

RESEARCH

Open Access



# Proteogenomic characterization of molecular and cellular targets for treatment-resistant subtypes in locally advanced cervical cancers

Do Young Hyeon<sup>1†</sup>, Dowoon Nam<sup>2†</sup>, Hye-Jin Shin<sup>3†</sup>, Juhee Jeong<sup>4†</sup>, Eunsoo Jung<sup>1†</sup>, Soo Young Cho<sup>3</sup>, Dong Hoon Shin<sup>3</sup>, Ja-Lok Ku<sup>5</sup>, Hye Jung Baek<sup>3</sup>, Chong Woo Yoo<sup>3</sup>, Eun-Kyung Hong<sup>3</sup>, Myong Cheol Lim<sup>3</sup>, Sang-Jin Lee<sup>3</sup>, Young-Ki Bae<sup>3</sup>, Jong Kwang Kim<sup>3</sup>, Jingi Bae<sup>2</sup>, Wonyoung Choi<sup>3</sup>, Su-Jin Kim<sup>2</sup>, Seunghoon Back<sup>2</sup>, Chaewon Kang<sup>2</sup>, Inamul Hasan Madar<sup>2</sup>, Hokeun Kim<sup>2</sup>, Suhwan Kim<sup>2</sup>, Duk Ki Kim<sup>4</sup>, Jihyung Kang<sup>4</sup>, Geon Woo Park<sup>4</sup>, Ki Seok Park<sup>4</sup>, Yourae Shin<sup>1</sup>, Sang Soo Kim<sup>3\*</sup>, Keehoon Jung<sup>4,6\*</sup>, Daehee Hwang<sup>1\*</sup>, Sang-Won Lee<sup>2\*</sup> and Joo-Young Kim<sup>3\*</sup>

## Abstract

We report proteogenomic analysis of locally advanced cervical cancer (LACC). Exome-seq data revealed predominant alterations of keratinization-TP53 regulation and O-glycosylation-TP53 regulation axes in squamous and adeno-LACC, respectively, compared to in early-stage cervical cancer. Integrated clustering of mRNA, protein, and phosphorylation data identified six subtypes (Sub1-6) of LACC among which Sub3, 5, and 6 showed the treatment-resistant nature with poor local recurrence-free survival. Elevated immune and extracellular matrix (ECM) activation mediated by activated stroma (PDGFD and CXCL1<sup>high</sup> fibroblasts) characterized the immune-hot Sub3 enriched with MUC5AC<sup>high</sup> epithelial cells (ECs). Increased epithelial-mesenchymal-transition (EMT) and ECM remodeling characterized the immune-cold squamous Sub5 enriched with PGK1 and CXCL10<sup>high</sup> ECs. We further demonstrated that CIC mutations could trigger EMT activation by upregulating ETV4, and the elevation of the immune checkpoint PVR and neutrophil-like myeloid-derived suppressive cells (FCN1 and FCGR3B<sup>high</sup> macrophages) could cause suppression of T-cell activation in Sub5. Increased O-linked glycosylation of mucin characterized adeno-LACC Sub6 enriched with MUC5AC<sup>high</sup> ECs. These results provide a battery of somatic mutations, cellular pathways, and cellular players that can be used to predict treatment-resistant LACC subtypes and can serve as potential therapeutic targets for these LACC subtypes.

<sup>†</sup>Do Young Hyeon, Dowoon Nam, Hye-Jin Shin, Juhee Jeong, Eunsoo Jung contributed equally to this work.

## \*Correspondence:

Sang Soo Kim  
sangsookim@ncc.re.kr  
Keehoon Jung  
keehoontjung@snu.ac.kr  
Daehee Hwang  
daehee@snu.ac.kr  
Sang-Won Lee  
sw\_lee@korea.ac.kr  
Joo-Young Kim  
jooyoungcasa@ncc.re.kr

Full list of author information is available at the end of the article



© The Author(s) 2025. **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/>.

**Keywords** Locally advanced cervical cancer, Proteogenomics, Cancer subtyping, And treatment resistance

## Introduction

Uterine cervical cancer (UCC) is the fourth most commonly diagnosed cancer and the fourth leading cause of cancer-related deaths in women [1]. Early-stage UCC can be effectively treated with surgery. Accordingly, the mortality has been ascribed mainly to locally advanced UCC (LACC). The incidence of LACC has been high in under-developed countries [2], and remains still substantial in some high-income countries, such as Korea, Japan, and Taiwan [3–5]. A standard treatment modality for LACC declared by National Cancer Institute (NCI) [6] has been concurrent cisplatin-based chemoradiotherapy (CCRT) for the past twenty years. However, it has been argued that CCRT should be further customized to reflect clinical characteristics of LACC. For example, combinations of CCRT with immuno-oncology agents, such as PD-1 or PD-L1 inhibitors, have been exploited as promising modalities [7, 8]. However, these trials showed heterogeneous treatment responses and differential benefits depending on the stage, immune checkpoint expression, and/or mutation burden [9]. Therefore, it is necessary to define LACC subtypes and then develop tailored treatments based on characteristics of the individual subtypes.

Several genomic and/or transcriptomic analyses, including the Cancer Genome Atlas (TCGA), have identified subtypes of UCC based on genomic alterations, mRNAs, and/or their associated cellular pathways that reflect pathophysiological characteristics of patients [10–13]. However, these studies have focused on surgically treated early-stage UCC. There has been thus a significant need for a comprehensive molecular characterization of LACC. The previous studies have investigated mainly genomic and mRNA alterations. However, signaling pathways (e.g., PI3K/AKT signaling) have been also shown to play significant roles in UCC pathogenesis, and proteins (e.g., phosphorylated AKT) central in these pathways have been proposed as therapeutic targets [14]. To explore these protein/phosphorylation signatures, the TCGA study [10] profiled the abundances or phosphorylation levels of proteins in UCC, but only for a limited set of 192 proteins using reverse phase protein arrays. Recent proteogenomic studies have demonstrated that mass spectrometry (MS)-based proteomic analysis can provide abundance and phosphorylation levels for larger proteomes, facilitating effective discovery of protein/phosphorylation signatures and their associated cellular and signaling pathways.

Here, we present proteogenomic analysis of LACC patients who underwent primary radiotherapy with or

without chemotherapy involving whole-exome sequencing (WES), RNA-sequencing, global proteomics, and phosphoproteomics, as well as single-cell RNA-seq (scRNA-seq) analysis. Our results provide 1) proteogenomic subtypes of LACC, 2) genomic, mRNA, and protein signatures defining these subtypes, 3) functional characteristics of the subtypes based on cellular pathways represented by the molecular signatures, and 4) subpopulations of epithelial, immune, and fibroblast cells associated with the subtypes through integration of scRNA-seq data. These results can contribute to subtype-dependent therapeutic stratification of LACC.

## Results

### Proteogenomic analysis of UCC

We collected treatment-naïve tumor tissues from 350 patients with LACC based on the following criteria, along with matched peripheral blood mononuclear cells from 251 of the patients: 1) stage IB2-IIA showing LACC characteristics [large tumor size (>3 cm in diameter) and/or lymph node metastasis] treated with primary CCRT or radiotherapy according to the NCI guideline [15] (107 patients); 2) stage IIB-IVA according to 2014 international federation of gynecology and obstetrics (FIGO) staging (212 patients); or 3) stage IVB treated by primary CCRT (31 patients) (Supplementary Table S1). Our LACC cohort included a larger percentage (70%) of patients at stage IIB or higher (36.2% in TCGA [10], and 9.1% in Huang et al. [16]) with the median cellularity of 75.0% (Supplementary Fig. S1A and S1B). How WES, RNA-seq, global proteomics, phosphoproteomics, and scRNA-seq analyses were performed for individual samples is described in Supplementary Fig. S1C.

