 0.014$ ), *MACF1* (Fisher's exact test,  $p = 0.014$ ), *ZNF469* (Fisher's exact test,  $p = 0.014$ ), and *FOXN1* (Fisher's exact test,  $p = 0.014$ ), were significantly associated with stages, especially prominent in the T2 and T3 stages of ESCC (Fig. 3c). In addition, the mutations of *FAT4*, *AKAP9*, and *PCDHB16*, were extensive co-occurrence (Fisher's exact test,  $p = 7.66E-5$  (*FAT4* & *AKAP9*),  $1.50E-7$  (*FAT4* & *PCDHB16*),  $2.00E-5$  (*AKAP9* & *PCDHB16*)), which were prevalent in the stages 6 – 9 (submucosa stage to T3 stage) (Fig. 3c).

The recurrent co-occurrence of genomic events helps to dissect the genomic complexity underlying tumor progression<sup>47</sup>, thus enabled us to explore the functional impacts of the co-occurrence of the mutations of *FAT4*, *AKAP9*, and *PCDHB16* in ESCC progression. Furthermore, the multi-dimensional omics data provided an excellent chance to explore the relationships between the genome and the proteome in the time resolved ESCC progression. We thus performed overrepresentation pathways analysis using the differentially expressed proteins ( $n = 927$ ) which were the overlapped between the differentially up-regulated proteins ( $n = 1,420$ , FC (Mut/WT)  $\geq 2$ ) and highly expressed phosphoproteins ( $n = 1,279$ , submucosa stage to T3 stage) (Fig. 3d). The results showed that these proteins participated in DNA repair, mTOR signaling, PI3K-AKT signaling, and glycolysis (Fig. 3e). Specifically, DNA repair (e.g., *LIG1* (S51, S141), *MED30* (S137), *MSH6* (S94), etc.), mTOR signaling (e.g., *MTOR* (S1261), *RPTOR* (S863), *TOR1AIP1* (S143, S157), etc.), and PI3K-AKT signaling (e.g., *AKT1* (S124), *AKT1S1* (T90/S92), *EIF4EBP1* (T70, S35/T46)), were overrepresented in the submucosa stages (stages 6 and 7), and glycolysis (e.g., *PGK1* (S203), *PGM3* (S64), *ENO3* (S37), etc.) was dominant in the T2 and T3 stages (stages 8 and 9) at the protein and phosphoprotein levels (Fig. 3f, g). The diverse regulations of co-occurrence (*FAT4*, *AKAP9*, and *PCDHB16*) were notably consistent with the dynamic waves in ESCC progression at the gene, protein, and phosphoprotein levels.

Taken together, a carcinogenesis path with eight dynamic waves in ESCC progression was revealed on the basis of the consistency among the genomic aberrations, proteomic alterations, and phosphoproteomic actions: metabolism (e.g., *ANXA1*, *CTSB*, and *HMGCS1*) – DNA damage (e.g., *PPP2R5A*, *IVL*, and *SERPINB3*) – cell proliferation (e.g., *ERBB2*, *NOTCH3*, and *MOV10*) – lesion

invasion (e.g., INSR, IRAK1, and IGFALS) – cell cycle (e.g., ORC3, RAD21, and FGFR1OP) – cell differentiation (e.g., BRCC3, MSH6, and MTOR) – tumor metastasis (e.g., TK1, VIM, and PIK3R4) – EC (e.g., PGK1, GSK3B, and CTNNB1) (Fig. 3h). Furthermore, these results also defined the substages-specific molecular characteristics and uncovered the potential candidates for ESCC malignancy.

### **Proteome clusters of ESCC progression**

Consensus clustering (Methods) identified two major proteomic clusters: the early stages (Cluster 1 (C1),  $n = 324$ ) included the substage 1, 2\_1, and 2\_2, and the later stages (Cluster 2 (C2),  $n = 432$ ) contained the rest of the substages, including T2 and T3 stages (Fig. 4a, Supplementary Fig. 5a, and Supplementary Table 6a). Comparative analysis of the later stage samples (especially the IEN phase samples) of C1 and C2 with the clinic features, disclosed that the later samples in C1 were prominent from the ESCC patients no more than 50 years old (Fisher's exact test,  $p = 3.2E-12$ ) (Supplementary Fig. 5b), indicating the potential impacts of ages in ESCC progression. In addition, the two clusters were greatly associated with 3 phases in ESCC progression, in which C2 contained the most of samples in the IEN phase (87.8%) and the ESCC phase (100%), reflecting more malignancy of C2 (Supplementary Fig. 5c). This distinct two-stage separation suggested that the irreversible fundamental proteome alterations happened as early as the Tis stage (stage 3). *Van der Schaaf* indicated that the Tis stage was associated with worse survival<sup>48</sup>.

The significance analysis of microarray (SAM)<sup>49</sup> analysis was performed to investigate the characteristics of the two clusters at the proteome level, which identified 3,405 differentially expressed proteins (DEPs) between C1 and C2 (t-test,  $p < 0.05$ , FC (C2/C1)  $\geq 2$  or  $\leq 0.5$ ), including 3,163 elevated (C2) and 242 descend proteins (C1) (Fig. 4b, Supplementary Fig. 5d, and Supplementary Table 6b). The results showed that the oncogenic pathways-related proteins, including TP53, ATM, CDK1, EGFR, CASP3/8, NFKB1/2, VCAN, THBS1, etc. were significantly overrepresented in C2 (t-test,  $p < 0.05$ , FC (C2/C1)  $\geq 2$ ). On the contrary, the overrepresented proteins of C1, including KLK12/13, CSTA, HMGCS1, ALOX15B, ECM1, ANXA1, and IL36A involved in keratinization and inflammatory responses, were associated with the primary biological function of normal esophagus (t-test,  $p < 0.05$ , FC (C2/C1)  $\leq 0.5$ ). In an earlier study, consistency has also reported the downregulation of ECM1 in squamous cell carcinomas tissues compared to their paired non-cancerous adjacent tissues (NATs)<sup>50</sup>. Furthermore, the specific biomarkers of normal esophagus annotated from Human Proteome Atlas (HPA,

<https://www.proteinatlas.org>), including GBP6 (t-test,  $p = 2.3E-53$ , FC (C2/C1) = 0.23), TGM1 (t-test,  $p = 3.2E-43$ , FC (C2/C1) = 0.17), TGM3 (t-test,  $p = 1.4E-45$ , FC (C2/C1) = 0.06), and S100A14 (t-test,  $p = 2.6E-43$ , FC (C2/C1) = 0.35), were highly expressed in C1 (Supplementary Fig. 5e), which gradually decreased from the normal stage (stage 1) to advanced stages (stages 8 and 9) of ESCC (Kruskal-Wallis test,  $p$  (GBP6) <  $2.2E-16$ ,  $p$  (TGM1) =  $5.7E-5$ ,  $p$  (TGM3) =  $4.5E-9$ ,  $p$  (S100A14) =  $3.8E-4$ ) (Supplementary Fig. 5f). Taken together, these results provided further confirmation in the dysregulation of metabolism and oncogenic pathways between genomic aberrations and proteomic alterations in ESCC progression.

### **The *AKAP9* mutation enhanced glycolysis in the ESCC phase**

Incorporation of genomic aberrations ( $n = 3,366$ ) and the corresponding proteomic alterations ( $n = 3,285$ , FC (C2/C1)  $\geq 2$ ) revealed 9 common mutations (e.g., *TP53*, *STAG2*, *ITGB5*, etc.) were both detected in C1 and C2 (Fig. 4c). Among of the C2 specific molecules ( $n = 516$ ) at the gene and protein levels, only the mutations of *AKAP9* (Fisher's exact test,  $p = 0.027$ ) and *MACF1* (Fisher's exact test,  $p = 0.014$ ) were significantly associated with 3 phases in ESCC progression. Furthermore, the mutation of *AKAP9* was closely related to the T2 and T3 stages (Fisher's exact test,  $p = 0.013$ ) (Fig. 4d and Supplementary Fig. 5g, h), which was associated with prognosis outcomes of ESCC patients (Log-rank test,  $p = 0.08$ ) (Supplementary Fig. 5i).

*AKAP9*, one of A-kinase anchoring proteins (AKAPs), binds to the regulatory subunit of AMP-dependent protein kinase (PKA) and achieved the activation of PKA<sup>51</sup>, which regulates multiple signaling cascade<sup>52</sup>, such as glucose metabolism including glycolysis<sup>53</sup>. In our cohort, we found the mutation of *AKAP9* upregulated the proteome-levels of *AKAP9* (FC (Mut/WT) = 1.77) and *PRKACA* (FC (Mut/WT) = 4.33) (Fig. 4e), and *PRKACA* was one of the PKA catalytic subunits<sup>54</sup>. Moreover, *AKAP9* was significantly positive correlated to *PRKACA* at the protein level (Pearson's R = 0.30,  $p = 0.042$ ) (Fig. 4f). Gene set enrichment analysis (GSEA) demonstrated that the *AKAP9* mutation positive-correlated proteins were converged on glycolysis (normalized enrichment score (NES) = 2.34, FDR = 0) (e.g., *HK1/3*, *GPI*, *PGK1*, etc.), pentose phosphate pathway (NES = 1.90, FDR = 0.091) (e.g., *G6PD*, *PGLS*, *TKT*, etc.), and TCA cycle (NES = 1.88, FDR = 0.059) (e.g., *MDH2*, *IDH1*, *SDHA/C*, etc.), which were predominant in the ESCC phase (Fig. 4g, h, and Supplementary Fig. 5j). Furthermore, we observed the positive correlation between *AKAP9* and *HK3* at the protein level (Pearson's R = 0.29,  $p = 0.051$ )

(Supplementary Fig. 5k), implying the activation of the transform process from glucose to glucose-6-phosphate.

Allosteric regulation refers to the process where the effect of binding of a ligand at one site of a protein is transmitted to another, often distant, functional site, and thus regulates biological processes including glucose metabolism<sup>55,56</sup>. As well as HK1 and GPI (Pearson's  $R = 0.58$ ,  $p = 2.5E-5$ ), and HK3 and GPI (Pearson's  $R = 0.42$ ,  $p = 3.2E-3$ ), G6PD as one of the activators in the transform process from glucose-6-phosphate to fructose-6-phosphate<sup>57</sup>, was significantly correlated to GPI ((Pearson's  $R = 0.51$ ,  $p = 3.1E-4$ ), further demonstrating the activation of glycolysis (Fig. 4i and Supplementary Fig. 5k). At the phosphoprotein level, we also detected the overrepresented phosphorylation of glycolysis (e.g., GPI T109, ALDOA S39, PKM S97, PGAM1 S14, etc.) in the ESCC phase (Supplementary Fig. 5l). Taken together, the mutation of *AKAP9*, activated PKA and enhanced the energy formation, elevating the process of glucose metabolism, especially glycolysis, in the T2 and T3 stages of ESCC (Fig. 4j). Thus, our study provided new insights into the proteomic characteristics with *AKAP9* mutation and revealed the impacts on glycolysis in ESCC progression at the multi-omics level.

### **Personalized trajectory revealed six major carcinogenesis tracks of early ESCC**

The diversity and tumoral heterogeneity of ESCCs remain challenging to decide precise clinical strategies for different ESCC patients who have diverse featured carcinogenesis tracks. To this end, we used the trajectory inference methods<sup>58</sup> (Methods) to trace the carcinogenesis lineages of early ESCC in the cohort. As a result, six major tracks were classified as track 1 – 6 (13.2%, 7.0%, 7.9%, 43.9%, 19.3%, and 8.7% of patients, respectively) (Fig. 5a, b, Supplementary Fig. 6a, and Supplementary Table 7a). Track enhanced proteins were determined by the expression trend along with 22 substages in ESCC (slope ( $K$ )  $> 0$ ,  $R^2 > 0.15$ ), and the dominant pathways of 6 major tracks were annotated as follows: (1) track 1 (T1,  $n = 15$ , 3,933 track enhanced proteins), biomaterial synthesis ( $p$  (NTPs metabolism) =  $5.1E-4$ ,  $p$  (Nitrogen compound biosynthesis) =  $4.3E-4$ ); (2) track 2 (T2,  $n = 8$ , non-drinking/smoking track, 3,731 track enhanced proteins), ECM signaling ( $p = 7.4E-5$ ); (3) track 3 (T3,  $n = 9$ , female track, 4,122 track enhanced proteins), cell cycle ( $p = 2.6E-5$ ); (4) track 4 (T4,  $n = 50$ , mainstream track (mainstream population of ESCC patients), 2,029 track enhanced proteins), DNA repair ( $p = 4.0E-4$ ); (5) track 5 (T5,  $n = 22$ , older track, 928 track enhanced proteins), glucose metabolism ( $p = 1.1E-5$ ); (6) track 6 (T6,  $n = 10$ , drinking/smoking track, 705 track enhanced proteins), immune response ( $p = 2.7E-5$ ) (Fig. 5c,

Supplementary Fig. 6b, c, and Supplementary Table 7b, c). The track descending proteins ( $K < 0$ ) involved in the primary biology of normal esophagus, such as epithelial cell differentiation ( $p = 2.9E-10$  (T1),  $7.5E-11$  (T2),  $8.3E-14$  (T3),  $6.3E-14$  (T4),  $2.3E-8$  (T5), and  $1.7E-8$  (T6)) (e.g., EVPL and AHNAK2) and keratinization ( $p = 2.1E-11$  (T1),  $8.3E-10$  (T2),  $1.9E-12$  (T3),  $7.4E-8$  (T4),  $4.3E-9$  (T5), and  $1.6E-3$  (T6)) (e.g., FLG, KLK12/13 and SPRP3), were observed during ESCC progression in 6 tracks (Fig. 5c).

Furthermore, we observed that different tracks were closely associated with various clinical features of early ESCC patients, which improved our understanding of tumor heterogeneity. For example, T3, featured with cell cycle, had the highest proportion of female patients (29%) (Fig. 5j), significantly greater than the natural proportion of female patients in ESCC (no more than 10%)<sup>2,59,60</sup>. T5 had the highest proportion of older patients (age > 50, 100%). T6, featured as immune response, had the highest proportion of patients with a habit of drinking/smoking (18%), which was associated with chronic inflammatory and regulated oxidative stress in various cancer types<sup>61</sup>, revealing the unique track in the drinking/smoking patients.

Investigation of the diversity of genomic alterations of tracks through somatic SNVs analysis revealed a prominent C>A transition in the non-drinking/smoking track (T2) (t-test,  $p$  (T2/T5) =  $5.8E-6$ ,  $p$  (T2/T6) =  $6.3E-4$ ), whereas C>T transition was significantly prevalent in the older track (T5) (t-test,  $p$  (T2/T5) =  $4.3E-4$ ) (Supplementary Fig. 6d). We identified 100 differential track mutations (> 5% mutation frequency), 6 of which were also observed in CAGs, including the mutations of *TP53*, *EPAS1*, *EPHA3*, *MACF1*, *STAG2*, and *USP6* (Fig. 5d). Furthermore, we found the mutations of *STAG2* and *USP6* were prevalent in the T5, which were all co-occurrence with the mutation of *AKAP9* (Fisher's exact test,  $p < 1.0E-4$ ), suggesting the correlation of the age and glycolysis in the ESCC patients (Fig. 5e). In addition, GSEA showed the positive impacts of the mutations of *STAG2* (NES = 2.08, FDR =  $7.0E-3$ ), and *USP6* (NES = 2.01, FDR =  $6.0E-3$ ) on pentose phosphate pathway (Supplementary Fig. 6f), which were the evidence of the characteristics of glucose metabolism in the T5.

Notably, the mutations of *TP53*, *EPAS1*, *EPHA3*, and *MACF1* were prominent in the T2 (Supplementary Fig. 6e). As showed in Fig. 2h, *TP53* mutation enhanced the proteome-level of ECM signaling, demonstrating the characteristics of the T2. GSEA displayed the positive association of the mutations of *EPAS1* (NES = 2.34, FDR = 0), *EPHA3* (NES = 1.98, FDR = 0.019), and *MACF1* (NES = 1.99, FDR = 0.053), on adherens junction (e.g., EGFR, RAC1/2, ITGA2/3, ITGB3, etc.) (Fig. 5f, g, and Supplementary Fig. 6g). Furthermore, *EPAS1* mutation was co-occurrence with the mutations of *TP53*,

*EPHA3*, and *MACF1* (Fisher's exact test,  $p < 0.05$ ), and was associated with the phases in ESCC progression (Fisher's exact test,  $p = 0.013$ ) (Fig. 5g). To elucidate the impacts of *EPAS1* genomic aberrations on proteomic alterations and phosphoproteomic actions in the T2, we integrated the overlapped molecules ( $n = 243$ ) between the *EPAS1* SUPs ( $n = 385$ , t-test,  $p < 0.05$ ,  $FC \geq 2$ ) and T2 highly expressed phosphoprotein ( $n = 3,025$ ), which were involved in focal adhesion and adherens junction (e.g., TNC, TNXB, EGFR, etc.) (Fig. 5h). Compared with other tracks, the phosphorylation of the proteins related in focal adhesion and adherens junction was overrepresented in the T2, such as TNC S72, TNXB S3651, COL1A1 S176, CAV2 S23, EGFR S991, etc. (Fig. 5i). In short, these findings illustrated the carcinogenesis lineages of ECM signaling in the non-drinking/smoking ESCC patients.

To profile the personalized dynamic waves of 6 major tracks, we performed Fishplot analysis<sup>62</sup> (Methods) with track enhanced proteins and kinetic driving pathways (Supplementary Fig. 6h). For example, the mainstream track (T4) was represented with DNA repair, and the driven wave axis was lipid-protein metabolism (e.g., ACOX1/3 and HMGCS1) – ECM/cytokine signaling (e.g., ALOX12B and ITGB8) – FGFR/NOTCH signaling (e.g., FGFR2 and NOTCH2) – p53 signaling/DNA repair (e.g., TP53 and ATR) – cell cycle/glycolysis (e.g., CDK2 and PGK1/2) – MAPK/Wnt signaling (e.g., MAPK3 and WNT4) – PI3K-AKT/Hedgehog signaling (e.g., AKT1 and YWHAE/G). The female track was featured with cell cycle, and the track enhanced proteins (e.g., CDK5 and MAP2K6) were overrepresented in T3.

For each track, we assessed the drug targets approved by the US Food and Drug Administration database (FDA)<sup>63</sup> (<https://www.fda.gov>). It showed a trend of elevated expression of proteins involved in mTOR signaling (e.g., MTOR and TSC2) (ANOVA,  $p = 0.013$ ), PI3K-AKT signaling (e.g., AKT1, PIK3CA, and PTEN) (ANOVA,  $p = 2.0E-4$ ), and RTK signaling (e.g., EGFR, ERBB2, and PDGFRB) (ANOVA,  $p = 0.013$ ) in T1, suggesting that temsirolimus/everolimus, AZD5363/taselisib/AZD8186, and gefitinib/Her2/imanitib as the potential therapeutic candidates to T1. The expression of proteins involved in cell cycle (e.g., CDK4, CDKN2A, and IDH1/2) (ANOVA,  $p < 1.0E-4$ ) and Ras signaling (e.g., A/BRAF, K/N/HRAS, and MAP2K1) (ANOVA,  $p = 0.011$ ) were elevated in the female track (T3), implying that olaparib/abemaciclib and cobimetinib/trametinib/cetuximab could be potential drugs for the female track (T3). As expected, the highest prevalence of BRCA2 and ERCC2 in DNA repair (ANOVA,  $p = 0.026$ ) were detected in the mainstream track (T4), indicating that olaparib and cisplatin targeting could be effective. Cancer vaccines and immune checkpoint inhibitors, including anti-CD276

mab, have recently been employed as immunotherapies for ESCC<sup>64</sup>; however, tumoral heterogeneity imposes obstacles. In the drinking/smoking track (T6), the increased expressions of CD4, CD8A, CD19, CD28, CD274, and CD276 (ANOVA,  $p = 1.6E-3$ ) were identified, reflecting responses to immunotherapy (Fig. 5j). Together, our study revealed six major carcinogenesis tracks and nominated the corresponding potential drug targets, providing a reference database for the personalized medicine of ESCC.

