

Integrative proteogenomic characterization of early esophageal cancer

Lingling Li

Institute of Biomedical Sciences, Human Phenome Institute, Fudan University

Dongxian Jiang

Department of Pathology, Zhongshan Hospital, Fudan University,

Qiao Zhang

Institute of Biomedical Sciences, Human Phenome Institute, Fudan University

Hui Liu

Henan Normal University, Xinxiang

Zhaoyu Qin

Institute of Biomedical Sciences, Human Phenome Institute, Fudan University

Fujiang Xu

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

Dan Shen

Institute of Biomedical Sciences, Human Phenome Institute, Fudan University

Haixing Wang

Department of Pathology, Zhongshan Hospital, Fudan University

Jinwen Feng

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

Yang Liu

Fudan University

Weijie Chen

Department of Pathology, Zhongshan Hospital, Fudan University

Xue Zhang

Zhongshan Hospital, Fudan University

Lin Bai

Fudan University

Sha Tian

Institute of Biomedical Sciences, Human Phenome Institute, Fudan University

Subei Tan

Fudan University

Chen Xu

Zhongshan Hospital, Fudan University

Qi Song

Zhongshan Hospital, Fudan University

Yalan Liu

Zhongshan Hospital, Fudan University

Yunshi Zhong

Endoscopy Center and Endoscopy Research Institute, Zhongshan Hospital, Fudan University

Tianyin Chen

Endoscopy Center and Endoscopy Research Institute, Zhongshan Hospital, Fudan University

Pinghong Zhou (✉ zhou.pinghong@zs-hospital.sh.cn)

Endoscopy Center and Endoscopy Research Institute, Zhongshan Hospital, Fudan University

Jian-Yuan Zhao (✉ zhaoyj@fudan.edu.cn)

Fudan University

Ying-Yong Hou (✉ hou.yingyong@zs-hospital.sh.cn)

Fudan University

Chen Ding (✉ chend@fudan.edu.cn)

Fudan University

Article

Keywords: Early esophageal cancer, carcinogenesis, proteogenomics, PGK1

Posted Date: March 1st, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-227745/v1>

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^{1,4}, Dongxian Jiang^{1,4}, Qiao Zhang^{1,4}, Hui Liu^{3,4}, Zhaoyu Qin^{1,4}, Fujiang Xu^{1,4}, Dan Shen^{1,4}, Haixing Wang^{1,4}, Jinwen Feng¹, Yang Liu¹, Weijie Chen¹, Xue Zhang¹, Lin Bai¹, Sha Tian¹, Subei Tan¹, Chen Xu¹, Qi Song¹, Yalan Liu¹, Yunshi Zhong^{1,2}, Tianyin Chen^{1,2}, Pinghong Zhou^{1,2,*}, Jian-Yuan Zhao^{1,*}, Yingyong Hou^{1,*}, Chen Ding^{1,5,*}

¹State 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;

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

³State 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.

⁴These authors contributed equally

⁵Lead Contact

*Correspondence: zhou.pinghong@zs-hospital.sh.cn (P.Z.), zhaojy@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 genomic, proteomic, and phosphoproteomic analysis of 756 trace-tumor-samples from 124 esophageal cancer (EC) patients, covering 9 major histopathological stages and 22 substages. The results revealed a significant diversity of proteome patterns in the 22 substages. The integrated multi-omics data identified genomic-proteomic aberrations revealing the association of *TP53*, *ATM*, and *EP400* mutations that affected cell cycle, DNA repair, and glycolysis, with poor prognosis. Proteome-based analysis elucidated the stage-specific molecular characterization and defined the cancer-driving waves along with the mutation accumulation in esophageal carcinogenesis and progression. Furthermore, the trajectory analysis identified 6 major tracks related to different clinical features during early EC progression. Growingly enhanced and hyperphosphorylated phosphoglycerate kinase 1 (PGK1, S203) was detected and considered as a drug target in EC progression. Collectively, this study provides insight into the understanding of the molecular mechanism of EC progression and a valuable resource for the development of therapeutic targets.

Keywords

Early esophageal cancer, 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^{1,2}. 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 countries³, whereas ESCC predominantly occurs in Eastern Asia, particularly in China and Japan⁴, indicating the diverse lifestyle of the countries as a major etiological factor of EC. The other etiological factors of EC include gender, age⁵, and the habit of drinking/smoking⁶⁻⁹. 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)¹⁰, 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*¹¹⁻¹³. 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 adventitia¹⁴. The depth of infiltration of the cancer cells, determining the stage of lesions, was measured at the deepest point of their penetration in corresponding layers¹⁵. 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%)¹⁶. Though the recent advances of endoscopic submucosal dissection (ESD)¹⁷ have achieved the early detection of ESCC patients (T1 stage) with higher QOL and significantly improved overall survival rate (> 90%)¹⁸⁻²⁰, 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 an integrated omics analyses, including genomic, proteomic, and phosphoproteomic approaches, on 756 trace samples dissected from early esophageal cancers of 22 substages in 124 patients (114 in early- and 10 in advanced- stage). The integrative multi-omics dataset established the relationships between the genome aberrations and proteome alterations of ESCC,

revealing *TP53*, *ATM*, and *EP400* mutations elevated cell cycle, DNA repair, and glycolysis, and elucidating the molecular characteristics of different oncogenesis stages. Additionally, we used the trajectory inference methods²¹ to trace the carcinogenesis lineages of 114 early ESCC patients, and resulted in 6 major tracks that were associated with different clinical features (age, gender, and risk factors) during esophageal carcinogenesis. We observed the significant dynamic changes of glucose metabolism in the whole process of carcinogenesis, in which PGK1 was growingly enhanced and hyperphosphorylated at the S203 which is associated with worse prognosis. We further showed that inhibition of PGK1 suppresses tumor growth *in vivo* and *in vitro*. Collectively, our study so far provided a comprehensive genomic, proteomic, and phosphoproteomic landscape of ESCC in whole progression stages, and revealed the molecular attributes associated with diverse tracks and brought a new insight of PGK1 in ESCC progression.

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 criteria²². The number of substages identified from the samples varied from 4–16. The tissue samples from the corresponding substages were separately dissected from the formalin-fixed, paraffin-embedded (FFPE) slides. In total, 114 normal samples (stage 1), 632 early-stage samples (stages 2 to 7), and 10 advanced-stage samples (stages 8 and 9) were collected (**Supplementary Fig. 1; Supplementary Table 1**; details shown **Methods**). Subsequently, we performed mass spectrometry (MS) profiling of all 756 samples, phosphoproteomic profiling of 54 samples (**Supplementary Table 1**), and whole-exome sequencing (WES) of 53 samples (**Fig. 1a** and **Supplementary Fig. 4a; Supplementary Table 2**).

WES profiling (> 100× coverage, > 12G raw data) identified 97,173 non-silent mutations in 11,877 genes (mutation frequency > 5%), and 449,169 silent mutations in the Fudan cohort (this study) (**Supplementary Table 3**). The most frequently mutated genes were *FADS6* (60.9%), *UBASH3B* (56.5%), *ACAD8* (54.3%), *NLRP14* (54.3%), *TP53* (52.2%), *DVLI* (50.0%), *FLRT2* (50.0%), *HHIP* (47.8%), *TGFBR2* (47.8%), and *TKI* (45.7%) (**Fig. 1b**). We observed that the number of mutations gradually

cumulated during ESCC carcinogenesis, ranging from 6,151 mutations in stage 2 to 8,374 mutations in stage 9 (**Fig. 1c; Supplementary Table 3**). The number of the novel mutations reached three peaks in stage 2 ($n = 6,151$, hyperplasia stage), stage 3 ($n = 3,894$, Tis stage), and stage 6 ($n = 1,089$, submucosa (sm) a stage) indicated the significant events during carcinogenesis (**Fig. 1c**). Comparison of the mutational signatures identified in the Fudan cohort with the other ESCC cohorts, including the TCGA cohort^{4,12,18}, revealed frequent cytosine to thymine (C>T) transition in ESCC. In the Fudan cohort, it was detected with an average frequency of 37.4% (**Fig. 1d and Supplementary Fig. 4b, c**), and especially, overrepresented in female patients (t-test, $p = 2.1E-4$, 48.8% (female patients) vs. 37.1% (male patients)), older patients (t-test, $p = 3.6E-4$, 36.6% (>70) vs. 33.0% (50~69)), and drinking/smoking patients (t-test, $p = 0.021$, 47.4% (drinking/smoking patients) vs. 37.3% (non-drinking/smoking patients)) (**Fig. 2a, b, and Supplementary Fig. 4d; Supplementary Table 4**). On the contrary, compared with other ESCC cohorts, a higher proportion of thymine to cytosine (T>C) transition (25.1%) was observed in the Fudan cohort (**Fig. 1d, and Supplementary Fig. 4c; Supplementary Table 3**), especially in male patients (t-test, $p = 1.2E-4$, 25.6% (male patients) vs. 7.7% (female patients)) (**Fig. 2a; Supplementary Table 4**). In addition, the C>A mutational signature, was more predominant in male patients (t-test, $p = 0.018$, 12.3% (male patients) vs. 12.2% (female patients) in the Fudan cohort (**Fig. 2a**). Furthermore, the comparative analysis of the most frequent genomic aberrations in ESCC also reflected significant differences between the Fudan cohort and other ESCC cohorts (Fisher's exact test, $p < 0.05$) (**Fig. 1e**). Higher mutation frequencies of cell cycle related genes were detected in the Fudan cohort, such as *FAM135B* (39.1%), *FAT3* (30.4%), *MYC* (15.2%), *EGFR* (13.0%), *ERBB2* (6.5%), etc., compared to other ESCC cohorts (Fisher's exact test, $p < 0.05$) (**Supplementary Table 3**). Collectively, these findings indicated the tumor heterogeneity between early and advanced stages of ESCC and suggested that the difference is dependent on the diverse lifestyles of the countries and territories.

The numbers of proteins per sample ranged from 5,000–7,000, with a total of 13,711 in 756 samples (**Methods**) (**Supplementary Fig. 3a, Supplementary Table 5**). With the advancement of the stages of ESCC, the numbers of identified proteins were slightly elevated from ~5,000 in stage 1 to ~7,000 in stages 8 and 9, indicating an enhanced proteome complexity during the process of carcinogenesis (**Supplementary Fig. 3b, c**). Proteomic analysis was performed using a label-free quantification strategy^{23,24}. Protein abundance was firstly calculated by intensity-based absolute quantification (iBAQ)^{25,26} and then normalized as a fraction of the total (FOT), allowing the comparison between

samples (mean (correlation coefficient) = 0.83) (**Supplementary Fig. 2b, 3d and 3e**).

Furthermore, a total of 34,395 phosphosites corresponding to 6,293 phosphoproteins were identified in 54 samples (**Fig. 1f and Supplementary 3f, g; Supplementary Table 5**). Whole-cell extract of HEK293T cells was used as the quality control (QC), which showed the MS was robust and consistent based on the large Spearman's correlation coefficients (mean = 0.91) between the QC samples (**Supplementary Fig. 2c; Supplementary Table 6**). In addition, the reference proteome was highly dynamic based on the protein abundance (FOT), which spanned over eight orders of magnitude. The top 10 proteins extensively expressed in our dataset were HBB, HBD, ALB, KRT5, HBG1, ANXA1, CSTB, HSPB1, FABP5, and S100A9 (**Supplementary Fig. 3d**). Low abundance proteins, including NEB, DD11, AK9, LIPE, SDK1, MYH8, HELZ, TRIO, KMT2C, and DNAH5 (**Supplementary Fig. 3d**), and common transcription factors, including BAZ1A, ATF7IP, BPTF, BCAS3, CDH24, and EP300 were also covered (**Supplementary Table 5**). Additionally, the ESCC biomarkers identified in previous studies (several ESCC tissues or cell lines), including ACTA2, TAGLN, ANXA1²⁷, HSPA9, PDIA4, PLEC, POSTN, PSAP, and THBS1²⁸ were also covered in the Fudan cohort (**Supplementary Table 5**). Overall, these findings established a comprehensive landscape of ESCC progression at the genome, proteome, and phosphoproteome levels.