### Predominant alteration of keratinization-TP53 regulation axis in LACC

We performed WES analysis for 251 patients for whom the matched PBMC samples were available. By comparing WES data from tumors of 251 patients and their matched PBMCs, we identified 55,397 somatic mutations affecting protein sequences (Supplementary Fig. S1D) and 14 significantly ( $q < 0.1$ ) mutated genes (SMGs) in our cohort using MutSigCV (Fig. 1A, 2nd panel). To investigate the SMGs predominant in LACC, we compared their mutation frequencies with those in three previous cohorts including more early-stage tumors (TCGA [10], Ojesina et al. [17], and Huang et al. [16]) and found that mutation frequencies of *STK11*, *CASP8*, *TGFBR2*, *KRT10*, and *DYNAP* were significantly ( $P < 0.05$ ) higher in

our LACC cohort (Fig. 1B, red-labeled genes). Next, we identified 12 genes frequently (>1% of patients) altered by genome integration of human papillomavirus (HPV; Fig. 1A, 3rd panel), and 10 of them (*KRT5/13/14/16*, *MUC4*, *UBC*, *S100A9*, *TPX2*, and *GANAB*) had significantly higher integration frequencies in our LACC cohort than in the three previous cohorts (Fig. 1C). Finally, we identified 16 genes having frequent copy number alterations (CNAs) (Fig. 1A, 4th panel) and found that 13 of them (*HRNR*, *FLG/FLG2*, *RPTN*, *TCHH*, *CRNN*, *RUFY2*, *DNA2*, *SLC25A16/20*, *TET1*, *PRKAR2A*, and *ARIH2*) showed significantly higher CNA frequencies in our LACC cohort than in the TCGA cohort (Fig. 1D).

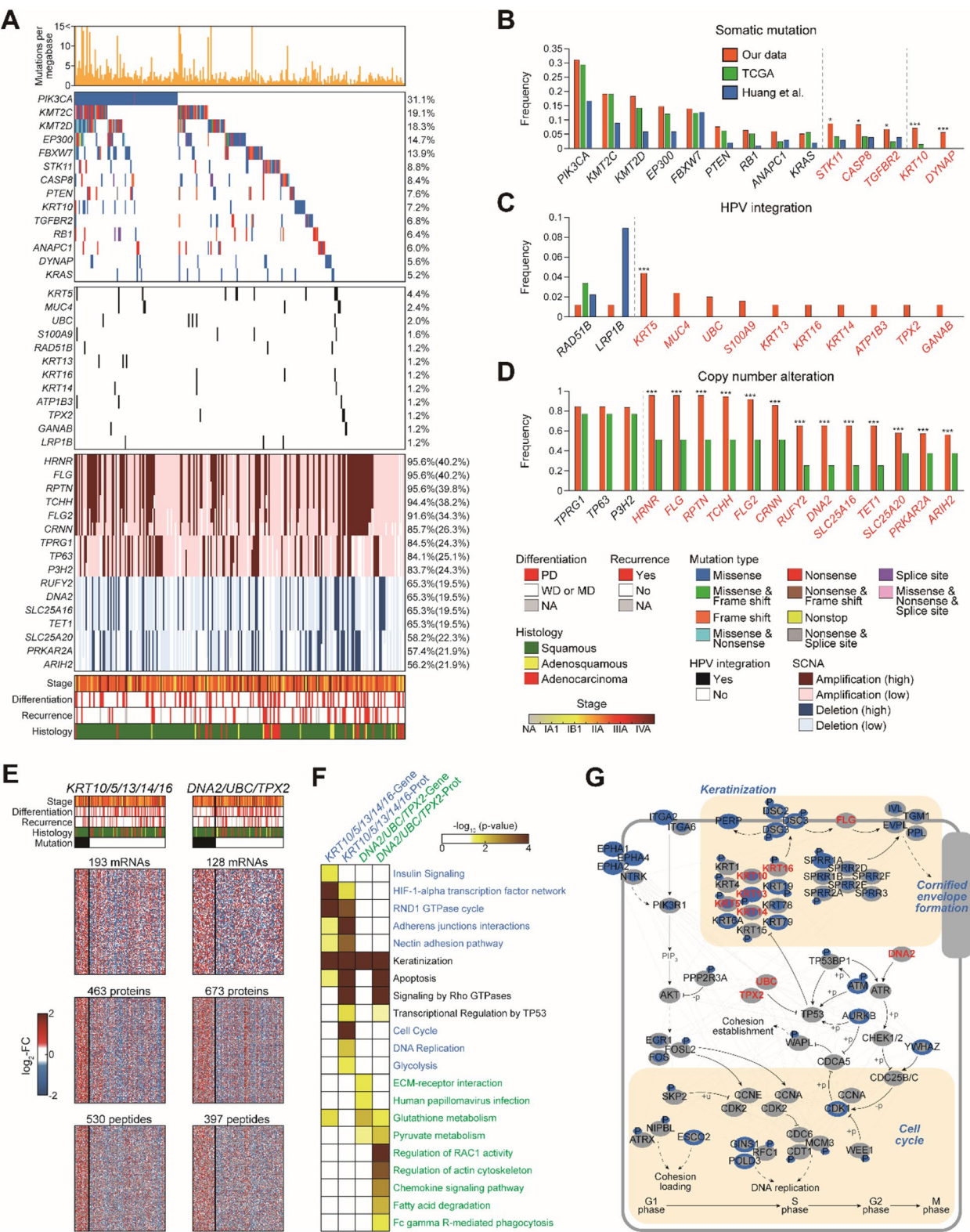
To explore functional associations of these predominantly altered genes in LACC (PAG-LACC; Fig. 1B-D) with LACC pathogenesis, we performed gene set enrichment analysis (GSEA) for them using ConsensusPathDB [18]. The PAG-LACC significantly represented keratinization, TP53 regulation, and toll-like receptor (TLR) cascade pathways (Supplementary Fig. S1E), suggesting that substantial dysregulation of these pathways are associated with development of LACC. In support, the absence of keratinization was reported to be suggestive for a higher tumor grade associated with LACC characteristics (larger tumor size and lymph node metastasis) [19], and the loss of TP53-dependent tumor suppression and innate immune response to HPV infection contributes to promoting tumor growth. Furthermore, we identified genes/proteins whose expression or phosphorylation levels correlated with alterations of (Fig. 1E). Interestingly, the alterations of these genes were significantly ( $P < 0.01$ ) enriched in SCC (Fig. 1E, Histology). To examine the functions of the genes/proteins correlated with the alterations of PAG-LACC, we performed GSEA for them (Supplementary Table S2). The genes/proteins correlated

with either *KRT5/10/13/14/16* or *DNA2/UBC/TPX2* alterations commonly represented the keratinization and apoptosis/TP53 regulation pathways (Fig. 1F, black labeled), supporting the importance of these pathways in LACC pathogenesis. On the other hand, the genes/proteins correlated with *KRT5/10/13/14/16* alterations uniquely represented the pathways related to hypoxia (Hif-1 and glycolysis), cell cycle (AP-1, DNA replication, and RND1 GTPase cycle), and cell adhesion (Nectin adhesion and adherens junctions) (Fig. 1F, blue labeled). The genes/proteins correlated with *DNA2/UBC/TPX2* alterations predominantly represented the pathways related to reactive oxygen species (ROS; glutathione and fatty acid degradation/pyruvate metabolism), phagocytosis, and actin cytoskeleton-extracellular matrix (ECM) (ECM-receptor interaction and regulation of RAC1 activity and actin cytoskeleton) (Fig. 1F, green labeled). These results suggest that the two sets of PAG-LACC can have the shared (keratinization and apoptosis/TP53 regulation pathways) and distinctive (blue and green labeled pathways in Fig. 1F) roles in LACC pathogenesis.

Interestingly, the genes/proteins correlated with *KRT5/10/13/14/16* or *DNA2/UBC/TPX2* alterations represented both keratinization and apoptosis/TP53 regulation pathways, suggesting potential links between these pathways. To explore such links, we built a network model describing interactions among the PAG-LACC and their correlated genes/proteins involved in these pathways. TP53/63 are known to suppress expression of keratins [20] and cornified envelop components [21] whose disruption facilitates HPV release. Consistent with this previous finding, the network model showed that the loss of TP53 functions by *DNA2/UBC/TPX2* mutations accompanied upregulation of cornified envelop components to prevent HPV release (Fig. 1G,

(See figure on next page.)