### **Aberrant glycolytic enzyme, PGK1, is a potential therapeutic target**

Abnormal glycolytic metabolism was observed in the whole process of early ESCC at the gene, protein, and phosphoprotein levels, which dramatically increased throughout carcinogenesis (Fig. 6a). PGK1, the first ATP-generating enzyme in glycolysis, was identified as the nominated drug-targetable protein in ESCC progression and overrepresented in all 6 tracks. In addition, we performed Cox regression analysis to assess the prognostic value of PGK1 expression, which was negatively correlated with the overall survival of ESCC in the TCGA dataset (log-rank test,  $p = 7.8E-3$ ) (Fig. 6b and Supplementary Table 8a). To further cross validate these results, our dataset confirmed that the expression of PGK1 was gradually enhanced in ESCC at the protein (Kruskal-Wallis test,  $p < 2.2E-16$ , FC (AS/S1) = 4.31) and phosphoprotein levels (PGK1 S203, Kruskal-Wallis test,  $p = 3.3E-6$ , FC (AS/S1) = 3794.85) (Fig. 6c), as demonstrated by immunohistochemistry of ESCC FFPE slides in which PGK1 was increasingly enhanced in the process from T0 stage (normal tissue) to the Tis stage, SM2 stage, and advanced stage (Fig. 6d). Specifically, we found the expression of PGK1 was gradually increased during the carcinogenesis progress of ESCC in the case of No. 020, No.105, and No.109 (Supplementary Fig. 7a). The only identified motif (sP) of PGK1 was ubiquitously (53/54) detected, and the expression of phosphorylation of PGK1 S203 was also elevated with ESCC progression (Fig. 6e).

To investigate the proteome-phosphoproteome regulation between the motif of PGK1 and the kinases, we applied motif extraction algorithm to MS phosphorylation dataset from 54 samples to the set of threonine phosphorylated peptides from the *Beausoleil, Jedrychowski et al.* dataset (Methods)<sup>65,66</sup>. The results revealed the association of ERK1/2, CDKs, and GSK-3 with the motif of PGK1 ([http://hprd.org/PhosphoMotif\\_finder/](http://hprd.org/PhosphoMotif_finder/)). These results indicated that the total PGK1 activity could increase significantly in both cytosol and mitochondria of tumor cell, as ERK1/2 could phosphorylate PGK1 S203 and result in the mitochondrial translocation of PGK1<sup>67</sup>. Therefore, these proteomic and

phosphoproteomic results indicated that the total PGK1 activity increased notably in both cytosol and mitochondrial in tumor cell. In addition, the SCNAs of *CDK2* had positive effects on the expression of PGK1 (Wilcoxon rank-sum test,  $p = 1.5E-10$ , FC (Mut/WT) = 1.25) (Fig. 6f), and the substrates (e.g., TP53BP1 (FC (Mut/WT) = 1.38), RPS6KA3 (FC (Mut/WT) = 1.32), MAST2 (FC (Mut/WT) = 1.38), and KLC2 (FC (Mut/WT) = 1.13)) of kinases of CDKs, ERK1/2, and GSK-3 (Fig. 6g), suggesting that *CDK2* mutation and ERK1/2 synergistically determined the total activity of PGK1 through increasing PGK1 expression both at the protein and phosphoprotein levels. Subsequent kinase-substrates analysis in all 54 samples revealed the elevated expression of the downstream substrates of kinases, including IRS2 (Kruskal-Wallis test,  $p = 1.8E-14$ ), BAD (Kruskal-Wallis test,  $p = 1.5E-3$ ), TP53BP1 (Kruskal-Wallis test,  $p = 4.5E-5$ ), RPS6KA3 (Kruskal-Wallis test,  $p = 3.3E-4$ ), MAP1B (Kruskal-Wallis test,  $p = 7.2E-3$ ), and GSK3B (Kruskal-Wallis test,  $p = 8.1E-3$ ), etc. in ESCC progression. Enhanced phosphorylation of these substrates, including IRS2 S577 (Kruskal-Wallis test,  $p = 1.1E-3$ ), BAD S99 (Kruskal-Wallis test,  $p = 1.9E-4$ ), TP53BP1 S831 (Kruskal-Wallis test,  $p = 9.6E-4$ ), RPS6KA3 S369 (Kruskal-Wallis test,  $p = 5.1E-3$ ), MAP1B S1785 (Kruskal-Wallis test,  $p = 3.8E-4$ ) and S1265 (Kruskal-Wallis test,  $p = 0.040$ ), and GSK3B Y216 (Kruskal-Wallis test,  $p = 0.028$ ), was also observed in ESCC progression (Fig. 6h), indicating their positive association with ESCC progression. Fig. 6i showed the summary of the regulatory network of the kinase-substrates associated with the motif of PGK1 (sP), which could be implicated in identifying the potential therapeutic targets and universal clinical strategies to manage ESCC.

Next, we investigated the roles of PGK1 in regulating glucose and serine metabolism. Overexpression of PGK1 in KYSE150 cells increased the levels of glycolytic-citrate cycle flux metabolites, including 3-PG, pyruvate, and lactate in glycolysis, and citrate, succinate, and fumarate in citrate cycle (Fig. 7a and Supplementary Table 8b). In addition, the levels of serine and glycine were also increased in PGK1-overexpressing cells. Conversely, knockdown of PGK1 decreased the concentration of metabolites in glycolysis and citrate cycle, as well as serine and glycine (Fig. 7b and Supplementary Table 8c). Furthermore, the overexpression of ERK2 led to increased Ser-phosphorylation level, but not Thr- or Tyr- phosphorylation levels of PGK1 in both KYSE150 cells and ECA109 cells (Fig. 7c). On the contrary, ERK2 could not increase the Ser-phosphorylation level of PGK1 S203 mutant (S203A) (Fig. 7c), indicating that ERK2 phosphorylated PGK1 S203 in ESCC cells. Moreover, the increased Ser-phosphorylation level of PGK1 led to mitochondrial translocation of PGK1 (Fig. 7d), which increased

the phosphorylation level of PDHK1 at T338 (Fig. 7e), and decreased PDH activity in ERK2 overexpressing cells (t-test,  $p$  (KYSE150) =  $7.7E-6$ ,  $p$  (ECA109) =  $1.3E-7$ ) (Fig. 7f and Supplementary Table 8d). It was also observed that the metabolites of glycolysis and serine metabolism were further increased, while citrate cycle metabolites were decreased in ERK2 overexpression cells (t-test,  $p < 0.05$ ) (Fig. 7a), suggesting that the overexpression of ERK2 could shut down the pyruvate dehydrogenase complex. Collectively, these results indicated that the increased expression and phosphorylation levels of PGK1 synergistically enhanced glycolysis and serine metabolism. Accordingly, we confirmed that co-overexpression of PGK1 and ERK2 in KYSE150 cells and ECA109 cells promoted their proliferation most profoundly, compared to cells overexpressing either PGK1 or ERK2 (t-test,  $p < 1.0E-4$ ) (Fig. 7g and Supplementary Table 8e). In contrast, the knockdown of PGK1 slowed down the cell proliferation, which was further inhibited by double-knock down of PGK1 and ERK2 in both KYSE150 cells and ECA109 cells (t-test,  $p < 0.001$ ) (Fig. 7h and Supplementary Table 8f). Furthermore, unlike PGK1, the overexpression of other glycolytic enzymes, including GAPDH and PGM1, which catalyzed the last and the next step reaction of PGK1, respectively, did not show pro-proliferation effects in KYSE150 cells (t-test,  $p$  (PGK1) =  $3.7E-7$ ,  $p$  (GAPDH) = 0.19,  $p$  (PGM1) = 0.026) and ECA109 cells (t-test,  $p$  (PGK1) =  $1.9E-6$ ,  $p$  (GAPDH) = 0.17,  $p$  (PGM1) =  $1.6E-3$ ) (Supplementary Fig. 7b and Supplementary Table 8g). Taken together, these results indicated that the glycolytic enzyme PGK1 specifically contributed to the progression of ESCC.

It has been reported that pyrimidine deoxynucleoside analog diphosphates (for example, gemcitabine) could be used as PGK1 inhibitors because L-nucleoside analog diphosphates were selectively phosphorylated by PGK1<sup>68</sup>. Therefore, we tested the potential of gemcitabine to inhibit ESCC tumor growth. First, we validated the significant inhibitory effects of gemcitabine on PGK1 (t-test,  $p = 4.8E-3$ ) with an IC<sub>50</sub> of 16.3 nM (Supplementary Fig. 7c, d, and Supplementary Table 8h, i) using an *in vitro* enzymatic assay. Second, treating the cultured cells with 50 nmol/L gemcitabine significantly decreased the glycolytic flux (such as pyruvate (t-test,  $p = 1.6E-3$ ) and lactate (t-test,  $p = 0.014$ )) and cell proliferation (Supplementary Fig. 7e and Supplementary Table 8j). Third, the xenograft growth-promoting ability of PGK1 overexpression was abolished in gemcitabine treated mice bearing either KYSE150 cells (t-test,  $p = 4.4E-9$ ) or ECA109 cells (t-test,  $p = 2.2E-9$ ) (Fig. 7i and Supplementary Table 8k), which was consistent with the effects of PGK1 knockdown (t-test,  $p$  (KYSE150) =  $8.6E-8$ ,  $p$  (ECA109) =  $3.4E-7$ ) (Fig. 7j and Supplementary Table 8l). Overall, these observations suggested that

upregulated PGK1 expression contributes to early ESCC occurrence and development.

## Discussion

ESCC is one of the most common malignancies, with a relatively low overall 5-year survival rate (less than 30%). Even though the whole genome sequences of ESCC patients have been obtained, the tumor heterogeneity and lack of understanding of the molecular mechanisms in ESCC progression imposes many challenging unmet clinical needs in ESCC. It was speculated that tracking the occurrence and development of early ESCC could provide direct evidence of cancer-driving pathways and molecules in each stage. It is generally agreed that carcinogenesis is a chronic process involving several genes and pathways in different stages. For the first time, our study detected precise temporal molecular switches promoting the progression of ESCC at the multi-omics level. The sequence of canonical cancer pathways was also disclosed, which involved ERBB2, NOTCH, IGF, cell cycle, DNA repair, PI3K-AKT, mTOR, glycolysis, and Wnt signaling. Furthermore, the gradual reduction in keratinization and lipid metabolism revealed that the loss of normal esophagus identity was an important event in initiating early ESCC. The biomarkers of ESCC tissues, such as ACTA2 (Kruskal-Wallis test,  $p < 2.2E-16$ , FC (AS/S1) = 5,163.61), TAGLN (Kruskal-Wallis test,  $p < 2.2E-16$ , FC (AS/S1) = 12.17), HSPA9 (Kruskal-Wallis test,  $p < 2.2E-16$ , FC (AS/S1) = 3.27), PDIA4 (Kruskal-Wallis test,  $p < 2.2E-16$ , FC (AS/S1) = 4.49), PLEC (Kruskal-Wallis test,  $p < 2.2E-16$ , FC (AS/S1) = 3.58), POSTN (Kruskal-Wallis test,  $p < 2.2E-16$ , FC (AS/S1) = 9,187.51), PSAP (Kruskal-Wallis test,  $p < 2.2E-16$ , FC (AS/S1) = 1,096.75), and THBS1 (Kruskal-Wallis test,  $p < 2.2E-16$ , FC (AS/S1) = 49.52)<sup>32,33</sup>, were also detected and significantly enhanced during ESCC progression (Supplementary Fig. 4d, e). These findings provided a novel temporal dimension and trans-omics dimension in understanding the ESCC progression.

The dual-peak of the novel mutations detected at the Tis stage and the T2 stage perfectly matched their corresponding pathological phenotypes observed in the clinic. These significant genomic aberrations were then translated to proteomic alterations in ESCC progression. The co-occurrence mutations of *FAT4*, *AKAP9*, and *PCDHB16* were prominent in the IEN phase (especially submucosa stages) and the ESCC phase (T2 and T3 stages) that had positive impacts on PI3K-AKT-mTOR signaling and DNA repair in the IEN phase, and glycolysis in the ESCC phase. In addition, the mutation of *AKAP9* enhanced the activation of PKA, and thus improved the transfer of ATP to ADP and elevated glycolysis

at the protein and phosphoprotein levels. The  $\text{Ca}^{2+}$  signal impacts nearly all aspect of cellular life<sup>69</sup>, and is implicated in a variety of processes important in tumor progression, such as proliferation and invasiveness<sup>70,71</sup>. In our cohort, we found the gain of chr3q was characterized in the IEN phase, and the amplifications of *GMPS*, *GFMI*, *ATP1B3*, and *ATP2C1*, had *cis* effect on their counterpart proteins and enhanced  $\text{Ca}^{2+}$  signal in cytoplasm. *MACF1*, a large crosslinker that contributes to cell integrity and cell differentiation<sup>72</sup>, was predominantly mutated in the ESCC phase, and activated Wnt signaling at the multi-omics level. These results had explored that all these pathways followed a very precise temporal order in the whole process of early ESCC.

One of the main limiting factors in personalized treatment is the lack of therapeutic strategies based on a patient's carcinogenesis lineage. In our study, 746 samples from 114 early ESCC patients allowed us to trace the carcinogenesis lineages of early ESCCs, resulting in 6 tracks closely related to the clinical feature, including gender, age, and risk habit of drinking/smoking, such as more female ESCC patients were observed in the track 3. ECM signaling, associated with cell proliferation and migration in ESCC<sup>73,74</sup>. Observation of the impacts of the top mutation of *TP53*, prominent in the track 2, showed the upregulation of ECM signaling. As well as the mutations of *EPHAS3* and *MACF1*, *EPAS1* mutation in the track 2, showed enhancement of adherens junction (e.g., TNC (S72), ITGA5 (S127), etc.) at the protein and phosphoprotein levels, confirming the characteristics of ECM signaling in the track 2. Consistent with the observation of *AKAP9* mutation, the co-occurrence mutations of *STAG2* and *USP6* elevated pentose phosphate pathway in the track 5. A large-scale, population-based cohort study has shown that drinking/smoking promotes ESCC carcinogenesis<sup>6</sup>. In consequence, our study disclosed the carcinogenesis lineage of the drinking/smoking patients and presented corresponding potential therapeutic strategies (e.g., PDL1 strategy) in the drinking/smoking track. Furthermore, the overrepresentation of EGFR and ERBB2 observed in the track 1 was shown to be prevalent in ESCC<sup>12</sup> and considered as therapeutic targets<sup>75</sup>, indicating the potential curative effects of Her2 and gefitinib. In brief, in addition to the driver waves, this study provided 6 carcinogenesis tracks as references for diverse ESCC clinical therapies.

Proliferative cancer demands a great deal of energy and building blocks, and cancer cells mainly rely on aerobic glycolysis to produce building blocks and energy, known as the Warburg effect<sup>76</sup>. In the Fudan cohort, aberrant glycolysis and alterations in its key enzyme, PGK1, which was negatively correlated with overall survival rate, were noticed at the gene, proteoin, and phosphoprotein levels and

promoted ESCC cell proliferation and tumor growth. Glycolysis is a sequence of ten enzyme-catalyzed reactions and links other parallel pathways, including the pentose phosphate pathway, serine *de novo* synthesis pathway, citrate cycle, etc. Several rate-limiting enzymes determine the overall glycolysis rate, and dysregulated glycolytic enzymes are frequently observed in various cancers<sup>77</sup>. In the current study, increased PGK1 both at the protein and phosphoprotein levels with a gradual elevation model through the tumor progress was observed. PGK1 works in the hub of glycolysis and serine/glycine synthesis, and meanwhile, phosphorylated PGK1 at Ser203 was able to inhibit metabolic flux from glycolysis to the citrate cycle. Therefore, this change in PGK1 activated glycolysis, serine synthesis, and inactivated pyruvate dehydrogenase complex leading to further accumulation of glycolysis metabolites. The present study also revealed that the upstream and downstream enzymes of PGK1 in glycolysis did not provide a strong pro-proliferation effect, unlike PGK1.

Furthermore, we inhibited PGK1 using gemcitabine, which was used as a DNA analog for inhibiting DNA synthesis<sup>78</sup>. However, its competitive ability to inhibit PGK1 enzymatic activity was ignored because PGK1 could selectively phosphorylate it. Additionally, the inhibitory effect of gemcitabine on PGK1 enzymatic activity and cell proliferation was validated through *in vitro* biochemistry assay, cultured ESCC cells and xenografts model. Collectively, our study indicated that PGK1 is an important drug target in ESCC; however, the association of the expression and phosphorylation levels of PGK1 with other kinds of cancer requires further investigations.

In summary, our study, for the first time, depicted the comprehensive genomic, proteomic, and phosphoproteomic map in ESCC progression. We discovered the kinetic waves of the dominant cancer pathways via integrative proteogenomic analysis in the whole process of carcinogenesis. We also revealed 6 major tracks and their molecular characteristics during the carcinogenesis of early ESCC. Specifically, we presented the preliminary exploration of multi-omics in the multistage carcinogenesis of ESCC, which can be used to model the key events in esophageal tumorigenesis (Supplementary Fig. 8). The chr3q gain was the key event in the transmit from the NT phase to the IEN phase, leading to the ESCC cell proliferation. *TP53* mutation was observed in the IEN phase and the ESCC phase, and enhanced cell cycle in the mucosa stages. *MACF1* mutation was prominent in the ESCC phase, and increased the Wnt signaling. The recurrent co-occurrence of the mutations of *FAT4*, *AKAP9*, and *PCDHB16* driven the PI3K-AKT-mTOR signaling in the submucosa stages, and exhibited positive

impacts on glycolysis in the advanced stages. Furthermore, we demonstrated and proposed the value of a drug-targetable protein, PGK1, especially the phosphoprotein, PGK1 S203, at the multi-omics map. We believe this study provides insights into understanding the architecture of ESCC progression and enables new advances in promoting the diagnostics and therapeutics to manage ESCC.

## Methods

**Patient samples of early ESCCs. Construction of the ESCC cohort.** Three hundred consecutive patients that were presumed to have esophageal lesions underwent ESD therapy from January 2018 to December 2018 at Zhongshan Hospital, Fudan University. There were no biases in selecting patients, and none of the patients had received any prior treatment, such as radiotherapy or chemotherapy. One hundred and fourteen early ESCC cases were eligible for the establishment of the intended study cohort. Among the 186 excluded patients, 21 were diagnosed with non-tumor lesions, 26 had stromal tumors, 86 patients were precluded due to unavailability of their normal tissue samples, and 53 samples failed to pass the pathological quality check, such as tissue cells rate < 80%. Subsequently, 10 advanced ESCC cases (n (T2) = 3, n (T3) = 7) were screened after surgical resection without neoadjuvant therapy. All cases were staged according to the 8<sup>th</sup> edition of the American Joint Committee on Cancer (AJCC) TNM staging system.

As for the pathology quality control, it was our primary concern for a strict pathology classification. The complex pathological staging was based on the morphological observation according to the 8<sup>th</sup> edition of the American Joint Committee on Cancer (AJCC) TNM staging system. Notably, all early-stage ESCC samples in our cohort were dissected with 3 mm thick and stood up one by one in the embedding, and were then marked in the hematoxylin and eosin (H&E)-stained sections. The hematoxylin and eosin (H&E)-stained sections were reviewed evaluated by two or three experienced gastrointestinal pathologists, who would mark them according to the proportion of tumor cells (Score 0 = 0%; 0% ≤ Score 1 < 20%; 20% ≤ Score 2 < 40%; 40% ≤ Score 3 < 60%; 60% ≤ Score 4 < 80%; 80% ≤ Score 5 ≤ 100%). The tumor purity of all samples was defined as the Score 5, indicating the high quality of all samples of our cohort. The present study was carried out in compliance with the ethical standards of Helsinki Declaration II and approved by the Institution Review Board of Fudan University Zhongshan Hospital (B2019-200R). All the patient samples were obtained with Zhongshan's approval

of the Research Ethics Committee. Written informed consent was provided by all participants before any study-specific investigation was performed. Each sample was assigned a new research ID, and the patient's name or medical record number used during hospitalization was de-identified. Clinical information of individual patients, including age, gender, smoking status, and substages, were listed in Supplementary Table 1.

According to the WHO and Japanese pathology diagnostic criteria, all the substages in our early ESCC cohort were contained in 4 TNM stages: T0 (normal epithelial, n = 114), T1 (T1a/b cancer, n = 114), T2 (n = 3), and T3 (n = 7). T1 was sub-classified into hyperplasia stage (2\_1, n = 114), mild and/or moderate dysplasia stage (2\_2, n = 92), Tis stage (3\_1 (n = 61), 3\_2 (n = 73), 3\_3 (n = 67), 3\_4 (n = 19), and 3\_5 (n = 39)), lamina propria cancer stage (including m2 stage (4\_1 (n = 61), 4\_2 (n = 18), and 4\_3 (n = 7), and muscularis mucosa stage (including m3 stage (5\_1 (n = 14), 5\_2 (n = 9), and 5\_3 (n = 9))), and submucosal invasion cancer stage (sm stage), namely sm stage a (6\_1 (n = 5), 6\_2 (n = 5), and 6\_3 (n = 7)), and sm stage b (7\_1 (n = 12), 7\_2 (n = 9), and 7\_3 (n = 11)). According to the infer of the peaks of novel mutation, all the samples were distributed to three phases: NT phase (normal tissue stage and hyperplasia stage), IEN phase (Tis stage to submucosa stage), and ESCC phase (T2 and T3 stage).