The impacts of genomic aberrations on the proteome level in ESCC progression

Genome aberrations were considered the cause of proteome alterations. To explore the impacts of genomic aberrations, we performed whole exome-based somatic copy number alterations (SCNAs) analyses and examined the regulatory effects of 23,172 SCNAs on protein expressions (**Methods**). Through the correlation of SCNAs and the corresponding identified proteins (**Fig. 2c; Supplementary Table 7**), more oncogenic pathways-related proteins involved in immune response ($p = 3.7E-4$) (e.g., CD248, CD274, and CD276), Ras signaling ($p = 2.7E-5$) (e.g., N/KRAS, ERBB2, and EGFR), cell cycle ($p = 1.7E-4$) (e.g., CDK1/6, RB1, and CCNB1/D1), and DNA damage ($p = 1.1E-3$) (e.g., TP53, ATM, and ATR) were identified (**Fig. 2d and Supplementary Fig. 4e; Supplementary Table 7**). The observation of the mutations in DNA damage was also disclosed in previous studies. *Abuzeid et al.* found *RAD50* mutation induces ESCC carcinogenesis²⁹, which is considered as a source of the immune response³⁰. Apart from glucose metabolism ($p = 3.7E-5$) (e.g., ALDOC, PFKM, and PDK1), genes involved in biomaterials synthesis had positive SCNAs-protein correlations, including nucleoside

triphosphates (NTPs) biosynthesis ($p = 3.4E-5$) (e.g., CAD, ATP5A1, and PPAT), carboxylic acid metabolism ($p = 4.6E-4$) (e.g., ACOX1, ACAA1, and BDH2) (**Fig. 2d** and **Supplementary Fig. 4e; Supplementary Table 7**). Metabolic reprogramming of the Warburg effect³¹ is associated with human ESCC cells and increases malignancies³². The dysregulation of the critical enzymes involved in the Warburg effect, a hallmark of nearly all tumors, observed in this study could help establish novel management strategies for ESCC therapies.

TP53, a prominent mutated gene in ESCC^{4,33}, was found to be mutated in all stages of ESCC in the Fudan cohort, in addition to *GSK3B* (34.8%), *FAM175A* (45.7%), and *SUMO1* (28.3%). Moreover, the mutations in genes associated with immune response (e.g., *C2* (26.1%), *HLA-A* (43.5%), and *FGF19* (30.4%)) and glucose metabolism (e.g., *PFKM* (45.7%), *PHKG2* (26.1%), and *ABHD5* (39.1%)) were identified in all stages of ESCC (**Supplementary Fig. 5; Supplementary Table 7**). The mutations of the genes in canonical cancer-related pathways, such as Wnt signaling (e.g., *FZD4* (30.4%), *LRP5* (43.5%), and *PRKCA* (30.4%)), extracellular matrix organization (ECM) (e.g., *NCAM1* (34.8%), *LAMB1* (28.3%), and *COL6A2* (30.4%)), and Hedgehog signaling (e.g., *HHIP* (47.8%), *EVC* (43.5%), and *IQCE* (43.5%)), were also observed in all stages of ESCC (**Supplementary Fig. 5; Supplementary Table 7**).

Mutually exclusive (ME) analysis is generally used to identify driver genes, pathways, and investigate their functional relationships³⁴. Therefore, we performed ME analysis of the identified mutations in the Fudan cohort to elucidate the mutation dynamics in ESCC. Specifically, the mutations in genes related to cell cycle (e.g., *SKP1* (2.2%), *ORC1* (2.2%), and *CDKN2C* (2.2%)) (Bartlett's test, $p = 8.4E-7$), and DNA repair (e.g., *POLD4* (2.2%), *MED1* (2.2%), and *RAD50* (2.2%)) (Bartlett's test, $p = 7.0E-5$) were ME from each other (**Supplementary Fig. 2e; Supplementary Table 7**), indicating the impacts of genomic aberrations of cell cycle and DNA repair in ESCC progression.

To disclose the impacts of genomic aberrations on their counterpart proteins, we incorporated somatic mutations to the corresponding proteins with normalized FOT (Z-score), and found that the mutations of immune-related genes, such as *HLA-A* (t-test, $p = 5.3E-3$, fold change (FC) (Mut/WT) = 0.55), *CD8A* (t-test, $p = 6.6E-3$, FC (Mut/WT) = 3.2E-4), and *IGFALS* (t-test, $p = 0.046$, FC (Mut/WT) = 0.53), were significantly associated with decreased protein expression (**Fig. 2f**). Recent research has provided evidence that inflammation is a strong risk factor in the microenvironment for developing digestive tumors³⁵, including ESCC³³. Compared with ESCC cohort in TCGA¹², the mutations of cell cycle related genes, such as *SUMO1* (t-test, $p = 0.028$, FC (Mut/WT) = 4.9E-3) and *EGFR* (t-test, $p =$

0.045, FC (Mut/WT) = 0.52), were negatively associated with their protein expressions, reflecting the distinct profile of Chinese ESCC patients.

Among the mutations, we found 315 genes were co-mutated in all stages in ESCC progression, including *TP53* mutation. To delineate the consequences of genomic aberrations on protein expressions, we performed the mutation-protein correlation analysis of most frequent mutated genes of those genes using \log_2 FC (Mut/WT). It was observed that the negative regulation proteins (n = 787) were enriched in protein metabolism (e.g., *ANLN* and *OXA1L*) ($p = 3.2E-3$), endocytosis (e.g., *CD36* and *CYTH2*) ($p = 0.019$), and immune response (e.g., *CD8A* and *C1QA*) ($p = 0.017$), and the positive regulation proteins (n = 1,067) were enriched in the regulation of cell growth (e.g., *NRG1* and *TGFB2*) ($p = 4.8E-4$), cell death (e.g., *TP53* and *EP300*) ($p = 2.6E-3$), apoptosis (e.g., *BAD* and *CASP3*) ($p = 4.5E-3$), cell cycle (e.g., *CHEK2* and *CDKN2A*) ($p = 0.012$), ECM signaling (e.g., *ECM2* and *COL5A2*) ($p = 0.029$), and insulin signaling (e.g., *IGF1R* and *IRS2*) ($p = 0.052$) (**Supplementary Fig. 2g; Supplementary Table 7**). Notably, *TP53* mutation was co-mutated with plentiful genes, including *HHIP* (47.8%), *TTN* (41.3%), *FAM175A* (45.7%), and *GSK3B* (34.8%) in the Fudan cohort. In addition, *TP53* mutation displayed a negative effect on the protein expression of *ERBB2* (t-test, $p = 0.065$, FC (Mut/WT) = 0.22), and showed a systemic positive effect on thymidine kinase (*TK1*) at the proteome level (t-test, $p = 0.013$, FC (Mut/WT) = 2.61), which has been considered as a valuable biomarker for predicting ESCC patients prognosis³⁶. We then investigated the impacts of *TP53* aberration on the alterations of driven pathways and the related protein expressions, and discovered the biological function of normal esophagus was down-regulated, such as protein metabolism (e.g., *ANXA1* (t-test, $p = 0.034$, FC (Mut/WT) = 0.50) and *PTK2B* (t-test, $p = 0.039$, FC (Mut/WT) = 0.38)) and immune response (e.g., *HLA-B* (t-test, $p = 5.2E-3$, FC (Mut/WT) = 0.42) and *CD8A* (t-test, $p = 9.1E-3$, FC (Mut/WT) = 0.03)). Conversely, several oncogenic pathways were elevated in the patients with *TP53* mutation, including p53 signaling (e.g., *TP53* (FC (Mut/WT) = 10.11) and *EP300* (FC (Mut/WT) = 321.38)), cell cycle (e.g., *CDK6* (FC (Mut/WT) = 14.31) and *E2F4* (FC (Mut/WT) = 29.29)), PI3K-AKT signaling (e.g., *PIK3CA* (FC = 56.92) and *AKT3* (FC (Mut/WT) = 64.38)), Wnt signaling (e.g., *WNT5A* (FC (Mut/WT) = 60.13) and *FZD1* (FC = 101.54)), Insulin signaling (e.g., *INSR* (FC (Mut/WT) = 4.79) and *EGFR* (FC (Mut/WT) = 2.07)), and ECM and epithelial-mesenchymal transition (EMT) signaling (e.g., *ITGA2* (FC (Mut/WT) = 2.49) and *MMP2* (FC (Mut/WT) = 3.03)). Additionally, *TP53* mutation showed positive effects on the expression of proteins, related to alcohol metabolism and glycolysis, especially *PGK1* (t-test, $p = 0.035$,

FC (Mut/WT) = 1.49) and PFKM (t-test, $p = 0.013$, FC (Mut/WT) = 1.79) (**Fig. 2h**). Thus, our study provided new insights into the proteomic characteristics with *TP53* mutation and revealed the impacts of genome aberrations on the proteome level.

Proteome clusters of ESCC progression

Visualization of the abundance of the most variable proteins ($n = 6,687$, E3, **Methods**) by principal component analysis (PCA) (**Methods**) differentiated the proteome profiles for all the 22 ESCC substages, which clearly discriminated the proteomes of the early (T0–T1) and the progressive (T2–T3) ESCC stages. Additionally, PCA of early ESCCs indicated obvious diversity among the stages, along with EC progression (**Fig. 3a**). 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. 3b** and **Supplementary Fig. 6; Supplementary Table 8**). This distinct two-stage separation suggested that the irreversible fundamental proteome alterations happened as early as the Tis carcinoma stage (stage 3), which was further supported by the results in the WES dataset from 53 ESCC samples, reflecting more malignancy of C2 (**Fig. 3b, d**). *Van der Schaaf* indicated that the Tis stage was associated with worse survival³⁷. The differential analysis of the two major clusters displayed 6,335 common mutations between the two clusters, including the frequent mutations of *TP53* (52.2%), *EP300* (8.7%), *EGFR* (13.0%), and *FAT3* (30.4%), and disclosed 5,542 specific mutations (e.g., *INSR* (6.5%), *ATM* (19.6%), and *VCAMI* (15.2%)) in C2. At the proteome level, 4,478 common proteins were identified in C1 and C2, whereas the remaining 2,480 proteins were exclusive to C2 (**Fig. 3d**).

The significance analysis of microarray (SAM)³⁸ 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) ≥ 2 or ≤ 0.5), including 3,163 elevated (C2) and 242 descend proteins (C1) (**Fig. 3c; Supplementary Table 8**). 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) ≥ 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) ≤ 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)³⁹. 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, $p = 3.6E-43$, 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. 7a – d; Supplementary Table 8**), 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. 7e – h; Supplementary Table 8**). Taken together, these results provided further confirmation in the dysregulation of metabolism and oncogenic pathways between genomic aberrations and proteomic alterations in ESCC progression.

Proteomic characteristics of ESCC progression with co-mutations of *ATM* and *TP53*, and *EP400* and *TP53*

Incorporation of genomic aberrations and the corresponding proteomic alterations revealed an overrepresentation of the canonical cancer-related pathways, including cell cycle (e.g., CDK1 (FC (C2/C1) = 488.27), and TP53 (FC (C2/C1) = 1.51)), DNA repair (e.g., ATM (FC (C2/C1) = 3.63, and EP400 (FC (C2/C1) = 1.16)), EMT signaling (e.g., MMP11 (FC (C2/C1) = 2.34), and VCAM1 (FC (C2/C1) = 4.75)), PI3K-AKT signaling (e.g., MTOR (FC (C2/C1) = 65.77), and PIK3C3 (FC (C2/C1) = 4.04)), Ras signaling (e.g., KRAS (FC (C2/C1) = 1.54), and RAF1 (FC (C2/C1) = 1.23)), and insulin signaling (e.g., INSR (FC (C2/C1) = 1.12) and IRAK1 (FC (C2/C1) = 2.93)) in C2, both at the genome and proteome levels (**Fig. 3e**). The mutations of *ATM* (19.6%) and *EP400* (23.9%) were exclusively detected in C2 and associated with poor survival (log-rank test, p (*ATM*) = 7.0E-4, p (*EP400*) = 8.9E-3) (**Fig. 3e**), suggesting their roles in ESCC progression. To investigate the effects of the mutations of driver genes on the protein expression levels and the related pathways, we examined significantly altered proteins ($n = 1,392$) in the patients with or without driver mutations (*ATM*, *EP400*, and *TP53*). As a result, we identified 716 and 676 proteins associated with *ATM*, *EP400*, and *TP53* mutations, positively and negatively, respectively (**Supplementary Table 8**). Of these, immune response related proteins ($p = 5.3E-4$, FDR = 0.029) (e.g., IL1RN, IL36A, and MUC21), and protein metabolism related proteins ($p = 3.3E-5$, FDR = 2.1E-3) (e.g., AACCS, ACOX3, and ALOX12), were negatively associated with *ATM*,

EP400, and *TP53* mutations (**Fig. 3f; Supplementary Table 8**). Wherein, proteins associated with glycolysis ($p = 7.2E-3$, FDR = 0.19) (e.g., HK3, ENO3, and PGM2L1), and canonical cancer-related pathways, including cell cycle ($p = 6.8E-3$, FDR = 0.19) (e.g., CSNK1A1, CSNK1D, and CSNK2B), DNA repair ($p = 3.9E-3$, FDR = 0.17) (e.g., XRCC1, TP63, and POLE3), apoptosis ($p = 0.031$, FDR = 0.31) (e.g., BAK1, CASP3, and BID), EGFR signaling ($p = 0.012$, FDR = 0.20) (e.g., EGFR, IGF2BP2/3, and PLCG1), and ECM signaling ($p = 2.7E-5$, FDR = 0.015) (e.g., ITGA3, ITGB3, and MMP1), were positively associated with *ATM*, *EP400*, and *TP53* mutations (**Fig. 3g; Supplementary Table 8**). Notably, a consistent reduction of protein metabolism- (t-test, p (protein metabolism) = $1.2E-3$) and immune-related proteins (t-test, p (immune response) = $2.6E-3$), and elevation of oncogenic proteins (t-test, p (glycolysis) = $9.4E-25$, p (cell cycle) = $5.2E-10$, p (DNA repair) = $3.9E-25$, p (apoptosis) = $5.7E-26$, p (EGFR signaling) = $3.4E-19$, and p (ECM signaling) = $6.6E-30$) were observed in ESCC progression (**Supplementary Fig. 8a – c; Supplementary Table 8**). These findings presenting the impacts of *ATM*, *EP400*, and *TP53* mutations in ESCC indicated the synergies among the mutations.

