

A proteogenomic analysis of cervical cancer reveals therapeutic and biological insights

Received: 10 November 2023

Accepted: 21 October 2024

Published online: 22 November 2024

 Check for updates

Jing Yu^{1,2,16}, Xiuqi Gui^{3,16}, Yunhao Zou^{4,5,16}, Qian Liu^{6,16}, Zhicheng Yang^{5,6}, Jusheng An¹, Xuan Guo^{3,5}, Kaihua Wang^{4,5}, Jiaming Guo^{5,6}, Manni Huang¹, Shuhan Zhou^{3,5}, Jing Zuo¹, Yimin Chen^{5,6}, Lu Deng¹, Guangwen Yuan¹, Ning Li¹, Yan Song⁷, Jia Jia⁷, Jia Zeng¹, Yuxi Zhao¹, Xianming Liu⁸, Xiaoxian Du⁸, Yansheng Liu⁹, Pei Wang¹⁰, Bing Zhang¹¹, Li Ding¹², Ana I. Robles¹³, Henry Rodriguez¹³, Hu Zhou^{5,6,14} , Zhen Shao^{3,5} , Lingying Wu¹  & Daming Gao^{4,5,15} 

Although the incidence of cervical cancer (CC) has been reduced in high-income countries due to human papillomavirus (HPV) vaccination and screening strategies, it remains a significant public health issue that poses a threat to women's health in low-income countries. Here, we perform a comprehensive proteogenomic profiling of CC tumors obtained from 139 Chinese women. Integrated proteogenomic analysis links genetic aberrations to downstream pathogenesis-related pathways and reveals the landscape of HPV-associated multi-omic changes. EP300 is found to enhance the acetylation of FOSL2-K222, consequently accelerating the malignant proliferation of CC cells. Proteomic stratification identifies three patient subgroups with distinct features in prognosis, genetic alterations, immune infiltration, and post-translational modification regulations. PRKCB is further identified as a potential radioresponse-related biomarker of CC patients. This study provides a valuable public resource for researchers and clinicians to delve into the molecular basis of CC, to identify potential treatments and to ultimately advance clinical practice.

Cervical cancer (CC) remains the fourth most common malignancy and the fourth leading cause of cancer mortality in women¹. Unlike numerous other cancers lacking direct etiological association with virus infection, persistent human papillomavirus (HPV) infection is the major cause of CC, and over 90% of CC patients are associated with HPV infection at the time of diagnosis². Although HPV vaccination and screening strategies have reduced the incidence of CC in high-income countries³, CC is still a major public health problem threatening women's health in low-income countries⁴. Patients with early-stage CC who are treated with radical surgery generally have a good prognosis. For locally advanced CC, cisplatin-based concurrent chemoradiotherapy is recommended as the standard treatment^{5–7}. However, about 30–50% of CC patients treated with standard care of

chemoradiotherapy experienced progression or recurrence within five years^{8–10}. Some patients do not benefit optimally from radiotherapy because of radio-resistance, and the prognosis for those with persistent, recurrent, or metastatic CC is poor. Therefore, exploring potential radio-resistance biomarkers and effective interventions to improve outcomes for locally advanced disease is pivotal. Investigating the molecular mechanism of CC development and its HPV association, and identifying candidate biomarkers or therapeutic targets could improve individualized therapy for CC, which would greatly improve patient survival and quality of life.

Previous genomic and transcriptomic studies revealed the genetic landscape of CC as well as patient stratification based on distinct molecular signatures^{11–15}. The study from The Cancer Genome Atlas

A full list of affiliations appears at the end of the paper.  e-mail: zhouhu@simm.ac.cn; shaozhen@sinh.ac.cn; wulingying@cscs.org.cn; dgao@sibcb.ac.cn

(TCGA) clustered 178 CC patients using mRNA, methylation, miRNA, and copy number data, and identified three subtypes: (i) Keratin-low squamous, (ii) Keratin-high squamous, and (iii) Adenocarcinoma-rich¹⁴. The CC multi-omics research by Fan et al. focused on exploring the functional consequences of HPV fusion transcripts on the biology and pathophysiology¹⁶. However, the previous studies either lacked or possessed only limited proteomic data. Consequently, a comprehensive characterization of the changes in the protein landscape within CC was absent. In contrast, recent research on various women's cancer types, including breast^{17,18}, endometrial^{19,20}, and ovarian cancers^{21,22}, conducted by the Clinical Proteomic Tumor Analysis Consortium (CPTAC), has demonstrated the indispensable role of integrating proteomic and post-translational modification (PTM) analyses with genomic strategies to reveal essential biological insights into complex diseases. This compelling precedent prompted our pursuit of proteogenomic analysis to delve deeper into the molecular mechanisms of CC, particularly in the context of HPV-induced carcinogenesis.

In this study, we performed integrated genomic, transcriptomic, proteomic, phosphoproteomic, and acetylproteomic analyses on Chinese CC patients. Our proteogenomic study discovered: 1) HPV-promoted phenotypic perturbations at multi-omics levels; 2) acetylation of FOSL2-K222 by EP300 accelerating malignant proliferation; 3) proteomic subgroups associated with distinct clinical and biological features; and 4) PRKCB as a potential radioresponse-related biomarker. The underlying data presented here serves as a valuable resource for advancing research in various aspects of CC, including biology, treatment strategies, and the exploration of potential therapeutic drugs.

Results

Proteogenomic landscape of a Chinese CC cohort

To comprehensively understand the molecular characteristics of CC in China and provide valuable biological insights to guide clinical treatments, we conducted a comprehensive analysis using whole exome sequencing (WES), RNA sequencing (RNA-seq), proteomics, phosphoproteomics, and acetylproteomics (Supplementary Fig. 1a). This analysis was performed on a total of 139 treatment-naïve tumor samples and 33 normal adjacent tissues (NATs) prospectively collected from the Peking Union Medical College (PUMC). Out of 139 patients, 112 were diagnosed with squamous cell carcinoma, 19 with adenocarcinoma, 2 with adenosquamous carcinoma, 4 with small cell neuroendocrine carcinoma of the cervix (NECC), and 2 with other mixed histology type (Supplementary Fig. 1b). Clinical parameters including tumor grade, stage, primary treatment modality, and progression-free survival (PFS) information were summarized in Supplementary Fig. 1c and Supplementary Data 1a, b. The analysis flow chart of this study is presented in Supplementary Fig. 1d.

A total of 31,047 non-silent point mutations and 1050 small insertions-deletions were detected in the WES data of the primary tumors of 134 CC patients (Supplementary Data 2a). The most frequently mutated genes in the PUMC cohort included *PIK3CA* (32.8%), *KMT2C* (18.7%), *KMT2D* (17.2%), *FAT1* (13.4%), *ARID1A* (11.2%), *FBXW7* (11.2%), *EP300* (8.2%), *TP53* (8.2%), *PTEN* (7.5%), and *ERBB3* (7.5%) (Fig. 1a). *PIK3CA* was the most frequently mutated gene in CC, as previously published^{14,15}. *KMT2C* and *KMT2D* were frequently mutated in this cohort, and also observed across various types of cancers^{23–25}. The mutation rate of *KRAS* was higher in cervical adenocarcinoma than in squamous cell carcinoma (25.0% vs 0%) in our cohort, consistent with the TCGA CC study (15.6% vs 2.9%)¹⁴ (Fig. 1b). In this study, 24,595 genes were quantified by RNA-seq; Isobaric tandem mass tags (TMT)-based proteomic analysis identified 9,600 proteins; Phosphoproteomic analysis identified 41,448 highly reliable phosphosites corresponding to 7721 phosphoproteins and acetylproteomics detected 5749 highly reliable acetylsites corresponding to 2456 acetylproteins

(Supplementary Data 3a–d). Besides, data-independent acquisition (DIA)-based proteomics identified 11,904 proteins and exhibited good consistency with TMT-proteomics data (Supplementary Data 3e, Supplementary Fig. 2a). Principal-component analysis (PCA) clearly discriminated tumors and NATs based on both TMT and DIA proteomics data, and no obvious batch effects were observed (Supplementary Fig. 2b). In addition, high data reproducibility and technical quality were demonstrated across the entire proteomics analysis (Supplementary Fig. 2c, d).

For 6089 genes whose mRNA and protein abundance were both measured across 132 CC samples, we observed an overall positive mRNA-protein correlation (81.7% showed significantly positive correlation at Benjamini-Hochberg (BH) adjusted $P < 0.05$) with a median Spearman's correlation coefficient of 0.40 (Supplementary Fig. 2e), which is similar to previous studies investigating ovarian cancer (median $r = 0.45$)²¹ and clear cell renal cell carcinoma (median $r = 0.44$)²⁶. Previous studies have revealed the intricate cellular heterogeneity of CC tissues^{27–29}. We leveraged transcriptomic data to estimate the immune and stromal infiltration scores³⁰ (Supplementary Data 1a). The immune infiltration scores were significantly higher in squamous cell carcinoma than in adenocarcinoma ($P = 2.2 \times 10^{-3}$, Supplementary Fig. 2f), consistent with the better prognosis observed in squamous cell carcinoma^{31,32}.

Multi-omics analysis of somatic copy number alternations

For 134 CC samples, somatic copy number alternations (SCNAs) were derived based on WES data. Overall, genome-wide focal alterations showed the most frequent gains in chromosomes 1q, 3q, 5p and focal losses in chromosome 2q, 3p, 11q, 13q (Supplementary Fig. 3a), consistent with previous reports^{12,14,15}. In addition to amplification of genes *CD274*, *PDCD1LG2*, *TP63* and *PIK3CA* reported in previous studies, we observed recurrent amplification events at *PIK3CB* (3q22.3, 55.97%), *CDK4* (12q14.1, 37.3%), *AKT1* (14q32.22, 39.55%), *AKT2* (19q13.2, 57.46%) (Supplementary Fig. 3b, Supplementary Data 4a), and also identified deletions of tumor suppressors such as *LATS1* (6q25.1, 15.67%), *ARID1B* (6q25.3, 14.93%), *BRCA2* (13q13.1, 23.88%) and *RBI* (13q14.2, 23.88%) ($|\log_2(\text{Tumor/Blood})| > 0.3$) (Supplementary Fig. 3c, Supplementary Data 4b).

The gains and losses of gene copies often *cis* or *trans*-regulate mRNA, protein, and phosphoprotein abundance, as shown by the diagonal and vertical patterns in Fig. 2a. The *cis* and *trans* association among copy number alternation (CNA)-protein (TMT and DIA) and CNA-phosphoprotein were more attenuated than that of CNA-RNA (Fig. 2a, Supplementary Fig. 3d, e). Furthermore, based on the TMT and DIA proteomic datasets, 1358 common attenuated proteins were identified, which were mainly enriched in biological processes such as mRNA processing, mRNA splicing, and ribosome biogenesis (Supplementary Fig. 3f–h). Thus, PTM may play a key role in determining protein half-lives, resulting in decreased correlation between gene dosage and protein abundance³³.

To identify hub CNA genes, we screened correlated CNA, mRNA, and protein levels across tumors and concordant protein-level changes between tumors and NATs as below. 178 genes were identified as copy number gain genes with *cis* effects (CNG-*Cis* genes, see Methods) and were enriched in proteasome, RNA transport, riboflavin metabolism, spliceosome, DNA replication and N-Glycan biosynthesis (Fig. 2b, Supplementary Data 5a), highlighting the functional impact of CNA on tumor proliferation and progression. It is noteworthy that tumors with positive lymph node exhibited higher average CNG-*Cis* protein abundances than tumors with negative lymph node (Fig. 2c). Meanwhile, 28 copy number loss genes with *cis* effects (CNL-*Cis* genes, see Methods) were identified and were involved in cell adhesion molecules and valine, leucine and isoleucine degradation pathways (Fig. 2d, Supplementary Data 5b). The average protein expression level of CNL-*Cis* genes was inversely

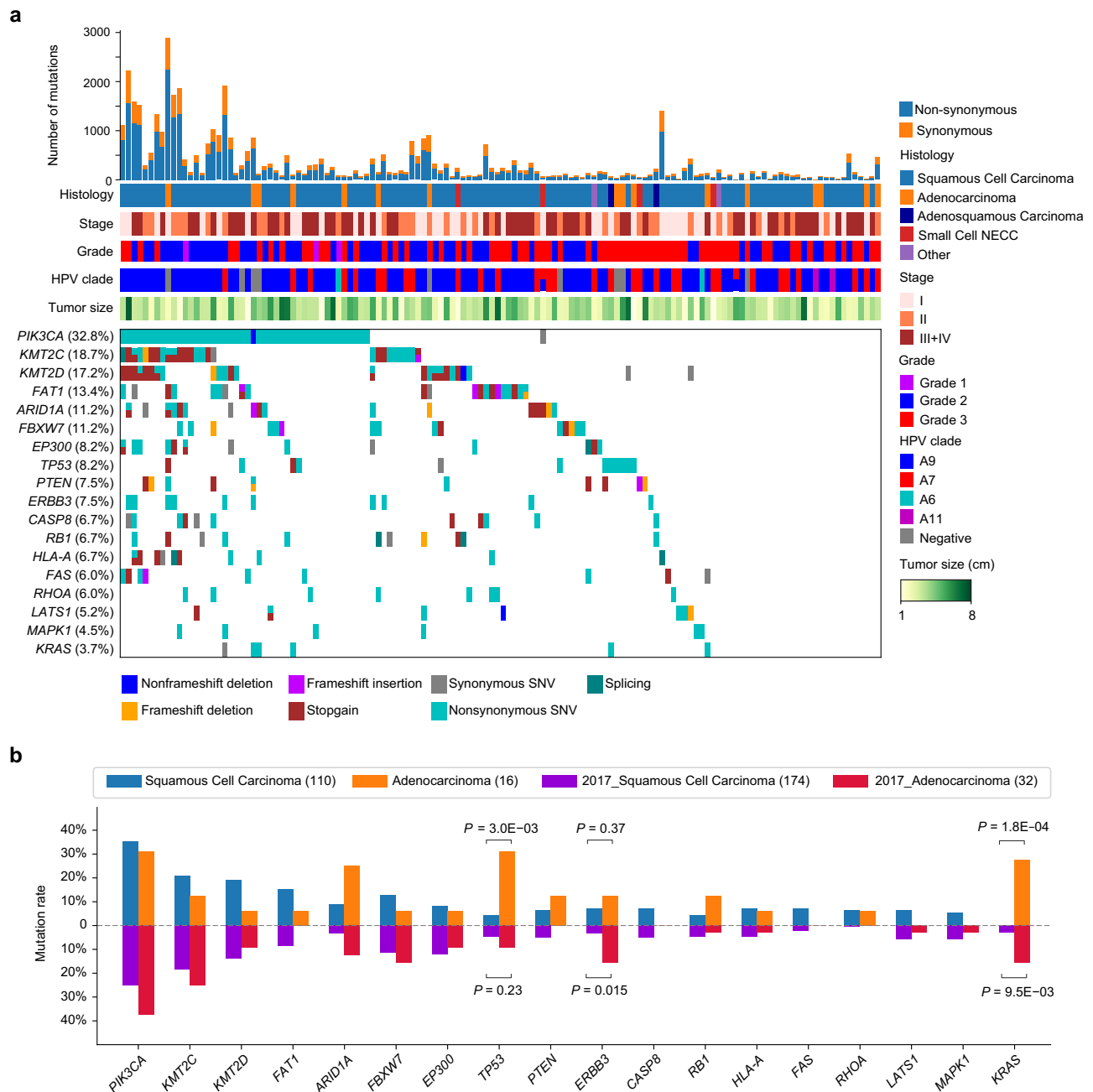


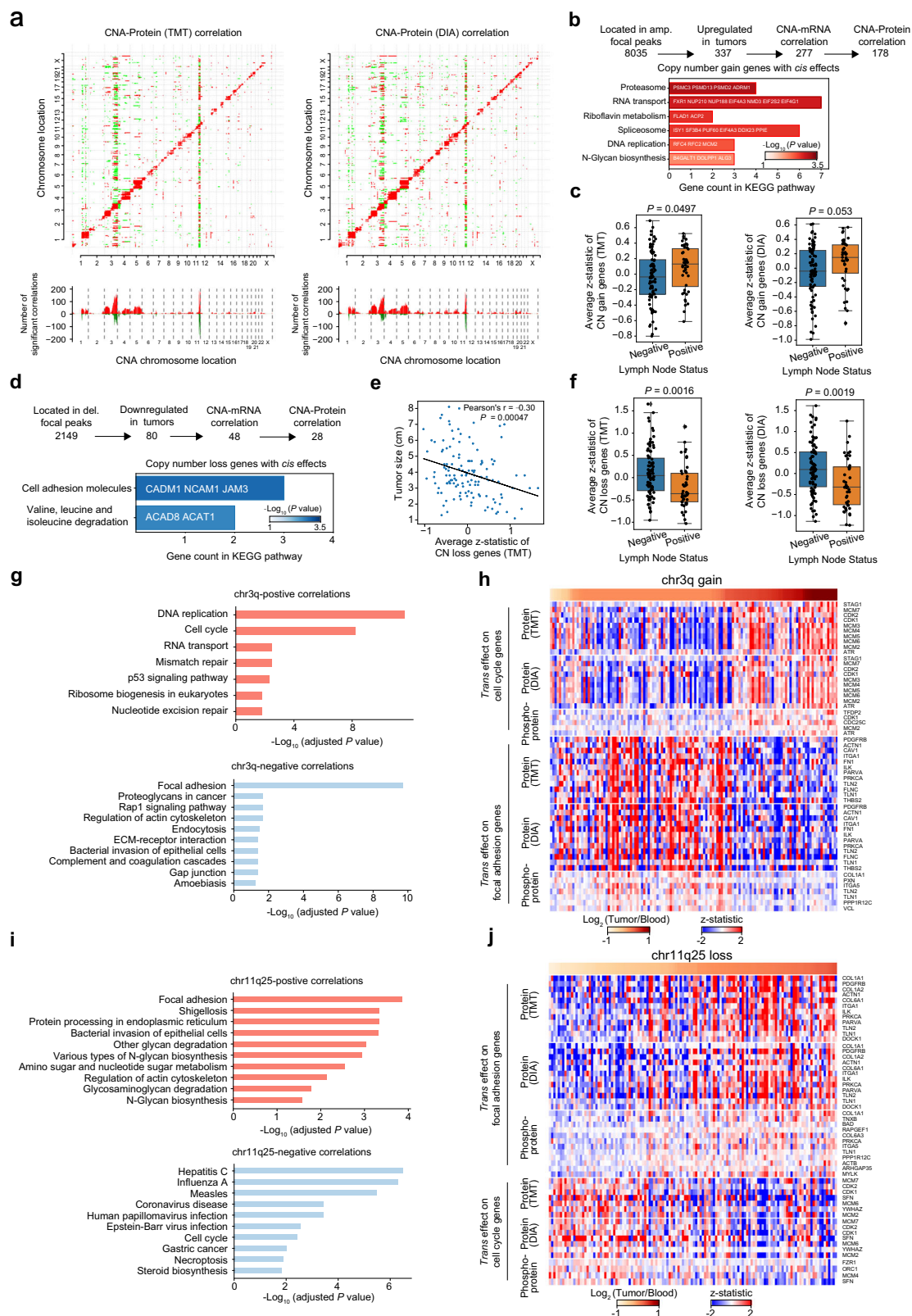
Fig. 1 | WES-based mutation profile of Chinese CC cohort. a Genetic profile of the 134 CC patients. **b** Comparisons of frequently mutated genes between cervical squamous cell carcinoma and adenocarcinoma in Chinese and TCGA CC cohorts (one-sided Fisher's exact test). Source data are provided as a Source Data file.

correlated with tumor size (Fig. 2e, Supplementary Fig. 3j). Additionally, tumors with positive lymph node displayed significantly lower CNL-*Cis* protein abundance than tumors with negative lymph node (Fig. 2f). Among the CNL-*Cis* genes, 50 (28.1%) were located on chromosome 3q, while 16 (57.1%) of the CNL-*Cis* genes were located on chromosome 11q (Supplementary Fig. 3i). Meanwhile, we found Chr3q and Chr11q25 CNAs had much more *trans* effects at protein and phosphoprotein levels (Fig. 2a, Supplementary Fig. 3e). Pathway enrichment analysis revealed that gain of Chr3q and loss of Chr11q25 demonstrated the same *trans* effects on up-regulated cell-cycle pathway and down-regulated focal adhesion pathway at the protein and phosphoprotein levels (Fig. 2g–j). Furthermore, co-occurrence of CNG on Chr3q and CNL on Chr11q25 was observed (Supplementary Fig. 3k). Notably, loss of Chr11q25 was associated with bigger tumor size (Supplementary Fig. 3l) and positive lymph

node metastasis (Supplementary Fig. 3m), also likely contributing to proliferation and metastasis of CC cells.