All 756 samples were subjected to proteomic profiling. Owing to the definite volume of the samples of early ESCC cohort, only 54 samples (from 19 ESCC patients) were adequate for phosphoproteomic profiling: normal tissue (n = 8), hyperplasia stage (n = 13), Tis stage (n = 10), lamina propria cancer stage (n = 5), muscularis mucosa stage (n = 3), sm stage a (n = 2), sm stage b (n = 8), T2 stage (n = 4), and T3 stage (n = 1). Additionally, only 53 samples (from 11 ESCC patients) covering 20 substages: stage 1 (n = 5), hyperplasia stage 2\_1 (n = 7), hyperplasia stage 2\_2 (n = 3), Tis stage 3\_1 (n = 1), Tis stage 3\_2 (n = 3), Tis stage 3\_3 (n = 4), Tis stage 3\_5 (n = 4), lamina propria cancer stage 4\_1 (n = 5), muscularis mucosa stage 5\_1 (n = 2), muscularis mucosa stage 5\_3 (n = 1), sm stage (a) 6\_1 (n = 1), sm stage (a) 6\_2 (n = 2), sm stage (a) 6\_3 (n = 1), sm stage (b) 7\_1 (n = 4), sm stage (b) 7\_2 (n = 2), sm stage (b) 7\_3 (n = 4), T2 stage (n = 2), and T3 stage (n = 2) were adequate for WES.

***Processing of FFPE specimens.*** All the FFPE specimens were prepared and provided by Zhongshan Hospital, Fudan University. For clinical sample preparation, slides (10 µm thick) from FFPE blocks were macro-dissected, deparaffinized with xylene and washed with ethanol. One 3 µm-thick slides from FFPE blocks were sectioned for hematoxylin and eosin (H&E) staining. All the selected specimens were

scraped according to the substages, which were evaluated and confirmed by two or three experienced and board-certified gastrointestinal pathologists, and materials were aliquoted and kept in storage at -80 °C until further processing.

**Whole-exome sequencing (WES).** WES was performed by Novogene Co., LTD. DNA from FFPE tumor tissue samples were collected and used for WES and matched germline DNA was obtained from non-tumor tissue samples. Fifty-three samples from 11 cases covering 20 substages were analyzed. Paired-end sequencing (PE150) was performed on an Illumina HiSeq (Illumina Novaseq 6000) with the mean coverage of the samples conducted in WES was 131× (range: 104–179×), and the mean volume of raw data was 14.0G (range: 10.6–19.3G), which was consistent with other literature studies<sup>80,81</sup>. The resulting sequence libraries (the paired-end sequence and insert DNA between two ends) were quantified with a Qubit 2.0 (Thermo Fisher), and the insert size was determined using an Agilent 2100 Bioanalyzer. The original fluorescence image files obtained from the Hiseq platform are transformed into short reads (raw data) by base calling. These short reads are recorded in FASTQ format, which contains sequence information and corresponding sequencing quality information. Base-calling was used to obtain the raw data (sequenced reads, mean (raw data) of all samples was no less than 12G) from the primary image data.

**DNA extraction and DNA qualification.** Fifty-three samples from 11 cases were analyzed by WES. All the samples were firstly dewaxing with dimethylbenzene, and then DNA degradation and contamination were monitored on 1% agarose gels. And DNA concentration was measured by Qubit® DNA Assay in Qubit® 2.0 Fluorometer (Invitrogen, USA, Catalog: 5190-8863). A total amount of at least 0.6 µg genomic DNA per sample was used as input for DNA sample preparation.

**Library preparation.** A total amount of 0.6 µg genomic DNA per sample was used as input for DNA sample preparation. Sequencing libraries were generated using Agilent SureSelect Human All Exon kit (Agilent Technologies, CA, USA, Catalog: 5190-8863) following the manufacturer's recommendations and index codes were added to each sample.

Fragmentation was carried out by hydrodynamic shearing system (Covaris, Massachusetts, USA) to generate randomly 180-280 bp fragments. Remaining overhangs were converted into blunt ends via exonuclease/polymerase activities. After adenylation of 3' ends of DNA fragments, adapter

oligonucleotides were ligated. DNA fragments with ligated adapter molecules on both ends were selectively enriched in a PCR reaction. After PCR reaction, libraries hybridize with liquid phase with biotin labeled probe, then use magnetic beads with streptomycin to capture the exons of genes. Captured libraries were enriched in a PCR reaction to add index tags to prepare for sequencing. Products were purified using AMPure XP system (Beckman Coulter, Beverly, USA) and quantified using the Agilent high sensitivity DNA assay on the Agilent Bioanalyzer 2100 system.

The clustering of the index-coded samples was performed on a cBot Cluster Generation System using Hiseq PE Cluster Kit (Illumina) according to the manufacturer's instructions. After cluster generation, the DNA libraries were sequenced on Illumina Hiseq platform and 150 bp paired-end reads were generated.

**QC of WES data processing and analysis.** The following criteria were used to ensure the high-quality clean data for the downstream bioinformatics analyses;

- (1) A paired read was discarded; if at least one read contains adapter contamination (>10 nucleotides aligned to the adapter), allowing  $\leq 10\%$  mismatches,
- (2) A paired read was discarded; if  $> 10\%$  of bases were uncertain in at least one read,
- (3) A paired read was discarded; if the proportion of low quality (Phred quality  $< 5$ ) bases was over 50% in either one read.

At the same time, QC statistics, including total reads number, raw data, raw depth, sequencing error rate, percentage of reads with Q30 (the percent of bases with a Phred-scaled quality score greater than 30 and GC content distribution were calculated and summarized.

**Reads mapping to the reference sequence.** Valid sequencing data were mapped to the reference human genome (UCSC hg19) by Burrows-Wheeler Aligner (BWA) software<sup>82</sup> to get the original mapping results stored in BAM format. If one or one paired read(s) were mapped to multiple positions, the strategy adopted by BWA was to choose the most likely placement. If two or more most likely placements were presented, BWA randomly picked one. Then, SAMtools<sup>83</sup> and Picard (<http://broadinstitute.github.io/picard/>) were used to sort BAM files and perform duplicate markings, local realignment, and base quality recalibration to generate final BAM file for computation of the sequence coverage and depth. The mapping step was very difficult due to mismatches, including true

mutation and sequencing errors, and duplicates resulted from PCR amplification. These duplicate reads were uninformative and should not be considered as evidence for variants. We used Picard to mark these duplicates for subsequent analysis.

**Detection and calling of somatic mutations.** BWA and Samblaster were used for genome alignment, and muTect Software<sup>84</sup> was used for identifying the Somatic SNV sites, whereas Strelka<sup>85</sup> was used to detect the Somatic InDels. Control-FREEC was used to detect SCNAs. SAMtools mpileup and bcftools were used for the variant calling and to identify the SNPs and InDels. Statistical analysis included two-tailed Student's test and Fisher's exact test.

**GISTIC and MutSig analysis.** To identify significantly amplified or deleted focal-level and arm-level events, we used The Genomic Identification of Significant Targets in Cancer (GISTIC) algorithm<sup>86</sup> to the genomic data in the Fudan cohort, and the Q value < 0.25 considered significant. Each gene of every sample is assigned a threshold copy number level that reflects the magnitude of its deletion or amplification. These are integer values ranging from -2 to 2, where 0 means no amplification or deletion of magnitude greater than the threshold parameters described above. Amplifications are represented by positive numbers: 1 means amplification above the amplification threshold; 2 means amplification larger than the arm level amplifications observed in the sample. Deletions are represented by negative numbers: -1 means deletion beyond the threshold, and -2 means deletions greater than the minimum arm-level copy number observed in the sample.

**The gain of novel mutations.** To investigate the mutation at all stages during the progression of ESCC, the numbers of total mutations and novel mutations at each stage were counted. The number of novel mutations at a certain stage could reflect the impacts of mutations in the progression of ESCC, demonstrating the genomic characteristics of the early-stage; therefore, we estimated the gain of novel mutations. The mutation frequency was estimated by the ratio of the number of mutated samples vs. the number of total samples<sup>12,23,28,80</sup>. Here, the novel mutation represented the mutations appearing at a certain stage, which has not been identified in earlier stages. For example, if *FAT3* was not mutated in stage 2 but mutated in stage 3, *FAT3* was considered the novel mutation in stage 3.

**Mutational signature analysis.** Based on the single nucleotide substitution and its' adjacent bases

pattern of samples, frequencies of 96 possible mutation types for each sample could be estimated. Non-negative matrix factorization (NMF) algorithm was used to estimate the minimal components that could explain maximum variance among samples. Then each component was compared to mutation patterns of 30 validated cancer signatures reported from the COSMIC database to identify cancer-related mutational signatures in the Fudan cohort and other ESCC cohorts. Cosine similarity analysis<sup>25,87</sup> was used to measure the similarity between component and signatures, which ranged from 0 to 1, indicating maximal dissimilarity to maximal similarity.

**Defining cancer-associated genes (CAGs).** CAGs were compiled from genes defined by Bailey et al.<sup>88</sup> and cancer-associated genes listed in Mertins et al.<sup>89</sup> and adapted from Vogelstein et al.<sup>90</sup>. The list of genes is provided in Supplementary Table 4c.

**Impacts of the detected mutations on the protein level. *Impacts of TP53 mutation.*** To compare the impact of mutations on the protein level across the 53 ESCC samples (n (early stages) = 49, n (T2/T3 stages) = 4), we firstly focused on the significantly enhanced proteins (t-test,  $p < 0.05$ ) by the mutation of *TP53* in the Fudan cohort, which were then overlapped with the proteins enhanced by the mutation of *TP53* in the IEN phase and ESCC phase. The overlapped proteins then were enriched into the GO/KEGG database. To compare the impacts of *TP53* mutation on the protein level and the associated biological pathways, fold change (FC) of the protein abundance (FOT) of the two groups (Mut group and WT group) was compared, and the  $p$  value was from by two-tailed Student's test.

***Analysis of SCNAs and the impacts on protein expressions.*** SCNAs analysis was performed using the WES-derived BAM files that were processed in the somatic mutation detection pipeline. These BAM files were further processed by the R package copywriteR (version 1.18.0), which used off-target WES read to infer copy number values. In this study, we used the multiomicsViz (version 1.6.0) in R (version 3.5.1) to perform the correlation of genomics and proteomics data. Correlations between SCNAs and proteome (with proteome data mapped to genes, by choosing the most variable protein as the gene-level representative) were determined using Spearman's correlation of common genes present in SCNA-proteome (3,474 mutations/proteins), which was the key event in esophageal carcinogenesis. Only genes or proteins with  $< 66.7\%$  NAs (missing values) were considered for the analysis, and protein IDs were mapped to gene names.

**Analysis of mutation spectra.** To compare the somatic SNVs of all samples in different tracks, we estimated the percentage of somatic SNVs in all samples. In this study, six types of SNVs were identified, and the percentage was used to investigate the difference in diverse groups of ESCC patients. Moreover, the percentages of the six types of SNVs were also used to compare the mutation spectra of the Fudan cohort with those of other published ESCC cohorts. When in SNVs analysis of diverse tracks and different clinical characteristics patients, two-tailed Student's test was applied.

**Protein extraction and trypsin digestion.** All samples of early and advanced ESCC patients were dissected with microdissection and then collected in 1.5 mL EP tubes, and then stored in -80°C refrigerator. The thickness of every FFPE piece is 10 µM, and every substage is no more than 10,000 cells.

50 µL TCEP buffer (2% deoxycholic acid sodium salt (Solarbio, Catalog: D8330), 40 mM 2-chloroacetamide (ALDRICH, Catalog: 22790-250G-F), 100 mM tris-phosphine hydrochloride (AMRESCO, Catalog: 0497), 10 mM (2-carboxyl)-phosphine hydrochloride (ALDRICH, Catalog: 4706-10G), 1 mM phenylmethylsulfonyl fluoride (AMRESCO, Catalog: M145-5G) mixed with MS water (J.T. Baker, Catalog: 4218-03), PH 8.8) was added into 1.5 mL EP tubes with prepared samples, and then heated in a 99°C metal bath for 30 min (minutes). Cool to room temperature, 3 µg trypsin (Promega, Catalog: V528A) was added into each tube and digested for 18 hours in a 37°C incubator. Then, 13 µL 10% formic acid (FA) (Sigma, Catalog: F0507) was added into each tube and made vortex for 3 min, and then sedimentation for 5 min (12,000 g). After that, a new 1.5 mL tube with 350 µL buffer (0.1% FA in 50% acetonitrile [ACN] (J.T. Baker, Catalog: 9830-03)) is needed for collecting the supernatant for extraction (vortex for 3 min, and then 12,000 g sedimentation for 5 min). And then the supernatant was transferred into a new tube for drying in a 60 °C vacuum drier. After drying, 100 µL 0.1% FA was needed for dissolving the peptides and vortex for 3 min, and then sedimentation for 3 min (12,000 g). The supernatant was picked into a new tube and then desalinated. Before desalination, the activation of pillars with 2 slides of 3M C8 disk is required, and the lipid is as follows: 90 µL 100% ACN twice, 90 µL 50% and 80% ACN once in turn, and then 90 µL 50% ACN once. After pillar balance with 90 µL 0.1% FA twice, the supernatant of the tubes was loading into the pillar twice, and decontamination with 90 µL 0.1% FA twice. Lastly, 90 µL elution buffer (0.1% FA in 50% ACN) was added into the pillar fir elution twice and only the effluent was collected for MS. And then the collection liquid was put in a 60 °C

vacuum drier for drying (~1.5 h).

**Proteome analysis in LC-MS/MS analysis.** For the proteomic profiling of samples, peptides were analyzed on a Q Exactive HF-X Hybrid Quadrupole-Orbitrap Mass Spectrometer (Thermo Fisher Scientific, Rockford, IL, USA) coupled with a high-performance liquid chromatography system (EASY nLC 1200, Thermo Fisher). Dried peptide samples re-dissolved in Solvent A (0.1% FA in water) were loaded to a 2-cm self-packed trap column (100- $\mu$ m inner diameter, 3  $\mu$ m ReproSil-Pur C18-AQ beads, Dr. Maisch GmbH) using Solvent A and separated on a 150- $\mu$ m-inner-diameter column with a length of 15 cm (1.9  $\mu$ m ReproSil-Pur C18-AQ beads, Dr. Maisch GmbH) over a 150 min gradient (Solvent A: 0.1% FA in water; Solvent B: 0.1% FA in 80% ACN) at a constant flow rate of 600 nL/min (0-150 min, 0 min, 4% B; 0-10 min, 4-15% B; 10-125 min, 15-30% B; 125-140 min, 30-50% B; 140-141 min, 50-100% B; 141-150 min, 100% B). The eluted peptides were ionized under 2.0 kV and introduced into mass spectrometer). MS was performed under a data-dependent acquisition mode. For the MS1 Spectra full scan, ions with m/z ranging from 300 to 1,400 were acquired by Orbitrap mass analyzer at a high resolution of 120,000. The automatic gain control (AGC) target value was set as 3E6. The maximal ion injection time was 80 ms. MS2 Spectra acquisition was performed in the ion trap mode at a rapid speed. Precursor ions were selected and fragmented with higher energy collision dissociation (HCD) with a normalized collision energy of 27%. Fragment ions were analyzed by the ion trap mass analyzer with the AGC target at 5E4. The maximal ion injection time of MS2 was 20 ms. Peptides that triggered MS/MS scans were dynamically excluded from further MS/MS scans for 12 s.

**Phosphopeptide enrichment and analysis.** All qualified profiling data were processed at firmiana platform against the human RefSeq protein database (updated on 04-07-2013) in the National Center for Biotechnology Information (NCBI). Owing to the definite volume of the samples of early ESCC cohort, only 54 samples (from 19 ESCC patients)—stage 1 (n = 8), stage 2 (n = 13), stage 3 (n = 10), stage 4 (n = 5), stage 5 (n = 3), stage 6 (n = 2), stage 7 (n = 8), stage 8 (n = 4), and stage 9 (n = 1) — were found to be adequate.

The phosphoproteome samples were prepared by Fe-NTA Phosphopeptide Enrichment Kit (Thermo, Catalog: A32992) according to the manufacturer's instruction. Briefly, 2 mg peptides were resuspended in 200  $\mu$ L binding/wash buffer and loaded to the equilibrated spin column. The resin was mixed with the

sample by gently tapping. The mixture was incubated for 30 min and centrifuged at  $1,000 \times g$  for 30 s to discard the flowthrough. The column was then washed by 200  $\mu\text{L}$  of binding/wash buffer and centrifuged at  $1,000 \times g$  for 30 s for 3 times and washed by 200  $\mu\text{L}$  of LC-MS grade water for one additional time. The phosphopeptide was eluted by adding 100  $\mu\text{L}$  of elution buffer and centrifuged at  $1,000 \times g$  for 30 s for 2 times. Phosphopeptides were dried down for LC-MS/MS analysis.

For the phosphoproteomic analysis, the phosphopeptides were analyzed on Orbitrap Fusion Lumos Tribrid Mass Spectrometer (Thermo Fisher Scientific, Rockford, IL, USA) equipped with an Easy nLC-1000 (Thermo Fisher Scientific, Rockford, IL, USA) and a Nanoflex source (Thermo Fisher Scientific, Rockford, IL, USA). Dried peptide samples re-dissolved in Solvent A (0.1% FA in water) were loaded to a 2-cm self-packed trap column using Solvent A and separated on a 150- $\mu\text{m}$ -inner-diameter column with a length of 30 cm over a 150 min gradient (buffer A: 0.1% FA in water; buffer B: 0.1% FA in 80% ACN) at a constant flow rate of 600 nL/min (0-150 min, 0 min, 4% B; 0-10 min, 4-15% B; 10-125 min, 15-30% B; 125-140 min, 30-50% B; 140-141 min, 50-100% B; 141-150 min, 100% B). The eluted phosphopeptides were ionized and detected. Mass spectra were acquired over the scan range of  $m/z$  350-1500 at a resolution of 120,000 (AUG target value of  $5E5$  and max injection time 50 ms). For the MS2 scan, the higher-energy collision dissociation fragmentation was performed at a normalized collision energy of 30%. The MS2 AGC target was set to  $1E4$  with a maximum injection time of 10 ms, Peptide mode was selected for monoisotopic precursor scan, and charge state screening was enabled to reject unassigned 1+, 7+, 8+, and  $> 8+$  ions with a dynamic exclusion time of 45 s to discriminate against previously analyzed ions between  $\pm 10$  ppm.

**Qualification of global proteome data.** The one-stop proteomic cloud platform (firmiana platform) was further employed for protein quantification. Identification results and the raw data from mzXML file were loaded. Then for each identified peptide, the XIC (extracted-ion chromatogram) was extracted by searching against the MS1 based on its identification information, and the abundance was estimated by calculating the area under the extracted XIC curve. For protein abundance calculation, the nonredundant peptide list was used to assemble proteins following the parsimony principle. Then, the protein abundance was estimated with a traditional label-free, intensity-based absolute quantification (iBAQ) algorithm, which divided the protein abundance (derived from identified peptides' intensities) by the number of theoretically observable peptides. Then the fraction of total (FOT), a relative quantification

value which was defined as a protein's iBAQ divided by the total iBAQ of all identified proteins in one experiment, was calculated as the normalized abundance of a particular protein among experiments. Finally, the FOT was further multiplied by 1E6 for the ease of presentation and FOTs less than 1E6 were replaced with 1E6 to adjust extremely small values.

At proteomic profiling of 756 samples, all the data was processed as following: E1 (13,711 GPs): all 13,711 gene products (GPs) (data not shown) identified in 756 samples (124 ESCC cases) on the basis of the match between runs algorithm<sup>91</sup>; E2 (9,741 GPs): we excluded keratins proteins of which the maximum FOT in all 756 experiments were less than 1.0E-5 in FOT, and all the proteins were required to have at least 1 unique strict peptide; E3 (6,687 GPs): GPs identified in more than one-sixth samples of each substage (Supplementary Table 3b).

For the quality control of the performance of mass spectrometry, the HEK293T cell (National Infrastructure Cell Line Resource) lysate was measured every four days as the quality-control standard. The quality-control standard was digested and analyzed using the same method and conditions as the early ESCC samples. The same as HEK293T cell samples, a pairwise Spearman's correlation coefficient of 114 substage 1 samples was used not only as QC but also as evaluation of the robustness of label-free quantification, and the results are shown in Supplementary Fig. 2g, h (Supplementary Table 3a). The mean value of correlation coefficient among the standards of 114 substage 1 samples (top 4,000 proteins) and 42 HEK293T cell samples (no NA values, 3,474 proteins) were 0.83 and 0.91 respectively, demonstrating the significantly consistent stability of the mass spectrometry platform.

**Batch effect analysis.** The hierarchical clustering and principal component analyses were implemented in R studio (version 3.5.1) to assess the batch effects in our proteome dataset with respect to the following two variables: batch identity and sample type (substage). For the hierarchical clustering analysis, the pair-wise Spearman's correlation coefficients of the same substage samples were first investigated. The samples in the same type exhibited a high similarity, whereas samples of different types clearly differed. There was no clear association between the batch intensity and correlation coefficients. Furthermore, we used the average linkage algorithm with one minus the Spearman's correlation coefficient as the dissimilarity measure. In the global heatmap in our study, each protein expression value in the global proteomic expression matrix was transformed into a Z-score across all samples. For the sample-wise and protein-wise clustering, the distance was set as 'euclidean' distance, and the weight method was

'complete'. The Z-score-transformed matrix was clustered using R package: pheatmap (version 1.0.12).