**Fig. 1** Predominant alteration of keratinization-TP53 regulation axis in LACC. **A** Predominantly altered genes in LACC (PAG-LACC). Mutations per megabase in each patient are shown in the top panel. SMG mutations (2nd panel), alterations by HPV genome integration (3rd panel), and CNAs (amplification in red and deletion in blue; 4th panel) detected in each patient for the indicated genes are shown. Alteration frequencies are shown on the right, and the indicated clinical parameters for each patient are shown in the bottom. **B-D** Comparisons of frequencies of SMG mutations (**B**), alterations by HPV genome integration (**C**), and CNAs (**D**) between our LACC cohort and the two previous early-stage tumor-enriched cohorts. Red labels indicate the genes (PAG-LACC) having significantly ( $p < 0.05$  by proportional test) higher alteration frequencies in our LACC cohort than in the early-stage tumor-enriched cohorts. \*,  $p < 0.05$ ; \*\*,  $p < 0.01$ ; \*\*\*,  $p < 0.001$ .  $n = 251$ , 289, and 102 patients for our, TCGA, and Huang et al. cohorts, respectively, for SMG mutations;  $n = 251$ , 178, and 45 patients for our, TCGA, and Huang et al. cohorts, respectively, for alterations by HPV genome integration and CNAs. **E** mRNAs, proteins, phosphorylated peptides up-regulated in tumors harboring the alterations (SMG, HPV integration, or CNA) in any of the indicated PAG-LACC. The colored bar represents the gradient of  $\log_2$ -fold-changes of abundances for mRNAs, proteins, or phosphorylated peptides relative to their median levels. **F** Cellular pathways enriched by mRNAs (Gene) or proteins + phosphoproteins (Prot) correlated with alterations of the indicated PAG-LACC. The heat map shows the enrichment significance ( $p$  value from ConsensusPathDB) of each pathway by the mRNAs or proteins + phosphoproteins as  $-\log_{10}(p)$ . **G** Network model describing interactions among the mRNAs (blue node center), proteins (blue node border), or phosphoproteins (blue circled P) that have significant correlation with alterations of the PAG-LACC (red labeled) and are involved in keratinization and apoptosis/TP53 regulation. Gray nodes indicate molecules added to the pathways to increase connections among the nodes. Solid arrows, direct activations; dotted arrows, indirect activations; gray lines, protein-protein interactions; thick lines, plasma membrane



**Fig. 1** (See legend on previous page.)



top right). However, the loss of keratin functions by *KRT5/10/13/14/16* could lead to the failure of such upregulation to increase the cornified envelop integrity, thereby rather promoting HPV release and tumor growth, consistent with upregulation of cell cycle genes/proteins (Fig. 1G, bottom) and its upstream signaling pathways (PI3K-AKT). These results suggest that the collective loss of keratinization-TP53 regulation axis by PAG-LACC can define the characteristics of LACC during development of SCC.

### Proteogenomic subtypes of LACC

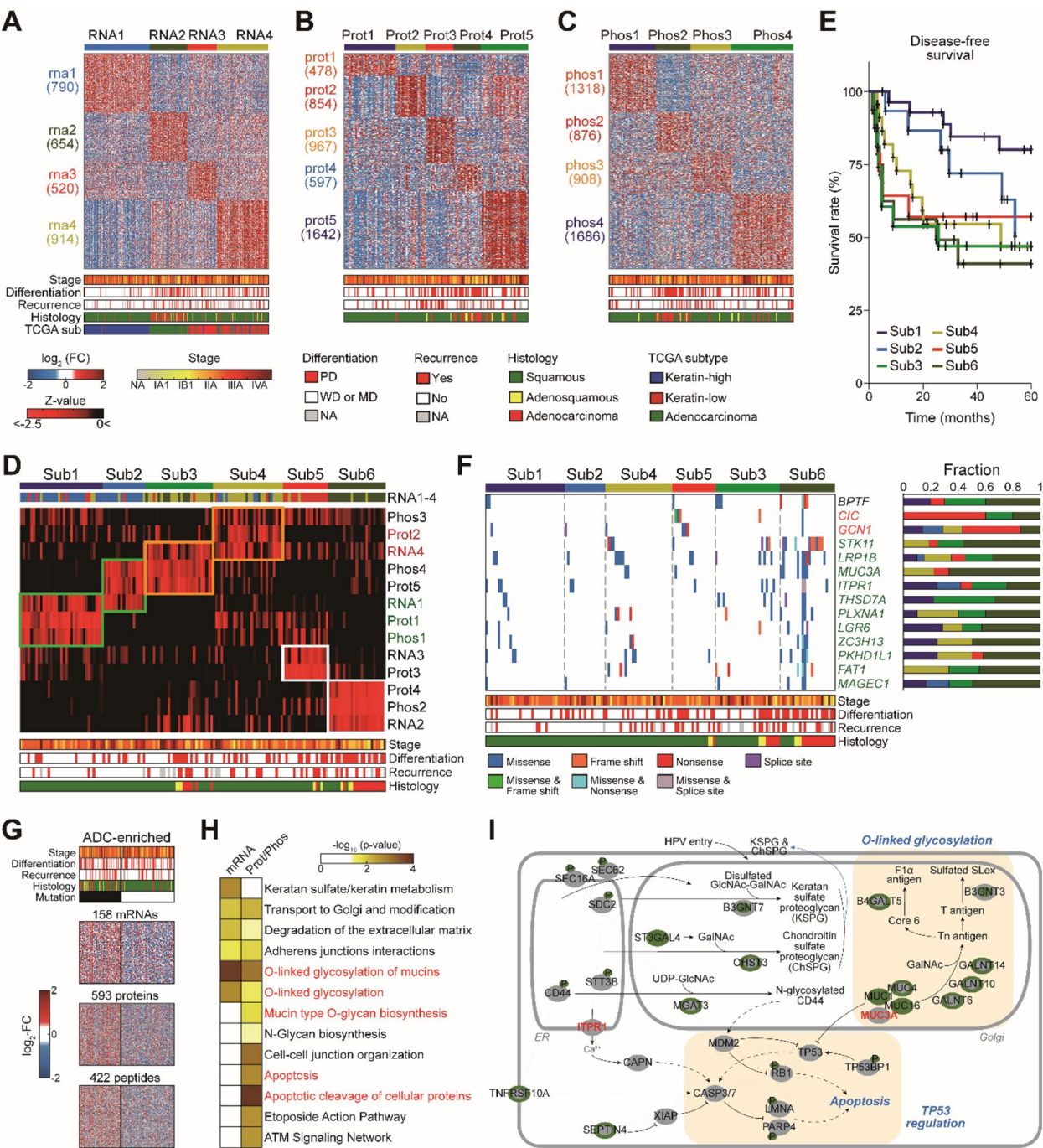
The TCGA study previously identified three subtypes of UCC based on mRNA data: Keratin-high and Keratin-low SCC, and ADC. However, the TCGA cohort included only 36.2% of stage IIB-IVA, and the TCGC subtypes were thus biased toward early-stage tumors. Moreover, how the whole proteome-level information, such as our global proteome and phosphoproteome data, can change the subtypes of the early-stage-enriched tumors remains elusive. Previously, global proteome profiling, not phosphoproteome profiling, has been performed in a few studies [22, 23] using liquid chromatography-tandem mass spectrometry (LC-MS/MS) analysis, but only for small cohort sizes (8 to 15) with low proportions of LACC. In this study, we generated global proteomes and phosphoproteomes from 146 LACC tumors with sufficiently large tumor wet weights to meet the required amounts of proteins and phosphopeptides using LC-MS/MS analysis, and also RNA-seq data for all 350 LACC tumors (Supplementary Fig. S1C). To identify the molecular subtypes of LACC, we performed clustering analysis on 1) 334 LACC tumors (272 SCC and 62 ADC) using mRNA data and 2) 146 LACC tumors (117 SCC and 29

ADC) using global proteome or phosphoproteome data. Because the different nature of SCC and ADC can confound the clustering, we employed a two-stage clustering that first clustered SCC only (1st stage) and then all tumors (2nd stage) based on the clustering results of SCC (Supplementary Fig. S2A).