Landscape of HPV-associated proteogenomic changes

HPV detection and genotyping were done based on RNA-seq data³⁴, the HPV gene transcriptome data were shown in Supplementary Data 6a. Most squamous cell carcinoma samples were HPV positive (98.2%), with 77.7% being infected with HPV A9 clade; 16.7% and 33.3% of adenocarcinoma samples were infected with A9 and A7 clade respectively, and 44.4% of adenocarcinoma samples were HPV-negative (Fig. 3a). Furthermore, the proportion of HPV-positive samples among all tumors was 92.6%, substantially higher than that among NATs (42.9%) (Fig. 3a). We also found that viral gene mRNA abundance (E1/2, E5/6/7, L1/2) in tumors was significant higher compared to NATs (Supplementary Fig. 4a), suggesting severe HPV infection was



primarily localized within cervical lesions. In addition, tumor samples infected with HPV A9 displayed higher immune infiltration scores compared with those with HPV A7 ($P = 0.008$, Fig. 3b).

Consistent with previous study that E6 and E7 could promote malignant transformation of host cells via forming a complex³⁵, there was a strong positive correlation observed between the mRNA abundance of E6 and E7 in tumor samples (Spearman's correlation = 0.98,

$P = 1.2 \times 10^{-94}$, Supplementary Fig. 4b). To further elucidate the relationship between viral gene and proteome and tumor phenotype, we performed a correlation analysis of E6, and discovered that proteins positively correlated with E6 expression were significantly enriched in the cell cycle pathway, whereas proteins negatively correlated were associated with focal adhesion, protein processing in the endoplasmic reticulum, and N-glycan biosynthesis (Supplementary Fig. 5a–c).

Fig. 2 | Effects of SCNAs. **a** Correlations of CNAs (x-axes) to TMT (left) and DIA (right) protein expression (y-axes) highlight CNA *cis* and *trans* effects. Significant positive (red) and negative (green) correlations (two-sided Spearman's correlation, BH adjusted $P < 0.01$) are indicated in top panels. The proteins positively associated with a particular CNA were presented as red bars underneath the respective panels and those negatively associated with a particular CNA were represented by dark green bars. **b** Approach schematic for identifying CNG-*Cis* genes. From 8035 genes located in GISTIC focal amplification peaks, 337 were up-regulated in tumors (two-sided Wilcoxon rank-sum test, BH adjusted $P < 0.05$), of which 277 had mRNA effects and 178 had protein level effects (Spearman's correlation > 0 , BH adjusted $P < 0.05$). **c** Lymph node-positive patients ($n = 44$) exhibited a higher level of protein expression of CNG-*Cis* genes compared to lymph node-negative patients ($n = 86$, two-sided Student's *t* test). **d** Flow chart for identification of CNL-*Cis* genes. From 2149 genes located in GISTIC focal deletion peaks, 80 were down-regulated in tumors (two-sided Wilcoxon rank-sum test, BH adjusted $P < 0.05$), of which 48 had mRNA effects and 28 had protein level effects (Spearman's correlation > 0 , BH

adjusted $P < 0.05$). **e** Protein abundance (TMT) of CNL-*Cis* genes was negatively correlated with tumor size (two-sided Pearson's correlation). **f** Lymph node-positive patients ($n = 44$) displayed lower protein expression levels of CNL-*Cis* genes than lymph node-negative patients ($n = 86$; two-sided Student's *t* test). **g** KEGG pathway enrichment analysis of proteins with positive ($n = 85$) or negative ($n = 64$) CNA-protein correlations in chromosome 3q (one-sided Fisher's exact test, BH adjusted P). **h** Heatmaps of CNG of chromosomes 3q and the protein/phospho-protein abundance of cell cycle and focal adhesion related proteins. **i** KEGG pathway enrichment analysis of proteins with positive ($n = 197$) or negative ($n = 178$) CNA-protein correlations in chromosome 11q25 (one-sided Fisher's exact test, BH adjusted P). **j** Heatmap of CNL of chromosomes 11q25 and the protein/phospho-protein abundance of cell cycle and focal adhesion related proteins. Data in (**c**, **f**) are shown using boxplots. Boxplots show the median (central line), the 25–75% IQR (box limits), and the $\pm 1.5 \times$ IQR (whiskers). Source data are provided as a Source Data file.

Simultaneously, we observed upregulation in the phosphoprotein of H2AX and WEE1 in tumors with higher E6 mRNA abundance, while their protein expressions did not show significant alteration (Fig. 3c–e). Previous studies have demonstrated that HPV-positive tumors responded remarkably to the treatment of WEE1 inhibitor Adavosertib^{36,37}. Phosphorylation of H2AX is a signature of DNA double-strand breaks³⁸. The two key HPV capsid genes L1 and L2 in tumor samples showed a high correlation (Spearman's correlation = 0.81, $P = 3.2 \times 10^{-34}$, Supplementary Fig. 4b). L1 and L2 together form the HPV's capsid, which is essential for the HPV's survival, transmission, and spread in the environment³⁹. Proteins positively associated with L1 mRNA abundance were enriched in DNA replication, while proteins negatively associated were involved in cell adhesion, platelet activation, Fc gamma receptor-mediated autophagy, and leukocyte trans-endothelial migration (Supplementary Fig. 5d). The proteins most correlated with L1 mRNA abundance were shown in Supplementary Fig. 5e (|Spearman's correlation| > 0.3). Interestingly, transcription levels of L1 gene negatively correlated with immune infiltration score and stromal score (Supplementary Fig. 5f).

The carcinogenic progression of lesions infected with HPV16 and HPV18 is typically linked to the integration of the viral genome into the host cell chromosome, and this process often results in the loss of E2 gene expression^{40–42}. Additionally, we observed the mRNA abundance of E2 and E5 exhibited a bimodal distribution in this cohort (Supplementary Fig. 4b), so we clustered the tumors into two groups with either high or low E2/E5 mRNA abundance. VRK2 and MTDH showed higher expression in HPV_G2 tumors (Fig. 3f). VRK2 has been identified as an unfavorable prognostic marker in liver cancer, renal cancer, and pancreatic cancer, and MTDH is an unfavorable prognostic marker in lung cancer, renal cancer, and pancreatic cancer (<https://www.proteinatlas.org>). Additionally, increased phosphorylation of SMIM13 and VANG1 was observed in HPV_G2 compared with HPV_G1 (Fig. 3f). VANG1 plays an important role in regulating the polarized cell behavior in cancer development⁴³. HPV_G2 contained a higher prevalence of HPV A7 infection (Fig. 3f). In addition, patients in HPV_G2 were significantly younger (Fig. 3g) and had a worse PFS (Fig. 3h) than those in HPV_G1. Together, our data revealed the landscape of HPV-associated proteogenomic changes and the impact of HPV on prognosis.

Proteomic alterations associated with tumorigenesis in CC

Tumors and paired NATs revealed remarkable differences in proteins, phosphosites, and acetylsites (Supplementary Data 7a–c). In total, 3268 (38.8%) proteins showed significantly differential expression (BH adjusted $P < 0.01$), with 483 proteins up-regulated and 1041 proteins down-regulated for more than 2-fold abundance changes (Fig. 4a, Supplementary Fig. 6a). The up-regulated proteins were enriched in DNA replication, mismatch repair, p53 signaling pathway, and cell

cycle pathway, while the down-regulated proteins were enriched in pathways such as cell adhesion molecules, focal adhesion, and complement and coagulation cascades (Fig. 4b). Enrichment analysis using altered phosphosites or acetylsites associated proteins, displayed similar results (Supplementary Fig. 6b, c). Additionally, we defined 23 proteins as CC-associated proteins that were elevated by more than 2-fold in at least 80% of all tumor-NAT pairs, and annotated its potential clinical utility by the Human Protein Atlas (Fig. 4c). Furthermore, the elevated tumor expression of these proteins was validated using DIA proteomic (Supplementary Fig. 6d).

Next, we compared the molecular differences between cervical adenocarcinoma and squamous cell carcinoma, by focusing on the molecules that exhibited increased abundance in tumors compared to NATs at multi-omics level (Fig. 4d, Supplementary Fig. 6e, f). For proteins, we found immune-related biological processes such as response to type I interferon, response to interferon-beta/gamma, and regulation of innate immune response were significantly up-regulated in squamous cell carcinoma, while several metabolism processes were enriched in adenocarcinoma (Fig. 4d). The expression of interferon-induced antiviral RNA-binding protein (IFIT1/2/3), interferon regulatory factor 6 (IRF6), interferon-induced GTP-binding protein (MX1/2), and STAT1/2 were only significantly up-regulated in tumors of squamous cell carcinoma (Fig. 4e). Furthermore, the expression of immune checkpoints and MHC molecules including CD274, CD44, CXCL10, and HLA-E/F/G were also up-regulated in tumors of squamous cell carcinoma (Fig. 4f). We thus speculate that squamous cell carcinoma has an active immune microenvironment and may be beneficial from immune therapies.

To explore the molecular dynamics of CC tumor progression, we performed differential protein expression analysis across different stages (I, II, III + IV). Enrichment analysis of KEGG pathways for these differentially expressed proteins suggested that several metabolic pathways including pyruvate metabolism, glycolysis/gluconeogenesis, carbon metabolism, and galactose metabolism were dysregulated among tumor progression (Fig. 4g). Notably, 14 proteins presented an overall positive regulation trend during CC progression and showed significantly up-regulated between tumors and NATs, while 9 proteins showed opposite trend and were down-regulated in tumors (Fig. 4h). Furthermore, TCGA CC cohort was utilized to better understand the clinical outcome for these metabolism-related proteins, and suggested that high transcription level, including B4GALT1, DHCR24, AGPS, ACACA, HK2 and SLC2A1, were positively associated with poor overall survival (OS) in CC (Fig. 4i–l and Supplementary Fig. 7a, b). It is worth noting that, HK2 expression is well known to be positively correlated with tumor size, pathological grade, and prognosis and served as a carcinogenic role in CC through AKT pathway^{44,45}. High SLC2A1 expression as an independent prognostic factor for OS may be associated with immunomodulation in CC^{46–49}. However, the relevance of

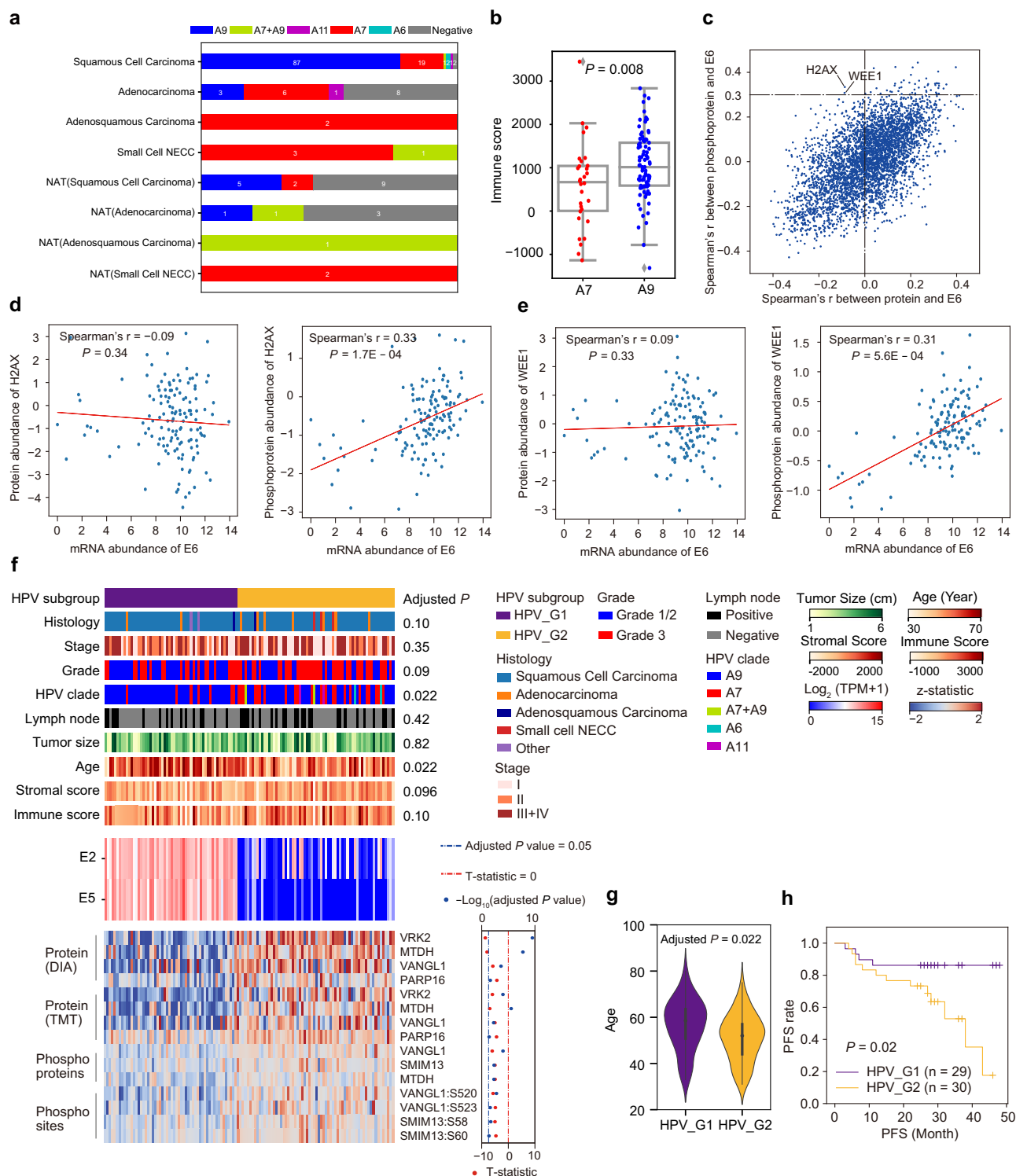
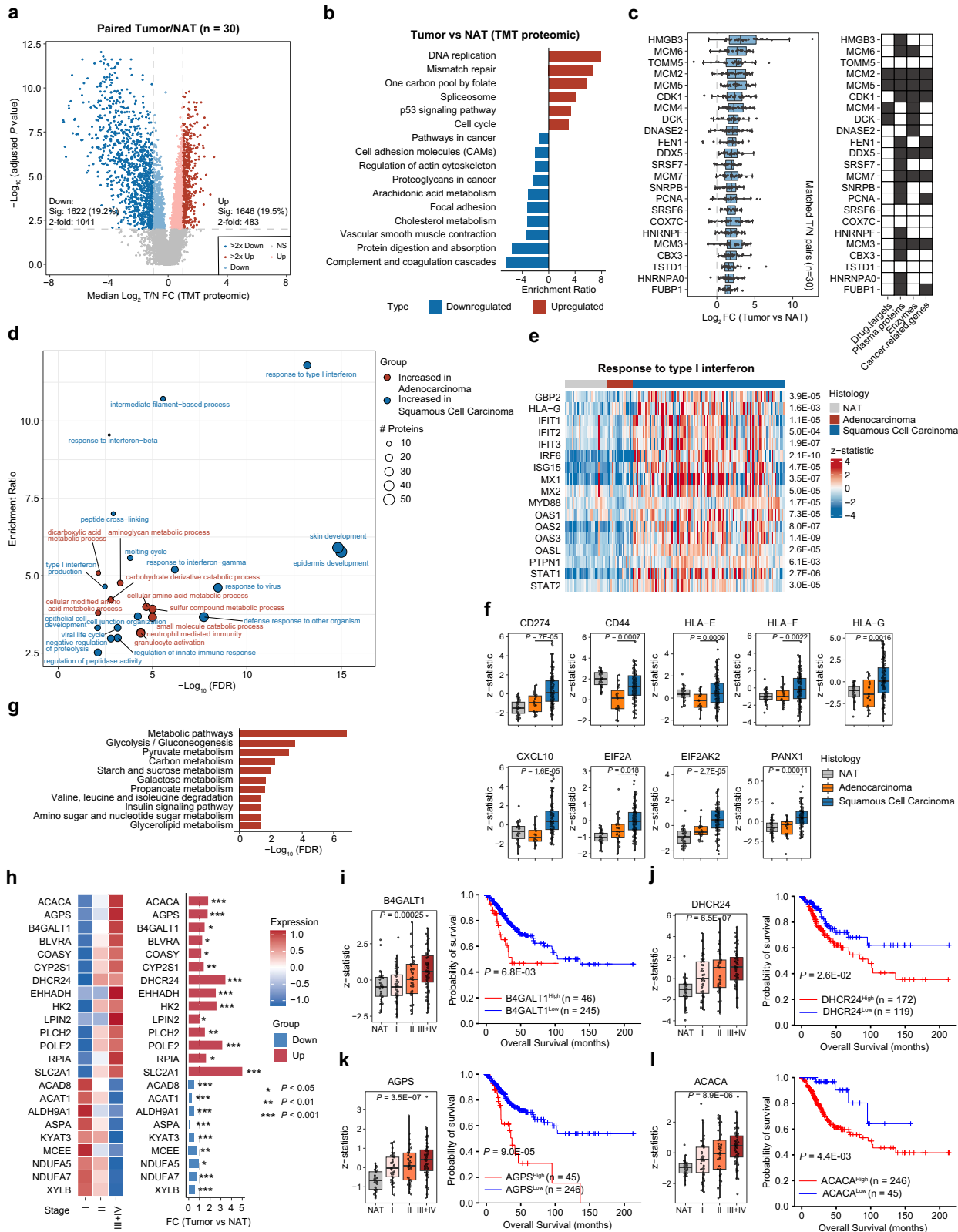


Fig. 3 | HPV-associated proteogenomic landscape and patient subgrouping based on HPV E2/E5 mRNA level. a The landscape of HPV infection types in Chinese CC cohort, with sample counts displayed. **b** Boxplot showing the comparison of immune score among different types of HPV clade (A9, $n = 89$; A7, $n = 30$, two-sided Wilcoxon rank-sum test). **c** Scatterplot displayed the Spearman's correlation between the protein (x-axis)/phosphoprotein (y-axis) and HPV E6 mRNA abundance across tumor samples (two-sided Spearman's correlation). **d** Scatterplot showing Spearman's correlation between HPV E6 mRNA and H2AX protein/phosphoprotein abundance (two-sided Spearman's correlation). **e** Scatterplot showing Spearman's correlation between HPV E6 mRNA and WEE1 protein/phosphoprotein

abundance (two-sided Spearman's correlation). **f** Patient subgrouping based on HPV E2/E5 mRNA level ($n = 122$), and the heatmap in the bottom panel displayed differentially expressed proteins, phosphoprotein, and phosphosites (two-sided Student's t test, BH adjusted P). **g** Age differences between patients in two HPV subgroups (HPV_G1, $n = 56$; HPV_G2, $n = 66$, two-sided Student's t test, BH adjusted P). **h** Kaplan-Meier curves for PFS of radical radiotherapy CC patients based on HPV subgroups (two-sided log-rank test). Numbers in parentheses represented the sample sizes for the involved groups. Boxplots in (**b**, **g**) show the median (central line), the 25–75% IQR (box limits), and the $\pm 1.5 \times$ IQR (whiskers). Source data are provided as a Source Data file.