**Differential proteomics analysis.** A SAM<sup>49</sup> analysis identified 3,405 differential proteins in C1 and C2, which were identified based on all 756 samples from 114 early and 10 advanced ESCC patients. Statistics analysis was performed with two-tailed Student's test on overlapping samples to determine the differential abundance of proteins between two clusters and diverse tracks, in which statistical significance (t-test,  $p < 0.05$ , and differential expression  $FC \geq 2$  or  $\leq 0.5$ ) was considered in the differential analysis. The student's t test was used for statistical analysis. Proteins with no missing value (NA) in at least 20% samples were considered in each substage, in which no missing values were then imputed with 1/2 the minimum value of the proteomic dataset ( $5.0E-5$ ), which was applied in other studies<sup>29,92</sup>.

When comparing the DEPs of 22 substages, we focused on the substages highly expressed proteins (one substage versus all other substage), which were then enriched by GO/KEGG database<sup>93</sup>. We then annotated the signaling pathways ( $p < 0.05$ ) and manually checked the pathway associated proteins, which were then estimated whether significantly associated with the 22 substages of ESCC (Kruskal-Wallis test).

In differential analysis of proteins in ESCC progression (gradually decreased or increased) at the protein level, the highly expressed proteins of each substage/track/panel were screened, in which FC and  $p$  value (Kruskal-Wallis test) were considered. Statistical analysis was performed in R (version 3.5.1).

**Pathway enrichment analysis.** To investigate the dominant signaling pathways of two clusters, six tracks and 22 substages, we used gene sets of molecular pathways from KEGG<sup>93</sup>/Reactome<sup>94</sup> databases to compute the pathway. For this analysis, pathways from the GO/KEGG/Reactome databases were considered, and as background, the full list of genes/proteins observed under each data type was utilized. Statistical significance was considered when  $p$  value was less than 0.05. The differential score (Q) is obtained as signed  $-\log_{10}(p\text{-value})$  derived from two-tailed Student's test.

For assess the impacts of the mutations, we applied gene set enrichment analysis (GSEA) for pathways enrichment analysis<sup>95</sup>. GSEA evaluated and determined whether a priori defined sets of genes show statistically significant, cumulative changes in gene expression that are correlated with a specific phenotype. To assess the impacts of the mutation of *AKAP9* and other track mutations (e.g., *EPAS1*,

*EPHA3*, *MACF1*, *STAG2*, *USP6*), the samples grouped were subjected to GSEA, respectively. The proteins input were all the identified proteins in those samples (FDR < 0.01, unique strict peptides  $\geq 1$ ). Molecular Signatures Database (MSigDB) of KEGG gene sets (C2) were used for enrichment analysis. An FDR value of 0.05 was used as a cutoff. The enrichment score (ES) in GSEA was calculated by first ranking the proteins from the most to least significant with respect to the two phenotypes (i.e., Mut and WT, the entire ranked list was then used to assess how the proteins of each gene set were distributed across the ranked list.

**Principal component analysis.** We performed PCA on a total of 6,687 proteins (FOT) in E3 (the most variable proteins) of 756 samples to illustrate the global proteomic difference between the 22 ESCC substages. The PCA function under the scikit-learn R package was implemented for unsupervised clustering analysis with the parameter ‘n\_components = 2’ on the expression matrix of global proteomic data. A colored ellipse represented the 95% confidence coverage for each group, calculated based on the mean and covariance of points in each specific substage.

**Consensus clustering analysis of proteomics data.** The protein expression matrix of the 756 samples was used to identify the proteomic subtypes using the consensus cluster method. Consensus clustering was performed using the ConsensusClusterPlus R package (ConsensusclusterPlus, version 1.46.0)<sup>96</sup> with the most variable proteins (E3, n = 6,687). Consensus Cluster Plus parameters were reps = 1,000, pItem = 0.8, pFeature = 1, clusterAlg = ‘km’, distance = ‘euclidean’, plot = ‘PDF’. Euclidean distance and 1,000 resampling repetitions in the range of 2 to 10 clusters. The consensus matrix of k = 10 showed clear separation among clusters. The empirical cumulative distribution function (CDF) plot initially showed optimal separation. Clustering by k = 2 appeared to have the clearest cut between clusters and showed a significant association with the pathological substages. Taken together, proteome clusters were defined using k-means consensus clustering with k = 2. As summarized in Supplementary Fig. 5a, the clustering analysis of the samples (vertical column) by protein abundance (horizontal rows) divided all samples into 2 proteomic clusters defined by silhouette analyses (Fig. 4a).

**Survival analysis.** To investigate the impacts of mutations on protein expressions and the development of ESCC carcinogenesis, identified common genes in genomic and proteomic data of biological pathways were screened to perform the survival analysis. The data associated with overall survival (OS)

information was referenced from previously published ESCC cohorts. In addition, Kaplan-Meier survival curves (log-rank test) were used for OS analysis.  $p$  value (less than 0.05) for significance was used. Owing to the lack of proteomics data of ESCC, the OS data (RNA-seq,  $n$  (ESCC) = 81) of PGK1 was downloaded from the TCGA database (<https://portal.gdc.cancer.gov>)<sup>12</sup>.

To assess the impacts of *AKAP9* mutation on the ESCC patients' survival, we used the OS information referenced from previously published Song 2014 cohort<sup>13</sup>.

**The estimate of stromal and immune scores.** Tumor purity, immune, stromal scores were inferred by R package ESTIMATE version 1.0.11<sup>97</sup> using proteome data. For the analysis of global proteomic data, immune and stromal scores were estimated based on global proteins (E3,  $n=6,687$ ) in all 756 samples, which were computed and estimated via xCell (<https://xcell.ucsf.edu>)<sup>46</sup>. Normalized Z-score of the scores and cellular signature component was applied, and the significantly enhanced immune score, stromal score, and microenvironmental score during the progression of ESCC were estimated, which was also approved by the cell signature component. Statistical analysis was performed in R (version 3.5.1).

**Trajectory inference methods and tracks analysis.** We used the monocle (version 2.10.1) and trajectory inference methods to trace the carcinogenesis lineages in 114 early ESCC patients. Firstly, the proteins (E2,  $n = 9,741$ ) of all 746 samples (114 early ESCC cohort) were used. The top 2,500 identified proteins of each sample ordering by the median were used to construct the dataset and trace each patient's lineage. In addition, the proteins with mean expression over  $1.0E-1$  were highlighted and screened. The dataset was clustered and pre-prepared by t-distributed stochastic neighbor embedding (t-SNE) using a Barnes-Hut implementation with Rtsne (version 0.15) in R (version 3.5.1). All the substages of each early ESCC patient were considered as the pseudotime to construct the trajectory of each early ESCC patient. To the end, the trajectory of each ESCC patient was revealed; and then, 9 groups were determined by the number of nodes and bifurcations, and finally, 6 tracks were determined.

Sequentially, the track enhanced proteins were determined by the expression trend ( $K > 0$ ,  $R^2 > 0.15$ ) by fitting curves with ggplot2 (version 3.3.0) in R (version 3.5.1). All the substages of each track were considered as the time, and the progressively enhanced/decreased proteins were screened as the track positive/negative proteins and then used to determine the lineage of tracks.

**Fishplot analysis of six tracks.** Fishplot (version 0.5) analysis was performed to present the dynamic

driver pathways waves of 6 tracks (Supplementary Fig. 6h), in which all track proteins and substages were considered. Substage-based supervised clustering analysis of the 6 tracks displayed the substages' highly expressed proteins, which were then enriched into the Reactome database, determining the dominant pathways of each substage/panel. In the fishplot matrix, the 20 substages were considered as the length of the timepoint. Fishplot parameters were shape = 'spline', cex.title = 0.5, vlines = c(0, 20), vlab = substages 1 to 7\_3, titile.btm = 'Early ESCCs', plot = 'PDF'. In addition, the percentages of patients' number and certain substage in 6 tracks were considered and determined the proportion of the panels.

**Kinase activity prediction and phosphopeptide analysis.** The phospho-proteome data of 54 ESCC samples were searched against the same database with MaxQuant. The phosphorylation of S or T or Y was set as variable modification, in which three mis-cleavages were allowed, with a minimum Andromeda score of 40 for spectra matches. The ratios of identified phosphorylation sites of all samples were used to estimate the kinase activities by Kinase-Substrate Enrichment Analysis (KSEA) algorithm<sup>98</sup>. The information of kinase-substrate relationships was obtained from publicly available databases, including PhosphoSite<sup>99</sup>, Phospho.ELM<sup>100</sup> and PhosphoPOINT<sup>101</sup>. The information of substrate motifs was obtained either from the literature<sup>66</sup> or from an analysis of the KSEA dataset with Motif (sP)<sup>102</sup>. PGK1 S203 was the only phosphosite, which was frequently detected (53/54) in ESCC progression. The motif (sP) was then matched to Human Protein Reference Database ([http://hprd.org/PhosphoMotif\\_finder](http://hprd.org/PhosphoMotif_finder)) and the kinase-substrate-motif network analysis was referenced from PhosphoSitePlus (PSP, <https://www.phosphosite.org/homeAction>)<sup>103</sup> and NetworKIN 3.0<sup>104</sup>. Statistical analysis was performed in R (version 3.5.1) with Kruskal-Wallis test.

**Immunohistochemistry (IHC) analysis.** To detect the expression of PGK1 in the tissue by IHC staining, 3- $\mu$ m-thick sections from each formalin-fixed, paraffin-embedded (FFPE) tissue block were de-waxed with xylene and rehydrated through a graded series of ethanol, prepared by Zhongshan Hospital, Fudan University.

Total PGK1 immunostaining was performed on representative samples from normal to progressive ESCC. The IHC assay using PGK1 rabbit antibody (Wuhan Fine Biotech Co., Ltd, Catalog: FNab06354, dilution 1:200) was performed with Ventana iView DAB Detection Kit on a BenchMark XT automated

staining system (Ventana Medical Systems, Tucson, AZ). Normal IgG from the same species of primary antibody diluted to match the concentration of the primary antibody was used as the negative control. For PGK1 negative cases, the experiment was repeated on the whole section in order to exclude heterogeneity. For assessment of staining, slides were scanned with the ScanScope System (Aperio, CA), and viewed with ImageScope (Aperio).

**Metabolite quantification.** The pyruvate, lactate, citrate, succinate, fumarate, and glycine levels were measured using NMR spectra as described before<sup>105</sup>. 3-phosphoglycerate and serine levels were measured using LC-MS/MS. Briefly,  $\sim 1 \times 10^7$  cells were treated with a cold aqueous methanol solution (80% v/v) to stop cell metabolism quickly. Samples were then centrifuged for 15 min at  $15,000 \times g$  and  $4^\circ\text{C}$ , after which the supernatants were collected. The supernatants were then lyophilized and reconstituted in 500  $\mu\text{L}$  methanol/water (10:90 v/v). The separated metabolites were acquired using high-performance liquid chromatography (HPLC) employing an LC-20AB pump (Shimadzu, Kyoto, Japan) and the Luna NH2 column (P/N 00B-4378-B0; 5  $\mu\text{m}$ ,  $50 \times 2.0$  mm; Phenomenex, Torrance, CA). The mobile phase comprised eluent A (0.77 g  $\text{NH}_4\text{OAc}$ , 1.25 mL  $\text{NH}_4\text{OH}$ , 25 mL ACN, and 300  $\mu\text{L}$  acetic acid [HAc] dissolved in 500 mL water) and eluent B (ACN). The elution program was as follows, 0.1 min, 85% B; 3 min, 30% B; 12 min, 2% B; 15 min, 2% B; and 16–28 min, 85% B. The flow rate of the pump was 0.3 mL/min, and the mass spectrometer used was the 4000 QTRAP system (AB Sciex, Framingham, MA) operated in multiple reaction monitoring (MRM) mode. The MS parameters were electrospray voltage, 5 kV; gas 1, 30 kPa; gas 2, 30 kPa; curtain gas, 25 kPa; and temperature,  $500^\circ\text{C}$ . The ions monitored for 3-phosphoglycerate and serine were at 185-79 and 106-60, respectively.

**Cell lines and cell culture.** The following ESCC cell lines were used in this assay: ECA109 cells (ATCC, Catalog: GCC-OE0002CS), KYSE150 cells (ATCC, Catalog: GCC-OE0004CS). All cells were cultivated in RPMI-1640 medium (HyClone, Logan, UT, USA) supplemented with 10% fetal bovine serum (HyClone, Logan, UT, USA), and incubated at  $37^\circ\text{C}$  in 5%  $\text{CO}_2$ .

**Gene overexpression and knockdown.** For transient gene overexpression, the whole-length cDNA of *PGK1*, *GAPDH* and *PGM1* were cloned into pcDNA3.1 (b)-Flag vector between the NheI and EcoRI sites. The whole-length cDNA of *ERK2* was cloned into pcDNA3.1(b)-Myc vector between the NheI and EcoRI sites, whereas the plasmid PGK1-S203A-Flag was generated by site-directed mutagenesis using

the Muta-nBEST kit (TaKaRa, Kyoto, Japan, Catalog: R401) according to the manufacturer's instructions. Then plasmids were transfected into KYSE150 cells and ECA109 cells using Lipofectamine 3000 (Invitrogen, Carlsbad, CA, USA) according to the manufacturer's instructions. For stable overexpression of *PGK1*, the whole-length cDNA of *PGK1* was cloned into vector pBABE puro between the BamHI and EcoRI sites; then, the plasmids were co-transfected with pCMV-VSV-G and pCMV-Gag-Pol plasmids into ECA109 cells and KYSE150 cells using the calcium phosphate method. For stable knockdown of *PGK1* and *ERK2*, pMKO.1-shRNA plasmids encoding specific shRNAs targeting human PGK1 (5'-CTGACAAGTTTGATGAGAATG-3') and human ERK2 (5'-CAAAGTTCGAGTAGCTATCAA-3') were transfected, together with pCMV-VSV-G and pCMV-Gag-Pol plasmids, into one HEK293T packaging cell line using the calcium phosphate method and the virus supernatants were collected from the medium for the subsequent infection of KYSE150 cells and ECA109 cells.

**Immunoprecipitation.** For immunoprecipitation of the FLAG-tagged proteins, cells were lysed with 0.1% NP-40 buffer containing 50 mM Tris-HCl (pH 7.5), 150 mM NaCl, 0.1% NONIDET P-40, 1 µg/mL aprotinin, 1 µg/mL leupeptin, 1 µg/mL pepstatin, and 1 mM PMSF. The whole-cell lysates were incubated with monoclonal anti-Flag antibody-conjugated M2 agarose beads (Sigma) for 4 h at 4°C. The bound proteins were triple-washed with 0.1% NP-40 buffer.

**Western blot analysis.** Standard procedures were followed for western blot analysis. Primary antibodies used in this study include anti-PGK1 antibody (Wuhan Fine Biotech Co., Ltd., China, Catalog: FNab06354, dilution 1:1000), Anti-β-actin (Genscript, Piscataway, NJ, USA, Catalog: A00702, dilution 1: 10000), Anti-p-Ser (Santa Cruz Biotechnology, CA, USA, Catalog: sc-81514, dilution 1: 500), Anti-phospho-Threonine (Cell Signaling Technology, Danvers, MA, USA, Catalog: 9386, dilution 1: 1000), Anti-phosphotyrosine (Catalog: 05-947, clone PY20, dilution 1: 1000), Anti-COX IV (Cell Signaling Technology, Danvers, MA, USA, Catalog: 4580, dilution 1: 1000), Anti-GAPDH (Yeasen, Shanghai, China, Catalog: 30201ES20, dilution 1: 10000), Anti-Thr-338 PDHK1 (Signalway Antibody, Nanjing, China, Catalog: C11596, dilution 1: 500), Anti-PDHK1 (Cell Signaling Technology, Danvers, MA, USA, Catalog: 3820, dilution 1: 1000), Anti-ERK2 (Cell Signaling Technology, Danvers, MA, USA, Catalog: 9108, dilution 1: 1000), Anti-Flag (Abmart, Shanghai, China, Catalog: M20008, dilution 1: 5000), Anti-

PGM1 (Cell Signaling Technology, Danvers, MA, USA, Catalog: 12098, dilution 1: 1000), and Anti-PHGDH (Novusbio, Littleton, CO, USA, Catalog: NBP1-87311, dilution 1: 1000). Western blot signals were obtained by detecting chemiluminescence by using a Typhoon FLA 9500 biomolecular imager (GE Healthcare).

**PGK1 purification and enzymatic assay.** Flag-tagged PGK1 protein immuno-precipitated from HEK293T cells was eluted with Flag peptide buffer. The eluent was further purified and concentrated using an Amicon Ultra Centrifugal Filter (10 kDa molecular weight cutoff, Millipore) in a buffer containing 50 mM Tris-HCl pH 7.5, 100 mM KCl, 5 mM MgCl<sub>2</sub> and 5% glycerol. PGK1 activity was measured using purified Flag-tagged PGK1 (0.2 µg/mL) mixed with DMSO or different concentrations of gemcitabine in the reaction buffer containing 50 mM Tris-HCl (pH 7.6), 8 mM MgCl<sub>2</sub>, 4 mM ATP, 0.2 mM NADH, 12 mM 3-phosphoglycerate, and 8 U of GAPDH at 25 °C. The change in absorbance at 340 nm owing to the decrease of NADH was measured.

**PDH activity assay.** The assays were carried out using Pyruvate dehydrogenase (PDH) Enzyme Activity Microplate Assay Kit (Abcam, Catalog: ab109902) according to the manufacture's guidance. The total intracellular PDH activities of ECA109 cells and KYSE150 cells were normalized with the protein expression levels of COX IV.

**Cell proliferation assay.** Cell proliferation was assessed using the Cell Counting Kit-8 (Dojindo Molecular Technologies, Inc, Kumamoto, Japan, Catalog: CK04). In brief, cells were seeded in a 96-well plate at  $4 \times 10^3$  cells per well and allowed to adhere. Cell Counting Kit-8 solution (10 µL) was added to each well, and the cells were cultured in 5% CO<sub>2</sub> at 37°C for 2 h. Cell proliferation was determined by measuring the absorbance at 450 nm.

**Mouse xenograft assay.** Four- to six-week-old Balb/C nude mice were obtained (Shanghai SLAC Laboratory Animal Co., Ltd, Shanghai, China) for *in vivo* xenografts. ESCC cells ( $\sim 1 \times 10^7$ ) were subcutaneously injected into nude mice (n = 10 mice/condition). When the tumor volumes reached approximately 100 mm<sup>3</sup>, PGK1-OE tumor-bearing mice were randomly separated into two groups (n = 10 per group) as follows: PGK1-OE group and PGK1-OE-inhibitor (gemcitabine (Sigma-Aldrich, Catalog: G6423, 6 mg/kg) group. Mice in the group of PGK1-OE-inhibitor were injected intravenously

every other day for eight times; while the mice in the other group were injected saline. Tumors were harvested and weighed after 30 days post-injection. This study was conducted under the guidelines of the animal care regulations of Fudan University and received ethical and scientific approval from Fudan University.

**Quantification and statistical analysis.** Statistical details of experiments and analyses can be found in the (supplementary)/figure legends and supplementary tables in this text. Statistical significance tests were as noted in each analysis, in which two-way ANOVA test and log-rank test were performed in GraphPad Prism (version 7.0), and Fisher's exact test, Bartlett's test, Kruskal-Wallis test and Spearman correlation test were performed using R (version 3.5.1). Data in the barplot were presented as mean  $\pm$  SEM. All statistical tests were two-tailed, and statistical significance was considered when  $p$  value  $<$  0.05. Kaplan–Meier plots (log-rank test) were used to describe the OS.

#### **Data and code availability**

All proteomics and phosphoproteomics raw data have been uploaded to the iProX Consortium (<https://www.iprox.org/>) with the subproject ID (IPX00021780001 and IPX00021780004, respectively). In addition, all proteomics and phosphoproteomics raw data have been deposited at the firmiana platform (a one-stop proteomic cloud platform) (<http://www.firmiana.org/login/>)<sup>79</sup>, and the qualified profiling datasets were processed at the platform. All the WES raw data have been deposited at NODE (<https://www.biosino.org/node/>) (accession ID: OEP001324) or through the URL: <https://www.biosino.org/node/project/detail/OEP001324>.