For mRNA data, the two-stage clustering identified three clusters of SCC (s-RNA1-3) at the 1st stage and four clusters (RNA1-4) of all LACC tumors (SCC + ADC) at the 2nd stage (Supplementary Fig. S2B; Supplementary Table S3A). RNA1-4 were characterized by up-regulation of 790 (rna1), 654 (rna2), 520 (rna3), and 914 (rna4) signature genes, respectively (Fig. 2A; Supplementary Table S3B). A majority of ADC (42 of 63) were found enriched in RNA2. We then used these mRNA signatures (rna1-4) to classify tumors in the TCGA cohort and found that 78.6% (198 of 252) could be categorized into RNA1-4 (Supplementary Fig. S2C, TCGA cohort). We further generated a merged cohort of patients reported by Medina-Martinez et al. [12], Espinosa et al. [11], and Shin et al. [13], which were also enriched with early-stage tumors, and found that 83.8% (62 of 74) could be also classified into RNA1-4 (Supplementary Fig. S2C, Merged cohort). The remaining tumors (21.4% in TCGA and 16.2% in the merged cohort) were not be classified into RNA1-4 (Supplementary Fig. S2C), which can make molecular subtypes different between early-stage tumors and LACC. For global proteome data, the two-stage clustering identified four clusters of SCC (s-Prot1-4) at the 1st stage and five clusters of all LACC tumors (Prot1-5) at the 2nd stage (Supplementary Fig. S2D). For phosphoproteome data, it also identified three clusters of SCC (s-Phos1-3) at the 1st stage and four clusters of all LACC tumors (Phos1-4) at the 2nd stage (Supplementary Fig. S2E).

(See figure on next page.)

**Fig. 2** Proteogenomic subtypes of LACC. **A** RNA1-4 clusters defined by mRNA signatures (rna1-4) in our LACC cohort. rna1-4 defining RNA1-4, respectively, are shown. Numbers of mRNAs in rna1-4 are indicated in parentheses. Clinical parameters for tumors are shown in the bottom, and the subtype bar plot shows subtypes predicted using molecular signatures defined by the TCGA [10]. **B** and **C**, Protein signatures (prot1-5 and phos1-4) defining Prot1-5 (**B**) and Phos1-4 (**C**) based on global proteome and phosphoproteome data. Numbers of proteins and phosphopeptides are indicated in parentheses. **D** Six subtypes (Sub1-6) identified via integrative clustering of mRNA, protein, and phosphorylation data. Each row in the heat map shows Z-values of the tumors with the corresponding molecular signature for each indicated cluster. Color bar, the gradient of Z-value representing distances from mean vectors with the signature. Colored bars at the top, Sub1-6; RNA1-4 clusters defined by rna1-4. **E** Disease-free survival of patients with tumors belonging to Sub1-6.  $n = 29, 15, 17, 23, 14$ , and  $16$  in Sub1-6, respectively. **F** Somatic mutations of frequently mutated genes enriched in Sub5 (red), Sub6 or Sub6 + ADC in Sub3 (green), or all Sub3, 5, and 6 (black). Fractions of tumors carrying mutations of each gene in Sub1-6 were shown in the right panel. **G** mRNAs, proteins, phosphorylated peptides up-regulated in tumors harboring ADC-enriched somatic mutations in the 11 genes (green labeled in **F**). The colored bar represents the gradient of  $\log_2$ -fold-changes of abundances for mRNAs, proteins, or phosphorylated peptides relative to their median levels. **H** Cellular pathways enriched by mRNAs or proteins + phosphoproteins (Prot/Phos) correlated with the ADC-enriched somatic mutations. The heat map shows the enrichment significance (p value from ConsensusPathDB) of each pathway by the mRNAs or Prot/Phos as  $-\log_{10}(p)$ . **I** Network model describing interactions among the mRNAs (green node center), proteins (green node border), or phosphoproteins (green circled P) that have significant correlation with the ADC-enriched somatic mutations and are involved in O-linked glycosylation and apoptosis/TP53 regulation. PAG-LACC are labeled in red. Gray nodes indicate molecules added to the pathways to increase connections among the nodes. Solid arrows, direct activations; dotted arrows, indirect activations; gray lines, protein-protein interactions; thick lines, plasma membrane



**Fig. 2** (See legend on previous page.)

Prot1-5 and Phos1-4 were characterized by up-regulation of the cluster-specific signature proteins (Fig. 2B, prot1-5) and phosphoproteins (Fig. 2C, phos1-4), respectively (Supplementary Table S3B). Finally, we performed an integrative clustering of 146 tumors with both mRNA and protein data and identified six subtypes (Sub1–6) of LACC (Fig. 2D). Prot1 and Phos1

profiles (Fig. 2D, green boxes) further divided RNA1 to Sub1 and 2, and Prot2 profiles (Fig. 2D, orange boxes) divided RNA4 to Sub3 and 4, indicating that the protein data further subdivide the mRNA-based subtypes due to their complementary nature. We next compared survival of LACC patients in Sub1–6. While molecular subtyping would not be problematic

because treatment-naïve samples were used, survival analysis can be problematic because distinct treatments can affect survival. We thus excluded 10 patients at stage IVB having distant metastases, and 9 patients who underwent surgery as the primary treatment modality, and used only patients who received primary CCRT for survival analysis. Sub3, 5, and 6 had relatively lower disease (local/regional recurrence or distant metastasis)-free survival (DFS) and local recurrence-free survival (LRFS) rates (Fig. 2E and Supplementary Fig. S2F) than Sub1 and 2, indicating their treatment-resistant nature. Interestingly, Sub4 showed better LRFS than the treatment-resistant Sub3, 5, and 6, but comparable DFS, suggesting a high rate of distant metastasis in Sub4 and thus a need for systematic chemotherapy. We next examined whether the SMGs and frequently mutated non-SMGs were significantly enriched in the treatment-resistant subtypes and found the following enrichments: 1) *BPTF* mutations in Sub3, 5, and 6 (Sub1 + 2 + 4 vs. Sub3 + 5 + 6), 2) *CIC* and *GCN1* mutations in Sub5, and 3) *STX11*, *LRP1B*, *MUC3A*, *ITPR1*, *THSD7A*, *PLXNA1*, *LGR6*, *ZC3H13*, *PKHD1L1*, *FAT1*, and *MAGEC1* mutations in Sub6 or in ADC of Sub3 and 6 (Fig. 2F). In SCC, we identified the keratinization-TP53 regulation axis as a potential cause for LACC. To identify a potential cause for LACC in ADC, we then identified the genes/proteins whose expression and phosphorylation levels correlated with the ADC-enriched alterations of the above 11 genes (Fig. 2G). The GSEA revealed that these genes/proteins represented N/O-linked glycosylation of mucins and apoptosis. The network model further demonstrated that 1) N-glycosylated molecules (e.g., CD44) could suppress TP53 via MDM2 activation [24]; and 2) O-glycosylation

of mucins could produce tumor-associated antigens (Tn) and inhibit TP53 transcription [25, 26], which could collectively facilitate tumor growth. These data suggest the collective aberration of N/O glycosylation-TP53 regulation axis can define the characteristics of LACC during development of ADC (Fig. 2H and I).

### Characteristics of proteogenomic subtypes of LACC

To investigate functional characteristics of Sub1-6, we identified cellular pathways represented by mRNA and protein (protein + phosphorylation) signatures of each subtype (Supplementary Fig. S3A) through GSEA. For SCC Sub1-2 (RNA1 in Fig. 2D), their mRNA signatures (S1/2-G, rna1) significantly ( $P < 0.05$ ) represented the keratinization-TP53 regulation axis (Fig. 3A, blue-labeled; Supplementary Table S4), which was suggested to be associated with PAG-LACC-driven development of LACC (Fig. 1G), as well as ROS-related pathways (biological oxidation and chemical carcinogenesis) to induce TP53 activation, which were represented by the genes/proteins correlated with *DNA2/UBC/TPX2* alterations (Fig. 1F). Sub1-2 can be thus considered to be early-stage squamous LACC, consistent with the findings that they had good DFS (Fig. 2E) and corresponded to the Keratin-high subtype in TCGA with good survival (RNA1 in Fig. 2a, TCGA cohort). The protein signatures defining Sub1 (S1-P; prot1 and phos1) consistently represented the same keratinization-TP53 regulation axis while those defining Sub2 (S2-P; prot5 and phos4) represented additionally platelet activation/complement cascade pathways [27] (Fig. 3A, light blue-labeled). Collectively, these data suggest that both Sub1 and 2 are treatment-sensitive

(See figure on next page.)