We further performed synergy analysis of *ATM*, *EP400*, and *TP53* mutations and discovered the significant correlation of co-mutations of *ATM* and *TP53* (*ATM* & *TP53*, Fisher's exact test, $p = 2.1E-3$) and *EP400* and *TP53* (*EP400* & *TP53*, Fisher's exact test, $p = 9.5E-3$) (**Fig. 3f, g**). Patterns of co-mutations in many tumors indicated functional synergies and, importantly, may reflect the prognosis of patients^{40,41}. We discovered co-mutations of *ATM* & *TP53* (log-rank test, $p = 5.0E-4$), and *EP400* & *TP53* (log-rank test, $p = 9.6E-3$), which were poor prognosis outcomes in patients (**Fig. 3h** and **Supplementary Fig. 8d**). To assess the functional impact of *ATM*, *EP400*, and *TP53* co-mutations, we divided the patients into four groups: *TP53* Mut only, *ATM* & *TP53* co-Mut, *EP400* & *TP53* co-Mut, and *ATM* & *EP400* & *TP53* co-Mut. We discovered overrepresentation of mTOR signaling ($p = 0.036$) (e.g., EIF4B, RPS6KA3, and EIF4E1B) in the group of *TP53* Mut only (**Fig. 3i, j**). DNA repair ($p = 3.7E-6$) (e.g., PCNA, POLE4, and DDB1), focal adhesion ($p = 1.5E-4$) (e.g., VAV1, CAV1, and PDGFRB), and glycolysis ($p = 3.1E-4$) (e.g., PGK1, ENO1, and PGM2) were elevated in the group of *ATM* & *TP53* co-Mut. In *EP400* & *TP53* co-Mut group, insulin signaling ($p = 2.2E-3$) and aminoacyl-tRNA biosynthesis ($p = 2.5E-3$) related proteins, including INSR, IRS2, PPP1CA, CARS, LARS, and VARS2, were overrepresented. Complement and coagulation cascade ($p = 1.2E-9$) related proteins (e.g., C3/C5/C6, FGA/FGG, and CD55) were dominant in the group of *ATM* & *EP400* & *TP53* co-Mut (**Fig. 3i, j; Supplementary Table 8**). Overall, our findings demonstrated the impacts of the *ATM*, *EP400*, and *TP53* mutations in ESCC

progression and prognosis via altering the expression of critical proteins in glycolysis, DNA repair, and complement and coagulation cascade.

The temporal driver pathway waves in ESCC progression

Recent studies have provided considerable evidence that tumor stroma plays a crucial role in the induction of immune tolerance to human cancers⁴². 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, $p < 2.2E-16$), stromal 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. 9a, b; Supplementary Table 9**).

Further, the substage-based supervised clustering analysis (**Methods**) displayed specific characteristics of the 22 substages. Based on the histopathological stages and temporal expression patterns, the substages were subsequently assigned to 8 panels (**Fig. 4a** and **Supplementary Fig. 10a; Supplementary Table 10**). We then performed a panel-based analysis to elucidate the molecular dynamic models that drove carcinogenesis from early to progressive ESCC. As shown in **Supplementary Fig. 10a**, 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, $p < 2.2E-16$)), particularly, ANXA1, a biomarker of ESCC²⁷, was gradually decreased during ESCC progression (FC (advanced stages (AS)/stage 1 (S1)) = 0.029) (**Supplementary Fig. 11a**). 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, was 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 PRKACA (Kruskal-Wallis test, $p < 2.2E-16$)) 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 LAMTOR (Kruskal-Wallis test, $p < 2.2E-16$)) were enhanced in substage 6_1 and substage 6_2 (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^{43,44}, and the associated protein markers, such as MMP2 and MMP8, were identified as specific indicators of high-grade malignancy^{45,46}. Overrepresentation of glycolysis (e.g., PGK1 (Kruskal-Wallis test, $p < 2.2E-16$), ENO1 (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. 10b**). PGK1 (FC (AS/S1) = 4.31), an important glycolytic enzyme, though not mutated at the genome level, was highly elevated at the proteome level, especially in the advanced 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 (<https://xcell.ucsf.edu>) (**Methods**)⁴⁷ 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. 9a, b; Supplementary Table 9**). In contrast, the innate/acquired immune response signatures, such as CD8⁺ 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 genome level, we observed extensive co-occurrence or exclusive mutations along with ESCC progression. Specifically, the mutations of *VEGFA* (2.2%), *IL36A* (2.2%), and *SERPINB1* (6.5%) in stage

2, were ME to the mutations of *VCAN* (6.5%), *INSR* (6.5%), *ITGAI* (32.6%), *ITGB6* (21.7%), *GYS2* (8.7%), *FGR* (10.9%), *IRAK1* (10.9%), *IGFBP1* (15.2%), *VWF* (13.0%), and *RPS6KA2* (8.7%) (Fisher exact test, $p < 0.05$) (**Fig. 4b; Supplementary Table 10**). On the contrary, the mutations of *ITGAI*, *VCAN*, *ITGB6*, *INSR*, *GYS2*, *FGR*, *VWF*, *IRAK1*, *IGFBP2*, and *RPS6KA2*, were co-occurred in the later stages (Fisher exact test, $p < 0.05$) (**Fig. 4b; Supplementary Table 10**). We then examined the impacts of these mutations in different stages on the proteome level during ESCC progression. As shown in **Fig. 4c**, the mutations of *VEGFA*, *IL36A*, and *SERPINB1*, positively affected lipid-protein metabolism (e.g., *ACOX3*, *ALOX12*, and *ANXA1*) (t-test, $p = 1.8E-7$) and inflammatory responses (e.g., *IVL*, *MVK*, and *SERPINB3*) (t-test, $p = 3.4E-14$) (**Fig. 4c**). The mutations of *VCAN*, *INSR*, *ITGAI*, and *ITGB6* observed in stages 3, 4, and 5 were positively associated with ERBB2 signaling (e.g., *EGFR*, *ERBB2*, and *PTPN1*) (t-test, $p = 0.036$) and insulin signaling (e.g., *IRF2BP2*, *CRK*, and *PPP1R21*) (t-test, $p = 6.4E-3$). The mutations of *GYS2*, *FGR*, *IRAK1*, *IGFBP1*, *VWF*, and *RPS6KA2*, in the later stages (stages 6, 7, 8, and 9) were positively associated with the elevated EMT signaling (e.g., *VCAN*, *MMP2*, and *ITGA5*) (t-test, $p = 1.5E-6$), Wnt signaling (e.g., *WNT5A*, *GSK3B*, and *CTBP2*) (t-test, $p = 4.4E-3$), and glycolysis (e.g., *PGK1*, *PGM1*, and *CALU*) (t-test, $p = 6.1E-3$).

Notably, the dynamic waves in ESCC progression were also detected at the phosphoproteome level (**Fig. 4d; Supplementary Table 10**). For example, elevated phosphorylation of *ANXA1* S37 and S170, and *FASN* S831 and S974 (t-test, $p = 2.6E-4$), was observed in stage 1. Phosphorylation of *CDH1* S793, *PPP2R5A* S2/S5, and *SERPINB3* T66 (t-test, $p = 5.6E-3$) was predominant in stage 2. ERBB2 signaling was the driven pathway in stage 3 (Tis stage), and phosphorylation of *ERBB2* S998, *EGFR* T693, S1166, and S1042/1046 (t-test, $p = 2.0E-4$) was also overrepresented in stage 3. In stage 4 (lamina propria stage), insulin signaling and IGF signaling were activated with elevated phosphorylation of the corresponding proteins, such as *IGFBP5* S125, *IRAK2* S144, *IGF2R* S2401, and *IRS1* S3 (t-test, $p = 9.1E-5$). Elevated phosphorylation of cell cycle related proteins, including *CDK1* S14/Y15, *ORC1* T230, and *RB1* S795 (t-test, $p = 6.1E-7$), was also observed in stage 5 (muscularis mucosa stage). The correlation of DNA repair and mTOR signaling in stage 6 (sm a stage) can be evidenced by the increased phosphorylation of *ATM* T1885, *MSH4* S7/S10/T23/S25, *MSH6* T139, and *MTOR* S1261 (t-test, $p = 9.5E-4$). Overrepresentation of EMT signaling and PI3K-AKT signaling related proteins (e.g., *MMP3*, *PIK3R4*, *TK1*, and *VIM*), and phosphorylation of the counterpart proteins (e.g., *MMP3* T466, *PIK3R4* T220, *K1* S231, and *VIM* S27 and S419) (t-test, $p = 2.0E-8$) were observed in stage 7 (sm b stage). Phosphorylation

of CTNNB1 S191, GSK3A T19, and GSK3B T390 and S215, and PGAM1 S118 (t-test, $p = 1.7E-8$) was observed in stages 8 and 9 (**Fig. 4d; Supplementary Table 10**), further indicating the consistency of impacts of glycolysis and Wnt signaling in ESCC progression at the genome, proteome, and phosphoproteome 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. 4e; Supplementary Table 10**). Furthermore, these results also defined the substages-specific molecular characteristics and uncovered the potential candidates for ESCC malignancy.

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 (**Methods**)²¹ 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, and Supplementary Fig. 12a**). 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 (**Supplementary Fig. 12b**) 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 and Supplementary Fig. 12c; Supplementary Table 11**). 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 (**Supplementary Fig. 13a**). Furthermore, we found the highest mutation frequencies of *KLK12* (85.7% (T2) vs. 45.0% (T5) and 20.0% (T6)) and *KLK13* (57.1% (T2) vs. 45.0% (T5) and 20.0% (T6)) in the non-drinking/smoking track (T2), associated with enhanced expressions of ECM signaling related proteins (e.g., *ITGA5*, *LAMB1*, and *VTN*) (ANOVA, $p = 4.9E-3$) (**Supplementary Fig. 13b; Supplementary Table 11**). In the older track (T5), the prevalent mutations of *SPRP3* (30.0% (T5) vs. 14.3% (T2) and 0% (T6)) and *EVPL* (25.0% (T5) vs. 14.3% (T2) and 6.7% (T6)) were associated with the overrepresentation of proteins in glucose metabolism (e.g., *PGK1*, *ALDOA*, and *ENO1/3*) (ANOVA, $p = 5.3E-3$) (**Supplementary Fig. 13b; Supplementary Table 11**). The mutation of *AHNAK2* (26.7% (T6) vs. 14.3% (T2) and 0%(T5)), associated with auto-immune disease⁴⁸, was prominent in the drinking/smoking track (T6) and related to the elevated expression of immune response related proteins (e.g., *C1QB*, *C1S*, and *SERPINH1*) (ANOVA, $p < 1.0E-4$) (**Supplementary Fig. 13b; Supplementary Table 11**).

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%), significantly greater than the natural proportion of female patients in ESCC (no more than 10%)^{2,49,50}. 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 types⁵¹, revealing the unique track in the drinking/smoking patients.

The consistency in the characteristics of 6 major tracks in ESCC between the proteome and phosphoproteome levels was also observed (**Fig. 5e**). The elevated phosphorylation of *GLI3* S295, *NCK1* S166, *NCBP1* T21, *SRF* S224, and *ATP2B4* S1165 (t-test, $p = 6.0E-3$) defined a key role of the biomaterial synthesis in T1. Increased phosphorylation of *COL1A1* S787, *SLC2A1* T478, *PLG* S45, and *CAV2* S20/S23 (t-test, $p = 0.036$) provided evidence that ECM signaling was prominent in the non-drinking/smoking track (T2). The phosphorylation of cell cycle related proteins such as *RB1* S780,

HDAC7 T366, MCM2 S41, CDCA2 S591, and CCNL1 S329/S342 (t-test, $p = 0.048$) was identified in the female track (T3), whereas phosphorylation of DNA repair related proteins, including TP53 S46, TP63 S477, ATM S1883, MSH6 S200, and RIPK1 S320 (t-test, $p = 0.014$) was predominant in the mainstream track (T4) (**Fig. 5e; Supplementary Table 11**). The consistent findings of elevated expression of glycolysis related phosphoproteins (e.g., PGK1 S203, PGK2 S393, and PGM3 S64, CALU Y47, and GAPDH S210) (t-test, $p = 0.011$) in the older track (T5) (**Fig. 5e; Supplementary Table 11**), further indicated the impacts of glycolysis in ESCC progression.