B4GALT1, DHCR24, AGPS and ACACA in CC has not reported in previous studies. In addition to correlation with CC stage, we found that the expression of the above four proteins was higher in patients with positive lymph nodes (Supplementary Fig. 7c-f). To further validate the driving role of these genes in CC, we performed CCK-8 and colony formation experiments. These results suggested that knockdown of B4GALT1, DHCR24, AGPS and ACACA by siRNAs significantly

decreased cell proliferation and colony formation capacity of the SiHa cells, further indicating their oncogenic roles in CC progression (Supplementary Fig. 7g-j).

Aberrant protein acetylation regulates CC progression

To date, very few attempts have been made to systematically describe and make use of the acetylome regulation in CC. In this

Fig. 4 | Proteomic alterations associated with CC tumorigenesis. **a** Volcano plot depicting differentially expressed proteins between paired (n = 30) tumors and NATs (two-sided Student's t test, BH adjusted). **b** Representative KEGG pathway terms for 2-fold upregulated and downregulated proteins. **c** Boxplot showing Log₂ FC between paired (n = 30) tumors and NATs for CC-associated proteins annotated with potential clinical utilities by the Human Protein Atlas (two-sided Student's t test, BH adjusted $P < 0.01$). **d** Scatterplot of significant enriched GO biological processes terms in cervical adenocarcinoma and squamous cell carcinoma based on differentially expressed proteins between their tumors and NATs. **e** Heatmap of the relative abundance of response to type I interferon-related proteins that were significantly upregulated between cervical squamous cell carcinoma and adenocarcinoma (two-sided Student's t test). **f** Boxplots showing the expression of immune checkpoints and MHC molecules at protein level in NATs (n = 33), tumors

from cervical adenocarcinoma (n = 19) and squamous cell carcinoma (n = 109) (two-sided Student's t test). **g** Enriched KEGG pathways for proteins that were differentially expressed among different stage. **h** Summary of metabolism-related proteins that were differentially expressed among stage and between tumors and NATs (two-sided Student's t test). **i–l** Boxplot showing comparison of protein expression among NATs and tumors from different stage of B4GALT1 (**i**), DHCR24 (**j**), AGPS (**k**) and ACACA (**l**) (one-way ANOVA test). Kaplan-Meier curves for OS based on mRNA abundance of B4GALT1, DHCR24, AGPS, and ACACA from TCGA CC cohort (n = 291, two-sided log-rank test). Patients were stratified by the optimal cutpoint using maximally selected rank statistics (maxstat) on mRNA abundance. Boxplots in (**c**, **f** and **i–l**) show the median (central line), the 25–75% IQR (box limits), and the $\pm 1.5 \times$ IQR (whiskers). Source data are provided as a Source Data file.

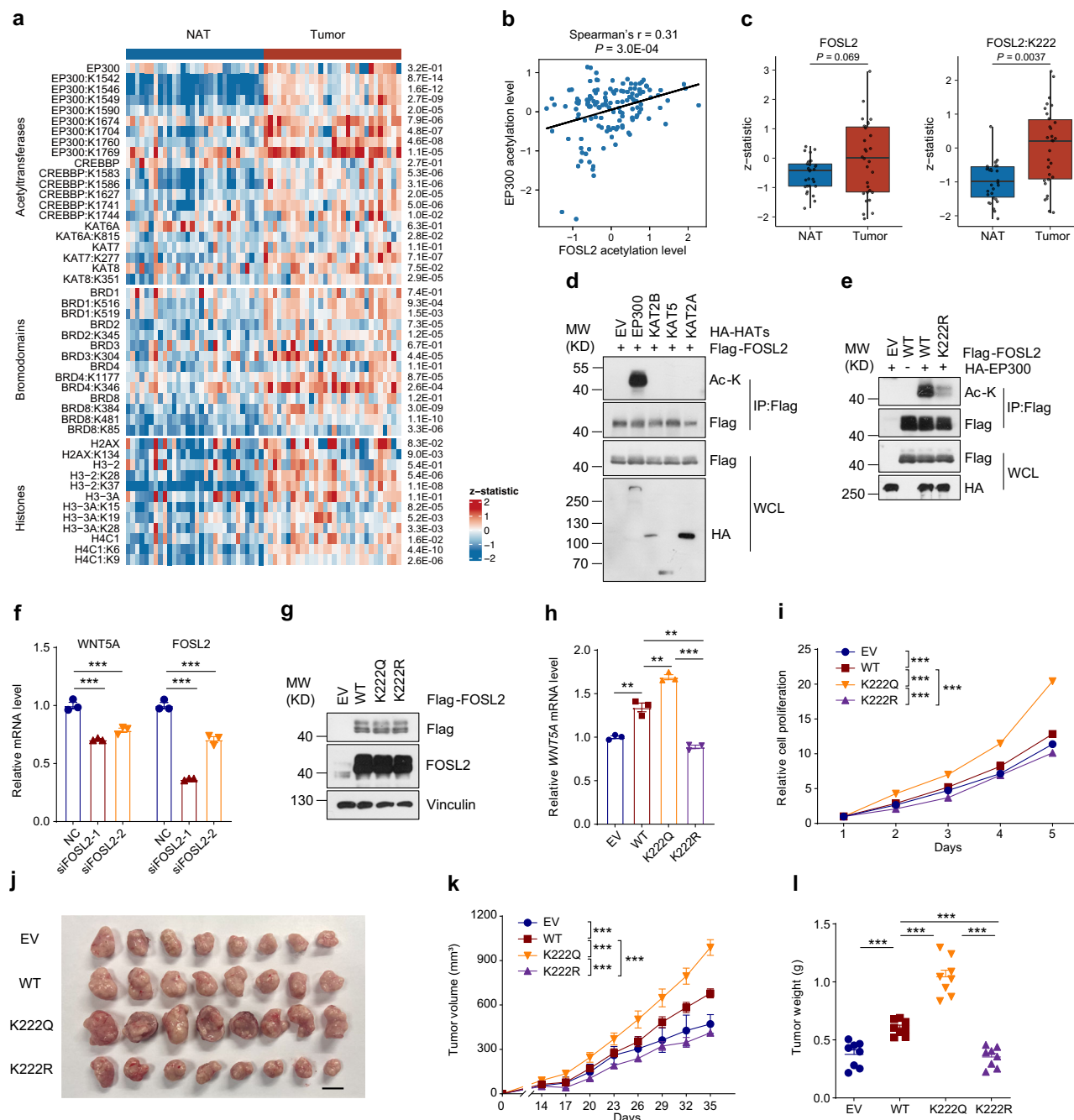


Fig. 5 | Aberrant protein acetylation regulates CC progression. **a** Heatmap illustrating the protein and acetylation abundance of KATs, BRDs, and histones between paired ($n = 30$) tumors and NATs (two-sided Student's t test). **b** Scatterplot showing Spearman's correlation of FOSL2 acetylation level (x-axis) and EP300 acetylation level (y-axis) (two-sided Spearman's correlation). **c** Boxplots showing comparison of protein or K222 acetylation expression of FOSL2 between paired ($n = 30$) tumors and NATs (two-sided Student's t test). Boxplots show the median (central line), the 25–75% IQR (box limits), and the $\pm 1.5 \times$ IQR (whiskers). **d** EP300 acetylated FOSL2. Flag-FOSL2 expression vector was co-transfected separately with expression vectors containing a variety of acetyltransferases into HEK293T cells. WCL and immunoprecipitates were detected by immunoblot with indicated antibodies. Empty Vector (EV) was used as a negative control. The experiments were repeated at least three times. **e** Mutation of K222 significantly decreased FOSL2 acetylation by EP300. HEK293T cells were co-transfected with indicated plasmids. Cell lysates were harvested for IP and immunoblotting. The experiments were repeated at least three times. **f** Relative mRNA level of WNT5A and FOSL2 in SiHa cells after FOSL2 siRNA transfection. Data are represented as mean \pm SEM ($n = 3$, one-way ANOVA test). WNT5A: *** $P = 7.3E-05$ (NC and siFOSL2-1), *** $P = 4.3E-04$ (NC and siFOSL2-2). FOSL2: *** $P = 1.3E-06$ (NC and siFOSL2-1), *** $P = 1.1E-04$ (NC and

siFOSL2-2). **g** SiHa cells ectopically expressing FLAG-tagged FOSL2-WT, FOSL2-K222R and FOSL2-K222Q mutants were generated and confirmed by immunoblot. The experiments were repeated at least three times. **h** Relative mRNA level of WNT5A was determined by qRT-PCR in the indicated SiHa cells. Data are represented as mean \pm SEM ($n = 3$, two-sided Student's t test). ** $P = 2.9E-03$ (EV and WT), *** $P = 4.0E-03$ (WT and K222Q), *** $P = 1.1E-03$ (WT and K222R), *** $P = 2.3E-05$ (K222Q and K222R). **i** Proliferation of indicated SiHa cells was measured. Data are represented as mean \pm SEM ($n = 3$, two-sided Student's t test). *** $P = 7.5E-06$ (EV and WT), *** $P = 4.5E-06$ (WT and K222Q), *** $P = 1.1E-05$ (WT and K222R), *** $P = 1.6E-06$ (K222Q and K222R). **j** Xenograft tumor pictures of nude mice in different groups. Scale bar, 1 cm. **k** Tumor growth curves of indicated SiHa cells subcutaneously injected into nude mice. Data are represented as mean \pm SEM ($n = 8$ mice per group, two-sided Student's t test). *** $P = 9.2E-05$ (EV and WT), *** $P = 1.7E-04$ (WT and K222Q), *** $P = 3.4E-06$ (WT and K222R), *** $P = 6.4E-08$ (K222Q and K222R). **l** Tumor weight of nude mice in different groups. Data are represented as mean \pm SEM ($n = 8$ mice per group, two-sided Student's t test). *** $P = 1.7E-04$ (EV and WT), *** $P = 5.6E-06$ (WT and K222Q), *** $P = 1.3E-05$ (WT and K222R), *** $P = 4.6E-08$ (K222Q and K222R). Source data are provided as a Source Data file.

study, we performed differential analysis of lysine acetyltransferases (KATs), bromodomain-containing proteins (BRDs), and histones with protein and acetylation abundances. Compared to NATs, the acetylation of these proteins was dramatically up-regulated in tumors relative to its protein abundance (Fig. 5a). Notably, we found that EP300 had considerable K1542, K1546, K1549 and K1590 hyperacetylation in its histone acetyltransferase domain (HAT)⁵⁰ (Fig. 5a). Among them, K1549, a key autoacetylation site in the activation loop that may activate EP300 activity was significantly up-regulated in tumors⁵¹.

We then examined the potential impact EP300 activity by measuring the acetylation of EP300 itself and performed correlation analysis with our acetylproteomics data. Consequently, high acetylation levels of EP300 were associated with high acetylation of FOSL2 (Fig. 5b), a transcription factor that belongs to the AP-1 family and was reported to be involved in cell proliferation and differentiation⁵². However, the acetylation or function of FOSL2 in CC has not been studied. Interestingly, we found a significant upregulation of FOSL2-K222 acetylation but not its protein abundance (Fig. 5c). Representative spectra of the acetylated peptides of FOSL2-K222 and EP300 were shown in Supplementary Fig. 8a–i.

To further explore the connection between EP300 and FOSL2-K222 in CC, we conducted a comprehensive set of experiments. First, co-transfection of EP300, but not KAT2B, KAT5 or KAT2A, promoted FOSL2 acetylation (Fig. 5d). The elevation of FOSL2-K222 acetylation upon EP300 co-transfection was further confirmed by mass spectrometry (MS) analysis (Supplementary Fig. 8j, k). Compared with FOSL2-WT (wild type), when the FOSL2-K222R was mutated (acetylation-deficient), the acetylation level of FOSL2 caused by co-transfection with EP300 was significantly reduced (Fig. 5e), suggesting that K222 was a major acetylation site of FOSL2 by EP300. FOSL2 has been implicated as an onco-protein in several cancers, including lung cancer⁵³, breast cancer⁵⁴ and hepatocellular carcinoma⁵⁵, and WNT5A has been shown as a transcription target gene of FOSL2⁵⁴. To explore the function of FOSL2-K222 acetylation in CC, we performed real-time quantitative RT-PCR (qRT-PCR) to validate the FOSL2-WNT5A regulation in SiHa cells (Fig. 5f). We established SiHa cells stably expressing FLAG-tagged FOSL2-WT, FOSL2-K222R and FOSL2-K222Q, respectively (Fig. 5g). Interestingly, we found the acetylation mimetic FOSL2-K222Q mutant was more potent than FOSL2-WT in elevating the transcription of WNT5A (Fig. 5h). Furthermore, compared with FOSL2-WT and FOSL2-K222R, cells expressing FOSL2-K222Q showed more potent proliferation rate (Fig. 5i), and formed the largest tumors in xenograft tumorigenesis assay (Fig. 5j–l), implying a strong tumor-promoting effect of FOSL2-

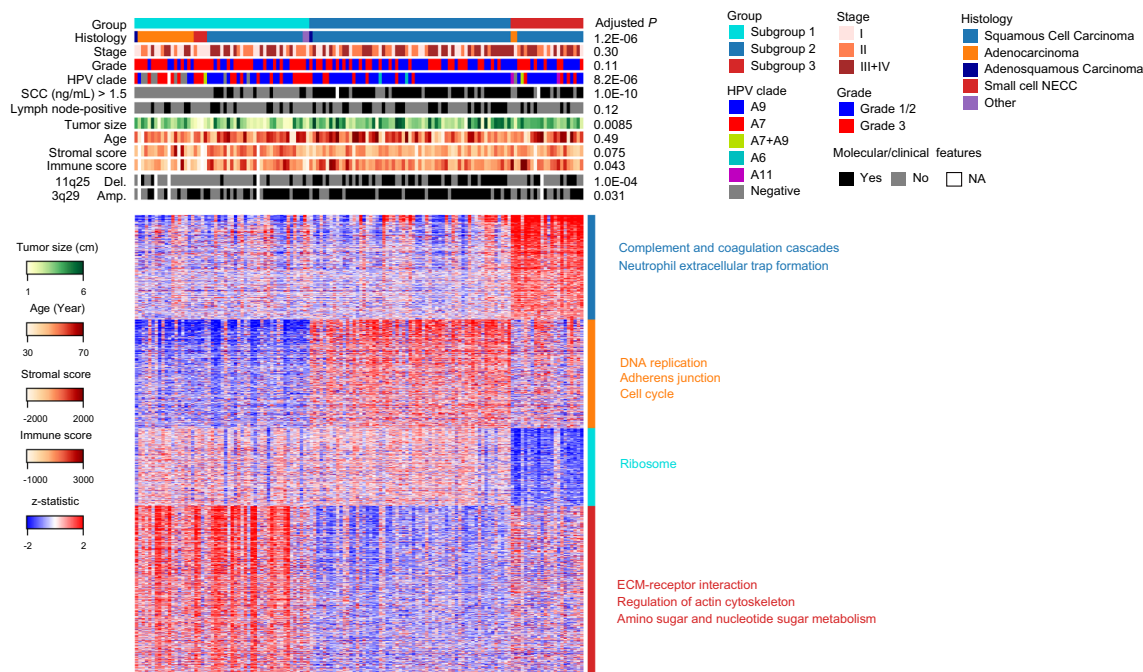
K222 acetylation. Therefore, these results together indicated that EP300 promoted CC cell proliferation at least impartially via FOSL2-K222 acetylation (Supplementary Fig. 8l).

Proteomic subgroups with distinct biological and clinical features

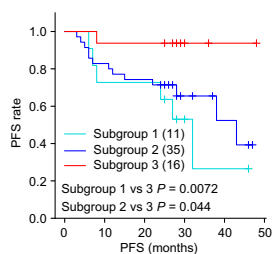
We performed unsupervised consensus clustering based on TMT proteomic data (see Methods), resulting in three proteomic subgroups (Fig. 6a, Supplementary Data 8a). Focusing further on radical radiotherapy patients with PFS information (see Methods, $n = 62$), a significant difference in PFS was observed among the three proteomic subgroups. Patients in Subgroup 3 exhibited the best PFS, even when further divided by known prognostic clinicopathological variables such as stage, lymph node status, grade, and histology type (Fig. 6b, Supplementary Fig. 9a–d). DIA-based proteomic and TMT-based phosphoproteomic/acetylproteomic data were also used to cluster CC patients into three subgroups respectively, consistent with the TMT proteomic subgroups, demonstrating the reliability and robustness of proteomic subgrouping (Supplementary Fig. 9e–h). Transcriptome-based subgroups showed lower concordance with TMT proteomic subgroups, and their PFS differences were less pronounced than those based on proteomic profiling (Supplementary Fig. 9i), highlighting the superiority of proteomic subgrouping.