**Fig. 3** Characteristics of proteogenomic subtypes of LACC. **A** Cellular pathways represented by mRNAs (S1-G to S6-G) and proteins (proteins + phosphoproteins) (S1-P to S6-P) defining Sub1-6. The heat map shows the enrichment significance (p value from ConsensusPathDB) of pathways by the mRNAs or proteins defining Sub1-6 as  $-\log_{10}(p)$ . **B, D, and F**, Network models showing interactions between the mRNAs and proteins involved in immune and ECM pathways upregulated in activated stroma for Sub3 (**B**); EMT signaling and ECM remodeling pathways for Sub5 (**D**); and O-linked glycosylation (**F**). Node colors (center and boundary) indicate whether the corresponding mRNA and protein were selected as signatures for the corresponding subtypes (green for Sub3, yellow for Sub4, orange for Sub5, and dark green for Sub6). Presence of a circled P on a node indicates that the corresponding phosphoproteins had phosphorylated peptides that defined the corresponding subtypes. Kinases that could control the pathways in the network models in Sub3 and 5 were identified as previously described [28–30] and included to the network models (blue-labeled nodes). Arrows, activation; inhibition symbols, inhibition; solid arrows, direct activation; dotted arrows, indirect activation; gray lines, protein–protein interactions. **C and E** Representative images for MMP2 (Sub3), ICOS (Sub4), and GPC4 (Sub6) (**C**) and BSG, VAV2, EGFR, and FOSL1 (Sub5) (**E**) and their quantification results in Sub1-6 from IHC analysis. Magnification 20 $\times$ . Scale bar = 100  $\mu$ m. Blue color: Hematoxylin counterstaining; Brown color: DAB positive staining.  $n = 17, 11, 16, 20, 13$ , and 16 in Sub1-6, respectively, for MMP2 and ICOS;  $n = 17, 10, 15, 20, 13$ , and 16 in Sub1-6, respectively, for GPC4, BSG, VAV2, EGFR, and FOSL1. \*,  $p < 0.05$ ; \*\*,  $p < 0.01$  by comparing H-scores (grades or density) between a subtype of interest and the other subtypes using one-sided Fisher's exact test (MMP2) or one-sided Student's t-test (ICOS, GPC4, VAV2, EGFR, and FOSL1). **G** Representative images of CIC, ETS transcription factors (ETV4), and cytoskeletal proteins (CDH1/2, CTNNA1, and VIM) in #6507A (Sub5) and #6595 (Sub3) cells from Western blotting analysis.  $\beta$ -actin was used as a loading control.  $n = 2$  independent experiments. **H** Quantification of indicated proteins relative to  $\beta$ -actin in the indicated cells. Band intensity from western blot images was quantified by ImageJ. Data are shown as mean  $\pm$  s.e.m.  $n = 2$  independent experiments. \*,  $p < 0.05$  by Student's t-test. **I**, Representative immunofluorescence images of CDH1 (red; epithelial marker) and VIM (green; mesenchymal marker) in the indicated cells grown in monolayer culture system ( $n = 2$  independent experiments). Magnification 40 $\times$ . Scale bar = 50  $\mu$ m







early-stage squamous LACC, but Sub2 may be a more advanced LACC with platelet/complement activation than Sub1.

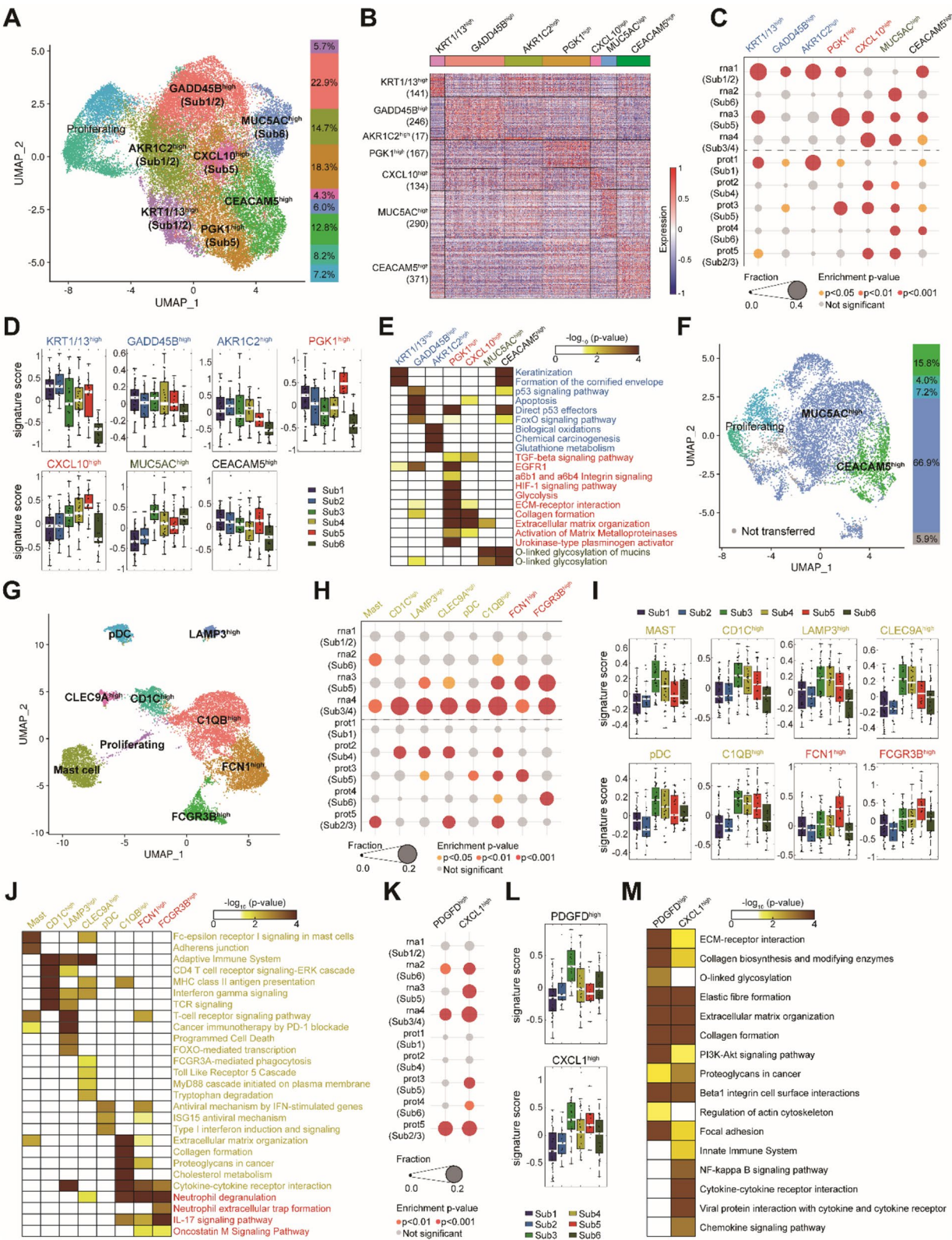
For Sub3 with mixed characteristics of SCC and ADC and SCC Sub4 (RNA4 in Fig. 2D), its mRNA signatures (S3/4-G, rna2) represented innate (leukocyte infiltration, phagocytosis, and antigen presentation) and adaptive (T/B-cell receptor signaling) immune pathways (Fig. 3A, yellow-labeled). Moreover, they also represented ECM pathways (collagen and proteoglycan; Fig. 3A, green-labeled). However, the protein signatures defining Sub4 (S4-P; prot5 and phos4) represented more strongly immune pathways while those defining Sub3 (S3-P; prot2 and phos3) represented further strongly ECM pathways, suggesting the different nature of immune responses between Sub3 and 4. To address this issue, we estimated the proportions of tumor, stromal, and immune cells in tumor microenvironment (TME) using the mRNA signatures previously defined for the cell types (Supplementary Fig. S3B). Confirming high immune activation in both Sub3 and 4 (ESTIMATE, Puleo, and Peng-Immune), this TME analysis further revealed that activated stroma to induce immune activation was found highly elevated in Sub3 compared to in Sub4 (Moffit, Puleo, Maurer, and Peng in Supplementary Fig. S3B), suggesting that they could be responsible for immune activation in Sub3. This can also explain strong activation of ECM pathways in Sub3, which was reported to be associated with poor survival via MMP2 activation [31]. These data account for the treatment-resistant nature in Sub3 compared to in Sub4 according to the LRFS (Fig. 2F). The network model for Sub3 supported these findings that activated stroma secreted cytokines (IL33, CCL21, and CXCL9/10/11/12) for immune activation and produced integrins (ITGA2/4/5/7/L/M), collagens (COL1A1-2/2A1/3A1/5A1-2/6A1-3/8A1), and proteases (MMP2 and ADAMS2) for ECM activation (Fig. 3B). In contrast, the network model for Sub4 showed activation of the aforementioned immune pathways in response to tumor antigen (Supplementary Fig. S3C). We further performed immunohistochemistry (IHC) analysis for the