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. 13c**). Consistent with the proteome alterations, the mutations of *FNI* (28.6% (T2) vs. 25.0% (T5) and 0%(T6)), *ITGAI* (57.1% (T2) vs. 40.0% (T5) and 13.3% (T6)), and *ITGAM* (14.3% (T2) vs. 10.0% (T5) and 0% (T6)) and their enhanced counterpart proteins (FC (Mut/WT) = 1.70 (*FNI*), 6.08 (*ITGAI*), and 1.93 (*ITGAM*)), were both observed in the non-drinking/smoking track (T2) (**Fig. 5f; Supplementary Table 11**). The mutations of glucose metabolism related genes (e.g., *ALDOA* (15.0% (T5) vs. 14.3% (T2) and 0% (T6)), *GPI* (15.0% (T5) vs. 0% (T2) and 0%(T6)), and *PGD* (30.0% (T5) vs. 0% (T2) and 0% (T6)), and their significantly increased corresponding proteins (*ALDOA*: t-test, $p = 0.011$, FC (Mut/WT) = 2.09; *GPI*: t-test, $p = 0.058$, FC (Mut/WT) = 2.06; *PGD*: t-test, $p = 5.2E-12$, FC (Mut/WT) = 9.84), were predominant in the older track (T5) (**Fig. 5f; Supplementary Table 11**). The mutations of immune response related genes (e.g., *MYOIG* (6.7% (T6) vs. 0% (T2) and 5.0% (T5)), *MAP4K4* (6.7% (T6) vs. 0% (T2) and 5.0% (T5)), and *TBCEL* (26.7% (T6) vs. 0% (T2) and 10.0% (T5)), and the overrepresentation of the corresponding proteins (FC (Mut/WT) = 3.27 (*MYOIG*), 4.54 (*MAP4K4*), and 68.17 (*TBCEL*)), were notably detected in the drinking/smoking track (T6) (**Fig. 5f; Supplementary Table 11**).

To profile the personalized dynamic waves of 6 major tracks, we performed Fishplot analysis⁵² (**Methods**) with track enhanced proteins and kinetic driving pathways (**Supplementary Fig. 13d**). 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 (**Fig. 5d**).

For each track, we assessed the drug targets approved by the US Food and Drug Administration database (FDA)⁵³ (<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 abemaciclib/palbociclib 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 ESCC⁵⁴; however, tumoral heterogeneity imposes obstacles. In the drinking/smoking track (T6), the increased expression of CD4, CD8A, CD19, CD28, CD274, and CD276 (ANOVA, $p = 1.6E-3$) was identified, reflecting responses to immunotherapy (**Fig. 5g**). 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 genome, proteome, and phosphoproteome 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**). To further cross validate these results, our dataset confirmed that the expression of PGK1 was gradually enhanced in ESCC at the proteome (Kruskal-Wallis test, $p < 2.2E-16$, FC (AS/S1) = 4.31) and

phosphoproteome levels (PGK1 S203, Kruskal-Wallis test, $p = 3.3E-6$, FC (AS/S1) = 3794.85) (**Fig. 6c**; **Supplementary Table 12**), 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**). 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**)^{55,56}. The results revealed the association of ERK1/2, CDKs, and GSK-3 with the motif of PGK1 (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 PGK1⁵⁷. Therefore, these proteomic and phosphoproteomic results indicated that the total PGK1 activity increased notably in both cytosol and mitochondrial in tumor cell. In addition, *CDK2* mutation had positive effects on the expression of PGK1 (t-test, $p = 2.0E-4$, FC (Mut/WT) = 2.21) (**Fig. 6f**; **Supplementary Table 12**), and the substrates (e.g., AKT1 (FC (Mut/WT) = 4.88), GSK3A (FC (Mut/WT) = 4.53), GSK3B (FC (Mut/WT) = 2.51), and MAP1B (FC (Mut/WT) = 5.15)) of kinases of ERK1/2, CDKs, and GSK-3 (**Fig. 6g**; **Supplementary Table 12**), suggesting that *CDK2* mutation and ERK1/2 synergistically determined the total activity of PGK1 through increasing PGK1 expression both at the proteome and phosphoproteome levels. Subsequent kinase-substrates analysis in all 54 samples revealed the elevated expression of kinases and corresponding downstream substrates, including RB1 (Kruskal-Wallis test, $p = 0.012$, FC (AS/S1) = 14.68), CDK1 (Kruskal-Wallis test, $p = 2.0E-5$, FC (AS/S1) = 290.70), CDK2 (Kruskal-Wallis test, $p = 2.3E-4$, FC (AS/S1) = 3,256.16), ARID1A (Kruskal-Wallis test, $p = 0.016$, FC (AS/S1) = 8.18), MCM2 (Kruskal-Wallis test, $p = 1.9E-4$, FC (AS/S1) = 27.89), GSK3B (Kruskal-Wallis test, $p = 1.4E-4$, FC (AS/S1) = 73.71), etc. in ESCC progression. Enhanced phosphorylation of these substrates, including RB1 T373 (Kruskal-Wallis test, $p = 0.038$, FC (AS/S1) = 273.16), CDK1 T14 (Kruskal-Wallis test, $p = 2.8E-3$, FC (AS/S1) = 3,131.92), CDK2 T160 (Kruskal-Wallis test, $p = 1.1E-3$, FC = 854.54), ARID1A S698 (Kruskal-Wallis test, $p = 0.034$, FC (AS/S1) = 59.59), MCM2 S139 (Kruskal-Wallis test, $p = 0.05$, FC (AS/S1) = 174.77), and GSK3B Y216 (Kruskal-Wallis test, $p = 0.028$, FC (AS/S1) = 59.19), was also

observed in ESCC progression (**Fig. 6h; Supplementary Table 12**), 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; Supplementary Table 13**). 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**). 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**). 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, $p < 0.0001$) (**Fig. 7g; Supplementary Table 13**). 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**). 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. 14a**). 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 PGK1⁵⁸. 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 IC₅₀ of 16.3 nM (**Supplementary Fig. 14b, c; Supplementary Table 13**) 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. 14d**). 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**), which was consistent with the effects of PGK1 knockdown (t-test, p (KYSE150) = $8.6E-8$, p (ECA109) = $3.4E-7$) (**Fig. 7j**). 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 challenges to meet the 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, ECM-EMT, glycolysis, and Wnt signaling, providing a novel temporal dimension and trans-omics dimension in understanding the ESCC progression.

The dual-peak of the novel mutations detected at stages 2, 3, and 6 perfectly matched their corresponding pathological phenotypes observed in the clinic. These significant genomic aberrations

were then translated to proteomic alterations in ESCC progression. In stage 2 (hyperplasia stage), the innate immune pathways of inflammation and interferon-gamma signaling were enhanced, revealing the typical immune response to external stress, such as virus infection, physical and chemical injury. The mutations of *VCAN*, *ITGA6*, and *INSR* were the novel mutations in stage 3 (Tis stage) that negatively affected inflammatory response and had positive effects on ERBB2 signaling at the multi-omics level. Furthermore, the prevalence of mTOR signaling and DNA repair in stage 6 (sm a stage), suggested the formation of an environment that promoted tumor immunity. In addition, the co-mutations of *TP53*, *ATM*, and *EP400* (C2 mutations) enhanced the expression of complement and coagulation cascade related proteins (e.g., C5/6, CD55, and FGA). ECM signaling, associated with cell proliferation and migration in ESCC^{59,60}, and the related proteins/phosphoproteins were overrepresented in stage 7 (sm b stage). Furthermore, the mutations and the consequent gradual reduction in keratinization and lipid metabolism at the multi-omics level 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)^{27,28}, were also detected and significantly enhanced during ESCC progression (**Supplementary Fig. 11a, b; Supplementary Table 10**). The findings of this study 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. The mutations of *FNI*, *ITGA1*, and *ITGAM* were prominent in the non-drinking/smoking track and positively impacted their corresponding protein expressions. The mutations of *ALDOA*, *GPI*, and *PGD*, enhanced the expression levels of their corresponding proteins, and the overrepresentation of phosphoproteins, including PGK1 S203, PGM2 S165, and PFKL S763, confirmed the impacts of glucose metabolism on ESCC progression in the older track. A large-scale, population-based cohort study has shown that drinking/smoking promotes ESCC

carcinogenesis⁶. 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 track 1 was shown to be prevalent in ESCC¹² and considered as therapeutic targets⁶¹, 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 effect³¹. 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 genome, proteome, and phosphoproteome 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 cancers⁶². In the current study, increased PGK1 both at the proteome and phosphoproteome 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 synthesis⁶³. 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, demonstrated 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. A drug-targetable protein, PGK1, especially the phosphoprotein, PGK1 S203, has demonstrated the value of this multi-omics map. We believe this study provides insights into understanding the mechanism of the ESCC progression and enables new advances in promoting the diagnostics and therapeutics to manage ESCC.

ACKNOWLEDGEMENTS

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.

DECLARATION OF INTERESTS

The authors declare no competing interests.

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 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).
- 19 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).
- 20 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).
- 21 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).
- 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 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).
- 24 Ge, S. *et al.* A proteomic landscape of diffuse-type gastric cancer. *Nat Commun* **9**, 1012, doi:10.1038/s41467-018-03121-2 (2018).
- 25 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).
- 26 Schwanhaussner, B. *et al.* Global quantification of mammalian gene expression control. *Nature* **473**, 337-342, doi:10.1038/nature10098 (2011).

- 27 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).
- 28 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).
- 29 Abuzeid, W. M. *et al.* Molecular disruption of RAD50 sensitizes human tumor cells to cisplatin-based chemotherapy. *J Clin Invest* **119**, 1974-1985, doi:10.1172/Jci33816 (2009).
- 30 Kang, C. *et al.* The DNA damage response induces inflammation and senescence by inhibiting autophagy of GATA4. *Science* **349**, doi:ARTN aaa561210.1126/science.aaa5612 (2015).
- 31 Racker, E. & Spector, M. Warburg Effect Revisited - Merger of Biochemistry and Molecular-Biology. *Science* **213**, 303-307, doi:DOI 10.1126/science.6264596 (1981).
- 32 Liu, L. *et al.* PDHA1 Gene Knockout In Human Esophageal Squamous Cancer Cells Resulted In Greater Warburg Effect And Aggressive Features In Vitro And In Vivo. *Oncotargets Ther* **12**, 9899-9913, doi:10.2147/Ott.S226851 (2019).
- 33 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).
- 34 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).
- 35 Chiba, T., Marusawa, H. & Ushijima, T. Inflammation-Associated Cancer Development in Digestive Organs: Mechanisms and Roles for Genetic and Epigenetic Modulation. *Gastroenterology* **143**, 550-563, doi:10.1053/j.gastro.2012.07.009 (2012).
- 36 Ji, Y. *et al.* Serum thymidine kinase 1 levels correlate with clinical characteristics of esophageal squamous cell carcinoma. *Int J Clin Exp Med* **8**, 12850-12857 (2015).
- 37 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).
- 38 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).
- 39 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).
- 40 Nissan, M. H. *et al.* Loss of NF1 in Cutaneous Melanoma Is Associated with RAS Activation and MEK Dependence. *Cancer Res* **74**, 2340-2350, doi:10.1158/0008-5472.Can-13-2625 (2014).
- 41 Sanchez-Vega, F. *et al.* Oncogenic Signaling Pathways in The Cancer Genome Atlas. *Cell* **173**, 321-+, doi:10.1016/j.cell.2018.03.035 (2018).
- 42 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).
- 43 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).
- 44 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).
- 45 Kalluri, R. EMT: When epithelial cells decide to become mesenchymal-like cells. *J Clin Invest* **119**, 1417-1419, doi:10.1172/Jci39675 (2009).
- 46 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).
- 47 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).
- 48 Akizuki, S. *et al.* PLD4 is a genetic determinant to systemic lupus erythematosus and involved in murine autoimmune phenotypes. *Ann Rheum Dis* **78**, 509-518, doi:10.1136/annrheumdis-2018-214116 (2019).
- 49 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).
- 50 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).