We found a higher prevalence of adenocarcinoma and HPV-A7 infection, along with a lower incidence of abnormal SCC level in Subgroup 1. Subgroup 2 showed the highest rates of 11q25 deletion and 3q29 amplification (Fig. 6a, Supplementary Fig. 9j). Supervised differential expression analysis revealed subgroup-specific proteins. In Subgroup 1, proteins linked to keratinocyte differentiation and KRT family KRT4, KRT5, KRT6A, KRT6B), as well as S100A family members (S100A7, S100A8, S100A9) were down-regulated (Supplementary Fig. 9k), suggesting that Subgroup 1 may be associated with previously reported keratin-low cluster¹⁴. In Subgroup 2, 395 proteins were up-regulated and enriched in DNA replication (MCM family), indicating proliferative characteristics for these tumors. In Subgroup 3, 381 proteins were up-regulated and enriched in complement and coagulation cascades, and neutrophil extracellular trap formation, indicating the inflammatory status of Subgroup 3^{56,57}. Pathway enrichment analysis also revealed that protein abundance in Subgroup 1 higher than in Subgroups 2 and 3 were enriched in ECM-receptor interaction, regulation of actin cytoskeleton, and amino sugar and nucleotide sugar metabolism pathways; Proteins up-regulated in Subgroup 1 and 2, were associated with ribosome, suggesting over-activated ribosome biogenesis^{58,59} of CC tumors in these

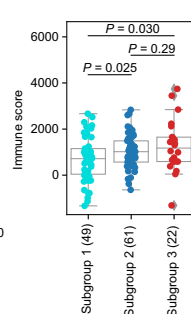
a



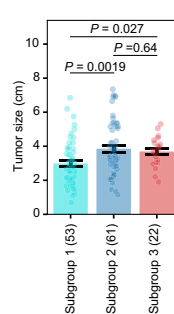
b



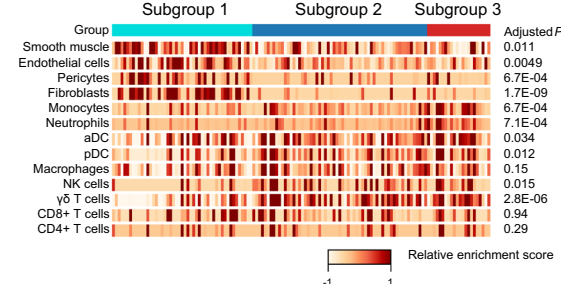
c



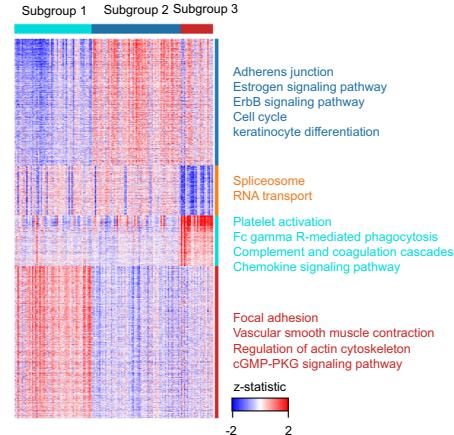
d



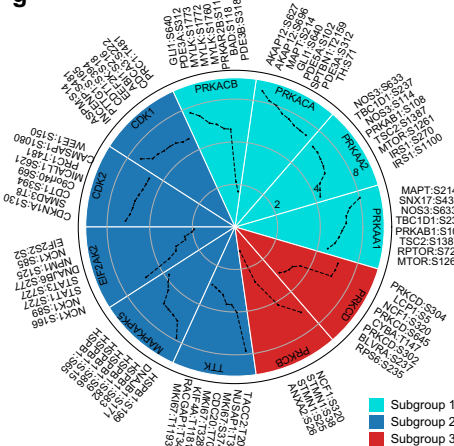
e



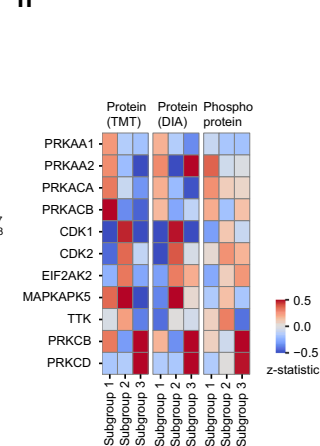
f



g



h



two subgroups (Fig. 6a, Supplementary Data 9a). The subgroup-related signatures from DIA-based proteomic data further corroborated the findings (Supplementary Fig. 9l, Supplementary Data 9b). Additionally, Subgroup 3 showed the highest immune infiltration scores, followed by Subgroup 2 and then Subgroup 1 (Fig. 6c). This could further explain the highest response rate to radiotherapy in Subgroup 3 patients. While patients in Subgroup 1 have smaller

tumors, their poorer PFS can be attributed to their significantly lower immune infiltration scores (Fig. 6d). Further comparison of the landscape of cellular heterogeneity revealed that Subgroup 1 patients exhibited the highest proportion of stromal cells and less NK and γδ T and DC cells, Subgroup 3 patients showed significantly higher proportions of monocytes and neutrophils (Fig. 6e), indicating Subgroup 3 patients might benefit from immunotherapy.

Fig. 6 | Proteomic stratification of the CC cohort and the corresponding protein pathways and subgroup-specific kinases. **a** Patient subgrouping analysis of proteomic profiling identified three proteomic subgroups: Subgroup 1 (cyan, $n = 53$), Subgroup 2 (blue, $n = 61$), Subgroup 3 (red, $n = 22$). Each column represented a tumor sample, the associations of proteomic subgroups with clinical characteristics and somatic mutations were annotated in the top panel (Chi-square test or one-way ANOVA test, BH adjusted P). Rows in the bottom panel indicated differentially expressed proteins (One-way ANOVA test, Bonferroni-adjusted $P < 0.05$), color of each cell showed TMT protein abundance. **b** Kaplan-Meier curves of PFS for radical radiotherapy patients in each TMT proteomic subgroup (two-sided log-rank test). The comparison of immune score (**c**) and tumor size (**d**) among TMT proteomic subgroups (two-sided Student's t test). Boxplots in **c** show the median

(central line), the 25–75% IQR (box limits), and the $\pm 1.5 \times$ IQR (whiskers). Data in **d** are represented as mean \pm SEM. **e** Transcriptome-based deconvolution of mRNA transcript cell signatures in tumors using Xcell (one-way ANOVA test, BH adjusted P). **f** Differential phosphorylation sites among three TMT proteomic subgroups (one-way ANOVA test, Bonferroni-adjusted $P < 0.05$). Unsupervised clustering and biological pathways significantly enriched are presented on the right (one-sided Fisher's exact test, BH adjusted $P < 0.05$). **g** Circular plot depicting the active kinases in each TMT proteomic subgroup compared to all other subgroups identified by KSEA (see Methods). The eight kinase-regulated phosphorylation sites with the highest t -statistics were indicated by black dots. **h** Heatmap showing the protein (TMT and DIA) and phosphorylation abundance of TMT subgroup-specific kinases. Source data are provided as a Source Data file.

To investigate the phosphoproteomic characteristics of CC subgroups, Unsupervised clustering was performed on the differentially expressed phosphosites, uncovering four phosphorylation modules corresponding to the four protein modules (Fig. 6f). Kinase-Substrate Enrichment Analysis (KSEA) was applied to further identify subgroup-specific kinases (Fig. 6g). PRKACA, PRKACB, PRKAA1, and PRKAA2 were identified as the Subgroup 1-specific kinases; CDK1, CDK2, EIF2AK2, MAPKAPK5 and TTK were identified as the Subgroup 2-specific kinases, associated with poor prognosis. PRKCB and PRKCD were identified as the Subgroup 3-specific kinases. PRKCB is involved in B cell activation, and apoptosis induction, and was considered as a tumor suppressor in colon cancer⁶⁰. Overexpression of PRKCD in SiHa cells enhances the radiation sensitivity, while silencing of PRKCD expression in ME180 cells by siRNA decreases the radiation sensitivity⁶¹. In TMT and DIA proteomic datasets, protein abundance of these subgroup-specific kinases showed similar expression patterns with phosphorylation, further demonstrating the strong association between phosphorylation function and protein expression (Fig. 6h).

Identification and validation of radioresponse-related biomarkers and risk-scoring model

By combining PFS data with protein expression in radical radiotherapy patients, we identified 92 radioresponse-related biomarkers (see Methods), including 37 favorable and 55 unfavorable candidates (Supplementary Fig. 10a). Then a risk scoring model was developed based on these biomarkers (see Methods). 62 radical radiotherapy patients were then divided into high-risk and low-risk groups based on the median risk score. Both TMT and DIA proteomic results showed that patients in the high-risk group had worse PFS than the low-risk group (Supplementary Fig. 10b, c). The prognostic predictive power of the risk scoring model was further validated in the TCGA CC cohort (Supplementary Fig. 10d).

It is noteworthy that the Subgroup 3-specific kinase PRKCB (Fig. 6g), identified as one of the favorable biomarkers, was down-regulated in tumors (Fig. 7a, Supplementary Fig. 10e). Immunohistochemistry (IHC) images also showed reduced PRKCB expression in tumors compared to NATs ($P = 4.4E-03$, Supplementary Fig. 10i). Additionally, PRKCB protein levels showed no significant association with clinical features, including different stages, lymph node metastasis and grade status (Supplementary Fig. 10f–h). Tumors with high PRKCB showed specific upregulation of Fc Gamma R-mediated phagocytosis, natural killer cell-mediated cytotoxicity, T cell receptor signaling pathway and B cell receptor signaling pathway, and down-regulation of cell cycle, DNA replication, spliceosome, and ErbB signaling pathway (Fig. 7b). Besides, radical radiotherapy patients with high PRKCB protein (TMT/DIA) expression have a better PFS (Fig. 7c, d). Moreover, increased PRKCB mRNA levels were positively correlated with a better OS in the TCGA CC cohort (Fig. 7e). The IHC results from an independent cohort of 124 patients (see Methods) who underwent radical radiotherapy further validated that higher PRKCB

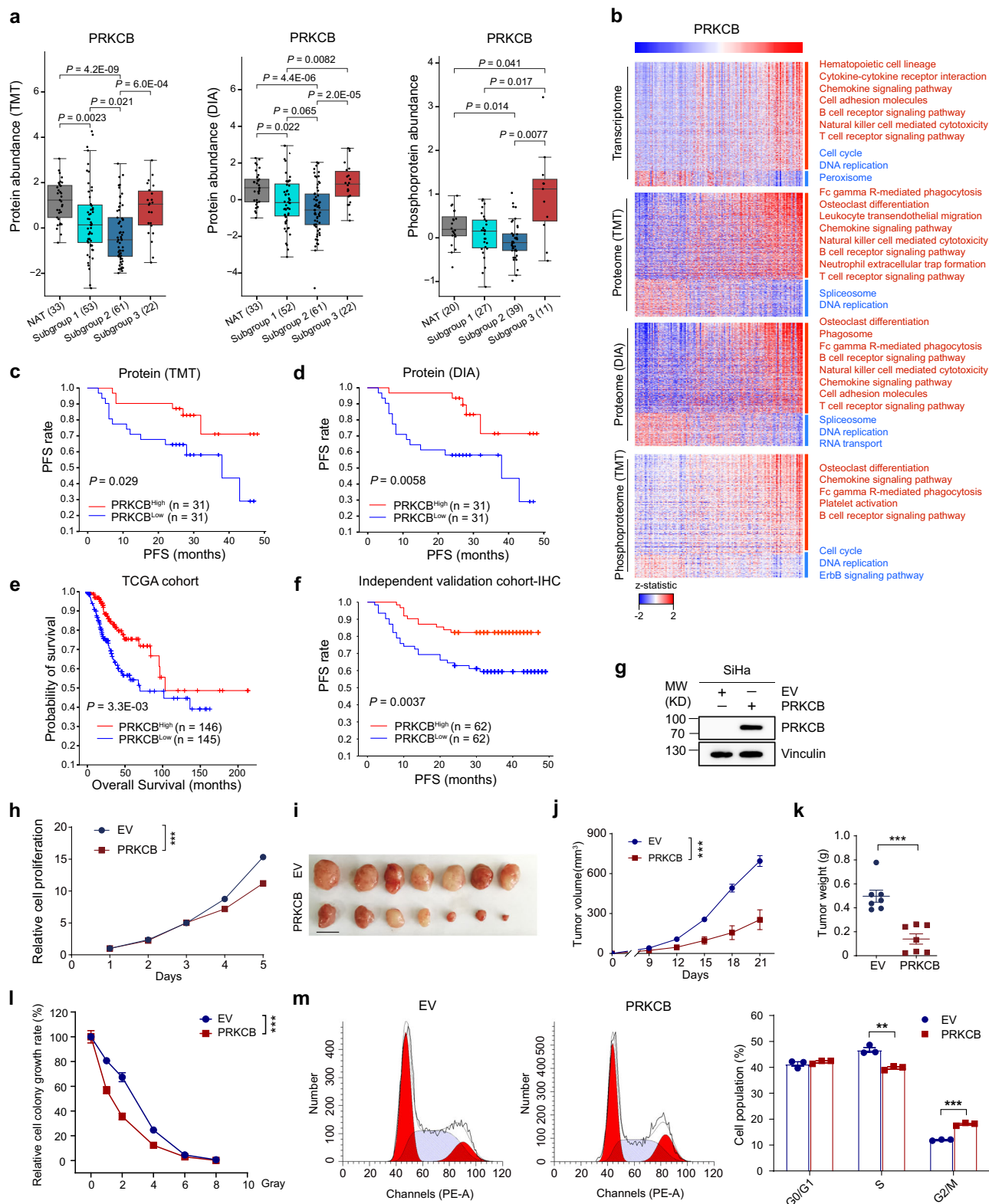
expression is positively correlated with better radiotherapy efficacy (Fig. 7f).

To explore the function of PRKCB in CC cell growth, we performed CCK-8 experiments and confirmed that overexpression of PRKCB in SiHa cells inhibited tumor cell proliferation (Fig. 7g, h). PRKCB-overexpressing cells formed smaller tumors than the control cells in xenograft models (Fig. 7i–k). It is worth noting that overexpression of PRKCB significantly enhanced the killing effect of radiation and induce G2/M cycle arrest in SiHa cells, implying that PRKCB may enhance radiosensitivity by regulating cell cycle progression (Fig. 7l, m). Altogether, our data indicated that PRKCB could be a promising biomarker for prediction of radiotherapy efficacy and may play a tumor suppressor role in CC.

Discussion

Comprehensive genomic and transcriptomic analysis of CC has significantly enhanced our comprehension of the molecular events of this malignancy^{11–15,62}. Unlike previous CC proteomic studies^{63,64}, this study presents global genomic, transcriptomic, proteomic, phosphoproteomic and acetylproteomic data, offering additional and more comprehensive insights into the clinical, biological, and therapeutic dimensions of CC. Our integrated analysis provided a global view of the multi-omic changes associated with HPV, revealed functional PTMs on the key proteins such as FOSL2-K222 acetylation involved in CC malignant proliferation, classified CC patients into three clinically relevant subgroups, and identified a candidate biomarker PRKCB and risk scoring model associated with radiotherapy response.

Previous studies reported that HPV viral protein interacted with different epigenetic modifiers, including CREBBP, KAT2B, EP300^{65–67}. However, effects of viral genes on multi-omic layers in CC remain unexplored. Our investigation was conducted to examine the alterations in the profiles of core proteins and phosphorylated proteins, with a particular emphasis on the upregulation of H2AX and WEE1 phosphorylation in tumors expressing elevated levels of E6 mRNA. It has been demonstrated that the WEE1 inhibitor (AZD1775) can be considered as a potential radiosensitizer in CC⁶⁸. Phosphorylation of the Ser 139 residue of H2AX, also known as γ H2AX, is an initial cellular response to DNA double-strand breaks⁶⁹. Phosphorylated H2AX is enriched in CC tumors with high expression of E6, which may be a consequence of HPV viral gene integration. We subsequently utilized high-quality MS data to uncover significant biological pathways and aberrant proteins that are closely associated with viral genes at the proteomic level, which is less studied in previous studies that primarily focused on genome and transcriptome. The integration of HPV viral genes usually results in the loss of viral E2 gene expression and a selection for the continued expression of E6 and E7^{40–42}. Additionally, we observed a bimodal distribution in the mRNA levels of E2 and E5 in tumors, along with low or absent expression of the E2 gene. Then we clustered the samples into two groups with either high or low E2/E5 mRNA abundance. The patients undergoing radical radiotherapy, who exhibit low mRNA abundance



of E2/E5 (HPV G₂), experience a worse PFS in comparison to patients with high mRNA abundance of E2/E5 (HPV G₁). We further analyzed the intrinsic motivation for the different prognosis from the perspective of aberrant proteins and phosphorylated proteins. These results not only expand and enhance the understanding that the deletion of the E2 gene promotes tumor progression, but also innovatively reveal the impact of HPV E2/E5 gene deletion expression on the prognosis of CC patients.