#### **References**

- 1 Siegel, R. L., Miller, K. D. & Jemal, A. Cancer Statistics, 2018. *Ca-Cancer J Clin* **68**, 7-30, doi:10.3322/caac.21442 (2018).
- 2 Bray, F. *et al.* Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* **68**, 394-424, doi:10.3322/caac.21492 (2018).
- 3 Brown, L. M., Devesa, S. S. & Chow, W. H. Incidence of adenocarcinoma of the esophagus among white Americans by sex, stage, and age. *J Natl Cancer I* **100**, 1184-1187,

- doi:10.1093/jnci/djn211 (2008).
- 4 Sawada, G. *et al.* Genomic Landscape of Esophageal Squamous Cell Carcinoma in a Japanese Population. *Gastroenterology* **150**, 1171-1182, doi:10.1053/j.gastro.2016.01.035 (2016).
- 5 Daly, J. M. Esophageal cancer: Results of an American College of Surgeons patient care evaluation study - Reply. *J Am Coll Surgeons* **190**, 573-573, doi:Doi 10.1016/S1072-7515(00)00275-1 (2000).
- 6 Ishiguro, S. *et al.* Effect of alcohol consumption, cigarette smoking and flushing response on esophageal cancer risk: A population-based cohort study (JPHC study). *Cancer Lett* **275**, 240-246, doi:10.1016/j.canlet.2008.10.020 (2009).
- 7 Ohashi, S. *et al.* Recent Advances From Basic and Clinical Studies of Esophageal Squamous Cell Carcinoma. *Gastroenterology* **149**, 1700-1715, doi:10.1053/j.gastro.2015.08.054 (2015).
- 8 Wu, A. H., Wan, P. & Bernstein, L. A multiethnic population-based study of smoking, alcohol and body size and risk of adenocarcinomas of the stomach and esophagus (United States). *Cancer Cause Control* **12**, 721-732, doi:Doi 10.1023/A:1011290704728 (2001).
- 9 Brown, L. M. *et al.* Excess incidence of squamous cell esophageal cancer among US black men: Role of social class and other risk factors. *Am J Epidemiol* **153**, 114-122 (2001).
- 10 Weinstein, J. N. *et al.* The Cancer Genome Atlas Pan-Cancer analysis project. *Nat Genet* **45**, 1113-1120, doi:10.1038/ng.2764 (2013).
- 11 Lin, D. C. *et al.* Genomic and molecular characterization of esophageal squamous cell carcinoma. *Nat Genet* **46**, 467-473, doi:10.1038/ng.2935 (2014).
- 12 Cancer Genome Atlas Research, N. *et al.* Integrated genomic characterization of oesophageal carcinoma. *Nature* **541**, 169-175, doi:10.1038/nature20805 (2017).
- 13 Song, Y. *et al.* Identification of genomic alterations in oesophageal squamous cell cancer. *Nature* **509**, 91-95, doi:10.1038/nature13176 (2014).
- 14 Holscher, A. H. *et al.* Prognostic Impact of Upper, Middle, and Lower Third Mucosal or Submucosal Infiltration in Early Esophageal Cancer. *Annals of Surgery* **254**, 802-808, doi:10.1097/SLA.0b013e3182369128 (2011).
- 15 Ono, H. *et al.* Endoscopic mucosal resection for treatment of early gastric cancer. *Gut* **48**, 225-229, doi:DOI 10.1136/gut.48.2.225 (2001).
- 16 Enzinger, P. C. & Mayer, R. J. Medical progress - Esophageal cancer. *New Engl J Med* **349**,

- 2241-2252, doi:DOI 10.1056/NEJMra035010 (2003).
- 17 Yamamoto, H. *et al.* Successful en bloc resection of a large superficial gastric cancer by using sodium hyaluronate and electrocautery incision forceps. *Gastrointest Endosc* **54**, 629-632, doi:DOI 10.1067/mge.2001.118643 (2001).
- 18 Nishizawa, T. & Yahagi, N. Endoscopic mucosal resection and endoscopic submucosal dissection: technique and new directions. *Curr Opin Gastroenterol* **33**, 315-319, doi:10.1097/MOG.0000000000000388 (2017).
- 19 Song, W. C., Qiao, X. L. & Gao, X. Z. A comparison of endoscopic submucosal dissection (ESD) and radical surgery for early gastric cancer: a retrospective study. *World J Surg Oncol* **13**, 309, doi:10.1186/s12957-015-0724-1 (2015).
- 20 Jiang, D. *et al.* A retrospective study of endoscopic resection for 368 patients with early esophageal squamous cell carcinoma or precancerous lesions. *Surg Endosc* **31**, 2122-2130, doi:10.1007/s00464-016-5208-z (2017).
- 21 Wani, S. *et al.* Comparison of endoscopic therapies and surgical resection in patients with early esophageal cancer: a population-based study. *Gastrointest Endosc* **79**, 224-232 e221, doi:10.1016/j.gie.2013.08.002 (2014).
- 22 Schlemper, R. J. *et al.* Differences in diagnostic criteria for gastric carcinoma between Japanese and Western pathologists. *Lancet* **349**, 1725-1729, doi:Doi 10.1016/S0140-6736(96)12249-2 (1997).
- 23 Liu, X. *et al.* Genetic Alterations in Esophageal Tissues From Squamous Dysplasia to Carcinoma. *Gastroenterology* **153**, 166-177, doi:10.1053/j.gastro.2017.03.033 (2017).
- 24 Lee, D. D. & Seung, H. S. Learning the parts of objects by non-negative matrix factorization. *Nature* **401**, 788-791, doi:10.1038/44565 (1999).
- 25 Mayakonda, A., Lin, D. C., Assenov, Y., Plass, C. & Koeffler, H. P. Maftools: efficient and comprehensive analysis of somatic variants in cancer. *Genome Res* **28**, 1747-1756, doi:10.1101/gr.239244.118 (2018).
- 26 Alexandrov, L. B. *et al.* Signatures of mutational processes in human cancer. *Nature* **500**, 415-421, doi:10.1038/nature12477 (2013).
- 27 Cerami, E. *et al.* The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov* **2**, 401-404, doi:10.1158/2159-

- 8290.CD-12-0095 (2012).
- 28 Jiang, Y. *et al.* Proteomics identifies new therapeutic targets of early-stage hepatocellular carcinoma. *Nature* **567**, 257-+, doi:10.1038/s41586-019-0987-8 (2019).
- 29 Ge, S. *et al.* A proteomic landscape of diffuse-type gastric cancer. *Nat Commun* **9**, 1012, doi:10.1038/s41467-018-03121-2 (2018).
- 30 Schwanhaussner, B. *et al.* Global quantification of mammalian gene expression control (vol 473, pg 337, 2011). *Nature* **495**, 126-127, doi:10.1038/nature11848 (2013).
- 31 Schwanhaussner, B. *et al.* Global quantification of mammalian gene expression control. *Nature* **473**, 337-342, doi:10.1038/nature10098 (2011).
- 32 Yazdian-Robati, R. *et al.* Comparative proteome analysis of human esophageal cancer and adjacent normal tissues. *Iran J Basic Med Sci* **20**, 265-271, doi:10.22038/ijbms.2017.8354 (2017).
- 33 Pawar, H. *et al.* Quantitative tissue proteomics of esophageal squamous cell carcinoma for novel biomarker discovery. *Cancer Biol Ther* **12**, 510-522, doi:10.4161/cbt.12.6.16833 (2011).
- 34 Berridge, M. J., Lipp, P. & Bootman, M. D. The versatility and universality of calcium signalling. *Nat Rev Mol Cell Biol* **1**, 11-21, doi:10.1038/35036035 (2000).
- 35 Ashby, M. C. & Tepikin, A. V. Polarized calcium and calmodulin signaling in secretory epithelia. *Physiol Rev* **82**, 701-734, doi:10.1152/physrev.00006.2002 (2002).
- 36 Reczek, C. R. & Chandel, N. S. ROS Promotes Cancer Cell Survival through Calcium Signaling. *Cancer Cell* **33**, 949-951, doi:10.1016/j.ccell.2018.05.010 (2018).
- 37 Tsherniak, A. *et al.* Defining a Cancer Dependency Map. *Cell* **170**, 564-+, doi:10.1016/j.cell.2017.06.010 (2017).
- 38 Whitfield, M. L., George, L. K., Grant, G. D. & Perou, C. M. Common markers of proliferation. *Nat Rev Cancer* **6**, 99-106, doi:10.1038/nrc1802 (2006).
- 39 Hua, X. *et al.* MEGSA: A Powerful and Flexible Framework for Analyzing Mutual Exclusivity of Tumor Mutations. *Am J Hum Genet* **98**, 442-455, doi:10.1016/j.ajhg.2015.12.021 (2016).
- 40 Chen, H. J. *et al.* The role of microtubule actin cross-linking factor 1 (MACF1) in the Wnt signaling pathway. *Genes Dev* **20**, 1933-1945, doi:10.1101/gad.1411206 (2006).
- 41 Cui, G. L. *et al.* Tumor-Associated Fibroblasts and Microvessels Contribute to the Expression of Immunosuppressive Factor Indoleamine 2, 3-Dioxygenase in Human Esophageal Cancers.

- Pathol Oncol Res* **24**, 269-275, doi:10.1007/s12253-017-0244-0 (2018).
- 42 Mani, S. A. *et al.* The epithelial-mesenchymal transition generates cells with properties of stem cells. *Cell* **133**, 704-715, doi:10.1016/j.cell.2008.03.027 (2008).
- 43 Tam, W. L. & Weinberg, R. A. The epigenetics of epithelial-mesenchymal plasticity in cancer. *Nat Med* **19**, 1438-1449, doi:10.1038/nm.3336 (2013).
- 44 Kalluri, R. EMT: When epithelial cells decide to become mesenchymal-like cells. *J Clin Invest* **119**, 1417-1419, doi:10.1172/Jci39675 (2009).
- 45 Dongre, A. & Weinberg, R. A. New insights into the mechanisms of epithelial-mesenchymal transition and implications for cancer. *Nat Rev Mol Cell Bio* **20**, 69-84, doi:10.1038/s41580-018-0080-4 (2019).
- 46 Aran, D., Hu, Z. & Butte, A. J. xCell: digitally portraying the tissue cellular heterogeneity landscape. *Genome Biol* **18**, 220, doi:10.1186/s13059-017-1349-1 (2017).
- 47 Manier, S. *et al.* Genomic complexity of multiple myeloma and its clinical implications. *Nat Rev Clin Oncol* **14**, 100-113, doi:10.1038/nrclinonc.2016.122 (2017).
- 48 van der Schaaf, M., Johar, A., Wijnhoven, B., Lagergren, P. & Lagergren, J. Extent of Lymph Node Removal During Esophageal Cancer Surgery and Survival. *Jnci-J Natl Cancer I* **107**, doi:ARTN djv04310.1093/jnci/djv043 (2015).
- 49 Tusher, V. G., Tibshirani, R. & Chu, G. Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci U S A* **98**, 5116-5121, doi:10.1073/pnas.091062498 (2001).
- 50 Yu, V. Z. Y. *et al.* Endoplasmic reticulum-localized ECM1b suppresses tumor growth and regulates MYC and MTORC1 through modulating MTORC2 activation in esophageal squamous cell carcinoma. *Cancer Lett* **461**, 56-64, doi:10.1016/j.canlet.2019.07.005 (2019).
- 51 Rubin, C. S. A Kinase Anchor Proteins and the Intracellular Targeting of Signals Carried by Cyclic-Amp. *Bba-Mol Cell Res* **1224**, 467-479 (1994).
- 52 Michel, J. J. & Scott, J. D. AKAP mediated signal transduction. *Annu Rev Pharmacol Toxicol* **42**, 235-257, doi:10.1146/annurev.pharmtox.42.083101.135801 (2002).
- 53 Lin, C. X. *et al.* Nobiletin inhibits cell growth through restraining aerobic glycolysis via PKA-CREB pathway in oral squamous cell carcinoma. *Food Sci Nutr* **8**, 3515-3524, doi:10.1002/fsn3.1634 (2020).

- 54 Bertherat, J. *et al.* Molecular and functional analysis of PRKAR1A and its locus (17q22-24) in sporadic adrenocortical tumors: 17q losses, somatic mutations, and protein kinase A expression and activity. *Cancer Res* **63**, 5308-5319 (2003).
- 55 Webb, B. A. *et al.* Structures of human phosphofructokinase-1 and atomic basis of cancer-associated mutations. *Nature* **523**, 111-+, doi:10.1038/nature14405 (2015).
- 56 Sengupta, U. & Strodel, B. Markov models for the elucidation of allosteric regulation. *Philos T R Soc B* **373**, doi:ARTN 2017017810.1098/rstb.2017.0178 (2018).
- 57 Weber, G., Convery, H. J. H., Lea, M. A. & Stamm, N. B. Feedback Inhibition of Key Glycolytic Enzymes in Liver - Action of Free Fatty Acids. *Science* **154**, 1357-&, doi:DOI 10.1126/science.154.3754.1357 (1966).
- 58 Saelens, W., Cannoodt, R., Todorov, H. & Saeys, Y. A comparison of single-cell trajectory inference methods. *Nat Biotechnol* **37**, 547-554, doi:10.1038/s41587-019-0071-9 (2019).
- 59 Daly, J. M. *et al.* Esophageal cancer: results of an American College of Surgeons Patient Care Evaluation Study. *J Am Coll Surg* **190**, 562-572; discussion 572-563 (2000).
- 60 Crocetti, E. Global, Regional, and National Cancer Incidence, Mortality, Years of Life Lost, Years Lived With Disability, and Disability-Adjusted Life-Years for 29 Cancer Groups, 1990 to 2017: A Systemic Analysis for the Global Burden of Disease Study (vol 5, pg 1749, 2019). *Jama Oncol* **6**, 789-789, doi:10.1001/jamaoncol.2020.1421 (2020).
- 61 Mansouri, A., Gattolliat, C. H. & Asselah, T. Mitochondrial Dysfunction and Signaling in Chronic Liver Diseases. *Gastroenterology* **155**, 629-647, doi:10.1053/j.gastro.2018.06.083 (2018).
- 62 Miller, C. A. *et al.* Visualizing tumor evolution with the fishplot package for R. *Bmc Genomics* **17**, doi:ARTN 88010.1186/s12864-016-3195-z (2016).
- 63 Momper, J. D. *et al.* Adolescent Dosing and Labeling Since the Food and Drug Administration Amendments Act of 2007. *Jama Pediatr* **167**, 926-932, doi:10.1001/jamapediatrics.2013.465 (2013).
- 64 Kojima, T. & Doi, T. Immunotherapy for Esophageal Squamous Cell Carcinoma. *Curr Oncol Rep* **19**, doi:ARTN 3310.1007/s11912-017-0590-9 (2017).
- 65 Beausoleil, S. A. *et al.* Large-scale characterization of HeLa cell nuclear phosphoproteins. *P Natl Acad Sci USA* **101**, 12130-12135, doi:10.1073/pnas.0404720101 (2004).

- 66 Schwartz, D. & Gygi, S. P. An iterative statistical approach to the identification of protein phosphorylation motifs from large-scale data sets. *Nat Biotechnol* **23**, 1391-1398, doi:10.1038/nbt1146 (2005).
- 67 Li, X. J. *et al.* Mitochondria-Translocated PGK1 Functions as a Protein Kinase to Coordinate Glycolysis and the TCA Cycle in Tumorigenesis. *Mol Cell* **61**, 705-719, doi:10.1016/j.molcel.2016.02.009 (2016).
- 68 Krishnan, P. *et al.* Phosphorylation of pyrimidine deoxynucleoside analog Diphosphates - Selective phosphorylation of L-nucleoside analog diphosphates by 3-phosphoglycerate kinase. *J Biol Chem* **277**, 5453-5459, doi:10.1074/jbc.M109025200 (2002).
- 69 Clapham, D. E. Calcium signaling. *Cell* **131**, 1047-1058, doi:10.1016/j.cell.2007.11.028 (2007).
- 70 Bong, A. H. L. & Monteith, G. R. Calcium signaling and the therapeutic targeting of cancer cells. *Bba-Mol Cell Res* **1865**, 1786-1794, doi:10.1016/j.bbamcr.2018.05.015 (2018).
- 71 Roderick, H. L. & Cook, S. J. Ca<sup>2+</sup> signalling checkpoints in cancer: remodelling Ca<sup>2+</sup> for cancer cell proliferation and survival. *Nat Rev Cancer* **8**, 361-375, doi:10.1038/nrc2374 (2008).
- 72 Baron, R. & Kneissel, M. WNT signaling in bone homeostasis and disease: from human mutations to treatments. *Nat Med* **19**, 179-192, doi:10.1038/nm.3074 (2013).
- 73 Senthebane, D. A. *et al.* The Role of Tumor Microenvironment in Chemoresistance: 3D Extracellular Matrices as Accomplices. *Int J Mol Sci* **19**, doi:ARTN 286110.3390/ijms19102861 (2018).
- 74 Fang, S. *et al.* Clinical significance and biological role of cancer-derived Type I collagen in lung and esophageal cancers. *Thorac Cancer* **10**, 277-288, doi:10.1111/1759-7714.12947 (2019).
- 75 Guo, M. Z., Liu, S., Herman, J. G., Zhuang, H. & Lu, F. M. Gefitinib-sensitizing mutation in esophageal carcinoma cell line Kyse450. *Cancer Biol Ther* **5**, 152-155, doi:DOI 10.4161/cbt.5.2.2318 (2006).
- 76 Racker, E. & Spector, M. Warburg Effect Revisited - Merger of Biochemistry and Molecular-Biology. *Science* **213**, 303-307, doi:DOI 10.1126/science.6264596 (1981).
- 77 Heiden, M. G. V., Cantley, L. C. & Thompson, C. B. Understanding the Warburg Effect: The Metabolic Requirements of Cell Proliferation. *Science* **324**, 1029-1033, doi:10.1126/science.1160809 (2009).
- 78 Mini, E., Nobili, S., Caciagli, B., Landini, I. & Mazzei, T. Cellular pharmacology of gemcitabine.

- Ann Oncol* **17**, V7-V12, doi:10.1093/annonc/mdj941 (2006).
- 79 Feng, J. *et al.* Firmiana: towards a one-stop proteomic cloud platform for data processing and analysis. *Nat Biotechnol* **35**, 409-412, doi:10.1038/nbt.3825 (2017).
- 80 Gao, Q. *et al.* Integrated Proteogenomic Characterization of HBV-Related Hepatocellular Carcinoma (vol 179, pg 561, 2019). *Cell* **179**, 1240-1240, doi:10.1016/j.cell.2019.10.038 (2019).
- 81 Xu, J. Y. *et al.* Integrative Proteomic Characterization of Human Lung Adenocarcinoma. *Cell* **182**, 245+, doi:10.1016/j.cell.2020.05.043 (2020).
- 82 Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-1760, doi:10.1093/bioinformatics/btp324 (2009).
- 83 Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079, doi:10.1093/bioinformatics/btp352 (2009).
- 84 Cibulskis, K. *et al.* Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat Biotechnol* **31**, 213-219, doi:10.1038/nbt.2514 (2013).
- 85 Saunders, C. T. *et al.* Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs. *Bioinformatics* **28**, 1811-1817, doi:10.1093/bioinformatics/bts271 (2012).
- 86 Mermel, C. H. *et al.* GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol* **12**, R41, doi:10.1186/gb-2011-12-4-r41 (2011).
- 87 Alexandrov, L. B., Nik-Zainal, S., Wedge, D. C., Campbell, P. J. & Stratton, M. R. Deciphering signatures of mutational processes operative in human cancer. *Cell Rep* **3**, 246-259, doi:10.1016/j.celrep.2012.12.008 (2013).
- 88 Bailey, M. H. *et al.* Comprehensive Characterization of Cancer Driver Genes and Mutations. *Cell* **174**, 1034-1035, doi:10.1016/j.cell.2018.07.034 (2018).
- 89 Mertins, P. *et al.* Proteogenomics connects somatic mutations to signalling in breast cancer. *Nature* **534**, 55-62, doi:10.1038/nature18003 (2016).
- 90 Vogelstein, B. *et al.* Cancer genome landscapes. *Science* **339**, 1546-1558, doi:10.1126/science.1235122 (2013).
- 91 Tyanova, S., Temu, T. & Cox, J. The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nat Protoc* **11**, 2301-2319, doi:10.1038/nprot.2016.136 (2016).