- 51 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).
- 52 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).
- 53 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).
- 54 Kojima, T. & Doi, T. Immunotherapy for Esophageal Squamous Cell Carcinoma. *Curr Oncol Rep* **19**, doi:ARTN 3310.1007/s11912-017-0590-9 (2017).
- 55 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).
- 56 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).
- 57 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).
- 58 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).
- 59 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).
- 60 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).
- 61 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).
- 62 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).
- 63 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).
- 64 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).
- 65 Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-1760, doi:10.1093/bioinformatics/btp324 (2009).
- 66 Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079, doi:10.1093/bioinformatics/btp352 (2009).
- 67 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).
- 68 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).
- 69 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).
- 70 Huang, D. W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4**, 44-57, doi:10.1038/nprot.2008.211 (2009).
- 71 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).
- 72 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).
- 73 Liberzon, A. *et al.* The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst* **1**, 417-425, doi:10.1016/j.cels.2015.12.004 (2015).
- 74 Croft, D. *et al.* The Reactome pathway knowledgebase. *Nucleic Acids Res* **42**, D472-477, doi:10.1093/nar/gkt1102 (2014).
- 75 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).
- 76 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).
- 77 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).
- 78 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).
- 79 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).
- 80 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).
- 81 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).
- 82 Hornbeck, P. V. *et al.* PhosphoSitePlus, 2014: mutations, PTMs and recalibrations. *Nucleic Acids Res* **43**, D512-D520, doi:10.1093/nar/gku1267 (2015).
- 83 Horn, H. *et al.* KinomeXplorer: an integrated platform for kinome biology studies. *Nat Methods* **11**, 603-604, doi:10.1038/nmeth.2968 (2014).
- 84 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).

Figure Legends

Fig. 1 | 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 in ESCC progression. Top: the mutation number and types 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 percentages of somatic SNVs in the Fudan cohort and other ESCC cohorts. **e**, Bar plot of mutation frequencies for genes among the Fudan cohort and other ESCC cohorts (Fisher's exact test). *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$. **f**, Overview of the phosphoproteomic profiles in ESCC progression. The cumulative number (left) and the identified number (right) of phosphosites in ESCC progression.

See also **Supplementary Fig. 1–3**, and **Supplementary Table 1–6**.

Fig. 2 | Integrative omics analyses of early ESCC samples. **a**, The percentages of somatic SNVs in the male (top) and female patients (bottom) in ESCC (t-test). **b**, Analysis of mutation spectra in different ages patients of ESCC (t-test). Top: analysis of mutation spectra of all samples for WES. Bottom: the percentages of SNVs among different ages patients of ESCC. **c**, 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 axe, respectively. **d**, The impacts of SCNAs on the biological pathways. **e**, Clustergrams of mutually exclusive mutated genes in cell cycle (left) and DNA repair (right) (Bartlett's test). **f**, Heatmap showing the direct effect of mutations on the expression levels of their corresponding proteins (t-test). **g**, Heatmap showing the impacts of the mutations of driver genes in ESCC (top) and the impacts of *TP53* mutation on *ERBB2* and *TK1* expression levels (bottom) (t-test). **h**, The impacts of *TP53* mutation on the dominant pathways (t-test). *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.

See also **Supplementary Fig. 4–5**, and **Supplementary Table 2, 4**, and **7**.

Fig. 3 | Proteomics characteristics of ESCC progression with *ATM*, *EP400*, and *TP53* mutations. **a**, Principal component analysis (PCA) of the Fudan cohort. Left: PCA of all 756 ESCC samples (T0 to T3); Right: PCA of 746 early ESCC samples (T0 to T1). **b**, 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. **c**, Volcano analysis of DEPs (left) and their associated biological pathways (right) in the two clusters (t-test). Biological pathways were analyzed from the Reactome database. **d**, Overlap of the two clusters for mutations and identified proteins. Left: the percentages of 53 samples for WES in ESCC progression. Right: overlap of mutations and identified proteins of the two clusters. **e**, The impacts of C2 high/specific mutations on their corresponding protein expressions (\log_2 FC (C2/C1)) (left), and the survival analysis (log-rank test) of patients with *ATM* and *EP400* mutations (right). The survival information was downloaded from other published ESCC cohorts (*Genta et al., 2016; Gao et al., 2014*). Heatmap analysis of the impacts of *ATM*, *EP400*, and *TP53* mutations (Fisher's exact test) on **f**, immune response, protein metabolism, and glycolysis (t-test); and **g**, canonical cancer-related pathways. *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$. **h**, Survival analysis (log-rank test) of patients with co-mutations of *ATM* & *TP53*, and *EP400* & *TP53*. The survival information was downloaded from other published ESCC cohorts (*Genta et al., 2016; Gao et al., 2014*). **i**, Representative pathways of the four groups. **j**, Summary of the biological pathways of the four groups.

See also **Supplementary Fig. 6–8**, and **Supplementary Table 8**.

Fig. 4 | The temporal driver pathway waves in ESCC progression. **a**, 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. **b**, The co-mutations and mutually exclusive mutations in ESCC progression (Fisher's exact test). *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$. **c**, The impacts of the co-mutations or mutually exclusive mutations (top) on the proteome level (bottom) in ESCC progression. **d**, The phosphorylation of representative proteins in the dynamic driver pathway waves in ESCC progression. **e**, A carcinogenesis path with 8 dynamic waves in ESCC progression at the multi-omics level.

See also **Supplementary Fig. 9–11**, and **Supplementary Table 9–10**.

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**, The track enhanced proteins in the dominant pathways of the 6 tracks. **e**, Representative phosphoproteins and their associated pathways of diverse tracks. **f**, The impacts of prominent mutations in diverse tracks on the expression of their corresponding proteins. The mutation frequencies and corresponding protein expressions of tracks were shown on the right. **g**, 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, *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.

See also **Supplementary Fig. 12–13**, and **Supplementary Table 11**.

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). Error bars represented mean \pm SEM. *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$. **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). **f**, *CDK2* mutation had positive effects on PGK1 expression (t-test). *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$. **g**, The impacts of *CDK2* mutation on the substrates expression of the kinases, associated with the PGK1 motif (sP). **h**, The expression of the kinases and the substrates in ESCC progression at phosphoproteome level. (I) A brief regulatory network of kinase-substrate, associated with the motif of PGK1 (sP).

See also **Supplementary Table 12**.

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, *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, ns > 0.05 .

See also **Supplementary Fig. 14**, and **Supplementary Table 13**.

Supplemental Figure Legends

Supplementary Fig. 1 | Hematoxylin and eosin (H&E) staining on Early ESCCs, related to Fig. 1.

a, Invasion layer in ESCC progression. T0–T1 indicated the early ESCC stages, ranging from substage 1 to substage 7_3. T2–T3 represented the progressive ESCC stages, including stages 8 and 9. Diversity sign was marked the distinctive invasion layer on the right. **b**, H&E staining analysis of different substages in early ESCC.

Supplementary Fig. 2 | A brief workflow of multi-omics analysis and quality assessment of proteomic data, related to Fig. 1.

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**, Correlation matrix of 114 (sub)stage 1 proteomes (Spearman's correlation coefficients). **c**, (Spearman's) correlation analysis of 42 HEK293T cell samples as MS quality control to evaluate the robustness of label-free quantification.

Supplementary Fig. 3 | Overview of the proteomic profiles of early and progressive ESCC cohorts, related to Fig. 1.

a, Cumulative protein number of 756 samples. The numbers of samples and cumulative identifications were represented on the x and y axes, respectively. **b**, The identified number of proteins in the 22 substages. Error bars represented mean \pm SEM. **c**, The identified protein numbers in the 756 samples. **d**, The dynamics of protein abundance identified in 22 substages. Proteins were quantified as normalized iBAQ value and transformed to log₁₀ FOT. The highest- and lowest- abundance proteins were shown in the box. **e**, Distribution of log₁₀-transformed FOT of identified proteins abundance in

756 samples that passed quality control. The cumulative number **f**, and the identified number **g**, of phosphoproteins in ESCC progression.

Supplementary Fig. 4 | Genomic characteristics of early ESCC cohort, related to Fig. 2. **a**, A brief workflow for WES. The numbers of samples and stages were shown in the box. **b**, The percentages of 6 somatic SNV types. **c**, Analysis of mutation spectra of all 53 samples for WES. **d**, The percentages of somatic SNVs in all samples (left) and in the ESCC non-drinking/smoking (right & top) patients and drinking/smoking (right & bottom) patients (t-test). *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$. **e**, The summary of the impacts of SCNAs on their counterpart protein expressions in ESCC.

Supplementary Fig. 5 | Clustergrams of 100% mutated genes and the associated biological pathways at all stages in ESCC progression, related to Fig. 2.

Supplementary Fig. 6 | Consensus clustering of 756 samples, related to Fig. 3. Heatmap of consensus matrix with $K =$ **a**, 2, **b**, 3, **c**, 4, **d**, 5, **e**, 6, **f**, 7, **g**, 8, **h**, 9, and **i**, 10. The input was the quantile-normalized iBAQ intensity (FOT) matrix of E3 data ($n = 6,687$). **j**, Consensus CDF. Consensus index and CDF were shown on x and y axes, respectively. **k**, Delta area. The K value and the relative change in area under the CDF curve were indicated on the x and y axes, respectively.

Supplementary Fig. 7 | The expression of specific molecules of esophageal tissue in ESCC, related to Fig. 3. The expression of **a**, GBP6, **b**, TGM1, **c**, TGM3, and **d**, S100A14 in two clusters (t-test). The expression of **e**, GBP6, **f**, TGM1, **g**, TGM3 and **h**, S100A14 in ESCC progression (Kruskal-Wallis test). Two clusters/Twenty substages and FOT were indicated on x and y axes, respectively. Error bars represented mean \pm SEM, **** $p < 0.0001$, *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.

Supplementary Fig. 8 | The impacts of *ATM*, *EP400*, and *TP53* mutations on the dominant pathways in ESCC progression, related to Fig. 3. **a**, The expression of the proteins in the dominant pathways affected by *ATM*, *EP400*, and *TP53* mutations in ESCC progression. The expression trend of the proteins in the negatively (decreasingly) **b**, and positively (increasingly) **c**, associated pathways affected by *ATM*, *EP400*, and *TP53* mutations in ESCC progression. Error bars represented mean \pm SEM.

Survival analysis of *TP53* mutation (log-rank test). *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$. **d**, The survival information was downloaded from other published ESCC cohorts (Gao et al., 2014, Genta et al., 2016).

Supplementary Fig. 9 | Immune-based panels of early ESCCs, related to Fig. 4. **a**, Cell signatures of 756 early ESCCs by xCell (Kruskal-Wallis test). *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$. **b**, The enhanced immune, stroma score and microenvironment score in ESCC progression. The samples and scores (Z-score) were indicated on x and y axes, respectively.

Supplementary Fig. 10 | The dynamic driver pathway waves of 8 panels in ESCC progression, related to Fig. 4. **a**, The dominant pathways of 8 panels in ESCC progression. The results of the k-means analysis were shown. **b**, A carcinogenesis path with 8 dynamic waves in ESCC progression.

Supplementary Fig. 11 | The expression trend of ESCC biomarkers in ESCC progression, related to Fig. 4. The expression of ESCC biomarkers from **a**, Yazdian-Robati et al.²⁷ and **b**, Pawar et al.²⁸ in ESCC progression. Twenty substages and log₁₀ (FOT) were indicated on x and y axes, respectively. Error bars represented mean ± SEM.

Supplementary Fig. 12 | The trajectory analysis of the early ESCC cohort, related to Fig. 5. **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 R² (left) and K value (right). **c**, The representative track enhanced proteins in the dominant pathways of 6 tracks. Twenty substages and FOT were indicated on x and y axes, respectively.

Supplementary Fig. 13 | The descending of epithelial cell differentiation and keratinization in the 6 tracks, Related to Fig. 5. **a**, The descending of the representation proteins in epithelial cell differentiation and keratinization in the 6 tracks. Twenty substages and FOT were indicated on x and y axes, respectively. **b**, Heatmap analysis of the impacts of tracks prominent mutations on the proteome level (Z-score) (two-way ANOVA test). Top: the prominent mutations in tracks. Bottom: heatmap analysis of the enhanced proteins in track dominant pathways. **c**, Analysis of somatic SNVs in diverse

tracks (t-test). Top: the percentages of all samples in tracks. Bottom: the somatic SNVs were indicated by a bar plot. *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$. **d**, Fishplot analysis performed the kinetic driver pathway waves of the 6 tracks.

Supplementary Fig. 14 | The impacts of GAPDH-OE or PGM1-OE on ESCC cell proliferation

and inhibiting PGK1 by gemcitabine attenuates cell proliferation, related to Fig. 7. a, 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. **b**, *In vitro* assay showing the effects of gemcitabine on inhibition of PGK1 activity (IC50: 16.3 nM). **c**, Gemcitabine inhibited cell proliferation. **d**, Gemcitabine decreased PGK1 mediated metabolic flux (t-test). *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, ns > 0.05 .

METHODS

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/>)⁶⁴, 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>.

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 in 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 tumor 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 8th edition of the American Joint Committee on Cancer (AJCC) TNM staging system. The hematoxylin and eosin (H&E)-stained sections were reviewed evaluated by two or three experienced gastrointestinal pathologists. The pathological characteristics were shown in **Supplementary Fig. 1**. 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), severe dysplasia-m1 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, namely sm a (6_1 (n = 5), 6_2 (n = 5), and 6_3 (n = 7)), and sm b (7_1 (n = 12), 7_2 (n = 9), and 7_3 (n = 11)). Finally, a total of 756 samples were collected and prepared for Fudan ESCC cohort.