Acetyltransferases mediated protein acetylation is an essential protein PTM that regulates various biological processes, including oncogenesis and tumor progression. Previous studies have applied acetylome analysis in different cancers to demonstrate the function of acetylation protein^{70–72}, but few of them are related to CC. Here, we systematically described the acetylome characteristics in CC to identify the putative regulatory mechanisms. As a major acetyltransferase and transcriptional co-activator, EP300 has been indicated in many

Fig. 7 | Identification and validation of the radioresponse-related biomarker, PRKCB. **a** Relative protein (TMT and DIA) and phosphoprotein abundance of PRKCB in NATs and TMT proteomic subgroups (two-sided Student's test). Boxplots show the median (central line), the 25–75% IQR (box limits), and the $\pm 1.5 \times$ IQR (whiskers). **b** Associations of PRKCB protein abundance (TMT) with multi-omics profiles (Spearman's correlation $|\rho| > 0.3$). **c** Kaplan-Meier curves for PFS based on PRKCB protein abundance (TMT) of radical radiotherapy patients (two-sided log-rank test). **d** Kaplan-Meier curves for PFS based on PRKCB protein abundance (DIA) of radical radiotherapy CC patients (two-sided log-rank test). **e** Kaplan-Meier curves of OS based on PRKCB mRNA levels of TCGA CC cohort (two-sided log-rank test). **f** Kaplan-Meier curves for PFS based on PRKCB protein expression investigated by IHC of independent validation cohort (two-sided log-rank test). **g** SiHa cells stably expressing PRKCB were generated and confirmed by immunoblot. EV was used as a negative control. The experiments were repeated at least three times. **h** Proliferation of indicated SiHa cells was measured. Data are represented as mean \pm SEM ($n = 6$, two-sided Student's t test). *** $P = 1.7E-06$ (EV and PRKCB). **i** Xenograft tumor pictures of nude mice in different groups. Scale bar,

1 cm. **j** Tumor growth curves of indicated SiHa cells subcutaneously injected into nude mice. Data are represented as mean \pm SEM ($n = 7$ mice per group, two-sided Student's t test). *** $P = 2.1E-04$ (EV and PRKCB). **k** Tumor weight of nude mice in different groups. Data are represented as mean \pm SEM ($n = 7$ mice per group, two-sided Student's t test). *** $P = 1.6E-04$ (EV and PRKCB). **l** Radiosensitization by PRKCB overexpression in SiHa cells in colony formation assays compared to EV. Clonogenic survival fraction were analyzed using SiHa cells in the relevant groups at radiation doses of 0, 1, 2, 4, 6, and 8 Gray. Data are represented as mean \pm SEM ($n = 6$, two-sided Student's t test). *** $P = 2.3E-05$ (EV and PRKCB). **m** Flow cytometry was used to assess the cell cycle distribution of SiHa cells in the relevant groups. Modfit LT 3.1 was used for data analysis. Forward scatter and side scatter were used to circle the cell population (excluding debris), and the fluorescence channels PE-A (area)/PE-W (width) were used to exclude adherent cells. Finally, the proportion of cells in each stage was fitted. Left: Representative results of the cell cycle analysis. Right: Histogram of cell cycle distribution. Data are represented as mean \pm SEM ($n = 3$, two-sided Student's t test). S: *** $P = 2.9E-03$ (EV and PRKCB). G2/M: *** $P = 7.6E-05$ (EV and PRKCB). Source data are provided as a Source Data file.

cancer types and plays important roles in cancer progression and poor prognosis^{71,73}. There are numerous instances where EP300 overexpression and inappropriate activation have been found to correlate with malignancy and drive cancer growth^{74–76}. Notably, small molecule inhibitors of EP300 have also shown considerable inhibition of tumor progression in certain types of cancer^{77–80}. C646, a selective inhibitor of EP300, inhibited the proliferation and induced apoptosis of CC cells⁸¹. We found that acetyltransferase EP300, consistent with a previous study in CC⁸², was hyperacetylation in CC tissues. Previous studies revealed that autoacetylation of the EP300 HAT domain enhances its HAT activity^{50,51}. Therefore, hyperacetylation of EP300, especially in its HAT domain, may cause aberrant protein acetylation change to drive CC progression. This study discovered and confirmed that hyperacetylated EP300 drives cell proliferation and accelerates malignant through the FOSL2-K222 acetylation axis. These findings may offer a theoretical foundation for targeting EP300 activity in the treatment of CC, but more precise mechanisms need to be investigated in future work.

In this study, we classified CC patients into three TMT proteomic subgroups which provided therapeutic and biological insights. Subgroup 2 consisted of sixty cases of squamous cell carcinoma and one case of adenosquamous carcinoma. However, the distribution of histological types of patients included in this study aligns with epidemiological statistics and is comparable to the TCGA CC patients with molecular subgroups^{14,83,84}. Both TCGA CC subgroups and our proteomic subgroups suggest a correlation between the histological type and molecular classification. Protein enriched in Subgroup (such as MCM2, MCM3, MCM4, MCM5, MCM6, MCM7) were involved in DNA replication, indicating proliferative characteristics for these tumors. Compared with subgroup 1 and subgroup 2, Subgroup 3 associated with the best radiotherapy effect and showed the highest immune infiltration scores, proportions of monocytes and neutrophils, indicating that Subgroup 3 patients might benefit the most from immunotherapy. Previous single-cell sequencing study has revealed radiochemotherapy-induced innate immune activation and MHC-II upregulation in CC⁸⁵, which could further explain the best radiotherapy effect in Subgroup 3 patients. Besides, the proteomic subgroups may also serve as a classifier to screen immune-benefiting populations and predict patient prognosis.

Notably, we identified the biomarker PRKCB and built a risk-scoring model associated with the response to radiotherapy at the proteomic level through high-throughput screening. This finding was also confirmed in both TMT and DIA proteomic datasets. PRKCB was down-regulated in CC tumor tissues, and high PRKCB expression was positively correlated with better prognosis, consistent with previous findings reported in lung adenocarcinoma⁸⁶. Other studies reported the implication of PRKCB in the regulation of mitochondrial integrity

and oxidative phosphorylation^{87,88}. Overexpression of PRKCB in HEK293 cells led to a significant downregulation of autophagy, as assessed by the decrease of endogenous LC3-II levels⁸⁹. PRKCB significantly enhanced the killing effect of radiation and induced G2/M cell cycle arrest in SiHa cells in our study. On one hand, we will elucidate the intricate molecular mechanism by which PRKCB enhances the sensitivity of CC to radiotherapy and its agonists. On the other hand, we plan to carry out multicenter cohort studies to facilitate the clinical implementation and utilization of PRKCB and risk scoring models, with the goal of enhancing the effectiveness of radiotherapy in treating locally advanced CC.

This study bears the following limitations: 1) Only some of the tumors had matching NATs. Besides, although the current data correlates with patients' radioresponse, however, we do not have tissue samples after radical radiotherapy. In the future, by expanding current primary treatment cohort and conducting a comparative analysis of samples before and after radiotherapy, it will be possible to find clues about important molecular events related to radioresponse. 2) The prognostic analysis of our current study was based on PFS data. And the analysis specifically focused on patients receiving radical radiotherapy, as surgery patients generally have a good prognosis. Extending the follow-up period in the future (up to 6 to 7 years post-treatment), we could consider including all patients and annotating patient prognosis from an OS perspective. 3) The proteogenomic data of tumors were collected from the same powdered sample in this study. This sample preparation methodology is good to keep the consistency when acquiring different omics data and to improve the accuracy of cross-omics analysis results, but it lost intra-tumor heterogeneity information in cellularity, especially the spatial structure information. Further integration of single-cell and spatial omics data will facilitate a more comprehensive analysis of the characteristics of CC.

In summary, we conducted a comprehensive and integrative proteogenomic analysis of Chinese CC, resulting in a valuable public resource for exploring the landscape features of the CC genome, transcriptome, and proteome. Our proteogenomic findings extended the biology and clinical aspects of CC, and they could lead to clues as well as opportunities to improve the treatment of CC patients.

Methods

Clinical sample of PUMC-CC cohort

Based on the sample selection parameters of TCGA CC cohort¹⁴, this study implemented the following criteria: 1) All tumors and NATs were collected before treatment, and were stored in liquid nitrogen within 30 min after isolation. 2) Each piece of tissue was sectioned and Hematoxylin and eosin (H&E) stained for tumor cellularity analysis before sequencing, and all H&E staining results were scanned and

pathologically analyzed. 3) All tumor samples were completed independently by two pathologists to ensure histological accuracy and the NATs contained no tumor cells. 4) The percent tumor nuclei and percent necrosis were also assessed. Tumors with at least 50% tumor cell nuclei and less than 10% necrosis were reserved. All patients signed separate informed consent forms for sampling, research, and publication. This study was conducted in accordance with the guidelines and regulations set forth by the Ethics Committee of National Cancer Center/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College (approval No. 22/374-3576). The treatment-naïve CC patients were recruited from Cancer Hospital Chinese Academy of Medical Sciences between June 2019 and July 2021. 139 tumors and 33 NATs (30 NATs with paired tumors) were collected, and each sample was homogenized via cryo-pulverization and aliquoted to subsequent genomic, transcriptomic, proteomic, phosphoproteomic, and acetylproteomic analyses.

Each sample was assigned a new research ID, and the patient's name or medical record number used during hospitalization was de-identified. The clinical information of these enrolled patients were collected, including SCC, FIGO stage, tumor size, tumor grade, histology type, age diagnosis, HPV type, primary treatment, lymph node status and PFS information. PFS was defined as the time from the start of treatment to disease progression, relapse, or death. The follow-up deadline for these enrolled patients was July 2023. 1 patient did not receive antitumor treatment. The surgical patients generally have a good prognosis, no progression was observed in 62 out of the 73 surgical patients at the time of the follow-up cutoff. In addition, the standard treatment for locally advanced CC is cisplatin-based concurrent chemoradiotherapy, while radiotherapy alone is indicated in older patients or those with insufficient renal function⁹⁰. Cisplatin remains the radiosensitizing agent for patients with locally advanced CC when used concomitantly with radiotherapy in the National Comprehensive Cancer Network guidelines⁹¹. In this study, 65 patients (46.8% of the cohort) received primary treatment with either concurrent chemoradiotherapy (n = 54) or radiotherapy alone (n = 11). These patients, collectively referred to as radical radiotherapy patients^{92,93}, were combined for the PFS analysis.

The 124 patients in the independent validation cohort were stage IIB to stage IVB CC patients who received concurrent chemoradiotherapy or radiotherapy alone at the Cancer Hospital of the Chinese Academy of Medical Sciences from November 2015 to December 2017. The clinicopathological parameters and FPS information for these cases were detailed in Supplementary Data 10.

Cell line

SiHa and HEK293T cells were kept in Dr. Daming Gao's lab. SiHa cells were cultured in MEM-NEAA with 10% FBS, 100 units of penicillin and 100 mg/mL streptomycin. HEK293T cells were cultured in DMEM with 10% FBS, 100 units of penicillin, and 100 mg/mL streptomycin.

Proteogenomic workflow

The proteogenomic research was carried out through the workflow shown in Supplementary Fig. 1a. To reduce the impact of intra-tumor heterogeneity on multi-omics analysis, tumors, and NATs were pulverized using the CryoPrepTM CP02 (Covaris) and then divided into two parts: the first part was snap-frozen in liquid nitrogen and stored at -80 °C for the following proteomic, phosphoproteomic and acetylproteomics analyses as well as DNA extraction; the second part was added to the RNeasy Lysis Solution (Qiagen, 70204) and kept in -80 °C, and then used for RNA extraction.

DNA/RNA extraction, WES, and RNA-seq

Genomic DNA was extracted from tumors and NATs using Magnetic Universal Genomic DNA Kit (TIANGEN, DP705) according to manufacturer's protocol. Matched blood DNA was also extracted using the

above kit. The quality of isolated genomic DNA was verified by Qubit® DNA Assay Kit in Qubit® 3.0 Fluorometer (Invitrogen) and NanoDrop 2000 (Thermo Fisher Scientific) and the integrity was assessed by TapeStation (Agilent Technologies). A total amount of 0.2 µg DNA per sample was used as input material for the DNA library preparations. Sequencing library was generated using NEB Next® Ultra™ DNA Library Prep Kit for Illumina (NEB) following manufacturer's recommendations and index codes were added to each sample. DNA libraries were sequenced on the Illumina NovaSeq 6000 platform and 150 bp paired-end reads were generated. WES was conducted with a mean coverage depth of 156X (range: 112-253X) for tumors, 157X (range: 125-216X) for NATs, and 126X (range: 83-205X) for blood samples.

Total RNA was extracted and purified from fresh frozen tissues using the TRIzol® reagent (Invitrogen, 15596026CN). RNA integrity was assessed using the RNA Nano 6000 Assay Kit of the Bioanalyzer 2100 system (Agilent Technologies). Briefly, mRNA was purified from total RNA using poly-T oligo-attached magnetic beads. Fragmentation was carried out using divalent cations under elevated temperature in First Strand Synthesis Reaction Buffer (5X). Paired-end libraries were sequenced on the NovaSeq 6000 Illumina platform, for each tumor and NAT sample, RNA-seq resulted in an average of 64.1 M and 64.4 M high-quality reads, respectively.

MS methods

Protein extraction and tryptic digestion. For protein extraction, ~50 mg of cryo-pulverized tumors or NATs were homogenized in 500 µL lysis buffer (4% SDS, 0.1 M DTT, 0.1 M Tris-HCl, pH 7.6) and sonicated at 20% amplitude for 2 min (5 s on, 5 s off) using Ultrasonic Homogenizer (NingBo Scientz Biotechnology Co., LTD, JY92-IIDN). The proteins were then denatured and reduced at 95 °C for 5 min. Lysates were centrifuged at 12,000 g for 10 min to remove the insoluble debris and protein concentrations of the clarified lysates were determined using tryptophan-based fluorescence quantification method⁹⁴. Protein lysates (1 mg) were diluted to equal concentration and alkylated with iodoacetamide (IAA) for 45 min at room temperature in the dark. Then, protein lysates were precipitated by adding ice-cold acidified acetone/ethanol buffer with five times the volume of lysis buffer mixed and put them overnight at -20 °C. Precipitated proteins were collected by centrifugation at 18,000 g for 40 min at 4 °C, washed with 1 mL ice-cold acetone and 1 mL 70% ethanol for 40 min at 4 °C. The proteins were then subjected to proteolytic digestion with sequencing grade modified trypsin (Promega) at 1:50 enzyme-to-substrate ratio overnight at 37 °C. The resulting digests were acidified with FA to achieve a final volumetric concentration of 1%, and then subjected to desalting using Waters Sep-Pak® Vac 1cc (50 mg) tC18 cartridges.

TMT 16-plex labeling of peptides

Desalted peptides from each sample were labeled with TMTpro 16-plex reagents (Thermo Fisher Scientific). Tumors and NATs were co-randomized to 12 TMT sets. The NATs were labeled using channels that closely matched the paired tumors within the same TMT set. In addition, paired samples were evenly distributed among the 12 TMT sets to ensure a balance of NATs across all sets. A mixed sample was prepared by pooling an aliquot from 30 CC tumors and NATs and labeled with channel 134N in all of the TMTpro 16-plex sets as internal reference. For each TMT labeling experiment, dried peptides (250 µg) from each sample were dissolved in 250 µL 100 mM TEAB. 5 mg of TMT reagent was dissolved in 205 µL anhydrous acetonitrile, and then 50 µL of each TMT reagent was added to the corresponding peptides. After 1-h incubation at room temperature, 12.5 µL 5% hydroxylamine was added to quench the labeling reaction for 15 min at room temperature. The labeled peptides were pooled, dried down via Speed-Vac, and subsequently desalted on a reversed-phase tC18 SepPak column (Waters).

Peptide fractionation

In the TMT-based proteome analysis, the labeled peptides were pooled together to create a highly complex mixture. To reduce sample complexity and increase the depth of protein identification, high-pH reverse phase liquid chromatography (RPLC) was used for peptide fractionation, as previously described³³. For each TMT set, about 4 mg TMTpro 16-plex labeled peptides were fractionated using a 4.6 mm × 250 mm Waters XBridge BEH300 C18 column with 3.5 μm size beads (Waters). Peptides were separated using an Agilent 1260 HPLC instrument via high-pH RPLC with solvent A (10 mM ammonium formate, pH 10) and a non-linear increasing concentration of solvent B (90% ACN, 10 mM ammonium formate, pH 10) at a flow rate of 0.7 mL/min. The 110-min separation gradient was set as follows: 1–5% B in 2 min; 5–25% B in 35 min; 25–40% B in 43 min; 40–55% B in 6 min; 55–95% B in 3 min; 95% B for 4 min; 95–1% B in 1 min; 1% B for 16 min. Peptides were separated and collected every minute for a total of 96 fractions from 3 min to 99 min, with fractions combined into 24 fractions by a stepwise concatenation strategy. 5% of each of the 24 fractions was allocated and dried down in a Speed-Vac for global proteome analysis. The remaining 95% sample was then utilized for phosphopeptides enrichment. In DIA-based proteome analysis, since the samples are analyzed individually, each sample is directly analyzed without prior fractionation.

Phosphopeptide enrichment

High-Select Fe-NTA kit (Thermo Fisher Scientific, A32992) was used for phosphopeptide enrichment according to the manufacturer's instructions. Briefly, the 24 fractionated peptides were dissolved in 200 μL 80% ACN/0.1% TFA and incubated with 50 μL Fe³⁺-NTA agarose beads for 20 min at room temperature. Then, the mixture was transferred into the filter tip (Axygen, TF-200-L-R-S) and clarified peptide flow-throughs with unbound peptides were collected by centrifugation. The resins with phosphopeptides were washed with 200 μL 80% ACN/0.1% TFA for 3 times and 200 μL H₂O for 3 times. The bound phosphopeptides were eluted twice with 200 μL 50% ACN/5% NH₃·H₂O and dried down via Speed-Vac. All centrifugation steps above were conducted at 50 g at room temperature. The eluted peptides were desalted using C18 stage tips and dried down.

Acetylpeptide enrichment

PTMScan Acetyl-Lysine Motif [Ac-K] Kit (CST, 13416) was used for acetylpeptide enrichment according to the manufacturer's instructions. In brief, tryptic peptides from the flow-through of IMAC were concatenated into 4 fractions and dried down via Speed-Vac. The dried peptides were reconstituted in 1.4 mL of the immunoaffinity purification (IAP) buffer (50 mM MOPS/NaOH pH 7.2, 10 mM Na₂HPO₄ and 50 mM NaCl) and incubated with freshly prepared acetyl-lysine motif antibody agarose beads for 2 h at 4 °C. After removing the supernatant, peptide-bound beads were washed 2 times with 1 mL of ice-cold IAP buffer followed by 3 times with 1 mL HPLC H₂O. Then, acetylated peptides were eluted by incubation 2 times with 55 μL of 0.15% TFA at room temperature for 10 min. The eluted peptides were desalted using C18 stage tips and dried down.

LC-MS/MS analysis for TMT-based proteomics

Peptides were resolved with 0.1% formic acid (FA) and separated on a nanoflow Easy nLC 1200 UHPLC system (Thermo Fisher Scientific) with an in-house packed 20 cm × 75 μm interbal diameter C18 column (3 μm ReproSil-Pur C18-AQ beads, Dr. Maisch GmbH). The column was heated to 50 °C using a home-made column heater. The flow rate was set at 300 nL/min with 0.1% FA in H₂O (Solvent A) and 0.1% FA in 80% acetonitrile (Solvent B). The 120-min separation gradient was used for proteome analysis and set as followings: 2–5% B in 1 min; 5–27% B in 94 min; 27–35% B in 15 min; 35–52% B in 3 min; 52–100% B in 1 min; and 100% B in 6 min. For phosphoproteome analysis, the same LC and

column setup was used except for a 70-min LC gradient (2–4% B in 1 min; 4–25% B in 51 min; 25–32% B in 8 min; 32–50% B in 3 min; 50–100% B in 1 min; and 100% B in 6 min). For acetylproteome analysis, the same LC and column setup was used except for a 180-min LC gradient (4–35% B in 152 min; 35–45% B in 20 min; 45–100% B in 2 min; and 100% B in 6 min).