- 92 Ni, X. *et al.* A region-resolved mucosa proteome of the human stomach. *Nat Commun* **10**, 39, doi:10.1038/s41467-018-07960-x (2019).
- 93 Kanehisa, M. & Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* **28**, 27-30, doi:10.1093/nar/28.1.27 (2000).
- 94 Croft, D. *et al.* The Reactome pathway knowledgebase. *Nucleic Acids Res* **42**, D472-477, doi:10.1093/nar/gkt1102 (2014).
- 95 Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* **102**, 15545-15550, doi:10.1073/pnas.0506580102 (2005).
- 96 Wilkerson, M. D. & Hayes, D. N. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. *Bioinformatics* **26**, 1572-1573, doi:10.1093/bioinformatics/btq170 (2010).
- 97 Yoshihara, K. *et al.* Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat Commun* **4**, 2612, doi:10.1038/ncomms3612 (2013).
- 98 Casado, P. *et al.* Phosphoproteomics data classify hematological cancer cell lines according to tumor type and sensitivity to kinase inhibitors. *Genome Biol* **14**, doi:ARTN R3710.1186/gb-2013-14-4-r37 (2013).
- 99 Hornbeck, P. V., Chabra, I., Kornhauser, J. M., Skrzypek, E. & Zhang, B. Phosphosite: A bioinformatics resource dedicated to physiological protein phosphorylation. *Proteomics* **4**, 1551-1561, doi:10.1002/pmic.200300772 (2004).
- 100 Dinkel, H. *et al.* Phospho.ELM: a database of phosphorylation sites-update 2011. *Nucleic Acids Research* **39**, D261-D267, doi:10.1093/nar/gkq1104 (2011).
- 101 Yang, C. Y. *et al.* PhosphoPOINT: a comprehensive human kinase interactome and phospho-protein database. *Bioinformatics* **24**, I14-I20, doi:10.1093/bioinformatics/btn297 (2008).
- 102 Casado, P. *et al.* Kinase-Substrate Enrichment Analysis Provides Insights into the Heterogeneity of Signaling Pathway Activation in Leukemia Cells. *Sci Signal* **6**, doi:ARTN rs610.1126/scisignal.2003573 (2013).
- 103 Hornbeck, P. V. *et al.* PhosphoSitePlus, 2014: mutations, PTMs and recalibrations. *Nucleic Acids Res* **43**, D512-D520, doi:10.1093/nar/gku1267 (2015).
- 104 Horn, H. *et al.* KinomeXplorer: an integrated platform for kinome biology studies. *Nat Methods*

11, 603-604, doi:10.1038/nmeth.2968 (2014).

105 An, Y. *et al.* High-fat diet induces dynamic metabolic alterations in multiple biological matrices of rats. *J Proteome Res* **12**, 3755-3768, doi:10.1021/pr400398b (2013).

### Acknowledgement

This work is supported by Chinese Ministry of Science and Technology (2016YFA0502500); National International Cooperation Grant (2014DFB30010); National Program on Key Basic Research Project (973 Program, 2012CB910300, 2014CBA02000), National High-tech R&D Program of China (863 program, 2014AA020201), National Natural Science Foundation of China (31270822), Beijing Natural Science Foundation (Z131100005213003), National Natural Science Foundation of China (No. 81702372), Shanghai Natural Science Foundation of China (No. 18ZR1406800), and Xiamen Science and Technology Project of Fujian Province, China (No. 3502ZZ20184003).

### Author contributions

Study Conception & Design, C.D., Y.H., J.Y.Z., Z.Q., L.L., and D.J.; Performed Experiment and Data Collection, L.L., D.J., Q.Z., F.X., D.S., H.L., and H.W.; Data analysis, L.L., D.J., Q.Z., Z.Q., F.X., D.S., and H.L.; Investigation, L.L., J.F., Yang Liu, L.B., S.T., and T.B.; Patient Sample Management and QC, Y.H., P.Z., T.C., D.J., H.W., Yalan Liu, Y.Z., C.X., Q.S., W.C., and X.Z.; Supervision, Y.H., P.Z., T.C., J.Y.Z., and C.D.; Writing – Review & Editing, C.D., J.Y.Z., L.L., and D.J.; Funding Acquisition, C.D. and Y.H.

### Competing interests

The authors declare no competing interests.

### Figure legends

**Fig. 1 The multi-omics landscape in ESCC progression. a** Overview of the experimental design and the number of samples for the genomic, proteomic, and phosphoproteomic analyses. **b** The genomic profiles of ESCC progression. Top: the mutation number and types of all the samples from early to

progressive ESCC. Middle: The somatic copy number alterations of all the samples from early to progressive ESCC. Bottom: the percentages of somatic single nucleotide variants (SNVs) of all samples. The mutation frequencies were shown by a bar plot at the right panel. **c** The gain of novel mutations at all stages in ESCC progression. **d** Analysis of the mutations loads of diverse cohorts. **e** The cosine similarity (top) and the description (bottom) of specific signatures of the IEN phase, ESCC phase, and the Fudan cohort. **f** Bar plot of the percentages of top mutations (left) and Spearman's correlation (right) among the Fudan cohort and other ESCC cohorts (Fisher's exact test). **g** The number of the identified proteins of independent 756 samples (left) and the phosphosites of 54 samples. \*\*\*\*  $p < 1.0E-4$ , \*\*\*  $p < 1.0E-3$ , \*\*  $p < 0.01$ , \*  $p < 0.05$ .

**Fig. 2 Integrative omics analyses of early ESCC samples.** **a** Profiling of absolute copy number alterations observed in the Fudan cohort. The square directed to a subset of patient samples used for WES. **b** Volcano plot showing the *cis*- correlation of the SCNA (x axis) and the associated  $-\log_{10}(p \text{ value})$  (y axis) on the genes at chr3q gain. **c** The *cis* SCNA-protein regulations of significantly correlated genes (top) on their corresponding proteins expression (bottom). **d** Dependency Map-supported (<https://depmap.org>) panels showing relative survival averaged across all available ESCC cell lines after depletion of the indicated genes by RNAi or CRISPR. The right showing the Pearson's correlation and  $p$  value of these genes with PCNA and MKI67 at the protein level. **e** A brief summary of the impacts of chr3q gain. **f** Venn diagram (left) and heatmap (right) depicting the number of the mutations in ESCC progression (Chi-square test). The mutations were detected both in the IEN phase and the ESCC phase and associated with cancer-associated genes (CAGs). The square directed to a subset of patient samples used for WES. **g** The number of the overlapped proteins (top) enhanced by *TP53* mutation and the representative pathways (bottom). **h** Heatmap showing the proteins and the associated biological pathways upregulated by *TP53* mutation in ESCC progression (Kruskal-Wallis test). The square directed to a subset of patient samples used for WES. **i** Summary of the impacts of *TP53* mutation in ESCC progression. **j** Volcano plot showing the impacts of the selected mutations on proteins expression ( $\log_2$ -transformed FOT) (x axis) and the associated  $-\log_{10}(p \text{ value})$  (y axis) (left) and heatmap depicting relative survival averaged across all available ESCC cell lines after depletion of MACF1 and FAT1 by RNAi or CRISPR (right). **k** Scatterplot showing the relationship between  $\log_{10}$  GSK3A (top)/GSK3B (bottom) and  $\log_{10}$  MACF1 expression at the protein level (Pearson's correlation). **l** Heatmap showing

the impacts the mutation of *MACF1* on the expression of Wnt signaling related proteins in ESCC progression (Kruskal-Wallis test). The right depicting the impacts of the depletion of these genes on ESCC lines. The square directed to a subset of patient samples used for WES. **m** Heatmap showing the impacts of *MACF1* mutation on the expression of Wnt signaling related phosphoproteins in ESCC progression. The square directed to a subset of patient samples used for phosphoproteome. **n** A brief summary of the impacts of the mutation of *MACF1*. \*\*\*\*  $p < 1.0E-4$ , \*\*\*  $p < 1.0E-3$ , \*\*  $p < 0.01$ , \*  $p < 0.05$ .

**Fig. 3 The temporal driver pathway waves in ESCC progression.** **a** Principal component analysis (PCA) of the Fudan cohort. Left: PCA of all 756 ESCC samples (including NT phase, IEN phase, and ESCC phase); Right: PCA of 746 early ESCC samples (NT phase and IEN phase). **b** Heatmap analysis of the dynamic switches during the carcinogenesis of ESCC. Left: heatmap analysis of DEPs of the 22 substages in ESCC progression. Right: the driver pathway waves of 8 panels in ESCC progression. **c** The mutations significantly associated with stages in ESCC progression (Fisher's exact test). The highlighted mutations (right) were exclusively co-mutations (Fisher's exact test). The square directed to a subset of patient samples used for WES. The number of the enhanced proteins regulated by **d** the co-mutations of *FAT4*, *AKAK9*, and *PCDHB16*, and **e** the associated biological pathways. **f** The impacts of the co-mutations of *FAT4*, *AKAK9*, and *PCDHB16* (middle), on the protein level (bottom) in ESCC progression. The proteins were from the enhanced pathways. The square directed to a subset of patient samples used for WES. **g** Heatmap showing the phosphorylation of representative phosphoproteins from the enhanced pathways in ESCC progression. The square directed to a subset of patient samples used for phosphoproteome. **h** A carcinogenesis path with 8 dynamic waves in ESCC progression at the multi-omics level. \*\*\*\*  $p < 1.0E-4$ , \*\*\*  $p < 1.0E-3$ , \*\*  $p < 0.01$ , \*  $p < 0.05$ .

**Fig. 4 Proteomic clusters and the impacts of *AKAP9* mutation in ESCC progression.** **a** Consensus clustering analysis of 756 samples. Left: the percentages of the two clusters in 22 substages; Right: 756 samples were classified into two clusters based on proteomic patterns. **b** Volcano analysis of DEPs (left) in the two clusters and their associated biological pathways (right) in the two clusters (t-test). Biological pathways were analyzed from the Reactome database. **c** Venn diagram depicting the number of the mutations in early-stage events (left) and in advanced-stages event (right) associated with stage in ESCC progression. The genes were detected both in C1 and C2, and were highly expressed in C2. **d** The

mutations in the advanced-stage event and their correlation with stages and phases (Fisher's exact test). The square directed to a subset of patient samples used for WES. **e** Heatmap showing the impacts of the mutation of *AKAP9* on the protein levels of AKAP9 and PRKACA. **f** Scatterplot showing the relationship between  $\log_{10}$  PRKACA and  $\log_{10}$  AKAP9 expression at the protein level (Pearson's correlation). **g** GSEA plots (KEGG gene sets) for glycolysis in *AKAP9* mutation and WT comparisons. **h** Heatmap depicting the impacts ( $\log_2$  FC) of *AKAP9* mutation on glucose metabolism (left) in ESCC progression. The square directed to a subset of patient samples used for WES. **i** Scatterplot showing the relationship between  $\log_{10}$  G6PD/HK1 and  $\log_{10}$  GPI expression at the protein level (Pearson's correlation). **j** A brief summary of the impacts of *AKAP9* mutation. \*\*\*\*  $p < 1.0E-4$ , \*\*\*  $p < 1.0E-3$ , \*\*  $p < 0.01$ , \*  $p < 0.05$ .

**Fig. 5 Personalized trajectory revealed six major carcinogenesis tracks of the early ESCCs.** **a** The trajectory of 746 samples (top) and 114 early ESCC cases were grouped into 9 (bottom). **b** Sankey diagram analysis of 114 early ESCC cases. Left: the 114 early ESCC cases; Middle: the 9 groups; Right: 6 tracks and their related dominant pathways. **c** Radar analysis of the negative (left,  $K < 0$ ) and positive pathways (right,  $K > 0$ ) of the 6 tracks. **d** Venn diagram showing the track mutations (top) and the CAGs (bottom). The overlapped genes were shown in the box. **e** The CAG associated track mutations in the early ESCCs. The co-mutations were highlighted in the left (Fisher's exact test), and the mutation frequency were shown in the right. The square directed to a subset of patient samples used for WES in early ESCCs. **f** GSEA plots (KEGG gene sets) for adherens junction in *EPAS1* mutation and WT comparisons. **g** Heatmap showing the represented protein in the adherens junction positive associated with *EPAS1* mutation (Fisher's exact test). The square directed to a subset of patient samples used for WES in early ESCCs. **h** Venn diagram depicting the number of the overlapped proteins upregulated by the mutation of *EPAS1* and T2 enhanced phosphoprotein (left), and the associated biological pathways (right). **i** Heatmap showing the phosphorylation of the phosphoproteins in the biological pathways. The square directed to a subset of patient samples used for phosphoproteome in early ESCCs. **j** Heatmap showing the expression of representative CD molecules and drug targets approved by the FDA in the 6 tracks (two-way ANOVA test). Error bars represented mean  $\pm$  SEM, \*\*\*\*  $p < 1.0E-4$ , \*\*\*  $p < 1.0E-3$ , \*\*  $p < 0.01$ , \*  $p < 0.05$ .

**Fig. 6 Aberrant glycolytic metabolism in ESCC and alterations in the activities of its key enzyme,**

**PGK1.** **a** Aberrant glycolytic metabolism in ESCC progression. **b** Highly expressed PGK1 was negatively correlated to prognosis (log-rank test). **c** Gradually enhanced PGK1 in ESCC at the proteome (black) and phosphoproteome (orange) levels (Kruskal-Wallis test). **d** Immunohistochemistry analysis of PGK1 expression in normal (T0), Tis, SM2, and advanced tissues. The zone with the dotted lines and red arrow represented PGK1 positive staining. **e** Analysis of the serine motif of PGK1 (sP). The top showed the sequence and phosphorylated sites of PGK1 (S203). The bottom represented that PGK1 S203 was detected at almost samples in ESCC (53/54) and the kinases associated with the motif of PGK1 S203 ('sP'). **f** The SCNAs of *CDK2* (middle) had positive effects on PGK1 expression (bottom) (Wilcoxon rank-sum test). **g** The impacts of the SCNAs of *CDK2* (middle) on the substrates expression of the kinases (bottom), associated with the PGK1 motif (sP). The square directed to a subset of patient samples used for WES. **h** The expression of the kinases and the substrates in ESCC progression at the phosphoprotein level. The square directed to a subset of patient samples used for phosphoproteome. **i** A brief regulatory network of kinase-substrate, associated with the motif of PGK1 (sP). Error bars represented mean  $\pm$  SEM. \*\*\*\*  $p < 1.0E-4$ , \*\*\*  $p < 1.0E-3$ , \*\*  $p < 0.01$ , \*  $p < 0.05$ .

**Fig. 7 PGK1 reprogrammed glucose metabolism and contributed to ESCC progression.** **a** Metabolite levels in KYSE150 cells transfected with PGK1 or co-transfected with PGK1 and ERK2 (t-test). **b** Metabolite levels in PGK1-knockdown KYSE150 cells and control cells (t-test). **c** Pan Serine/Threonine/Tyrosine-phosphorylation levels of PGK1 in KYSE150 cells and ECA109 cells. **d** PGK1 levels in mitochondria and cytosol fraction of KYSE150 cells and ECA109 cells. **e** The impacts of PGK1 and/or ERK2 on PDHK1 T338 phosphorylation levels in KYSE150 cells and ECA109 cells. **f** The impacts of PGK1 and/or ERK2 on PDH activity in KYSE150 cells and ECA109 cells (t-test). **g** The impacts of PGK1-overexpression (OE) and/or ERK2-OE on cell proliferation in KYSE150 cells and ECA109 cells (t-test). **h** Cell proliferation in KYSE150 cells and ECA109 cells with PGK1 knockdown and/or ERK2 knockdown (t-test). **i** Tumor weight (g) of KYSE150 cells (left) and ECA109 cells (right) xenografts in three groups (t-test): control group, PGK1-OE group, and PGK1-OE-inhibitor (gemcitabine) group. **j** Tumor weight (g) of KYSE150 cells (left) and ECA109 cells (right) xenografts in two groups (t-test): the control group and PGK1-knockdown group. Error bars represented mean  $\pm$  SEM. \*\*\*\*  $p < 1.0E-4$ , \*\*\*  $p < 1.0E-3$ , \*\*  $p < 0.01$ , \*  $p < 0.05$ .

## Supplementary Figure legends

**Supplementary Fig. 1 Hematoxylin and eosin (H&E) staining on early ESCCs.** (a) Invasion layer in ESCC progression. T0 – T1 indicated the early ESCC stages, ranging from substage 1 (normal tissue stage) to substage 7\_3 (submucosa stage). T2 – T3 represented the progressive ESCC stages, including stages 8 (T2 stage) and 9 (T3 stage). Diversity sign was marked the distinctive invasion layer on the right. (b) The procedure of dissection and embedding. (c) H&E staining analysis of different substages in early ESCC. (d) The tumor purity of 756 samples in ESCC progression.

**Supplementary Fig. 2 A brief workflow of multi-omics analysis and overview of the proteomic profiles of early and progressive ESCC cohorts.** (a) The workflow of sampling, processing, and analysis of all ESCC samples at the multi-omics level. The 756 samples were collected from 124 ESCC patients, including those from the surgery samples (n = 10, T2–T3) and biopsy samples after ESD (n = 746, T0–T1). (b) The top ten mutations in 11 cases of the Fudan cohort. The top 10 mutations were referred in the Fig. 1b. (c) The exclusively co-mutations of the top 25 mutations in the genomics data of the Fudan cohort. (d) The top mutations of No.020 in ESCC progression. (e) Table chart showing the mutation signatures of the three phases in the Fudan cohort. (f) Table chart showing the mutation signatures of other EC cohorts. (g) (Spearman's) correlation analysis of 42 HEK293T cell samples as MS quality control to evaluate the robustness of label-free quantification. (h) Correlation matrix of 114 (sub)stage 1 proteomes (Spearman's correlation coefficients). (i) The number of identified proteins in No.020 (green), No.105 (red), and No.109 (orange). (j) Distribution of log<sub>10</sub>-transformed FOT of identified proteins abundance in 756 samples that passed quality control. (k) The identified protein numbers of the 756 samples. (l) The dynamics of protein abundance identified in 22 substages. Proteins were quantified as normalized iBAQ value and transformed to log<sub>10</sub> FOT. The highest- and lowest-abundance proteins were shown in the box. The cumulative number of the phosphosites (m) and phosphoproteins (n) of 54 samples in ESCC progression.

**Supplementary Fig. 3 Integrative analyses of genomics, proteomics, and phosphoproteomics data in ESCC progression.** (a) A brief workflow for WES. The numbers of samples and stages were shown in the box. (b) Functional effects of somatic copy number alterations (SCNAs) on protein expressions. Positive and negative correlations/impacts were indicated in red and green plots, respectively. Genes

were ordered by chromosomal location on x and y axis, respectively. **(c)** Venn diagram showing the genes identified in the three phases at the gene, protein, and phosphoprotein levels. **(d)** Clustergrams of mutually exclusive mutated genes in cell cycle (left) and DNA repair (right) (Bartlett's test). **(e)** Venn diagram depicting the number of mutations detected in the 53 samples for WES. **(f)** Volcano plot showing the  $\log_2$  fold change (x axis) between *TP53* mutation and WT comparisons, and the associated  $\log_{10} p$  value (t-test). **(g)** Scatterplot showing the relationship between  $\log_{10}$  (CTNNB1 S191) and  $\log_{10}$  (GSSK3B Y216) expression at the phosphoprotein level (Pearson's correlation). \*\*\*\*  $p < 1.0E-4$ , \*\*\*  $p < 1.0E-3$ , \*\*  $p < 0.01$ , \*  $p < 0.05$ .

**Supplementary Fig. 4 The Immune-based panels and dynamic driver pathway waves of 8 panels in ESCC progression.** **(a)** Cell signatures of 756 early ESCCs by xCell (Kruskal-Wallis test). **(b)** The dominant pathways of 8 panels in ESCC progression. The results of the k-means analysis were shown. **(c)** A carcinogenesis path with 8 dynamic waves in ESCC progression. The expression of ESCC biomarkers from **(d)** Yazdian-Robati et al.<sup>32</sup> and **(e)** Pawar et al.<sup>33</sup> in ESCC progression in ESCC progression. Twenty-two substages and  $\log_{10}$ FOT were indicated on x and y axis, respectively. Error bars represented mean  $\pm$  SEM. \*\*\*\*  $p < 1.0E-4$ , \*\*\*  $p < 1.0E-3$ , \*\*  $p < 0.01$ , \*  $p < 0.05$ .

**Supplementary Fig. 5 Proteomic characteristics of two clusters and the impacts of AKAP9 mutation.** **(a)** Heatmap showing consensus matrix with K= 2, 3, and 4. The input was the quantile-normalized iBAQ intensity (FOT) matrix of variable proteomics data (n = 6,687). **(b)** Column diagram showing the difference of the two clusters in diverse gender, smoking/drinking, and ages ESCC patients. **(c)** Column chart showing the proportion of the phases in the two clusters (Fisher's exact test). **(d)** The DEPs and the representative pathways in the Cluster 1 (top) and 2 (bottom) (t-test). **(e)** The expression of specific molecules of esophageal tissue in the two clusters (t-test). **(f)** The expression of specific molecules of esophageal tissue in ESCC progression. **(g)** The distribution of the mutations of *AKAP9* and *TP53* in ESCC progression. The square directed to a subset of patient samples used for WES. **(h)** Column chart showing the proportion of *AKAP9* mutation in the three phases in ESCC progression (Fisher's exact test). **(i)** Survival analysis of patients with *AKAP9* mutation versus WT comparisons of ESCC (log-rank test). The survival information was referred the ESCC cohort (Song et al)<sup>13</sup>. **(j)** GSEA plots (KEGG gene sets) for pentose phosphate pathway (top) and TCA cycle (bottom) in *AKAP9* mutation and WT comparisons.