representative markers MMP2 and ICOS for ECM activation [31] and T-cell activation [32], respectively. The analysis confirmed upregulation of MMP2 in stromal cells of Sub3 tumors and upregulation of ICOS in the total area of tumor tissues of Sub4 tumors (Fig. 3C; see **Methods** for tumor-stromal distribution of these markers). Collectively, these results suggest that Sub3 and 4 are activated stroma-enriched treatment-resistant LACC and typical immunogenic treatment-sensitive squamous LACC, respectively.

For SCC Sub5 (RNA3 in Fig. 2D), both gene (S5-G, rna3) and protein (S5-P, prot3) signatures represented EMT-related 1) signaling (TGFB/EGFR/Integrin/Basigin/HIF signaling) and 2) ECM remodeling (metalloproteinase/urokinase) pathways (Fig. 3A, red-labeled). The network model for Sub5 confirmed upregulation of these pathways (Fig. 3D), and IHC analysis further confirmed upregulation of representative markers BSG [33], VAV2 [34], EGFR [35], and FOSL1 [36] for EMT-related signaling pathways in the tumor area of tissues in Sub5 (Fig. 3E). Sub3-5 (RNA3-4 in Fig. 2D) corresponded to the Keratin-low subtype with poor survival in TCGA, but such characteristics were much stronger in Sub5 (RNA3) than in Sub3-4 (RNA4) (Fig. 2A). Moreover, Sub5 had stronger basal-like characteristics (Puleo and Peng in Supplementary Fig. S3B) and weaker immune activation (ESTIMATE, Puleo, and Peng) than Sub3-4. These EMT-high, keratin-low, basal-like, and immune-cold characteristics collectively contribute to the treatment-resistant nature of Sub5 (Fig. 2E). For ADC-enriched Sub6, genes (S6-G, rna2) and/or protein (S6-P, prot4 and phos2) signatures represented N/O-glycosylation (Golgi transport-modification) and mucin-related pathways (mucin glycosylation; Fig. 4A, dark green-labeled). Mucin-type O-glycosylated structures including Tn antigens are linked to metastasis and poor survival [37], contributing to the treatment-resistant nature of Sub6. In addition, ciliopathy pathway was also found upregulated in Sub6, consistent with a previous finding that the loss or abnormality of cilia structure was frequently observed in glandular cells of ADC [38]. The network model for

(See figure on next page.)

**Fig. 4** Cellular heterogeneity associated with LACC Sub1-6. **A** and **G** Uniform manifold approximation and projection (UMAP) plot showing subclusters of ECs (**A**) or myeloid cells (**G**). Associations of the individual subclusters with Sub1-6 are indicated in parenthesis. The stacked bar graph shows percentages of ECs belonging to individual subclusters. **B** Marker genes upregulated predominantly in each subcluster. Numbers in parenthesis represents the number of marker genes in the corresponding subcluster. The color bar represents the gradient of  $\log_2$ -FC with respect to their mean values. **C**, **H**, and **K** Dot plot showing the enrichment  $p$ -value of marker genes for each subcluster of ECs (**C**), myeloid cells (**H**), or CAFs (**K**) in the indicated mRNA and protein signatures for Sub1-6 (see legend in bottom right). The associated subtype for each signature is denoted in parenthesis. Dot size represents the fraction of overlapping marker genes as shown in legend (bottom left). **D**, **I**, and **L** Signature scores of marker genes for each subcluster of ECs (**D**), myeloid cells (**I**), or CAFs (**L**) in Sub1-6.  $n = 33, 17, 27, 28, 18$ , and  $23$  in Sub1-6, respectively. **E**, **J**, and **M** Cellular pathways enriched by marker genes for each subcluster of ECs (**E**), myeloid cells (**J**), or CAFs (**M**). The color bar represents the gradient of  $-\log_{10}$ (enrichment  $p$ -value). **F** UMAP plot showing ADC ECs with transferred labels of the subclusters of SCC ECs. The stacked bar graph shows percentages of ADC ECs having individual transferred labels



Sub6 confirmed upregulation of these pathways (Fig. 3F), and IHC analysis further confirmed upregulation of a representative marker GPC4 for N/O-glycosylation in the tumor area of tissues in Sub6 (Fig. 3C). These results suggest that Sub5 is treatment-resistant Keratin-low/immune-cold squamous LACC while Sub6 is treatment-resistant ADC-enriched LACC, respectively.

#### EMT-inducing CIC mutations in treatment-resistant LACC Sub5

Deficiency of CIC in the nucleus was shown to promote cell growth and invasion by derepressing expression of ETV4 that can facilitate EMT [39, 40]. We showed that CIC mutations were enriched in the treatment-resistant Sub5 (Fig. 2G), which mainly occurred on the C-terminal region containing its nuclear localization signal (Supplementary Fig. S3D) and could contribute to the loss of nuclear CIC. Moreover, the mRNA level of ETV4 was upregulated in Sub5 as an mRNA signature of Sub5 (rna3 in Fig. 2GA and D). These data collectively propose a hypothesis that CIC mutations can cause the decrease of nuclear CIC, which leads to derepression of ETV4 mRNA expression and then contributes to EMT, a representative feature of Sub5. To test this hypothesis, we generated #6507A tumor cells carrying a frame shift deletion mutation in the C-terminal region of CIC from a treatment-resistant squamous Sub5 tumor and #6595 tumor cells from another treatment-resistant squamous Sub3 tumor as a control. Western blot analysis showed the significant decrease of nuclear CIC and also the substantial increase of cytosolic/nucleus ETV4 in #6507A (Sub5) cells compared to in #6595 (Sub3) cells (Fig. 3G and H). Moreover, the expression of epithelial cell (EC) markers CDH1 and CTNNA1 was decreased in #6507A cells compared to in #6595 cells while the expression of mesenchymal cell markers CDH2 and VIM was increased (Fig. 3G-I), indicating activation of EMT in #6507A (Sub5) cells. These data collectively support the hypothesis that CIC mutations can lead to EMT via the decreased nuclear CIC and the increased ETV4.

#### Cellular heterogeneity associated with LACC Sub1-6

We next investigated what cellular subpopulations are associated with LACC Sub1-6. To this end, we newly collected the 16 samples (12 SCC and 4 ADC samples) for scRNA-seq analysis such that at least one sample was collected for each RNA subtype to model the cellular heterogeneity across RNA1-4. RNA-seq was also performed for these samples, and RNA subtypes were predicted based on rna1-4 signatures as described for TCGA and merged cohorts (Supplementary Fig. S2C). Nonetheless, the sample collection was biased toward some RNA subtypes (e.g., RNA4). To resolve this bias problem and to

improve statistical power for characterization of cellular heterogeneity, we further integrated our scRNA-seq data with four previously reported datasets (GSE208653 [41] with 2 SCC and 1 ADC samples; GSE236738 [42] with 3 SCC samples; E-MTAB-12305 [43] with 4 SCC samples; and S-BSS1035 [44] with 3 SCC and 3 ADC samples). Using the quality control criteria in Supplementary Fig. S4A, we selected a total of 194,703 cells from our and previous data.