For proteomic and phosphoproteome analysis, samples were analyzed with a Q-Exactive HF mass spectrometer (Thermo Fisher Scientific) equipped with a nanoflow ionization source. Data-dependent acquisition was performed using Xcalibur software in positive ion mode at a spray voltage of 2300 V. The MS1 spectra were measured with a resolution of 120,000 @ m/z 200, an AGC target of 3e6, a maximum injection time of 50 ms and a mass range of 300 to 1700 m/z. The data-dependent mode cycle was set to trigger MS2 scan on up to the top 15 most abundant precursors per cycle at an MS2 resolution of 60,000 @ m/z 200, an AGC target of 1e5, a maximum injection time of 120 ms, an isolation window of 0.7 m/z, an high collision dissociation (HCD) collision energy of 30, and a fixed first mass of 110.0 m/z. The dynamic exclusion time was set as 40 s (30 s for phosphoproteome) and precursor ions with charge 1, 7, 8 and >8 were excluded for MS2 analysis.

For acetylproteome analysis, samples were analyzed with a Q-Exactive HF-X mass spectrometer (Thermo Fisher Scientific) equipped with a nanoflow ionization source. Data-dependent acquisition was performed using Xcalibur software in positive ion mode at a spray voltage of 1800 V. The MS1 spectra was measured with a resolution of 120,000 @ m/z 200, an AGC target of 3e6, a maximum injection time of 50 ms and a mass range of 350 to 1700 m/z. The data-dependent mode cycle was set to trigger MS2 scan on up to the top 15 most abundant precursors per cycle at an MS2 resolution of 45,000 @ m/z 200, an AGC target of 1e5, a maximum injection time of 120 ms, an isolation window of 0.7 m/z, an HCD collision energy of 30, and a fixed first mass of 110.0 m/z. The dynamic exclusion time was set as 30 s and precursor ions with charge 1, 7, 8 and >8 were excluded for MS2 analysis.

DIA analysis

Unlabeled, digested peptides from each sample were separated on a nanoElute LC system (Bruker) with an analytical column (20 cm × 75 μm, 1.6 μm C18 resin, IonOpticks). The column was heated to 50 °C using a column heater (Bruker). The flow rate was set at 300 nL/min with 0.1% FA in H₂O (Solvent A) and 0.1% FA in acetonitrile (Solvent B). The 90-min separation gradient was used for proteome analysis and set as followings: 2–22% B in 75 min; 22–37% B in 5 min; 37–80% B in 5 min; and 80% B in 5 min. Samples were analyzed using a timsTOF Pro mass spectrometer (Bruker) equipped with a nano-electrospray ion source (Bruker). Source capillary voltage was set to 1500 V in positive ion mode, and dry gas flow to 3 L/min at 180 °C. The DIA MS data were acquired using the diaPASEF method⁹⁵ consisting of 14 cycles, which including a total of 28 mass-width windows (25 Da width, from 452 to 1152 Da) with 4 mobility windows each, making a total of 56 windows covering the ion mobility range (1/KO) from 0.76 to 1.29 Vs/cm². The MS and MS/MS spectra were acquired from 100 to 1700 m/z. The TIMS-MS survey scan was acquired between 0.7 and 1.3 Vs/cm². The TIMS-MS survey scan was acquired between 0.7–1.3 Vs/cm² and 100–1700 m/z. The acquisition time of each PASEF scan was set to 100 ms with a near 100% duty cycle, which led to a total cycle time of around 1.59 s.

Database searching of MS data

For TMT-based proteomics, all MS raw files were searched against the human Swiss-Prot database containing 20,360 sequences plus 277 Swiss-Prot HPV protein sequences using MaxQuant (version 1.6.17.0)⁹⁶. TMTpro 16-plex based MS2 reporter ion quantification was chosen with reporter mass tolerance set as 0.003 Da. The purities of TMT

labeling channels were corrected according to the kit LOT number (WC320807). The PIF (precursor intensity fraction) filter value was set at 0.5. Enzyme digestion specificity was set to Trypsin and maximum 2 missed cleavages were allowed. Oxidized methionine, protein N-term acetylation, lysine acetylation, asparagine, and glutamine (NQ) deamidation were set as variable modifications. Carbamidomethyl cysteine was set as fixed modification. For phosphorylation data analysis, phospho (STY) was also chosen as a variable modification. The tolerances of first search and main search for peptides were set at 20 ppm and 4.5 ppm, respectively. A cutoff of 1% FDR was applied at the peptide, protein, and site level. A minimum of 7 amino acids was required for peptide identification. For phosphosite and acetylsite localization, the localization probability >0.75 was considered as a confident identification. Contaminants, reverse sequences, and only considered identified sequences identified from the MaxQuant runs were removed except for KRT family proteins since KRT is known to be highly expressed in squamous cell carcinoma.

For DIA proteomics, raw data files were analyzed using DIA-NN software (version 1.8.0)⁹⁷. The Swiss-Prot database was used for a library-free search with precursor FDR set as 1%. Deep learning-based spectra and retention time prediction were enabled. The fragment m/z was set from 200 to 1800 and the peptide length was set from 7 to 30. Trypsin/P was enabled and maximum number of missed cleavages set to 1. N-terminal methionine excision was enabled and cysteine carbamidomethylation was set as a fixed modification. The median number of points measured across chromatographic peak is 7.

Somatic mutation calling and filtering

WES sequencing reads after exclusion of low-quality reads were mapped to GRCh38 reference genome with BWA (Version: 0.7.17, <http://bio-bwa.sourceforge.net/>)⁹⁸. PCR duplicates were removed by Picard (version 2.27.2, <http://broadinstitute.github.io/picard/>), the base quality score was recalibrated by the BaseRecalibrator tool from GATK (version v4.2.6.1, <https://software.broadinstitute.org/gatk/>). For patients with blood WES data, somatic variants were detected using Mutect2 on tumor exome data and matched blood data. For those without matched blood data, the genetic variants database of the 1000 Genome Project was used as the panel of normal. Specifically, germline variants were filtered from the database of the 1000 Genome Project, NHLBI-ESP 6500 Exome Sequencing Project, Exome Aggregation Consortium (EXAC), and Genome Aggregation Database (gnomAD). Further filtering was done to obtain high-confidence somatic mutations using the criteria: a minimum of 8X coverage, Variant Allele Fraction (VAF) ≥ 5% and at least 3 variant supporting reads in the sample. Then, FilterMutectCalls. Annovar (version 2017 Jul 17, <https://annovar.openbioinformatics.org/en/latest/>) was used functionally annotate genetic variants based on the annotation of the human genome (GRCh38), version 32 (Ensembl 98)⁹⁹, and mutations in the non-coding regions (3'UTR, 5'UTR, Intron, gene intergenic etc.) were removed.

Analysis of significantly mutated genes

The filtered mutations (including SNV and indel) were further used to identify significantly mutated genes by MutSigCV (version 1.41 <http://www.broadinstitute.org/cancer/cga/MutSig>) with default parameters. Genes with $P < 0.05$ were identified as significantly mutated genes¹⁰⁰.

CNAs analysis

CNAs were called using cnvkit (version 0.9.7, <https://cnvkit.readthedocs.io/en/stable/pipeline.html>)¹⁰¹ on exome data with default parameters. All blood samples were generated to build the copy-number reference. Confident focal SCNAs across all tumor samples were identified by GISTIC2 with parameters amplifications threshold: 0.3, deletions threshold: -0.3, focal length cutoff: 0.5, value threshold: 0.25, genetic: yes, run broad analysis: yes, focal length

cutoff: 0.5, confidence level: 0.99, arm peel: yes, joined segment size: 100, gene collapse method: extreme and max sample segs: 3000. Significantly amplified and deleted chromosome arms were identified with a threshold of $FDR < 0.25$ ¹⁰².

Comparisons of frequently mutated genes between cervical squamous cell carcinoma and adenocarcinoma

The somatic mutation data called by Mutect from GDC TCGA CC were obtained from <https://xenabrowser.net/datapages/>, and the cancer type information was retrieved from TCGA CC paper¹⁴. Fisher's exact test was used to evaluate the statistical significance of the difference in mutation rates between patient groups.

RNA-seq data analysis

After removal of adaptor contamination and low-quality reads, sequencing reads were aligned using STAR (version 2.7.2a) to human reference sequence (GRCh38 assembly) and featureCounts was used to produce a matrix of read counts across all genes based on the GENCODE gene annotation (version 32)¹⁰³. An average of 68 million paired-end reads were sequenced per sample. The ratios of uniquely aligned reads exceeded 88% in all samples. Then, Transcripts Per Million (TPM) normalized read count of each gene was calculated based on the gene length and read count mapped to this gene. Subsequently, the Log₂ transformed TPM values were used for downstream analysis.

HPV detection

After being aligned to hg38 reference genome using STAR, the unmapped RNA-seq reads were then aligned to the reference genome of 440 HPV subtypes. HPV detection and genotyping algorithm HPV-EM was used to identify the exact genotypic makeup of each sample³⁴. Here subtypes with mapped read count less than 10 were not considered. The dominant infected HPV subtype was determined based on the criterion that its mapped read count was more than ten times higher than the sum of the other subtypes. For the three tumor samples (TN33, TB103-1, and TB98-1) lacking transcriptome data, HPV infection status was determined based on clinical detection records. HPV-EM was also utilized to measure HPV gene expression, providing read counts for HPV genes. HPV gene expression was subsequently normalized to TPM using the total read count aligned to the human genome.

Inference of cell type score

ESTIMATE was used to infer the stromal and immune scores from the RNA expression data¹⁰⁴. In addition, the signature-based method Xcell was used to dissect the relative infiltration levels of different immune cell types¹⁰⁵.

Proteomic data analysis

Analysis of TMT quantitative proteomic data. Here we used reverse-zMAP developed based on our previous MAP model¹⁰⁶ to process cancer proteomics data generated using iTRAQ/TMT platforms with internal reference samples. Briefly, all the clinical samples were separately compared to the internal reference sample generated in the same MS run using a modified MAP algorithm and a rescaled Log₂ ratio of MS intensities, termed z-statistic, was calculated for each protein to describe its relative abundance change between the two samples. In details, for each MS run, we systematically performed pairwise comparison for each tissue sample against the reference sample and the Log₂ ratio of its protein intensities is calculated as

$$M_{ij} = \text{Log}_2 \left(\frac{S_{ij}}{R_{ik}} \right), \quad (1)$$

in which S_{ij} denotes the MS intensity of protein i in sample j and R_{ik} represents its intensity in the reference sample of this MS run, i.e., run

k. Then, a similar sliding window analysis to that performed in MAP was applied to derive a linear model for the Log_2 ratios in each window.

$$M_j = \mu_j + \sigma_j \cdot \hat{q} \quad (2)$$

Next, natural cubic spline fitting was separately applied to characterize the dependence of the fitted μ_j and σ_j^2 on the Log_2 protein intensity across all windows. Finally, for each protein, μ_{ij} and σ_{ij} were calculated based on the fitted natural cubic spline function, respectively, and then, similar to the z-statistic defined in MAP, we defined the z-statistic of protein i in sample j as

$$Z_{ij} = \frac{M_{ij} - \mu_{ij}}{\sigma_{ij}}, \quad (3)$$

in which the dependence of observed Log_2 ratios on the protein intensities was effectively mitigated and the contribution of systematic and technical errors was rescaled to follow the standard normal distribution across all samples. Thus, it enables a better characterization of relative protein expression abundance and can be reliably used for downstream analyses.

Analysis of DIA quantitative proteomic data. For DIA proteomic data, protein expression profiles from all samples were normalized together, to minimize the impact of missing values, and we used the proteins that were detected in all samples to calculate the normalization factors. Since there are no internal reference samples in the DIA dataset, we take the geometric mean of all DIA proteomic profiles as a pseudo-reference sample. Then, a similar parallel pairwise comparison analysis to that performed on TMT data was applied to compare each DIA proteomic profile against the pseudo reference sample and calculate a z-statistic for each protein as the basis for downstream analysis.

Analysis of phosphoproteomic and acetylproteomic data. The phosphoproteomic and acetylproteomic data were normalized using the median centering method across adjacent phosphorylation/acetylation sites. Then, similar to the above analysis, we also utilized a sliding window approach on the M-A plot to first calculate the median modification intensity change and average modification intensity within each window. Subsequently, we applied natural cubic spline fitting to fit the dependence between the median modification intensity change and the average modification intensity across all windows. Then, the Log_2 ratio of each protein's modification intensities was adjusted by the estimated median modification intensity derived from the fitted natural cubic spline function. The median ratio of all the modification sites of each protein was calculated as a measure of its modification level.

Imputation of missing values

DreamAI ensemble algorithm¹⁰⁷ (implemented using the DreamAI R package, <https://github.com/WangLab-MSSM/DreamAI>) was applied to impute the missing values in proteomic, phosphoproteomic and acetylproteomic data. Imputation was only performed on the proteins, phosphosites and acetylsites with a missing rate <50%.

Batch effect and data quality analysis of proteomic data

The batch effect in TMT data after transforming the original MS intensities into z-statistics was assessed by performing unsupervised PCA. The leading PCs of the global proteomic data clearly separated the tumors from NATs, and no obvious batch effect was observed among the 12 TMT batches. We also performed unsupervised PCA on the DIA proteomic data, and no obvious batch effect was observed among the 3 DIA batches, too.

Subgrouping analysis of transcriptomic, proteomic, phosphoproteomic and acetylproteomic data

For TMT proteomic data, we selected the foremost 3000 proteins demonstrating the highest variability across tumor samples. Among these, 1674 proteins were identified without any missing values across all tumor samples. Subsequently, consensus clustering was performed on this set of 1674 highly variable proteins using the ConsensusClusterPlus R package¹⁰⁸. The detailed parameter settings were number of repetitions = 1000 bootstraps; pitem = 0.8 (resampling 80% of any sample); pFeature = 1; distance = "euclidean"; and k-means clustering with up to 5 clusters.

For phosphoproteomic and acetylproteomic data, top 3000 most variable phosphoproteins/phosphoproteins within tumor samples were selected based on the protein-level modification abundance matrix, and among them, phosphoproteins/phosphoproteins without missing values were used as input for the ConsensusClusterPlus R package to perform sample clustering, with the same parameters applied as above.

For RNA-seq data, we applied HyperChIP to model the mean-variance curve and subsequently utilized scaled variances that consider the mean-variance relationship for gene ranking¹⁰⁹. Before constructing the model, we excluded genes with an average expression value of $\text{Log}_2(\text{TPM} + 1) < 2$ in the tumor samples. Standardized z-score matrix of $\text{Log}_2(\text{TPM} + 1)$ for the top 3000 hyper-variable genes were used to perform consensus clustering using identical parameters as above. In summary, the tumor samples were classified into three clusters at transcriptome and PTMs levels as well.

Differential protein expression/modification analysis between patient subgroups

ANOVA test was used to detect differentially expressed proteins, phosphosites and acetylsites among the three subgroups, with Bonferroni-adjusted P -values < 0.05 as the cutoff. Then, to detect proteins/PTM sites specifically expressed/modified in each subgroup, hierarchical clustering was applied to divide the differentially expressed proteins and PTM sites into 3 or 4 clusters using seaborn.clustermap (python module) with parameter metric = "correlation"; method = "average" and scipy.cluster.hierarchy.fcluster (Python module) with parameter criterion = "maxclust". Pathway enrichment analysis of the proteins/PTM sites in each cluster was performed using GSEAPy.enrichr (Python module) with the parameter: gene_sets = "KEGG_2021_Human". Pathways with BH adjusted $P < 0.05$ were considered as significantly regulated.

Association between proteomic subgroup and clinical outcome

To compare the survival outcomes among the three proteomic subgroups, log-rank test was employed. Then, Kaplan-Meier survival curves were generated using the KaplanMeierFitter function in Python to visualize the survival difference between three subgroups of patients.

Tumor-NAT samples differential expression analysis

Differential expression analysis was performed for 30 paired CC tumors and NATs using the Student's t test. P -values were adjusted using the BH method. Each feature was required to be non-missing in at least 50% of the paired samples. Proteins or phosphosites/acetylsites (collapsed into gene-level) differentially expressed between tumors and NATs (BH adjusted $P < 0.01$, $\text{Log}_2 \text{FC} > 1$ or < -1) were further used for over-representation analysis by WebGestalt¹¹⁰.

Multi-omics data integration

Analysis of mRNA-protein expression correlation. Spearman's correlation coefficient was used to measure the correlation between each mRNA-protein pair across all 132 tumors (using the RNA-seq and TMT

proteomic data). In addition, a *P*-value was calculated for each correlation coefficient, which was then adjusted for multiple testing using the BH method. The correlation between a mRNA-protein pair was called significant if the adjusted *P*-value was found to be <0.05. A total number of 6089 mRNA-protein-matched genes were examined, with a median Spearman's correlation coefficient of *r* = 0.40. Moreover, the mRNA and protein levels were found to be positively correlated for most (95.9%) mRNA-protein pairs, and 81.7% showed significant positive correlation.

Analysis of the *cis/trans* effect of CNAs. The effects of SCNAs on mRNA, protein, and phosphoprotein abundance levels in either *cis* (within the same aberrant locus) or *trans* (remote locus) mode were visualized by multiOmicsViz (R package). Spearman's correlation coefficients and the associated *P* values after adjustment for multiple testing were calculated for all possible CNA-mRNA/protein/phosphoprotein pairs. BH adjusted *P* < 0.01 was used to identify the significant correlations. A total of 130 tumor samples with both mRNA, WES, and proteomic data were included in this analysis.

Identification of CNG-*Cis* and CNL-*Cis* genes

Among the 8035 genes located within amplification foci, 337 genes showed significantly higher abundance in tumor samples compared with NATs at both mRNA and protein levels (Wilcoxon rank-sum test, BH adjusted *P* < 0.05), and 227 genes also represented significant correlation of copy number with corresponding RNA levels, including 178 that displayed concordant protein expression (Spearman's correlation > 0, BH adjusted *P* < 0.05). These 178 genes were identified as CNG-*Cis* genes. The same screening method was used to identify 28 CNL-*Cis* genes that were downregulated in tumor samples.

Analysis of the effects of arm-level CNAs

Arm-level CNAs were detected using Log₂ (tumor/blood) = 0.3/−0.3 (corresponding to arm-gain/loss, respectively) as the cutoff. Then, for each detected arm gain/loss event, the tumor samples were accordingly divided into two subgroups and the proteins whose expression was significantly associated with this event were identified using Student's *t* test (with FDR < 0.05 as the cutoff). KEGG pathway enrichment analysis of the arm gain/loss-associated proteins supported by both TMT and DIA data was performed using GSEapy.enricher (python module) with the parameter `gene_sets = "KEGG_2021_Human"`¹³. Pathways with FDR < 0.05 were regarded as arm gain/loss-associated pathways.

Identification of radioresponse-related biomarkers

For each candidate protein, we stratified the 62 radical radiotherapy CC patients into two groups using its median expression level as the cutoff. Then, Cox proportional hazards regression for PFS data was performed to identify radioresponse-related biomarkers of CC. Proteins detected in at least more than half of the samples are used here. The filter criteria for biomarker were: *P*-value of Cox proportional hazards regression < 0.05; Log (HR) upper 95% < 0 and Log (HR) lower 95% > 0 for candidate favorable and unfavorable proteins, respectively. Finally, 37 favorable and 55 unfavorable radioresponse-related biomarkers were identified, supported by both TMT and DIA proteomic data. Kaplan-Meier curves (log-rank test) were used to visualize the prognosis difference between radical radiotherapy patients with high and low protein expression.