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: 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). Additionally, only 53 samples (from 11 ESCC patients) covering 20 substages: substage 1 (n = 5), substage 2_1 (n = 7), substage 2_2 (n = 3), substage 3_1 (n = 1), substage 3_2 (n = 3), substage 3_3 (n = 4), substage 3_5 (n = 4), substage 4_1 (n = 5), substage 5_1 (n = 2), substage 5_3 (n = 1), substage 6_1 (n = 1), substage 6_2 (n = 2), substage 6_3 (n = 1), substage 7_1 (n = 4), substage 7_2 (n = 2), substage 7_3 (n = 4), (sub)stage 8 (n = 2), and (sub)stage 9 (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 a 100× target depth and 12G volume of most raw data (> 90%). 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) software⁶⁵ 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, SAMtools⁶⁶ 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 Software⁶⁷ was used for identifying the Somatic SNV sites, whereas Strelka⁶⁸ 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.

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^{12,23,33,69}. Here, the novel mutation represented the mutations appearing at a certain stage, which has not been identified in earlier stages. For example, if *TP53* was not mutated in stage 2 but mutated in stage 3, *TP53* was considered the novel mutation in stage 3.

Impacts of the detected mutations on the proteome level

Impacts of *TP53* mutation

To compare the impact of mutations on proteins across the 53 ESCC samples (n (early stages) = 49, n (advanced stages) = 4), we focused on mutations and global proteome with 11,877 qualified mutations (mutation frequency (all samples) > 5%, and mutation frequency (advanced stages) > 25%). To compare the impacts of *TP53* mutation on the proteome 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. Then, the DEPs of the two groups were enriched into

the GO/KEGG database.

Impacts of *ATM*, *EP400*, and *TP53* mutations

All the samples with *ATM*, *EP400*, and *TP53* mutations were divided into four groups—*TP53* Mut only, *ATM* & *TP53* co-Mut, *EP400* & *TP53* co-Mut, and *ATM* & *EP400* & *TP53* co-Mut. Then, the DEPs of the four groups were subjected to the GO/KEGG pathway enrichment analysis in DAVID⁷⁰. To select the biological pathways of each group, we used the p value < 0.05 for differential behavior and the differential score (Q) was obtained from assigned $-\log_{10}(p\text{-value})$.

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.

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 (1,159 mutations/proteins) of stage 3 (Tis stage), which was the key event in esophageal carcinogenesis. Only genes or proteins with < 50% NAs (missing values) were considered for the analysis, and protein IDs were mapped to gene names. Then, the positively and negatively correlated genes ($n = 1,159$) were subjected to the GO/KEGG database in DAVID ($p < 0.05$, $FDR < 0.3$)⁷⁰ to investigate the impacts of SCNAs on the proteome level. The consequences of SCNAs on the proteome level were estimated from the mutation frequency (SCNA) and the \log_2 FC (Mut/WT).

Analysis of mutation spectra

To compare the somatic SNVs of all samples in different cohorts, tracks, and risk factors, 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-µm inner diameter, 3 µm ReproSil-Pur C18-AQ beads, Dr. Maisch GmbH) using Solvent A and separated on a 150-µm-inner-diameter column with a length of 15 cm (1.9 µm ReproSil-Pur C18-AQ beads, Dr. Maisch GmbH) over a 75 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 (0-75 min, 0 min, 4% B; 0-10 min, 4-15% B; 10-60 min, 15-30% B; 60-69 min, 30-50% B; 69-70 min, 50-100% B; 70-75 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 μ 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 μ L of binding/wash buffer and centrifuged at $1,000 \times g$ for 30 s for 3 times and washed by 200 μ L of LC-MS grade water for one additional time. The phosphopeptide was eluted by adding 100 μ 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- μ 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 ± 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 (**Supplementary Table 5**): 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 algorithm⁷¹; E2 (9,741 GPs): we excluded keratins and proteins whose maximum FOT in all 756 experiments were less than 10E5 in FOT, and all the proteins were required to have at least 1 unique strict peptides; E3 (6,687 GPs): GPs identified in more than one-sixth samples of each substage.

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. 2b, c (Supplementary Table 6)**. 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 SAM³⁸ 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 ≤ 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 ($5.0E-5$). When comparing the DEPs of 22 substages, we focused on the average of protein

abundance at each substage. The highest one was considered the substage specific protein and was used to elucidate the molecule dynamics in ESCC progression.

In differential analysis of proteins in ESCC progression (gradually decreased or increased) at the proteome 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).

Pathways enrichment analysis

To investigate the dominant signaling pathways of two clusters, six tracks and 22 substages, we used gene sets of molecular pathways from KEGG⁷² /Hallmark⁷³ /Reactome⁷⁴ 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.

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)⁷⁵ 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. 6**, 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. 3b**).

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>)¹².

The estimate of stromal and immune scores

Tumor purity, immune, stromal scores were inferred by R package ESTIMATE version 1.0.11⁷⁶ 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>)⁴⁷. 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.

In the analysis of diverse tracks at the phosphoproteome level, the highly expressed phosphoproteins were used, and the statistical analysis was performed with two-tailed Student's test (Fig. 5e). In Fig. 5f, the impacts of track prominent mutations on their counterpart protein expressions were performed to investigate the features of tracks with two-tailed Student's test and fold changes (certain stage/other stages) of FOT. In Supplementary Fig. 13b, the statistical analysis was performed with two-way ANOVA test in GraphPad Prism (version 7.0).

Fishplot analysis of six tracks

Fishplot (version 0.5) analysis was performed to present the dynamic driver pathways waves of 6 tracks (Supplementary Fig. 13d), 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 miscleavages 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) algorithm⁷⁷. The information of kinase-substrate relationships was obtained from publicly available databases, including PhosphoSite⁷⁸, Phospho.ELM⁷⁹ and PhosphoPOINT⁸⁰. The information of substrate motifs was obtained either from the literature⁵⁶ or from an analysis of the KSEA dataset with Motif (sP)⁸¹. 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) and the kinase-substrate-motif network

analysis was referenced from PhosphoSitePlus (PSP, <https://www.phosphosite.org/homeAction>)⁸² and NetworkKIN 3.0⁸³. 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- μ 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).

Metabolites quantification

The pyruvate, lactate, citrate, succinate, fumarate, and glycine levels were measured using NMR spectra as described before⁸⁴. 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°C , after which the supernatants were collected. The supernatants were then lyophilized and reconstituted in 500 μL 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 μm , 50×2.0 mm; Phenomenex, Torrance, CA). The mobile phase comprised eluent A (0.77 g NH_4OAc , 1.25 mL NH_4OH , 25 mL ACN, and 300 μL 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°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°C in 5% CO₂.

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 MgCl₂ and 5% glycerol. PGK1 activity was measured using purified Flag-tagged PGK1 (0.2 μ g/mL) mixed with DMSO or different concentrations of gemcitabine in the reaction buffer containing 50 mM Tris-HCl (pH 7.6), 8 mM MgCl₂, 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 4×10^3 cells per well

and allowed to adhere. Cell Counting Kit-8 solution (10 μ L) was added to each well, and the cells were cultured in 5% CO₂ 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 mm³, 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 Fisher's exact test, Bartlett's test, two-way ANOVA test, and log-rank test were performed in GraphPad Prism (version 7.0), and 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.

Figures

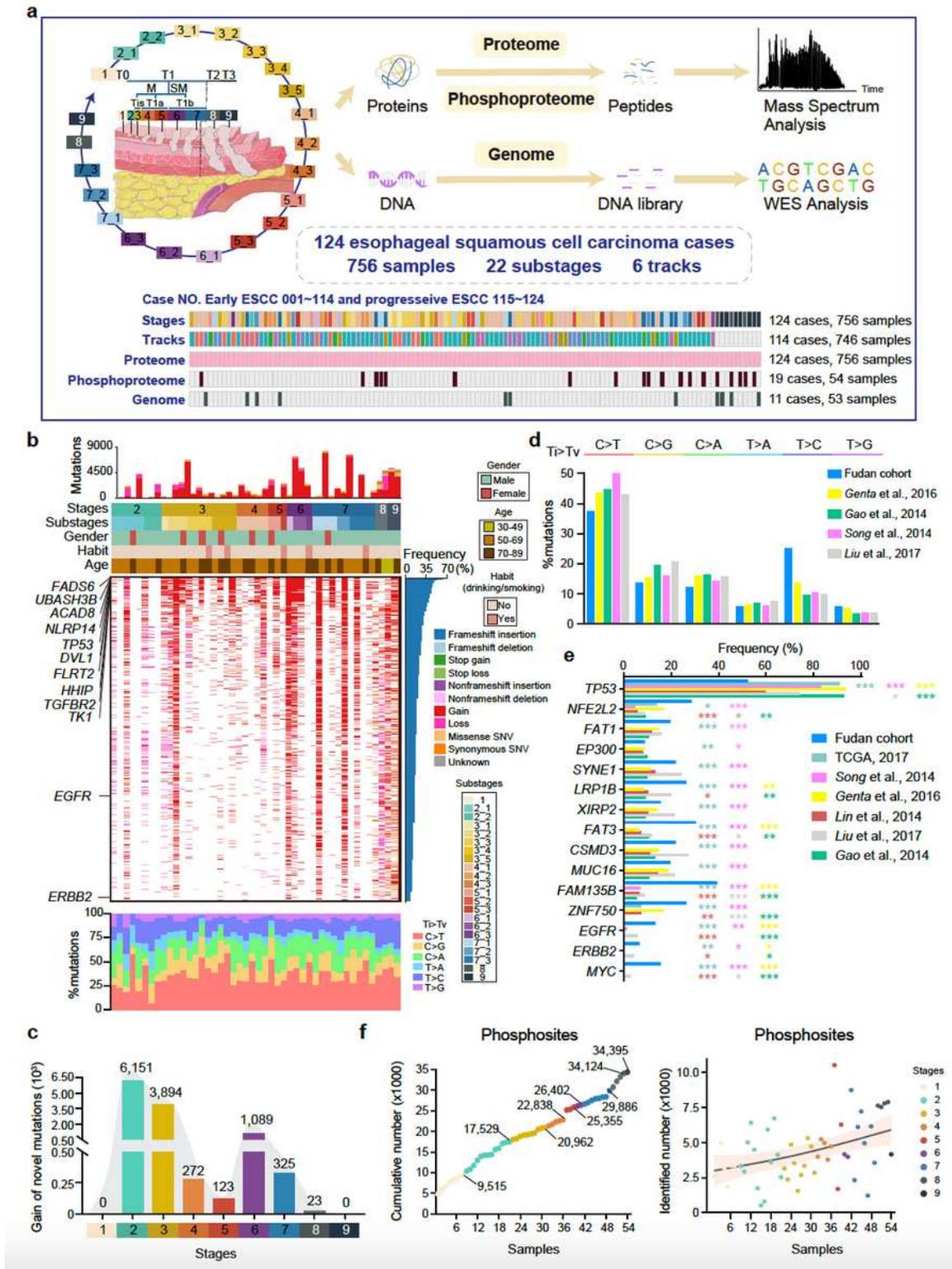


Figure 1

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 in ESCC progression. Top: the mutation number and types 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 percentages of somatic SNVs in the Fudan cohort and other ESCC cohorts. e, Bar plot of mutation frequencies for genes among the Fudan cohort and other ESCC cohorts (Fisher's exact test). *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$. f, Overview of the phosphoproteomic profiles in ESCC progression. The cumulative number (left) and the identified number (right) of phosphosites in ESCC progression. See also Supplementary Fig. 1–3, and Supplementary Table 1–6.