We performed clustering analysis for these selected cells and identified 14 clusters and their marker genes (Supplementary Fig. S4B and S4C). For detailed understanding of epithelial cells (ECs), we further performed subclustering of cells in the epithelial cluster. To avoid potential artifacts that can arise by mixing ECs from SCC and ADC, we first performed the subclustering only using ECs from SCC samples and identified 7 EC subclusters (KRT1/13, GADD45B, AKR1C2, PGK1, CXCL10, MUC5AC, and CEACAM5<sup>high</sup>; Fig. 4A) and their marker genes (Fig. 4B). The marker genes of KRT1/13, GADD45B, AKR1C2<sup>high</sup> subclusters showed 1) the largest overlaps with S1/2-G (rna1) and S1-P (prot1) (Fig. 4C); 2) the highest signature scores in Sub1 or 2 (Fig. 4D); and 3) significantly represented keratinization, TP53 signaling, and ROS pathways, the key pathways of Sub1-2 (Fig. 3A), respectively (Fig. 4E), indicating their strong association with Sub1-2. The marker genes of PGK1 and CXCL10<sup>high</sup> subclusters showed the largest or significant overlaps with S5-G (rna3) and S5-P (prot3) (Fig. 4C); the highest signature scores in Sub5 (Fig. 4D); and represented EMT signaling (TGFB/EGFR/Integrin/HIF) and ECM remodeling (metalloproteinase/urokinase) pathways (Fig. 4E), the key pathways of Sub5 (Fig. 3A), indicating their associations with Sub5. The marker genes of MUC5AC<sup>high</sup> subcluster showed the largest overlap with S6-G (rna2) and significant overlaps with protein signatures [S2/3-P (prot5) and S6-P (prot4)] of ADC-associated Sub3 and 6 (Fig. 4C); the highest signature scores in Sub3 and 6 (Fig. 4D); and represented the representative pathway of Sub6, O-linked glycosylation of mucin (Fig. 4E), suggesting its associations with Sub3 and 6. These results suggest that distinct EC subpopulations appear to dominate Sub1-6 (Fig. 4A, parenthesis).

To examine the relationships between ECs in SCC and ADC, we next transferred the labels of the above 7 SCC EC subclusters into ECs in ADC samples (Fig. 4F). The majority (82.7%) of ADC ECs were labeled with MUC5AC<sup>high</sup> (66.9%) and CEACAM5<sup>high</sup> (15.8%) subclusters (Fig. 4F) whose marker genes represented O-linked glycosylation of mucin (Fig. 4E). These results indicate that ADC ECs are present even in SCC samples, but in a relatively smaller proportion [e.g., 6.0% of MUC5AC<sup>high</sup> in SCC ECs (Fig. 4A) vs. 66.9% in ADC ECs



(Fig. 4F)], and there appear to be no EC subpopulations unique to ADC, as indicated by a small proportion (5.9%) of ECs with no SCC EC subcluster labels transferred. Of note, however, the marker genes of CEACAM5<sup>high</sup> subcluster, unlike the MUC5AC<sup>high</sup> subcluster, showed low signature scores for Sub6 (Fig. 4D), due to the significant overlaps with S1/2-G (rna1) and S5-G (rna3), besides S6-P (prot4), which caused to dilute their signature scores for Sub6.

To examine what subpopulations of myeloid cells are associated with Sub1-6, we next performed subclustering of macrophage, dendritic cells (DC), and mast cells in Supplementary Fig. S4B and identified 8 subclusters [mast cell; CD1C/LAMP3/CLEC9A<sup>high</sup> DC and plasmacytoid DC (pDC); and C1QB/FCGR3B/FCN1<sup>high</sup> macrophage] (Fig. 4G) and their marker genes (Supplementary Fig. S4D). The marker genes of the mast cell and CD1C/LAMP3/CLEC9A<sup>high</sup> DC/pDC and C1QB<sup>high</sup> macrophage subclusters showed 1) the largest overlaps with mRNA [S3/4-G (rna4)] and protein [S2/3-P (prot5) or S4-P (prot2)] signatures of immunogenic Sub3-4 (Fig. 4H); 2) the highest signature scores in Sub3-4 (Fig. 4I); and 3) represented innate (phagocytosis and antigen presentation) and adaptive (CD4/8<sup>+</sup> T-cell recruitment and activation/inhibition) immune pathways (Fig. 4J), consistent with their previously reported functions [45–48], suggesting their associations with the immunogenic Sub3-4. Of note, although the protein signatures of Sub4 represented more strongly the immune pathways (Fig. 3A), infiltrated proportions of these immune cells appeared to be comparable between Sub3 and 4, consistent with the findings from the TME analysis (Supplementary Fig. S3B). The marker genes of the FCGR3B and FCN1<sup>high</sup> macrophage subclusters showed the largest overlaps with S5-G (rna3) and/or S5-P (prot3) of Sub5; the highest signature scores in Sub5; and represented IL-17 and oncostatin M signaling pathways, the key immune pathways of Sub5 (Fig. 3A), as well as

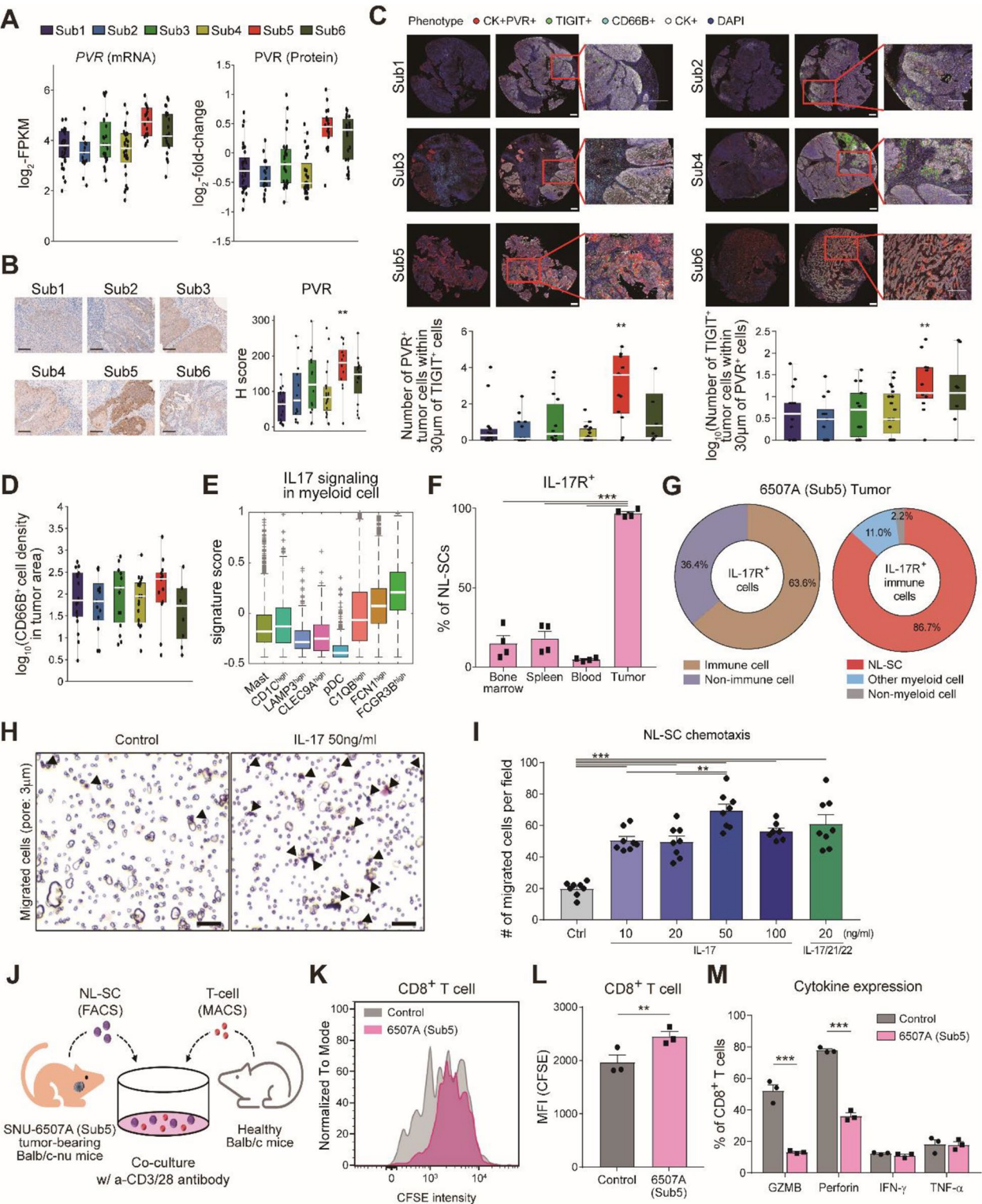
neutrophil activation pathways (neutrophil extracellular trap formation, NETosis, which can be induced by Th17-mediated inflammation and suppress T-cell activation), suggesting their associations with the treatment-resistant immune-cold Sub5. Finally, we performed the same subclustering analysis for cancer-associated fibroblasts (CAF) and smooth muscle cell (SMC) clusters in Supplementary Fig. S4B (Supplementary Fig. S4E–S4I). Among the subclusters, the marker genes of PDGFD and CXCL1<sup>high</sup> subclusters showed the largest overlaps with S3/4-G (rna4) and S2/3-P (prot5) of Sub3; the highest signature score in Sub3; and represented ECM pathways more strongly upregulated in Sub3 than in Sub4, as well as immune pathways (Fig. 4K–M), suggesting that they mainly constitute activated stroma in the treatment-resistant Sub3.