Risk scoring model based on radioresponse-related biomarkers

We identified 92 radioresponse-related biomarkers, comprising 55 prognostically unfavorable proteins and 37 favorable ones. Unfavorable score and favorable score of each sample were calculated

according to the following formulas.

$$\text{unfavorable score}_j = \frac{\sum_{i=1}^{m_j} (Z_{ij} - \text{median}(Z_{i,j=1,2,\dots,n_i}))}{m_j} \quad (4)$$

$$\text{favorable score}_j = \frac{\sum_{i=1}^{p_j} (Z_{ij} - \text{median}(Z_{i,j=1,2,\dots,q_i}))}{p_j} \quad (5)$$

Z_{ij} donate z-statistic of protein *i* in sample *j*. m_j and p_j represent the number of prognostically unfavorable or favorable proteins detected in sample *j*, respectively. n_i and q_i represent the number of samples in which unfavorable or favorable protein *i* was detected, respectively.

Then we calculated the risk score for each sample, the risk scoring model based on above formulas was successfully constructed.

$$\text{risk score}_j = \text{unfavorable score}_j - \text{favorable score}_j \quad (6)$$

62 radical radiotherapy patients were divided into high-risk and low-risk groups based on the median of the risk scores. Next, we validated the prognostic predictive ability of the risk scoring model using the TCGA CC. Firstly, we performed z-score transformation on the FPKM matrix at the gene level. Subsequently, risk scores were calculated for TCGA samples. 291 TCGA CC samples were divided into high-risk and low-risk groups based on the median of the risk scores. Kaplan-Meier analysis (log-rank test) was utilized to illustrate the prognosis disparity between the two groups.

Identification of subgroup-specific kinases

We applied KSEA to estimate the change in a kinase's activity based on the collective phosphorylation changes of its identified substrates using the method proposed in KSEA app¹². Here all Kinase-Substrate annotations from PhosphoSitePlus and from NetworkKIN were used. To identify subgroup-specific kinases, we first performed a Student's *t* test for each phosphosite between its phosphorylation levels in the samples of a specific subgroup and all other samples. Then, a kinase's normalized KSEA score is calculated as:

$$\text{KSEA Score} = \frac{(\bar{t} - \bar{p})\sqrt{m}}{\delta}, \quad (7)$$

in which \bar{t} denotes the mean t-statistic of the known substrate phosphosites of this kinase, \bar{p} represents the mean t-statistic of all phosphosites in the dataset, *m* denotes the total number of known substrate phosphosites of this kinase, δ denotes the standard deviation of the t-statistics across all phosphosites in the dataset. Subsequently, the corresponding *P*-value is determined by assessing the one-tailed probability of having a more extreme score than the one measured based on the normal distribution. Finally, subgroup-specific kinases were selected based on: protein expression levels were found to be significantly higher in a specific subgroup than in other tumor samples (Student's *t* test, *P*-value < 0.05); *P*-value of KSEA score < 0.05.

Pathway enrichment of HPV gene-associated proteins

Spearman's correlation coefficient was calculated between HPV E6/L1 mRNA level and protein abundance of human genes, which was subsequently used to rank the human genes. Then, the resulting ranked gene list was used as input for the GSEapy.prerank (Python module) to perform pathway enrichment analysis using the following detailed settings: `gene_sets = "KEGG_2021_Human"`; `min_size = 5`; `max_size = 1000`; and `permutation_num = 1000`. Pathways with FDR < 0.05 were regarded as HPV gene-associated pathways.

Subgrouping analysis based on HPV gene expression

CC tumors were stratified into two subgroups based on the mRNA expression levels of HPV E2 and E5 genes by hierarchical clustering (using the Python function `scipy.cluster.hierarchy.linkage` with parameters `method = "ward"` and `metric = "euclidean"`). The resulting linkage matrix was then used as input for `scipy.cluster.hierarchy.fcluster` to create flat clusters, with the criterion parameter set to "maxclust" and "t = 2". Then, we employed either Chi-square test or Fisher's exact test to test the association between the detected subgroups and the clinical features of patients, depending on the number of categories in each clinical feature. For continuous clinical features, we used the Student's t test. *P*-values were adjusted for multiple testing using the BH method.

Functional experiments

siRNA Interference. The siRNAs were synthesized by GenePharma. All siRNA transfections were performed with Lipofectamine 2000 (Invitrogen, 11668019) at 50 nM final concentration according to the manufacturer's protocol. The siRNA sequence information was shown in Supplementary Data 11.

Plasmids

The HA-tagged coding sequence of human HATs was cloned into pCDNA3.1 vector. The Flag-tagged coding sequence of human FOSL2-WT or the relevant FOSL2-K222R, FOSL2-K222Q mutant were cloned into the lentiviral vector pLEX-MCS-CMV-puro (Addgene) to generate corresponding expression plasmids. The Flag-tagged coding sequence of human PRKCB were cloned into pLVX-IRES-Neo (Clontech) vector.

Cell transfection and lentiviral production

Cells were transfected with plasmid vectors using Lipofectamine 2000 according to the manufacturer's instructions. To generate the lentivirus, HEK293T cells were co-transfected with psPAX2, pMD2.G, and recombinant lentiviral vectors. Forty-eight hours after transfection, the lentiviral supernatants were harvested and passed through a 0.45 µm filter. The lentiviruses were added to media supplemented with 10 µg/mL polybrene (Beyotime, C0351-1 mL) to transduce SiHa cells following the manufacturer's instructions.

Constructions of stable cell lines

To establish FOSL2-WT or FOSL2-K222R, FOSL2-K222Q mutant SiHa cell line models, SiHa cells were infected with respective lentivirus and further selected with 0.6 µg/mL puromycin (Beyotime Biotechnology, ST551-10mg). To establish inducible PRKCB-overexpressing SiHa cell strains, SiHa cells were infected with pLVX-IRES-Neo-PRKCB lentivirus and further selected with 0.8 mg/mL G418 (Gibco, 10131035).

Cell proliferation assay

For cell growth assays, experimental and control cells were plated in 96-well plate (2×10^3 cells per well). For measurement, CCK-8 solution (Beyotime Biotechnology, C0039) at the final concentration of 10% was added to the wells, and absorbance at 450 nm was measured 2 h after incubation to represent the relative cell numbers.

Colony formation assay

To evaluate the clonogenic capacity of experimental and control cells, 500 cells were plated in six-well plate in triplicated. After 10–12 days of continuous culture, the cells were fixed with methanol and stained with 1% crystal violet (Sigma, C0775). The colonies were counted, and each assay was repeated at least three times independently.

Cell cycle analysis

For cell cycle assays, experimental and control cells were seeded in 6-well plate (2×10^5 cells per well). The cells were collected by

trypsinization and fixed in ice-cold 70% ethanol overnight at 4 °C. The cell cycle detection kit (Nanjing KeyGen Biotech, KGA512) and BD LSR II flowcytometer (BD Biosciences) were applied to detect cell cycle distribution. The cell cycle result profiles were analyzed using ModFit LT 3.1 (Verity Software House).

IHC

The fresh tumors and NATs were fixed in 10% neutral buffered formalin, embedded in paraffin, and used to prepare sections that were 4 µm thick. The sections were deparaffinized, rehydrated, and then immersed in a 3% hydrogen peroxide solution for 10 min. For antigen retrieval, the sections were heated in citrate buffer (pH 6.0) at 95 °C for 25 min and then cooled at room temperature for 60 min. After each incubation step, the sections were washed with PBS (pH 7.4). Subsequently, the sections were incubated overnight at 4 °C with the anti-PRKCB antibody (Abcam, ab195039). Immunostaining was carried out following the instructions provided by the manufacturer (PV-9000 Polymer Detection System, Zhongshan Golden Bridge). The sections were counterstained with hematoxylin, dehydrated using graded ethanol, and sealed with neutral resin. Two investigators independently assessed the IHC staining of PRKCB on the sections.

Immunoblot and immunoprecipitation

Cells were lysed in EBC lysis buffer (50 mM Tris HCl, pH 8.0, 120 mM NaCl, 0.5% NP-40) supplemented with protease inhibitors (Selleck Chemicals) and phosphatase inhibitors (Selleck Chemicals), followed by pulse sonication for 10 s. 30 mg total protein were separated by 10% SDS-PAGE gel and blotted with indicated primary antibodies. For immunoprecipitation (IP), the whole cell lysates (WCL) were immunoprecipitated with anti-Flag M2 agarose beads for 2 h in the presence of 2 µM TSA and 10 mM nicotinamide (NAM). The pellet was then washed with NETN buffer (20 mM Tris-HCl, pH 8.0, 100 mM NaCl, 0.5% NP-40, 1 mM EDTA) for four times and analyzed by immunoblotting or MS.

qRT-PCR

Total RNA was extracted following the manufacturer's instructions using the RNeasy Tissue & Cell Kit (Qiagen, RNeasy9605). Subsequently, the isolated RNA served as a template for reverse-transcription reactions employing the HiFiScript cDNA Synthesis Kit (Cwbiotech, CW2569M). qRT-PCR analysis was conducted using the TB Green® Premix Ex Taq™ II (TaKaRa, RR820A) and CFX96 Real-Time System (Bio-Rad). β-actin was served as an internal control. The relative quantification of gene expression was analyzed by the $2^{-\Delta\Delta Ct}$ method. The primers used for qRT-PCR analyses are as following:

WNT5A Forward: 5'-GCCAGTATCAATTCGACATCG-3',
Reverse: 5'-TCACCGGTATGTGAAGGC-3'
FOSL2 Forward: 5'-CAGAAATTCGGGTAGATATGCC-3'
Reverse: 5'-GGTATGGGTTGGACATGGAGG-3'
β-actin Forward: 5'-AGAGCTACGAGCTGCCTGAC-3',
Reverse: 5'-AGCACTGTGTTGGCGTACAG-3'

Xenograft tumorigenesis assay

Five-week-old female BALB/c nude mice (HFK Bioscience) were maintained in pathogen-free conditions. All animals were acclimated for 1 week before experiments. 3×10^6 experimental and control SiHa cells in 100 µL PBS were subcutaneously inoculated at the flank of randomly grouped nude mice. Tumor size was measured every 3 days with a caliper and tumor volumes were calculated by the formula: volume = (width)² × length × 0.52. The maximum tumor burden allowed by the ethics committee did not exceed 1500 mm³. When the tumor burden reached 1500 mm³, the mice were euthanized, and the tumors were dissected for further analysis. After the mice were euthanized, the transplanted tumors were weighed and

photographed. All animal procedures were performed according to the guidelines approved by the National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College and adhered to the National Institutes of Health Guide for the care and use of laboratory animals.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The WES and RNA-seq raw data have been deposited in the Genome Sequence Archive in National Genomics Data Center, China National Center for Bioinformation/Beijing Institute of Genomics, Chinese Academy of Sciences^{113,114} under accession number [HRA005516](https://www.genome.gov/27532016). The raw WES and RNA-seq data are available for research use only with restricted access, which can be obtained through the Data Access Committees of the GSA-human database. Data access is open to all non-profit researchers in compliance with the guidelines set by GSA-human. Alternatively, the user may also directly contact the corresponding author. Once access has been approved, the data will be available to download for 2 months. The proteomics, phosphoproteomics, and acetylproteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE¹¹⁵ partner repository with the dataset identifier [PXD055203](https://www.ebi.ac.uk/PRIDE/archive). Transcriptomics and survival data of TCGA CC were downloaded from Xena [<https://xenabrowser.net/datapages/>]. For more information in this study, please contact the corresponding authors. Source data are provided with this paper.

Code availability

The codes are available at https://github.com/guixiuqi/CC_multiomics.

References

- Sung, H. et al. Global Cancer Statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **71**, 209–249 (2021).
- zur Hausen, H. Papillomavirus infections—a major cause of human cancers. *Biochim. Biophys. Acta* **1288**, F55–F78 (1996).
- Brisson, M. et al. Impact of HPV vaccination and cervical screening on cervical cancer elimination: a comparative modelling analysis in 78 low-income and lower-middle-income countries. *Lancet* **395**, 575–590 (2020).
- Arbyn, M. et al. Estimates of incidence and mortality of cervical cancer in 2018: a worldwide analysis. *Lancet Glob. Health* **8**, e191–e203 (2020).
- McNeil, C. New standard of care for cervical cancer sets stage for next questions. *J. Natl. Cancer Inst.* **91**, 500–501 (1999).
- Rose, P. G. et al. Concurrent cisplatin-based radiotherapy and chemotherapy for locally advanced cervical cancer. *N. Engl. J. Med.* **340**, 1144–1153 (1999).
- Whitney, C. W. et al. Randomized comparison of fluorouracil plus cisplatin versus hydroxyurea as an adjunct to radiation therapy in stage IIB-IVA carcinoma of the cervix with negative para-aortic lymph nodes: a Gynecologic Oncology Group and Southwest Oncology Group study. *J. Clin. Oncol.* **17**, 1339–1348 (1999).
- Shrivastava, S. et al. Cisplatin chemoradiotherapy vs radiotherapy in FIGO stage IIIB squamous cell carcinoma of the uterine cervix: a randomized clinical trial. *JAMA Oncol.* **4**, 506–513 (2018).
- Potter, R. et al. MRI-guided adaptive brachytherapy in locally advanced cervical cancer (EMBRACE-I): a multicentre prospective cohort study. *Lancet Oncol.* **22**, 538–547 (2021).
- Mileshkin, L. R. et al. Adjuvant chemotherapy following chemoradiotherapy as primary treatment for locally advanced cervical cancer versus chemoradiotherapy alone (OUTBACK): an international, open-label, randomised, phase 3 trial. *Lancet Oncol.* **24**, 468–482 (2023).
- Shi, Y. et al. A genome-wide association study identifies two new cervical cancer susceptibility loci at 4q12 and 17q12. *Nat. Genet.* **45**, 918–922 (2013).
- Ojesina, A. I. et al. Landscape of genomic alterations in cervical carcinomas. *Nature* **506**, 371 (2014).
- Hu, Z. et al. Genome-wide profiling of HPV integration in cervical cancer identifies clustered genomic hot spots and a potential microhomology-mediated integration mechanism. *Nat. Genet.* **47**, 158–163 (2015).
- Cancer Genome Atlas Research, N. et al. Integrated genomic and molecular characterization of cervical cancer. *Nature* **543**, 378–384 (2017).
- Gagliardi, A. et al. Analysis of Ugandan cervical carcinomas identifies human papillomavirus clade-specific epigenome and transcriptome landscapes. *Nat. Genet.* **52**, 800–810 (2020).
- Fan, J. et al. Multi-omics characterization of silent and productive HPV integration in cervical cancer. *Cell Genom.* **3**, 100211 (2023).
- Mertins, P. et al. Proteogenomics connects somatic mutations to signalling in breast cancer. *Nature* **534**, 55–62 (2016).
- Krug, K. et al. Proteogenomic landscape of breast cancer tumorigenesis and targeted therapy. *Cell* **183**, 1436–1456.e1431 (2020).
- Dou, Y. et al. Proteogenomic characterization of endometrial carcinoma. *Cell* **180**, 729–748.e726 (2020).
- Dou, Y. et al. Proteogenomic insights suggest druggable pathways in endometrial carcinoma. *Cancer Cell* **41**, 1586–1605.e15 (2023).
- Zhang, H. et al. Integrated proteogenomic characterization of human high-grade serous ovarian. *Cancer Cell* **166**, 755–765 (2016).
- Chowdhury, S. et al. Proteogenomic analysis of chemo-refractory high-grade serous ovarian cancer. *Cell* **186**, 3476–3498.e3435 (2023).
- Froimchuk, E., Jang, Y. & Ge, K. Histone H3 lysine 4 methyltransferase KMT2D. *Gene* **627**, 337–342 (2017).
- Fagan, R. J. & Dingwall, A. K. COMPASS ascending: emerging clues regarding the roles of MLL3/KMT2C and MLL2/KMT2D proteins in cancer. *Cancer Lett.* **458**, 56–65 (2019).
- Mendiratta, G. et al. Cancer gene mutation frequencies for the US population. *Nat. Commun.* **12**, 5961 (2021).
- Clark, D. J. et al. Integrated proteogenomic characterization of clear cell renal cell carcinoma. *Cell* **179**, 964–983.e931 (2019).
- Li, C., Guo, L., Li, S. & Hua, K. Single-cell transcriptomics reveals the landscape of intra-tumoral heterogeneity and transcriptional activities of ECs in CC. *Mol. Ther. Nucleic Acids* **24**, 682–694 (2021).
- Ou, Z. et al. Single-nucleus RNA sequencing and spatial transcriptomics reveal the immunological microenvironment of cervical squamous cell carcinoma. *Adv. Sci.* **9**, e2203040 (2022).
- Liu, C. et al. Single-cell dissection of cellular and molecular features underlying human cervical squamous cell carcinoma initiation and progression. *Sci. Adv.* **9**, eadd8977 (2023).
- Carter, S. L. et al. Absolute quantification of somatic DNA alterations in human cancer. *Nat. Biotechnol.* **30**, 413–421 (2012).
- Chen, R. J. et al. Influence of histologic type and age on survival rates for invasive cervical carcinoma in Taiwan. *Gynecol. Oncol.* **73**, 184–190 (1999).
- Jung, E. J. et al. Cervical adenocarcinoma has a poorer prognosis and a higher propensity for distant recurrence than squamous cell carcinoma. *Int. J. Gynecol. Cancer* **27**, 1228–1236 (2017).
- Gao, Q. et al. Integrated proteogenomic characterization of hbv-related hepatocellular carcinoma. *Cell* **179**, 561–577.e522 (2019).