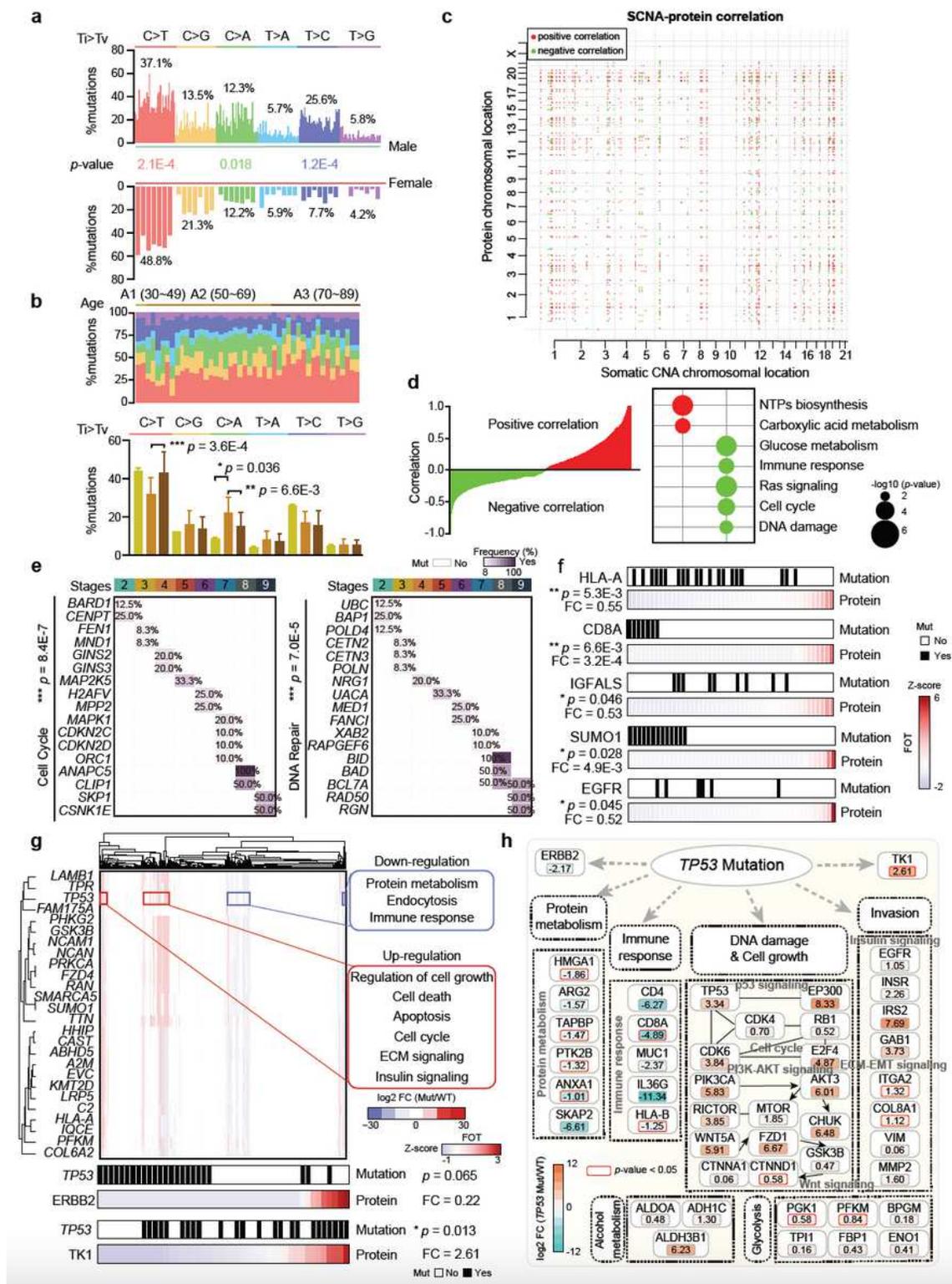


Figure 2

Integrative omics analyses of early ESCC samples. a, The percentages of somatic SNVs in the male (top) and female patients (bottom) in ESCC (t-test). b, Analysis of mutation spectra in different ages patients of ESCC (t-test). Top: analysis of mutation spectra of all samples for WES. Bottom: the percentages of SNVs among different ages patients of ESCC. c, 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. d, The impacts of SCNAs on the biological pathways. e, Clustergrams of mutually exclusive mutated genes in cell cycle (left) and DNA repair (right) (Bartlett's test). f, Heatmap showing the direct effect of mutations on the expression levels of their corresponding proteins (t-test). g, Heatmap showing the impacts of the mutations of driver genes in ESCC (top) and the impacts of TP53 mutation on ERBB2 and TK1 expression levels (bottom) (t-test). h, The impacts of TP53 mutation on the dominant pathways (t-test). *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$. See also Supplementary Fig. 4–5, and Supplementary Table 2, 4, and 7.

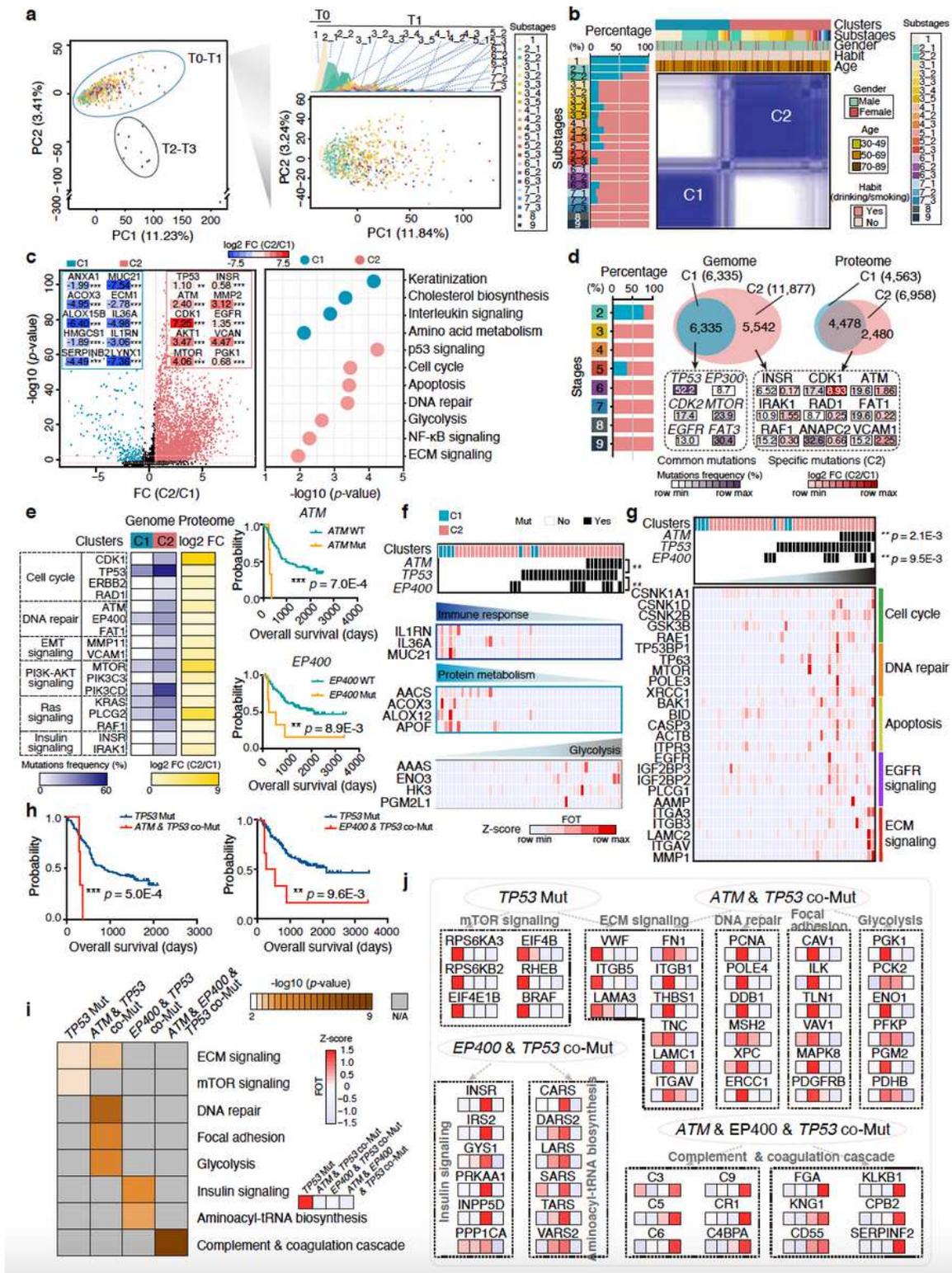


Figure 3

Proteomics characteristics of ESCC progression with ATM, EP400, and TP53 mutations. a, Principal component analysis (PCA) of the Fudan cohort. Left: PCA of all 756 ESCC samples (T0 to T3); Right: PCA of 746 early ESCC samples (T0 to T1). b, 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. c, Volcano analysis of DEPs (left) and their associated biological pathways

(right) in the two clusters (t-test). Biological pathways were analyzed from the Reactome database. d, Overlap of the two clusters for mutations and identified proteins. Left: the percentages of 53 samples for WES in ESCC progression. Right: overlap of mutations and identified proteins of the two clusters. e, The impacts of C2 high/specific mutations on their corresponding protein expressions (\log_2 FC (C2/C1)) (left), and the survival analysis (log-rank test) of patients with ATM and EP400 mutations (right). The survival information was downloaded from other published ESCC cohorts (Genta et al., 2016; Gao et al., 2014). Heatmap analysis of the impacts of ATM, EP400, and TP53 mutations (Fisher's exact test) on f, immune response, protein metabolism, and glycolysis (t-test); and g, canonical cancer related pathways. *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$. h, Survival analysis (log-rank test) of patients with co-mutations of ATM & TP53, and EP400 & TP53. The survival information was downloaded from other published ESCC cohorts (Genta et al., 2016; Gao et al., 2014). i, Representative pathways of the four groups. j, Summary of the biological pathways of the four groups. See also Supplementary Fig. 6–8, and Supplementary Table 8.

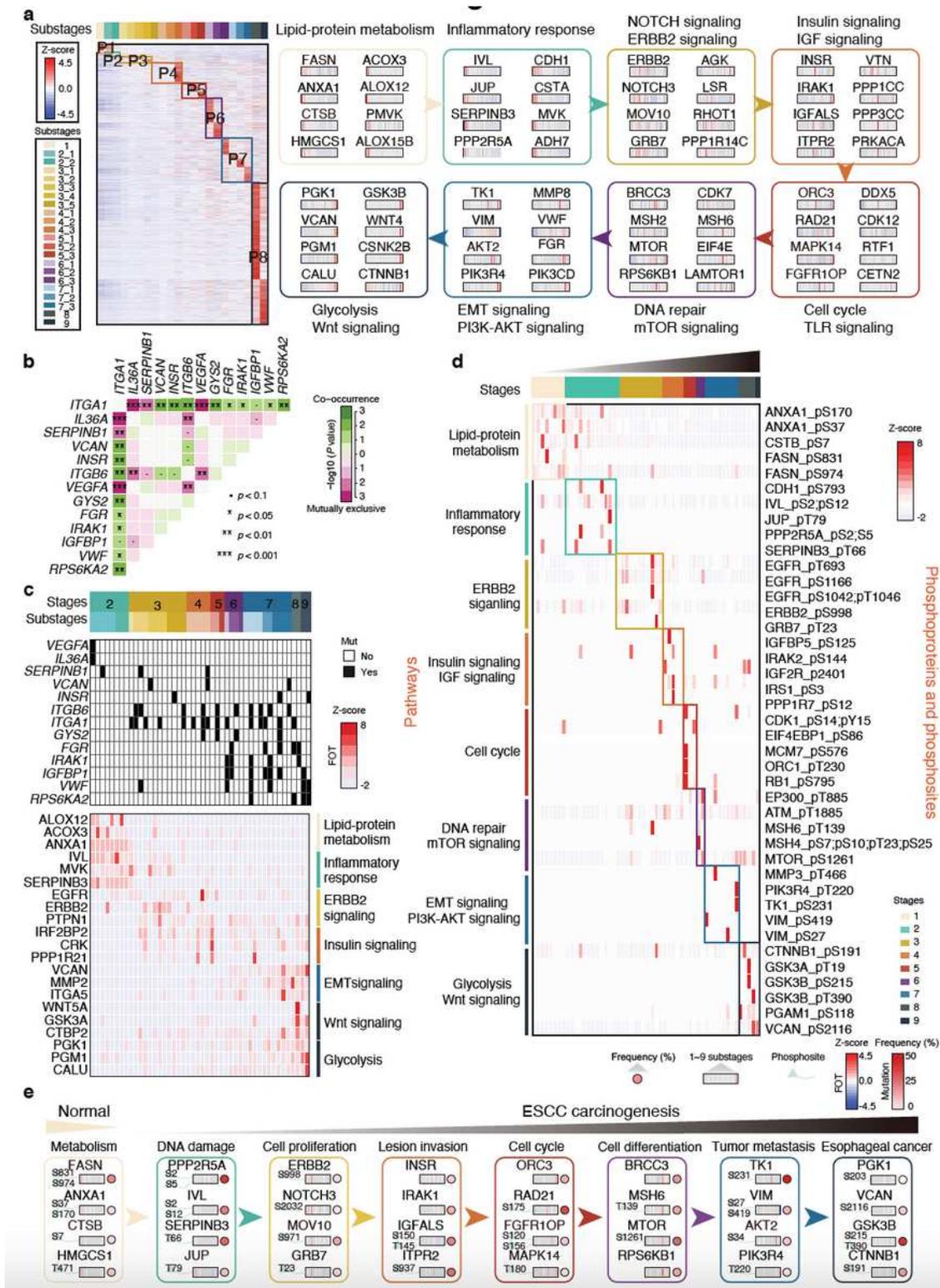


Figure 4

The temporal driver pathway waves in ESCC progression. a, 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. b, The co-mutations and mutually exclusive mutations in ESCC progression (Fisher's exact test). *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$. c, The impacts of the co-mutations or mutually exclusive mutations (top) on the proteome level

(bottom) in ESCC progression. d, The phosphorylation of representative proteins in the dynamic driver pathway waves in ESCC progression. e, A carcinogenesis path with 8 dynamic waves in ESCC progression at the multi-omics level. See also Supplementary Fig. 9–11, and Supplementary Table 9–10.