(k) Scatterplot showing the relationships between  $\log_{10}$  AKAP9 and  $\log_{10}$  HK3 expression (top), and between  $\log_{10}$  HK3 and  $\log_{10}$  GPI expression (bottom), at the protein level (Pearson's correlation). (l) Heatmap showing the phosphorylation of the phosphoproteins in the dominant pathways regulated by *AKAP9* mutation in ESCC progression. The square directed to a subset of patient samples used for phosphoproteome. \*\*\*\*  $p < 1.0E-4$ , \*\*\*  $p < 1.0E-3$ , \*\*  $p < 0.01$ , \*  $p < 0.05$ .

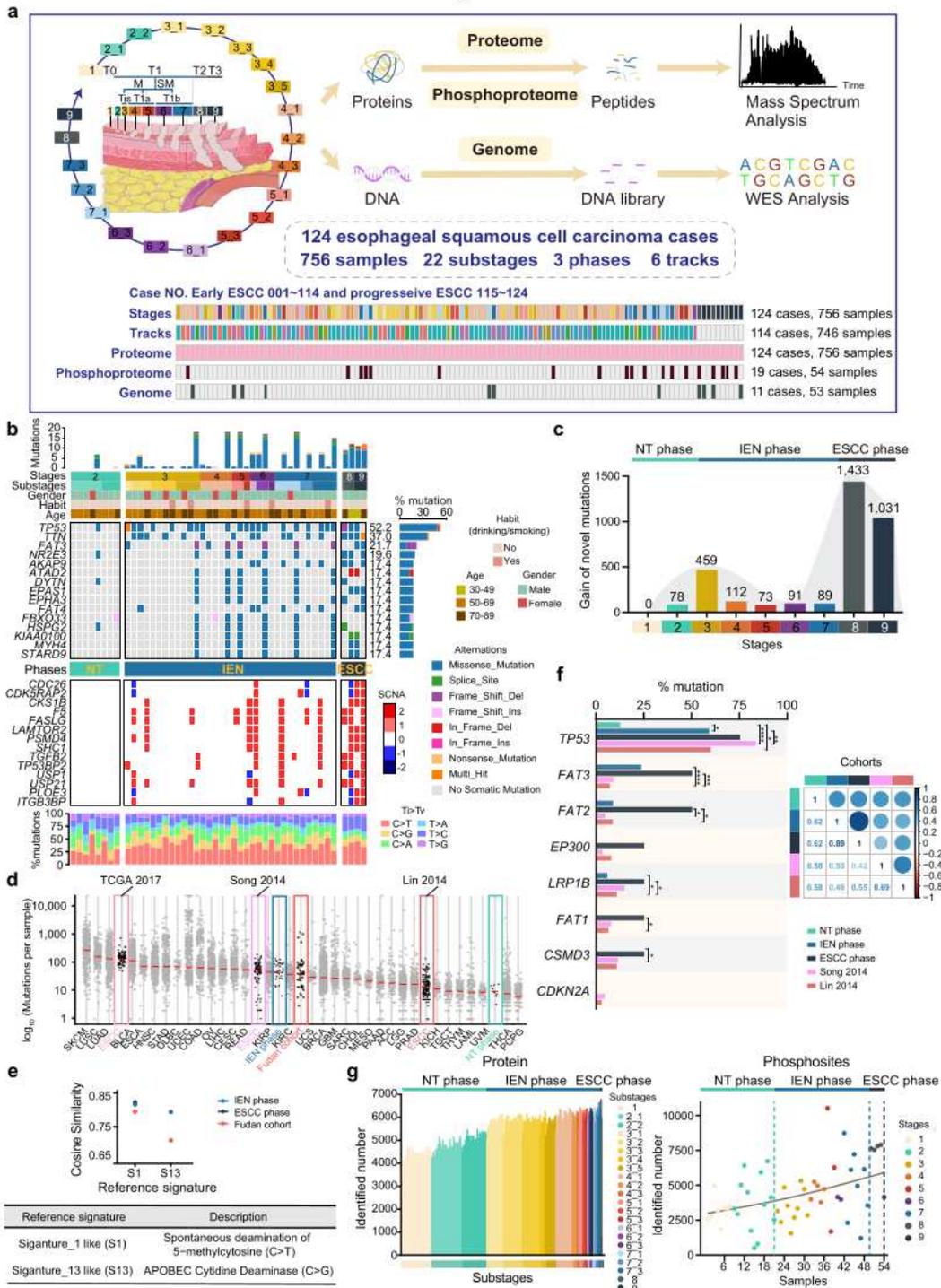
**Supplementary Fig. 6 The trajectory analysis of the early ESCC cohort.** (a) The screening and variance of proteins identified from the 746 samples in the early ESCC cohort. The proteins used were of E2 ( $n = 9,741$ ). The mean-expression of proteins and dispersion were shown on the x and y axes, respectively. (b) The expression of track enhanced proteins and their associated biological pathways of the 6 tracks. Shown were the  $R^2$  (top) and K value (bottom). (c) The track enhanced proteins in the dominant pathways of the 6 tracks. (d) Column chart analysis of somatic SNVs in diverse tracks (t-test). Left: the percentages of all samples in tracks. Right: the somatic SNVs were indicated by a bar plot. (e) Column chart depicting the mutations frequency of the selected genes in the track 2, track 5, and track 6 (Fisher's exact test). (f) GSEA plots (KEGG gene sets) for pentose phosphate pathway in *STAG2* mutation and WT comparisons (left) and *USP6* mutation and WT comparisons (right). (g) GSEA plots (KEGG gene sets) for pentose phosphate pathway in *EPHA3* mutation and WT comparisons (left) and *MACF1* mutation and WT comparisons (right). (h) Fishplot analysis representing the kinetic driver pathway waves of the 6 tracks. Error bars represented mean  $\pm$  SEM. \*\*\*\*  $p < 1.0E-4$ , \*\*\*  $p < 1.0E-3$ , \*\*  $p < 0.01$ , \*  $p < 0.05$ .

**Supplementary Fig. 7 The impacts of GAPDH-OE or PGM1-OE on ESCC cell proliferation and inhibiting PGK1 by gemcitabine attenuates cell proliferation.** (a) The expression of PGK1 in the carcinogenesis progress of ESCC in the case of N0.020 (blue), No.105 (red), and No.109 (yellow). (b) The impacts of GAPDH-OE or PGM1-OE on cell proliferation in KYSE150 cells (left) and ECA109 cells (right) (t test). PGK1-OE in KYSE150 cells (left) and ECA109 cells (right) was the positive control. (c) *In vitro* assay showing the effects of gemcitabine on inhibition of PGK1 activity (IC<sub>50</sub>: 16.3 nM). (d) Gemcitabine inhibited cell proliferation. (e) Gemcitabine decreased PGK1 mediated metabolic flux (t-test). Error bars represented mean  $\pm$  SEM. \*\*\*\*  $p < 1.0E-4$ , \*\*\*  $p < 1.0E-3$ , \*\*  $p < 0.01$ , \*  $p < 0.05$ .

**Supplementary Fig. 8 The model of the key events in ESCC progression.** The preliminary progression model of 3 phases (nontumor phase, intraepithelial neoplasia phase, and ESCC phase) covering 9 histological stages.

# Figures

**Fig. 1**

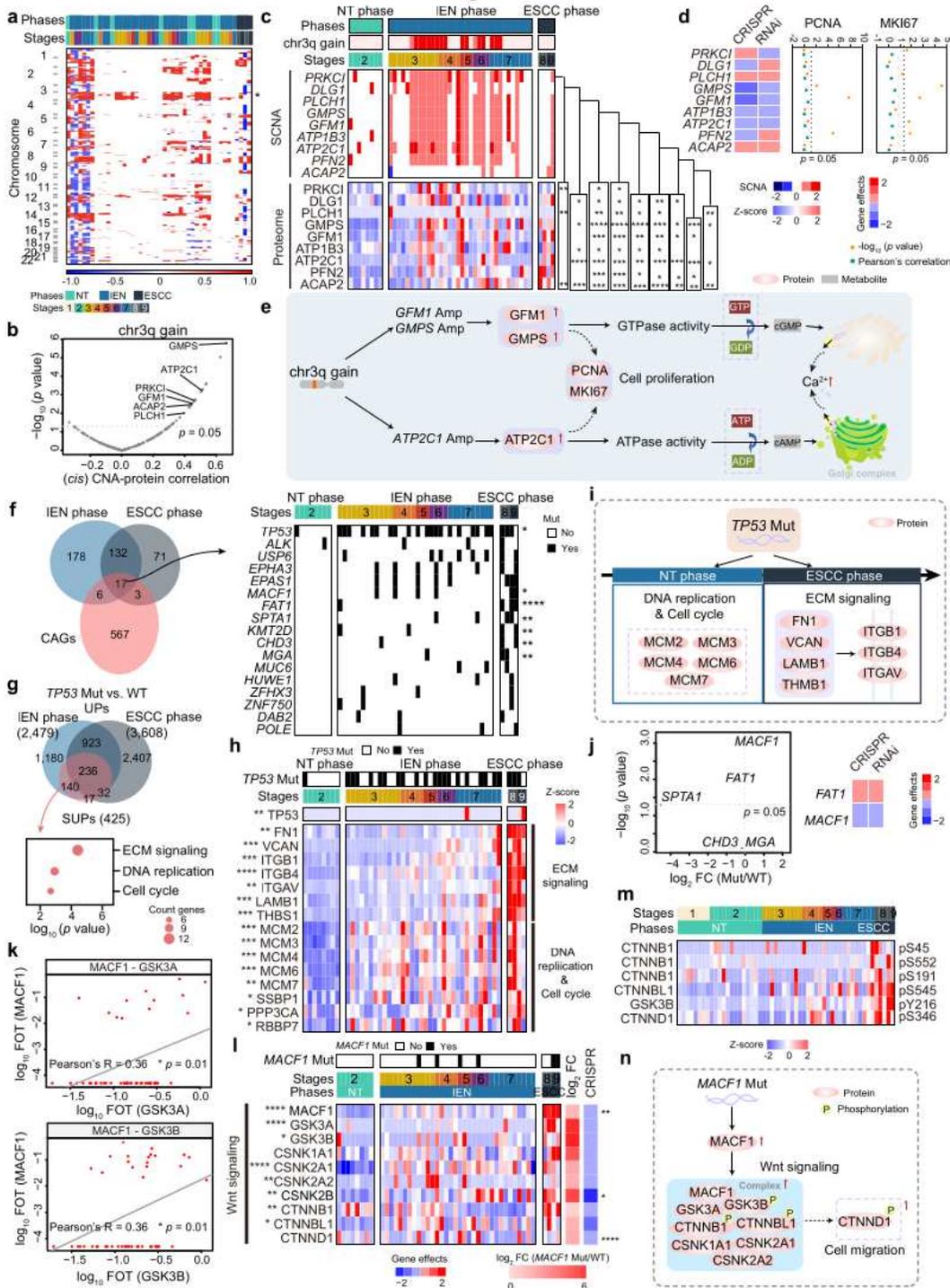


**Figure 1**

The multi-omics landscape in ESCC progression. a Overview of the experimental design and the number of samples for the genomic, proteomic, and phosphoproteomic analyses. b The genomic profiles of ESCC progression. Top: the mutation number and types of all the samples from early to progressive ESCC.

Middle: The somatic copy number alterations of all the samples from early to progressive ESCC. Bottom: the percentages of somatic single nucleotide variants (SNVs) of all samples. The mutation frequencies were shown by a bar plot at the right panel. c The gain of novel mutations at all stages in ESCC progression. d Analysis of the mutations loads of diverse cohorts. e The cosine similarity (top) and the description (bottom) of specific signatures of the IEN phase, ESCC phase, and the Fudan cohort. f Bar plot of the percentages of top mutations (left) and Spearman's correlation (right) among the Fudan cohort and other ESCC cohorts (Fisher's exact test). g The number of the identified proteins of independent 756 samples (left) and the phosphosites of 54 samples. \*\*\*\*  $p < 1.0E-4$ , \*\*\*  $p < 1.0E-3$ , \*\*  $p < 0.01$ , \*  $p < 0.05$ .

**Fig. 2**

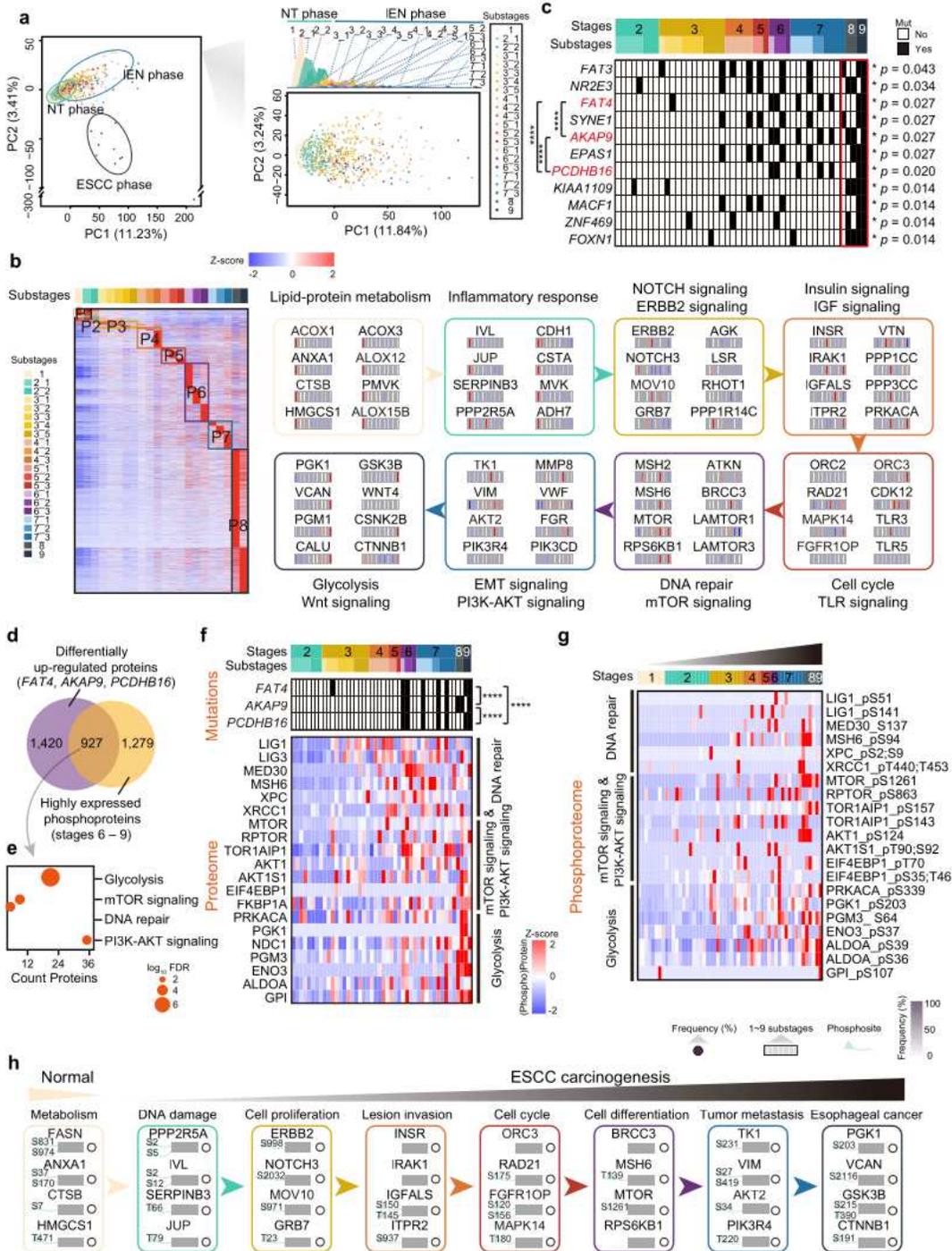


**Figure 2**

Integrative omics analyses of early ESCC samples. a Profiling of absolute copy number alterations observed in the Fudan cohort. The square directed to a subset of patient samples used for WES. b Volcano plot showing the cis-correlation of the SCNA (x axis) and the associated  $-\log_{10}(p \text{ value})$  (y axis) on the genes at chr3q gain. c The cis SCNA-protein regulations of significantly correlated genes (top) on their corresponding proteins expression (bottom). d Dependency Map-supported (<https://depmap.org>)

panels showing relative survival averaged across all available ESCC cell lines after depletion of the indicated genes by RNAi or CRISPR. The right showing the Pearson's correlation and p value of these genes with PCNA and MKI67 at the protein level. e A brief summary of the impacts of chr3q gain. f Venn diagram (left) and heatmap (right) depicting the number of the mutations in ESCC progression (Chi-square test). The mutations were detected both in the IEN phase and the ESCC phase and associated with cancer-associated genes (CAGs). The square directed to a subset of patient samples used for WES. g The number of the overlapped proteins (top) enhanced by TP53 mutation and the representative pathways (bottom). h Heatmap showing the proteins and the associated biological pathways upregulated by TP53 mutation in ESCC progression (Kruskal-Wallis test). The square directed to a subset of patient samples used for WES. i Summary of the impacts of TP53 mutation in ESCC progression. j Volcano plot showing the impacts of the selected mutations on proteins expression (log<sub>2</sub>-transformed FOT) (x axis) and the associated -log<sub>10</sub> (p value) (y axis) (left) and heatmap depicting relative survival averaged across all available ESCC cell lines after depletion of MACF1 and FAT1 by RNAi or CRISPR (right). k Scatterplot showing the relationship between log<sub>10</sub> GSK3A (top)/GSK3B (bottom) and log<sub>10</sub> MACF1 expression at the protein level (Pearson's correlation). l Heatmap showing the impacts the mutation of MACF1 on the expression of Wnt signaling related proteins in ESCC progression (Kruskal-Wallis test). The right depicting the impacts of the depletion of these genes on ESCC lines. The square directed to a subset of patient samples used for WES. m Heatmap showing the impacts of MACF1 mutation on the expression of Wnt signaling related phosphoproteins in ESCC progression. The square directed to a subset of patient samples used for phosphoproteome. n A brief summary of the impacts of the mutation of MACF1. \*\*\*\* p < 1.0E-4, \*\*\* p < 1.0E-3, \*\* p < 0.01, \* p < 0.05.