34. Inkman, M. J. et al. HPV-EM: an accurate HPV detection and genotyping EM algorithm. *Sci. Rep.* **10**, 14340 (2020).
35. Hawley-Nelson, P., Vousden, K. H., Hubbert, N. L., Lowy, D. R. & Schiller, J. T. HPV16 E6 and E7 proteins cooperate to immortalize human foreskin keratinocytes. *EMBO J.* **8**, 3905–3910 (1989).
36. Tanaka, N. et al. Wee-1 kinase inhibition sensitizes high-risk HPV+ HNSCC to apoptosis accompanied by downregulation of MCL-1 and XIAP antiapoptotic proteins. *Clin. Cancer Res.* **21**, 4831–4844 (2015).
37. Diab, A. et al. FOXM1 drives HPV+ HNSCC sensitivity to WEE1 inhibition. *Proc. Natl. Acad. Sci. USA* **117**, 28287–28296 (2020).
38. Celeste, A. et al. Histone H2AX phosphorylation is dispensable for the initial recognition of DNA breaks. *Nat. Cell. Biol.* **5**, 675–679 (2003).
39. McBride, A. A. Human papillomaviruses: diversity, infection and host interactions. *Nat. Rev. Microbiol.* **20**, 95–108 (2022).
40. Baker, C. C. et al. Structural and transcriptional analysis of human papillomavirus type 16 sequences in cervical carcinoma cell lines. *J. Virol.* **61**, 962–971 (1987).
41. Durst, M., Croce, C. M., Gissmann, L., Schwarz, E. & Huebner, K. Papillomavirus sequences integrate near cellular oncogenes in some cervical carcinomas. *Proc. Natl. Acad. Sci. USA* **84**, 1070–1074 (1987).
42. Schwarz, E. et al. Structure and transcription of human papillomavirus sequences in cervical carcinoma cells. *Nature* **314**, 111–114 (1985).
43. Feng, D. et al. Regulation of Wnt/PCP signaling through p97/VCP-KBTBD7-mediated Vangl ubiquitination and endoplasmic reticulum-associated degradation. *Sci. Adv.* **7**, eabg2099 (2021).
44. Liu, C., Wang, X. & Zhang, Y. The roles of HK2 on tumorigenesis of cervical cancer. *Technol. Cancer Res. Treat.* **18**, 1533033819871306 (2019).
45. Chen, Q. et al. Hexokinases 2 promoted cell motility and distant metastasis by elevating fibronectin through Akt1/p-Akt1 in cervical cancer cells. *Cancer Cell Int.* **21**, 600 (2021).
46. Kim, B. H. & Chang, J. H. Differential effect of GLUT1 over-expression on survival and tumor immune microenvironment of human papilloma virus type 16-positive and -negative cervical cancer. *Sci. Rep.* **9**, 13301 (2019).
47. Kanjanapan, Y. et al. Glut-1 expression in small cervical biopsies is prognostic in cervical cancers treated with chemoradiation. *Clin. Transl. Radiat. Oncol.* **2**, 53–58 (2017).
48. Reyna-Hernandez, M. A. et al. GLUT1, LDHA, and MCT4 expression is deregulated in cervical cancer and precursor lesions. *J. Histochem. Cytochem.* **70**, 437–446 (2022).
49. Priego-Hernandez, V. D. et al. Expression of HIF-1 α and genes involved in glucose metabolism is increased in cervical cancer and HPV-16-positive cell lines. *Pathogens* **12**, 33 (2022).
50. Delvecchio, M., Gaucher, J., Aguilar-Gurrieri, C., Ortega, E. & Panne, D. Structure of the p300 catalytic core and implications for chromatin targeting and HAT regulation. *Nat. Struct. Mol. Biol.* **20**, 1040–1046 (2013).
51. Thompson, P. R. et al. Regulation of the p300 HAT domain via a novel activation loop. *Nat. Struct. Mol. Biol.* **11**, 308–315 (2004).
52. Birnhuber, A., Biasin, V., Schnoegl, D., Marsh, L. M. & Kwapiszewska, G. Transcription factor Fra-2 and its emerging role in matrix deposition, proliferation and inflammation in chronic lung diseases. *Cell. Signal.* **64**, 109408 (2019).
53. Sarode, P. et al. Reprogramming of tumor-associated macrophages by targeting beta-catenin/FOSL2/ARID5A signaling: a potential treatment of lung cancer. *Sci. Adv.* **6**, eaaz6105 (2020).
54. Wan, X. et al. FOSL2 promotes VEGF-independent angiogenesis by transcriptionally activating Wnt5a in breast cancer-associated fibroblasts. *Theranostics* **11**, 4975–4991 (2021).
55. Song, L. N. et al. Hsa_circ_0003998 promotes epithelial to mesenchymal transition of hepatocellular carcinoma by sponging miR-143-3p and PCBP1. *J. Exp. Clin. Cancer Res.* **39**, 114 (2020).
56. Faridi, M. H. et al. CD11b activation suppresses TLR-dependent inflammation and autoimmunity in systemic lupus erythematosus. *J. Clin. Investig.* **127**, 1271–1283 (2017).
57. Faridar, A. et al. Ex vivo expanded human regulatory T cells modify neuroinflammation in a preclinical model of Alzheimer's disease. *Acta Neuropathol. Commun.* **10**, 144 (2022).
58. Wahl, M. C. & Luhrmann, R. SnapShot: spliceosome dynamics I. *Cell* **161**, 1474–e1471 (2015).
59. Chen, W. et al. Transcriptome-wide interrogation of the functional intronome by spliceosome profiling. *Cell* **173**, 1031–1044.e1013 (2018).
60. Dowling, C. M. et al. Protein kinase C beta II suppresses colorectal cancer by regulating IGF-1 mediated cell survival. *Oncotarget* **7**, 20919–20933 (2016).
61. Ke, G. et al. MiR-181a confers resistance of cervical cancer to radiation therapy through targeting the pro-apoptotic PRKCD gene. *Oncogene* **32**, 3019–3027 (2013).
62. Bowden, S. J. et al. Genetic variation in cervical preinvasive and invasive disease: a genome-wide association study. *Lancet Oncol.* **22**, 548–557 (2021).
63. Qing, S. et al. Proteomic identification of potential biomarkers for cervical squamous cell carcinoma and human papillomavirus infection. *Tumour Biol.* **39**, 1010428317697547 (2017).
64. Guzel, C. et al. Proteomic alterations in early stage cervical cancer. *Oncotarget* **9**, 18128–18147 (2018).
65. Kelly, A. D. et al. Pan-cancer landscape of CD274 (PD-L1) rearrangements in 283,050 patient samples, its correlation with PD-L1 protein expression, and immunotherapy response. *J. Immunother. Cancer* **9**, e003550 (2021).
66. Howie, H. L. et al. Beta-HPV 5 and 8 E6 promote p300 degradation by blocking AKT/p300 association. *PLoS Pathog.* **7**, e1002211 (2011).
67. Ou, H. D., May, A. P. & O'Shea, C. C. The critical protein interactions and structures that elicit growth deregulation in cancer and viral replication. *Wiley Interdiscip. Rev. Syst. Biol. Med.* **3**, 48–73 (2011).
68. Lee, Y. Y. et al. Anti-tumor effects of Wee1 kinase inhibitor with radiotherapy in human cervical cancer. *Sci. Rep.* **9**, 15394 (2019).
69. Mah, L. J., El-Osta, A. & Karagiannis, T. C. gammaH2AX: a sensitive molecular marker of DNA damage and repair. *Leukemia* **24**, 679–686 (2010).
70. Satpathy, S. et al. A proteogenomic portrait of lung squamous cell carcinoma. *Cell* **184**, 4348–4371.e4340 (2021).
71. Wang, Z. et al. Acetylation of PHF5A modulates stress responses and colorectal carcinogenesis through alternative splicing-mediated upregulation of KDM3A. *Mol. Cell* **74**, 1250–1263.e1256 (2019).
72. Chai, X. et al. Quantitative acetylome analysis reveals histone modifications that may predict prognosis in hepatitis B-related hepatocellular carcinoma. *Clin. Transl. Med.* **11**, e313 (2021).
73. Dou, C. et al. P300 acetyltransferase mediates stiffness-induced activation of hepatic stellate cells into tumor-promoting myofibroblasts. *Gastroenterology* **154**, 2209–2221.e2214 (2018).
74. Li, M. et al. High expression of transcriptional coactivator p300 correlates with aggressive features and poor prognosis of hepatocellular carcinoma. *J. Transl. Med.* **9**, 5 (2011).
75. Debes, J. D. et al. p300 in prostate cancer proliferation and progression. *Cancer Res.* **63**, 7638–7640 (2003).

76. Wang, L. et al. The leukemogenicity of AML1-ETO is dependent on site-specific lysine acetylation. *Science* **333**, 765–769 (2011).
77. Ogiwara, H. et al. Targeting p300 addiction in CBP-deficient cancers causes synthetic lethality by apoptotic cell death due to abrogation of MYC expression. *Cancer Discov.* **6**, 430–445 (2016).
78. Lasko, L. M. et al. Discovery of a selective catalytic p300/CBP inhibitor that targets lineage-specific tumours. *Nature* **550**, 128–132 (2017).
79. Giotopoulos, G. et al. The epigenetic regulators CBP and p300 facilitate leukemogenesis and represent therapeutic targets in acute myeloid leukemia. *Oncogene* **35**, 279–289 (2016).
80. Zhong, J. et al. p300 acetyltransferase regulates androgen receptor degradation and PTEN-deficient prostate tumorigenesis. *Cancer Res.* **74**, 1870–1880 (2014).
81. He, H. et al. Selective p300 inhibitor C646 inhibited HPV E6-E7 genes, altered glucose metabolism and induced apoptosis in cervical cancer cells. *Eur. J. Pharmacol.* **812**, 206–215 (2017).
82. Zhang, L. et al. Identification of lysine acetylation in cervical cancer by label-free quantitative proteomics. *Cancer Cell Int.* **20**, 182 (2020).
83. Young, R. H. & Scully, R. E. Invasive adenocarcinoma and related tumors of the uterine cervix. *Semin. Diagn. Pathol.* **7**, 205–227 (1990).
84. Cheng, Y. et al. The role of high-risk human papillomavirus-related long non-coding RNAs in the prognosis of cervical squamous cell carcinoma. *DNA Cell Biol.* **39**, 645–653 (2020).
85. Liu, C. et al. Single-cell RNA-sequencing reveals radiochemotherapy-induced innate immune activation and MHC-II upregulation in cervical cancer. *Signal Transduct. Target. Ther.* **8**, 44 (2023).
86. Wang, J. et al. PRKCB is relevant to prognosis of lung adenocarcinoma through methylation and immune infiltration. *Thorac. Cancer* **13**, 1837–1849 (2022).
87. Lin, G., Brownsey, R. W. & MacLeod, K. M. Regulation of mitochondrial aconitase by phosphorylation in diabetic rat heart. *Cell. Mol. Life Sci.* **66**, 919–932 (2009).
88. Kowalczyk, J. E. et al. Protein kinase C beta in postischemic brain mitochondria. *Mitochondrion* **12**, 138–143 (2012).
89. Patergnani, S. et al. PRKCB/protein kinase C, beta and the mitochondrial axis as key regulators of autophagy. *Autophagy* **9**, 1367–1385 (2013).
90. Yang, X. et al. Prognostic nomograms predicting survival in patients with locally advanced cervical squamous cell carcinoma: the first nomogram compared with revised figo 2018 staging system. *Front. Oncol.* **10**, 591700 (2020).
91. Abu-Rustum, N. R. et al. NCCN Guidelines(R) insights: cervical cancer, Version 1.2024. *J. Natl. Compr. Canc. Netw.* **21**, 1224–1233 (2023).
92. Serarslan, A., Gursel, B., Meydan, D. & Ozbek Okumus, N. Radical radiotherapy in patients with cervix uteri carcinoma: experience of Ondokuz Mayıs University. *BMC Cancer* **19**, 1208 (2019).
93. Chai, Y. et al. Radical hysterectomy with adjuvant radiotherapy versus radical radiotherapy for FIGO stage IIB cervical cancer. *BMC Cancer* **14**, 63 (2014).
94. Thakur, S. S. et al. Deep and highly sensitive proteome coverage by LC-MS/MS without prefractionation. *Mol. Cell. Proteomics* **10**, M110 003699 (2011).
95. Meier, F. et al. diaPASEF: parallel accumulation-serial fragmentation combined with data-independent acquisition. *Nat. Methods* **17**, 1229–1236 (2020).
96. Tyanova, S., Temu, T. & Cox, J. The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nat. Protoc.* **11**, 2301–2319 (2016).
97. Demichev, V., Messner, C. B., Vernardis, S. I., Lilley, K. S. & Ralser, M. DIA-NN: neural networks and interference correction enable deep proteome coverage in high throughput. *Nat. Methods* **17**, 41–44 (2020).
98. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
99. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38**, e164 (2010).
100. Lawrence, M. S. et al. Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* **499**, 214–218 (2013).
101. Talevich, E., Shain, A. H., Botton, T. & Bastian, B. C. CNVkit: genome-wide copy number detection and visualization from targeted dna sequencing. *PLoS Comput. Biol.* **12**, e1004873 (2016).
102. Mermel, C. H. et al. GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol.* **12**, R41 (2011).
103. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).
104. Yoshihara, K. et al. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat. Commun.* **4**, 2612 (2013).
105. Aran, D., Hu, Z. & Butte, A. J. xCell: digitally portraying the tissue cellular heterogeneity landscape. *Genome Biol.* **18**, 220 (2017).
106. Li, M. et al. MAP: model-based analysis of proteomic data to detect proteins with significant abundance changes. *Cell Discov.* **5**, 40 (2019).
107. Ma, W. et al. DreamAI: algorithm for the imputation of proteomics data. *bioRxiv*, <https://doi.org/10.1101/2020.07.21.214205> (2020).
108. Wilkerson, M. D. & Hayes, D. N. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. *Bioinformatics* **26**, 1572–1573 (2010).
109. Chen, H. et al. HyperChIP: identification of hypervariable signals across ChIP-seq or ATAC-seq samples. *Genome Biol.* **23**, 62 (2022).
110. Liao, Y., Wang, J., Jaehnig, E. J., Shi, Z. & Zhang, B. WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs. *Nucleic Acids Res.* **47**, W199–W205 (2019).
111. Fang, Z., Liu, X. & Peltz, G. GSEAPy: a comprehensive package for performing gene set enrichment analysis in Python. *Bioinformatics* **39**, btac757 (2023).
112. Wiredja, D. D., Koyuturk, M. & Chance, M. R. The KSEA app: a web-based tool for kinase activity inference from quantitative phosphoproteomics. *Bioinformatics* **33**, 3489–3491 (2017).
113. Chen, T. et al. The genome sequence archive family: toward explosive data growth and diverse data types. *Genom. Proteom. Bioinform.* **19**, 578–583 (2021).
114. CNCB-NGDC Members and Partners Database resources of the national genomics data center, China National Center for bioinformation in 2022. *Nucleic Acids Res.* **50**, D27–D38 (2022).
115. Perez-Riverol, Y. et al. The PRIDE database resources in 2022: a hub for mass spectrometry-based proteomics evidences. *Nucleic Acids Res.* **50**, D543–D552 (2022).

Acknowledgements

This study was supported by the National Key Research and Development Program of China (2022YFA1302900 to H.Z.; 2020YFA0803203, 2019YFA0802102 to D.G.), the Shanghai Young Excellent Academic Leader Program (20XD1424900 to H.Z.), the National Natural Science Foundation of China (82173334 to L.W.; 32370705 to Z.S.; 81925029, 82230098, 32221002 to D.G.; 22425703 to H.Z.), the Strategic Priority

Research Program of Chinese Academy of Sciences (XDB0850000 to H.Z.; XDB38040100 to Z.S.), the CAS Project for Young Scientists in Basic Research (YSBR-014 to D.G.) and the Shanghai Municipal Science and Technology Major Projects. We would like to thank the Institutional Technology Service Center of Shanghai Institute of Materia Medica for all the technical support. This work was done under the auspices of the US National Cancer Institute's International Cancer Proteogenome Consortium. CPTAC collaborates with international organizations/institutions to accelerate the understanding of the molecular basis of cancer through the application of proteogenomics, standards development, and publicly available datasets. The authors acknowledge the use of Biorender for creating Supplementary Fig. 1a.

Author contributions

The project was conceived and supervised by D.G., L.W., Z.S., and H.Z. Inclusion of patients, clinical sampling and sample prep was conducted by J.Y., L.W., J.A., M.H., J.Zuo, G.Y. and N.L. Pathological evaluation and IHC were performed by Y.S., J.J., J.Y., and L.W. Clinical sampling, inclusion of patients and clinical data review for the validation cohort were conducted by J.Y., L. Deng, J. Zeng and Y. Zhao. Genomics and transcriptomics data generation was performed by J.Y. and X.Gui. MS sample preparation, MS data generation and searching was conducted by Q.L., Z.Y., J.Y., J.G., and Y.C. DIA proteomics MS data was acquired by J.Y., X.L. and X.D. Analysis of the sequencing data was performed by X.Gui, Q.L., J.Y., X.Guo., S.Z., D.G., and Z.S. Proteogenomics analysis was conducted by X.Gui, Q.L., J.Y., Z.S. and H.Z. In vitro cell line experiments, animal model and biological experiments were coordinated and performed by J.Y., Y.Zou, K.W. and D.G. Resource and scientific advice was provided by A.R., H.R., B.Z., P.W., L. Ding and Y.L. Funding support was provided by H.Z., L.W., Z.S. and D.G. The proteogenomics data upload was conducted by J.Y., Q.L., and Z.Y. The paper was written by J.Y., X.Gui, Y.Zou, Q.L., H.Z., and D.G.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-024-53830-0>.

Correspondence and requests for materials should be addressed to Hu Zhou, Zhen Shao, Lingying Wu or Daming Gao.

Peer review information *Nature Communications* thanks the anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2024

¹Department of Gynecologic Oncology, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China. ²State Key Laboratory of Molecular Oncology, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China. ³CAS Key Laboratory of Computational Biology, Shanghai Institute of Nutrition and Health, Chinese Academy of Sciences, Shanghai, China. ⁴Key Laboratory of Multi-Cell Systems, Shanghai Institute of Biochemistry and Cell Biology, Center for Excellence in Molecular Cell Science, Chinese Academy of Sciences, Shanghai, China. ⁵University of Chinese Academy of Sciences, Beijing, China. ⁶Analytical Research Center for Organic and Biological Molecules, State Key Laboratory of Drug Research, Shanghai Institute of Materia Medica, Chinese Academy of Sciences, Shanghai, China. ⁷Department of Pathology, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China. ⁸Bruker (Beijing) Scientific Technology Co., Ltd, Shanghai, China. ⁹Department of Pharmacology, Cancer Biology Institute, Yale University School of Medicine, West Haven, CT, USA. ¹⁰Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA. ¹¹Department of Molecular and Human Genetics, Lester and Sue Smith Breast Center, Baylor College of Medicine, One Baylor Plaza, Houston, TX, USA. ¹²Department of Medicine, McDonnell Genome Institute, Siteman Cancer Center, Washington University, St. Louis, MI, USA. ¹³Office of Cancer Clinical Proteomics Research, National Cancer Institute, National Institutes of Health, Rockville, MD, USA. ¹⁴School of Pharmaceutical Science and Technology, Hangzhou Institute for Advanced Study, University of Chinese Academy of Sciences, Hangzhou, China. ¹⁵Key Laboratory of Systems Health Science of Zhejiang Province, School of Life Science, Hangzhou Institute for Advanced Study, University of Chinese Academy of Sciences, Hangzhou, China. ¹⁶These authors contributed equally: Jing Yu, Xiuqi Gui, Yunhao Zou, Qian Liu. ✉ e-mail: zhouhu@sim.ac.cn; shaozhen@sinh.ac.cn; wulingying@cscs.org.cn; dgao@sibcb.ac.cn