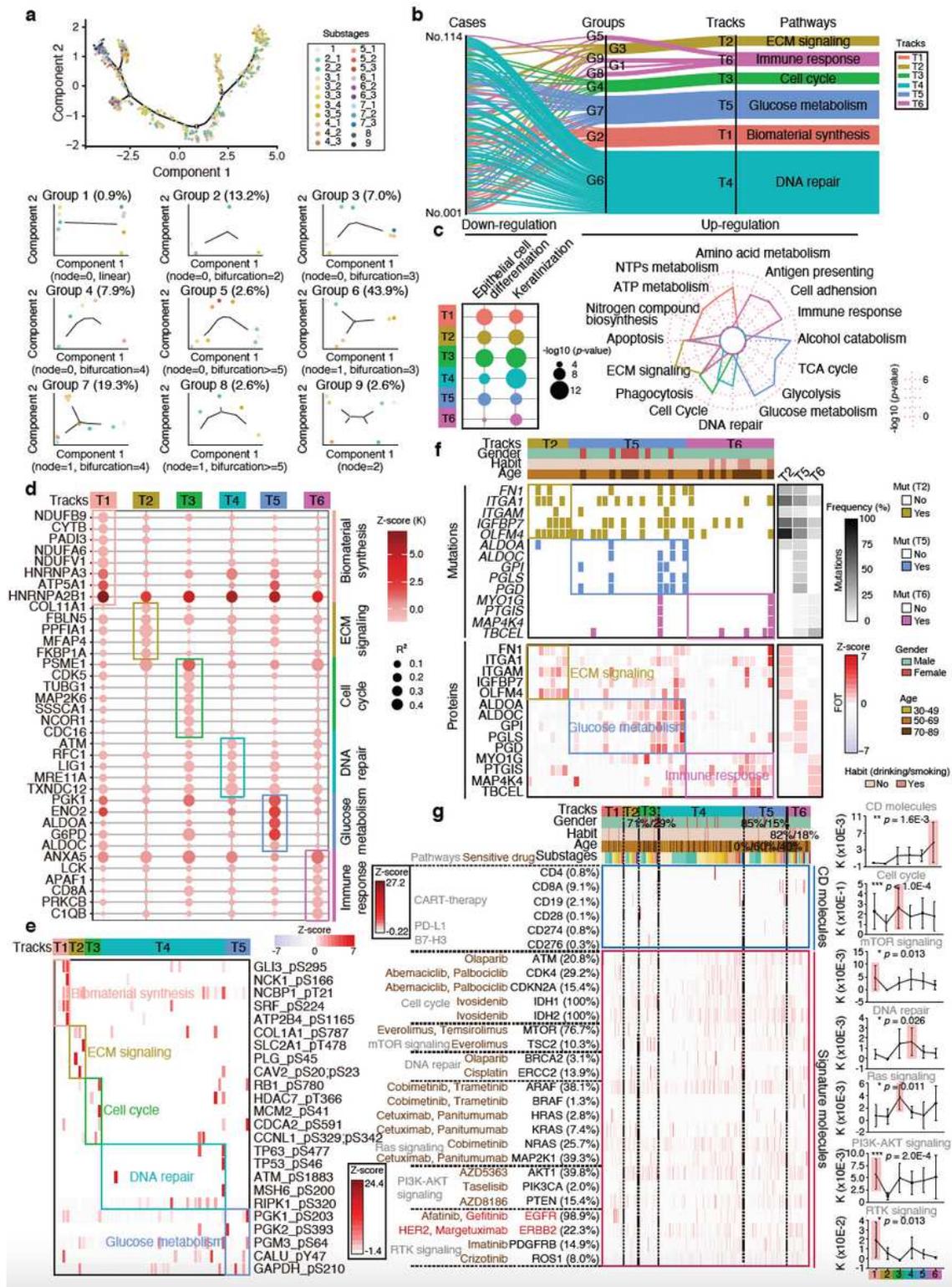


Figure 5

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, The track enhanced proteins in the dominant pathways of the 6 tracks. e, Representative phosphoproteins and their associated pathways of diverse tracks. f, The impacts of prominent mutations in diverse tracks on the expression of their corresponding proteins. The mutation frequencies and corresponding protein expressions of tracks were shown on the right. g, The expression of representative CD molecules and drug targets approved by the FDA in the 6 tracks (twoway ANOVA test). Error bars represented mean \pm SEM, *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$. See also Supplementary Fig. 12–13, and Supplementary Table 11.

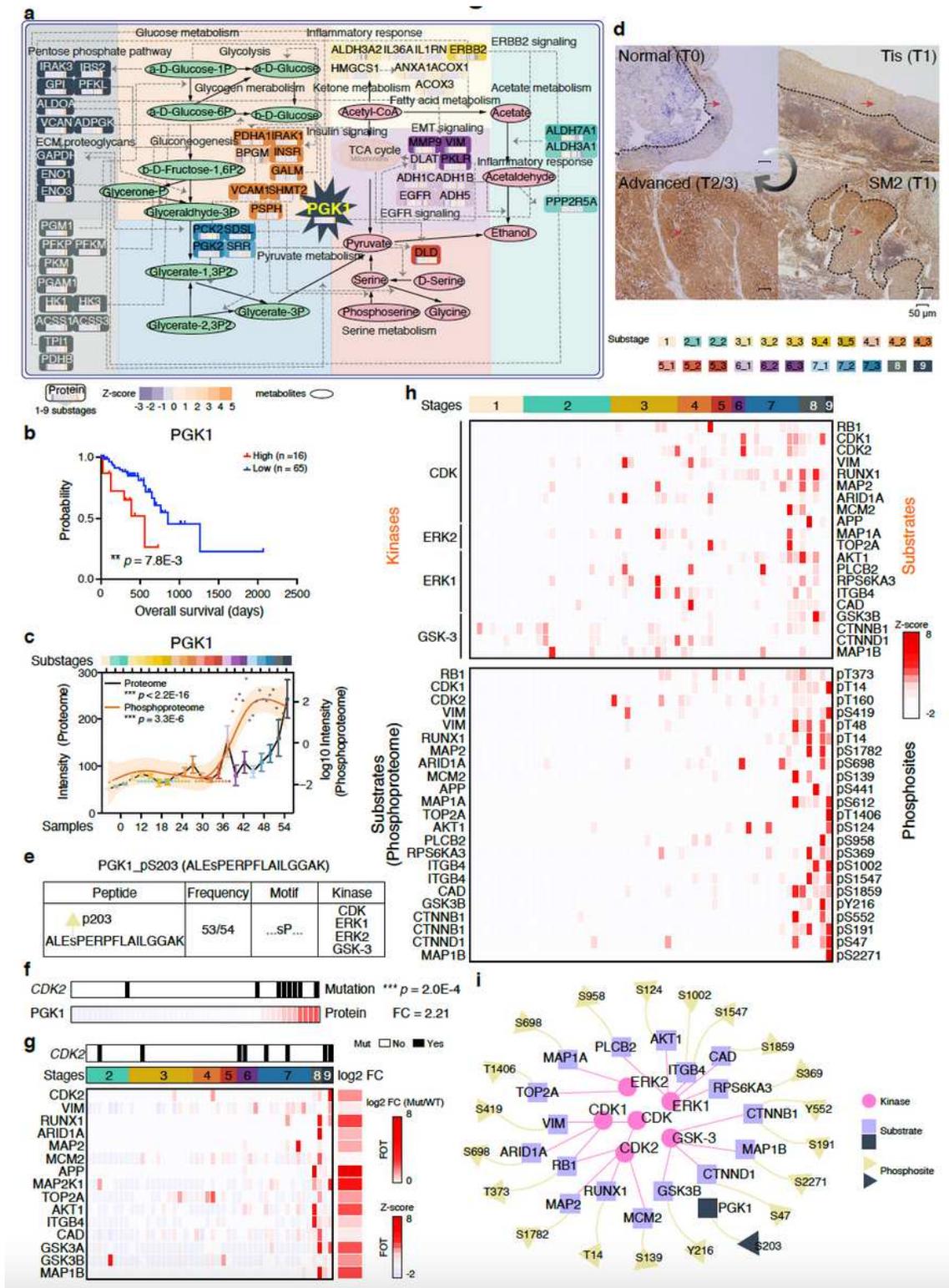


Figure 6

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). Error bars represented mean \pm SEM. *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$. 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). f, CDK2 mutation had positive effects on PGK1 expression (t-test). *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$. g, The impacts of CDK2 mutation on the substrates expression of the kinases, associated with the PGK1 motif (sP). h, The expression of the kinases and the substrates in ESCC progression at phosphoproteome level. (i) A brief regulatory network of kinase-substrate, associated with the motif of PGK1 (sP). See also Supplementary Table 12.

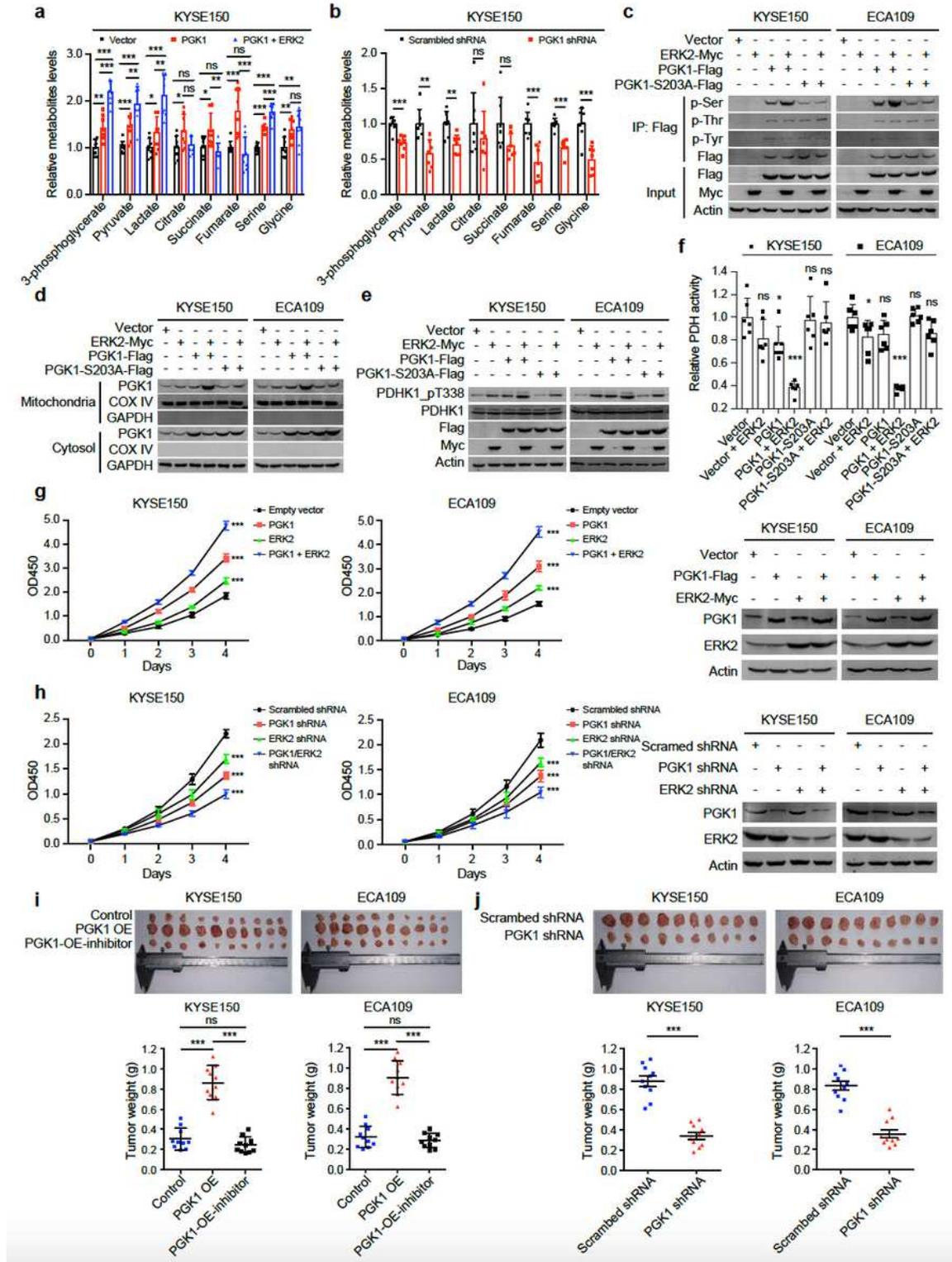


Figure 7

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, *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, ns > 0.05 . See also Supplementary Fig. 14, and Supplementary Table 13.

Supplementary Files

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

- [SupplementaryFigure.pdf](#)