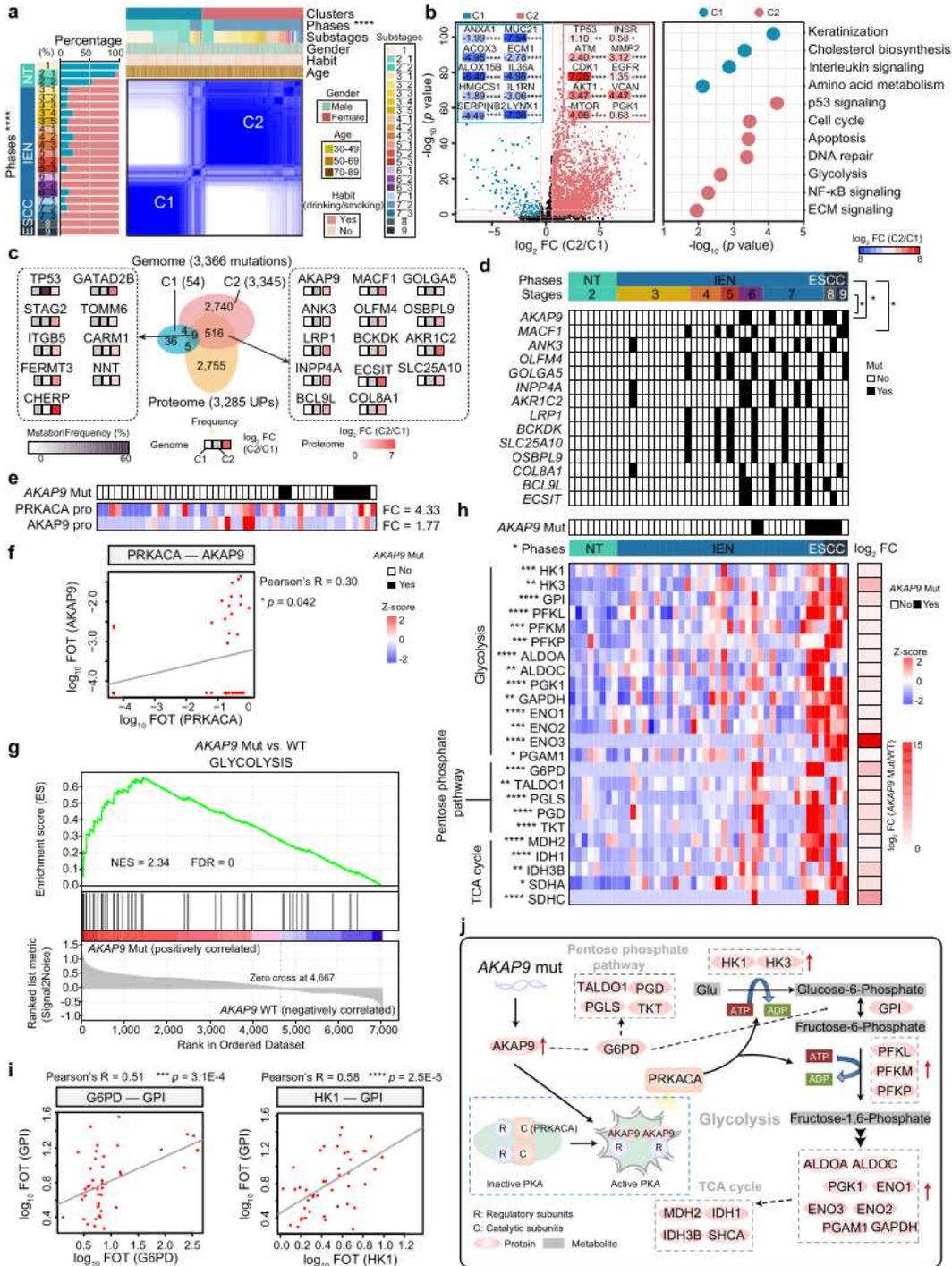
**Fig. 3**



**Figure 3**

The temporal driver pathway waves in ESCC progression. **a** Principal component analysis (PCA) of the Fudan cohort. Left: PCA of all 756 ESCC samples (including NT phase, IEN phase, and ESCC phase); Right: PCA of 746 early ESCC samples (NT phase and IEN phase). **b** Heatmap analysis of the dynamic switches during the carcinogenesis of ESCC. Left: heatmap analysis of DEPs of the 22 substages in ESCC progression. Right: the driver pathway waves of 8 panels in ESCC progression. **c** The mutations

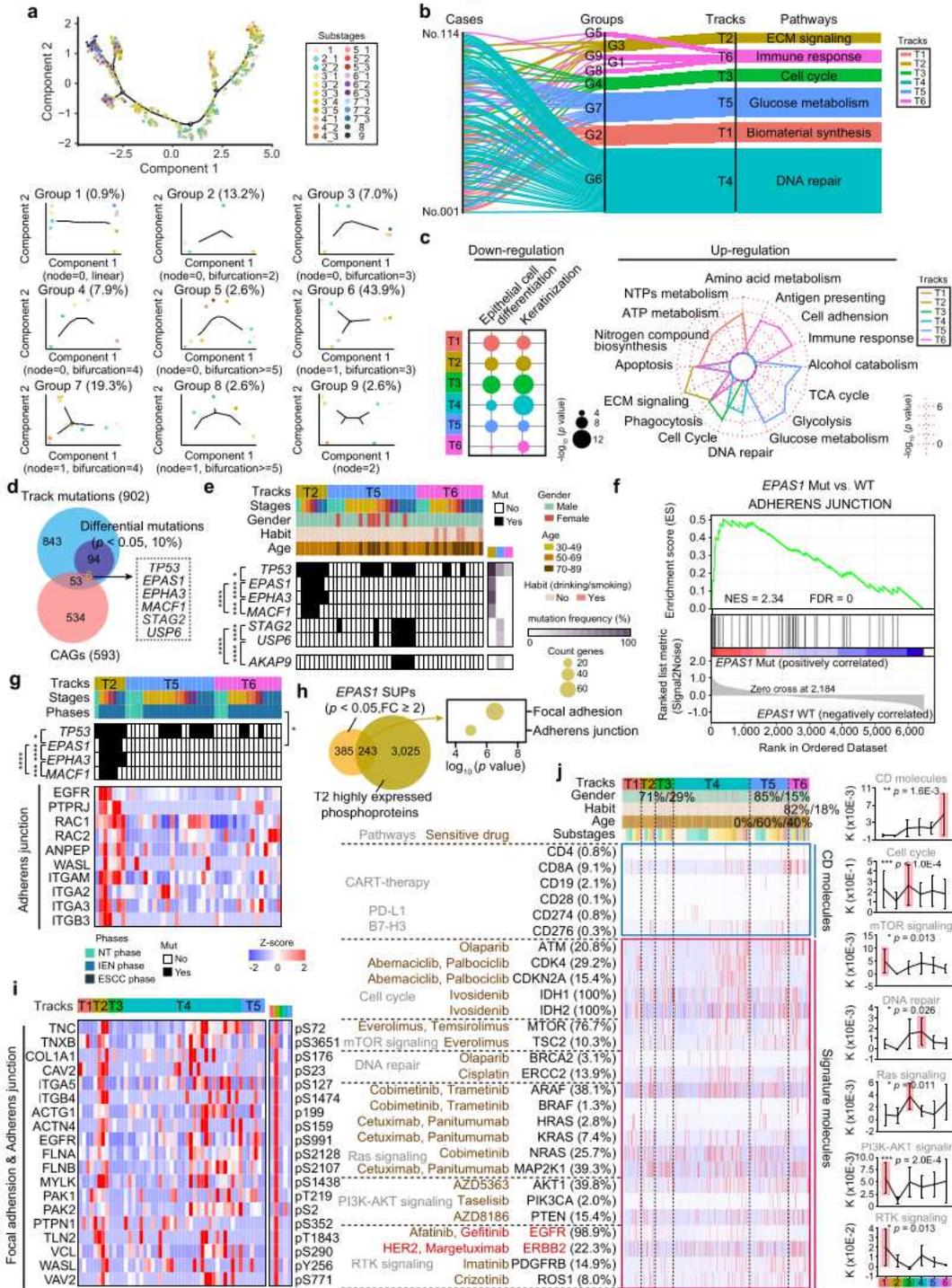
significantly associated with stages in ESCC progression (Fisher's exact test). The highlighted mutations (right) were exclusively co-mutations (Fisher's exact test). The square directed to a subset of patient samples used for WES. The number of the enhanced proteins regulated by d the co-mutations of FAT4, AKAK9, and PCDHB16, and e the associated biological pathways. f The impacts of the comutations of FAT4, AKAK9, and PCDHB16 (middle), on the protein level (bottom) in ESCC progression. The proteins were from the enhanced pathways. The square directed to a subset of patient samples used for WES. g Heatmap showing the phosphorylation of representative phosphoproteins from the enhanced pathways in ESCC progression. The square directed to a subset of patient samples used for phosphoproteome. h A carcinogenesis path with 8 dynamic waves in ESCC progression at the multiomics level. \*\*\*\*  $p < 1.0E-4$ , \*\*\*  $p < 1.0E-3$ , \*\*  $p < 0.01$ , \*  $p < 0.05$ .

**Fig. 4****Figure 4**

Proteomic clusters and the impacts of AKAP9 mutation in ESCC progression. **a** Consensus clustering analysis of 756 samples. Left: the percentages of the two clusters in 22 substages; Right: 756 samples were classified into two clusters based on proteomic patterns. **b** Volcano analysis of DEPs (left) in the two clusters and their associated biological pathways (right) in the two clusters (t-test). Biological pathways were analyzed from the Reactome database. **c** Venn diagram depicting the number of the

mutations in early-stage events (left) and in advanced-stages event (right) associated with stage in ESCC progression. The genes were detected both in C1 and C2, and were highly expressed in C2. d The mutations in the advanced-stage event and their correlation with stages and phases (Fisher's exact test). The square directed to a subset of patient samples used for WES. e Heatmap showing the impacts of the mutation of AKAP9 on the protein levels of AKAP9 and PRKACA. f Scatterplot showing the relationship between log<sub>10</sub> PRKACA and log<sub>10</sub> AKAP9 expression at the protein level (Pearson's correlation). g GSEA plots (KEGG gene sets) for glycolysis in AKAP9 mutation and WT comparisons. h Heatmap depicting the impacts (log<sub>2</sub> FC) of AKAP9 mutation on glucose metabolism (left) in ESCC progression. The square directed to a subset of patient samples used for WES. i Scatterplot showing the relationship between log<sub>10</sub> G6PD/HK1 and log<sub>10</sub> GPI expression at the protein level (Pearson's correlation). j A brief summary of the impacts of AKAP9 mutation. \*\*\*\* p < 1.0E-4, \*\*\* p < 1.0E-3, \*\* p < 0.01, \* p < 0.05.

**Fig. 5**

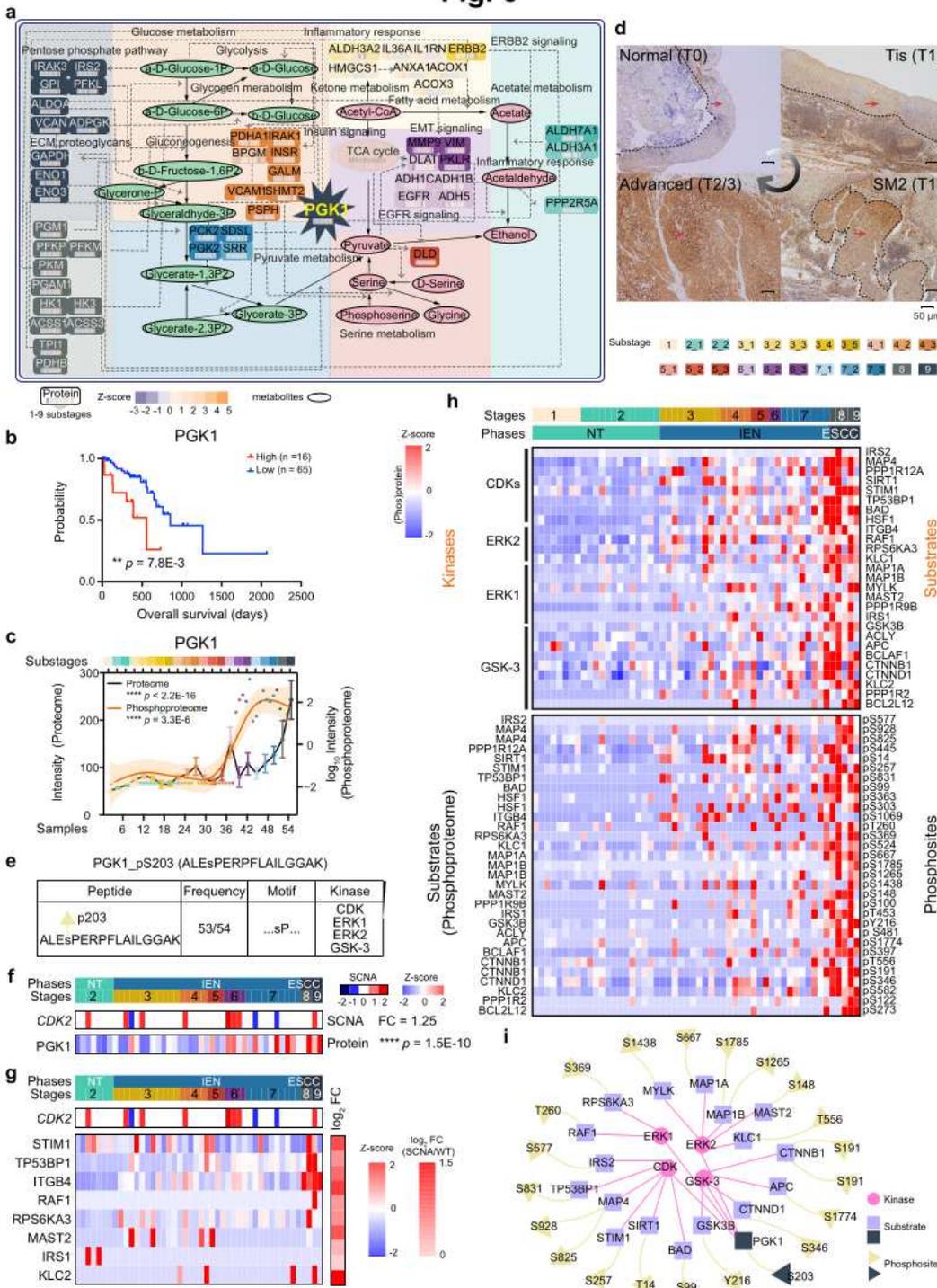


**Figure 5**

Personalized trajectory revealed six major carcinogenesis tracks of the early ESCCs. a The trajectory of 746 samples (top) and 114 early ESCC cases were grouped into 9 (bottom). b Sankey diagram analysis of 114 early ESCC cases. Left: the 114 early ESCC cases; Middle: the 9 groups; Right: 6 tracks and their related dominant pathways. c Radar analysis of the negative (left,  $K < 0$ ) and positive pathways (right,  $K > 0$ ) of the 6 tracks. d Venn diagram showing the track mutations (top) and the CAGs (bottom). The

overlapped genes were shown in the box. e The CAG associated track mutations in the early ESCCs. The co-mutations were highlighted in the left (Fisher's exact test), and the mutation frequency were shown in the right. The square directed to a subset of patient samples used for WES in early ESCCs. f GSEA plots (KEGG gene sets) for adherens junction in EPAS1 mutation and WT comparisons. g Heatmap showing the represented protein in the adherens junction positive associated with EPAS1 mutation (Fisher's exact test). The square directed to a subset of patient samples used for WES in early ESCCs. h Venn diagram depicting the number of the overlapped proteins upregulated by the mutation of EPAS1 and T2 enhanced phosphoprotein (left), and the associated biological pathways (right). i Heatmap showing the phosphorylation of the phosphoproteins in the biological pathways. The square directed to a subset of patient samples used for phosphoproteome in early ESCCs. j Heatmap showing the expression of representative CD molecules and drug targets approved by the FDA in the 6 tracks (two-way ANOVA test). Error bars represented mean  $\pm$  SEM, \*\*\*\*  $p < 1.0E-4$ , \*\*\*  $p < 1.0E-3$ , \*\*  $p < 0.01$ , \*  $p < 0.05$ .

**Fig. 6**

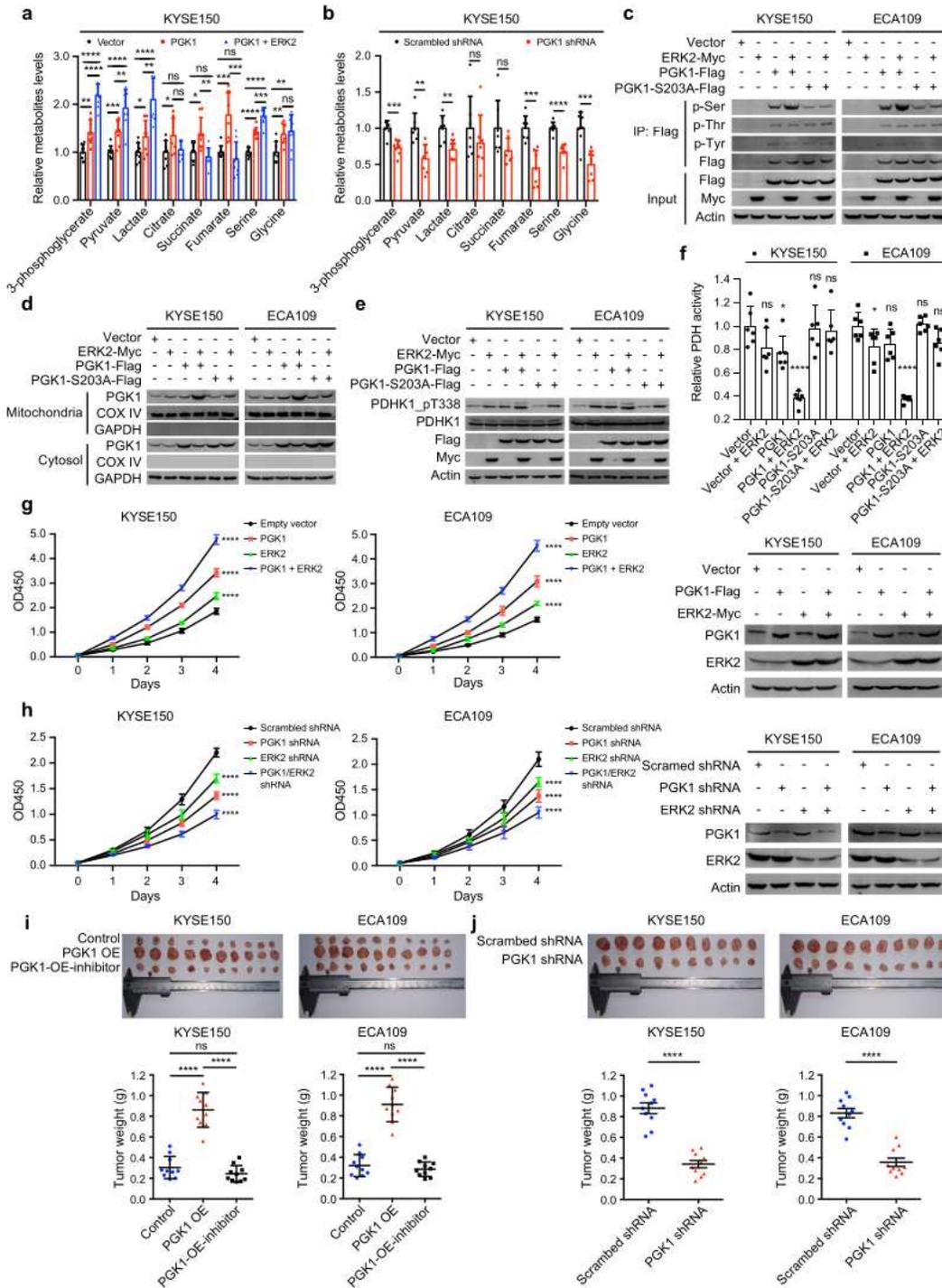


**Figure 6**

Aberrant glycolytic metabolism in ESCC and alterations in the activities of its key enzyme, PGK1. a Aberrant glycolytic metabolism in ESCC progression. b Highly expressed PGK1 was negatively correlated to prognosis (log-rank test). c Gradually enhanced PGK1 in ESCC at the proteome (black) and phosphoproteome (orange) levels (Kruskal-Wallis test). d Immunohistochemistry analysis of PGK1 expression in normal (T0), Tis, SM2, and advanced tissues. The zone with the dotted lines and red arrow

represented PGK1 positive staining. e Analysis of the serine motif of PGK1 (sP). The top showed the sequence and phosphorylated sites of PGK1 (S203). The bottom represented that PGK1 S203 was detected at almost samples in ESCC (53/54) and the kinases associated with the motif of PGK1 S203 (sP). f The SCNAs of CDK2 (middle) had positive effects on PGK1 expression (bottom) (Wilcoxon rank-sum test). g The impacts of the SCNAs of CDK2 (middle) on the substrates expression of the kinases (bottom), associated with the PGK1 motif (sP). The square directed to a subset of patient samples used for WES. h The expression of the kinases and the substrates in ESCC progression at the phosphoprotein level. The square directed to a subset of patient samples used for phosphoproteome. i A brief regulatory network of kinase-substrate, associated with the motif of PGK1 (sP). Error bars represented mean  $\pm$  SEM. \*\*\*\*  $p < 1.0E-4$ , \*\*\*  $p < 1.0E-3$ , \*\*  $p < 0.01$ , \*  $p < 0.05$ .

**Fig. 7**



**Figure 7**

PGK1 reprogrammed glucose metabolism and contributed to ESCC progression. a Metabolite levels in KYSE150 cells transfected with PGK1 or co-transfected with PGK1 and ERK2 (t-test). b Metabolite levels in PGK1-knockdown KYSE150 cells and control cells (t-test). c Pan Serine/Threonine/Tyrosine-phosphorylation levels of PGK1 in KYSE150 cells and ECA109 cells. d PGK1 levels in mitochondria and cytosol fraction of KYSE150 cells and ECA109 cells. e The impacts of PGK1 and/or ERK2 on PDHK1

T338 phosphorylation levels in KYSE150 cells and ECA109 cells. f The impacts of PGK1 and/or ERK2 on PDH activity in KYSE150 cells and ECA109 cells (t-test). g The impacts of PGK1-overexpression (OE) and/or ERK2-OE on cell proliferation in KYSE150 cells and ECA109 cells (t-test). h Cell proliferation in KYSE150 cells and ECA109 cells with PGK1 knockdown and/or ERK2 knockdown (t-test). i Tumor weight (g) of KYSE150 cells (left) and ECA109 cells (right) xenografts in three groups (t-test): control group, PGK1-OE group, and PGK1-OE-inhibitor (gemcitabine) group. j Tumor weight (g) of KYSE150 cells (left) and ECA109 cells (right) xenografts in two groups (t-test): the control group and PGK1-knockdown group. Error bars represented mean  $\pm$  SEM. \*\*\*\* p < 1.0E-4, \*\*\* p < 1.0E-3, \*\* p < 0.01, \* p < 0.05.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryFigures.pdf](#)
- [SupplementaryTable1.xlsx](#)
- [SupplementaryTable2.xlsx](#)
- [SupplementaryTable3.xlsx](#)
- [SupplementaryTable4.xlsx](#)
- [SupplementaryTable5.xlsx](#)
- [SupplementaryTable6.xlsx](#)
- [SupplementaryTable7.xlsx](#)
- [SupplementaryTable8.xlsx](#)