### Immune-suppressing pathways in treatment-resistant LACC Sub5

Among the treatment-resistant subtypes of LACC, the immune-cold squamous LACC Sub5 frequently exhibited aggressive local tumor growth even during the course of radiotherapy. To examine the modality of immunotherapy for this immune-cold Sub5, we first examined whether there could be immune checkpoints associated with suppression of T-cell activation in Sub5. Among the immune checkpoints, PVR, which induces immune evasion of tumor cells through its binding with TIGIT on T-cells, showed elevation of both mRNA and protein levels in Sub5 (Fig. 5A) and also higher expression levels in PGK1 and CXCL10<sup>high</sup> ECs upregulated in Sub5 than in the other EC subpopulations (Supplementary Fig. S5A). The IHC analysis further confirmed upregulation of PVR proteins in Sub5 (Fig. 5B). We then examined the interactions between PVR and TIGIT using multiplex fluorescence IHC stain and spatial proximity distance analysis and found that co-localization of PVR-expressing tumor cells and TIGIT-expressing cells within 30  $\mu$ m was most

(See figure on next page.)

**Fig. 5** Immune-suppressing pathways in treatment-resistant LACC Sub5. **A** PVR mRNA (left) and protein (right) expression patterns across Sub1-6. In violin plots, the line indicates the median value. **B** Representative images for PVR and its quantification results in Sub1-6 from IHC analysis. **C** Representative images for each subtype of 5-color mIHC staining scanned with PhenolmagerTM HT at 20 $\times$  magnification: PVR (red), TIGIT (yellow), NL-SC marker CD66B (cyan), cytokeratin (CK, white) and nucleus marker DAPI (blue). Scale bar = 200  $\mu$ m. In the original and magnified images of spatial cell-to-cell distance analysis, TIGIT-expressing cells were represented as green dots while PVR-expressing tumor cells within the 30  $\mu$ m radius from TIGIT-expressing cells were denoted as red dots. Box plots (right) showed the quantification results analyzed from 90 tumor tissues on TMA. **D** Quantification results in Sub1-6 from IHC analysis of CD66B. **E** Signature scores of mRNA/protein signatures of Sub5 involved in IL-17 signaling in the indicated myeloid subclusters. **F** IL-17R expressions on NL-SCs in bone marrow, spleen, blood, and tumor samples from #6507A (Sub5) tumor-bearing mice ( $n=4$ ). **G** Proportions of different IL-17R<sup>+</sup> cell types in #6507A (Sub5) tumors. **H** Representative images showing migration of NL-SCs (black triangles) in control and IL-17-treated groups. Scale bar = 100  $\mu$ m. **I** Number of migrated NL-SCs per field under indicated cytokine conditions. **J** Co-culture of NL-SCs and T-cells isolated from #6507A (Sub5) tumor models and naïve Balb/c mice, respectively. **K** Distribution of CFSE intensity for CD8<sup>+</sup> T-cells measured at Day 4 after the co-culture. **L** MFI of CFSE intensity for CD8<sup>+</sup> T-cells ( $n=3$ /group). **M** Effector cytokine expression (GZMB: Granzyme B, Perforin, IFN- $\gamma$ , TNF- $\alpha$ ) in CD8<sup>+</sup> T cells ( $n=3$ /group). Data are shown as the mean  $\pm$  SEM. \*,  $p<0.05$ ; \*\*,  $p<0.01$ ; \*\*\*,  $p<0.001$  from Student's t-test (**L** and **M**) and one-way ANOVA with Tukey's posthoc correction (**F** and **I**)



**Fig. 5** (See legend on previous page.)

evident in Sub5 compared to in the other subtypes (Fig. 5C), suggesting the increased PVR-TIGIT interactions in Sub5. These results collectively support the possibility that PVR can serve as a potential target for the treatment-resistant immune-cold Sub5.

Several myeloid subpopulations, such as suppressive neutrophils and macrophages, have been also reported to play critical roles in suppression of T-cell activation [49]. Thus, we next investigated whether these cells could contribute to suppression of T-cell activation in Sub5. Both mRNA and protein signatures for Sub5 significantly represented IL-17 signaling pathway (Fig. 3A), suggesting the presence of Th17 that can recruit these suppressive cells via IL-17 in Sub5. To check the presence of Th17 cells in Sub5, we performed subclustering of T-cell and NK cell clusters in Supplementary Fig. S4B and identified 8 T/NK cell subclusters and their marker genes (Supplementary Fig. S5B and S5C). The marker genes of CD4<sup>+</sup> T-cell subcluster, including Th17 cells, had a significant overlap with S5-G (rna3); relatively higher signature scores in Sub5 than in Sub1-2; and significantly represented Th17 cell differentiation (Supplementary Fig. S5D-S5F). Also, this CD4<sup>+</sup> T-cell subcluster included a substantial proportion of Th17 cells expressing IL-17 (Supplementary Fig. S5B, inset). These data collectively support the presence of Th17 cells in Sub5. Moreover, the marker genes of FCGR3B and FCN1<sup>high</sup> macrophage subclusters significantly represented IL-17 signaling, as well as neutrophil activation (Fig. 4J). These data collectively suggest a hypothesis that these macrophages with neutrophil-like characteristics may serve as the above suppressive cells activated by IL-17 released from Th17 cells in Sub5. To examine this hypothesis, we first checked the presence of these neutrophil-like suppressive cells (NL-SCs) in Sub5 tumors by performing IHC analysis of CD66B, an established marker for NL-SCs [50], across tumors in Sub1-6. The analysis revealed that Sub5 tumors had the highest density of CD66B<sup>+</sup> cells, compared to the tumors in the other subtypes, suggesting the presence of NL-SCs (Fig. 5D). Moreover, we also calculated the signature scores using the Sub5 mRNA/protein signatures for IL-17 signaling across the myeloid subclusters and found that the signature score (i.e., IL-17 signaling activation) was significantly elevated in FCGR3B and FCN1<sup>high</sup> macrophages, compared to in the other myeloid subclusters (Fig. 5E), supporting the above hypothesis.

To further test this hypothesis, we next developed a preclinical mouse model with the #6507A (Sub5) cells used in Fig. 3G and H (Supplementary Fig. S5G). Of note, after trying various models including orthotopic cervix tumor models using the cells, we were successful only for intratongue implantation models that reflect cancer clinical phenotypes as previously described [51–53].

Flow cytometry analysis showed higher percentages of IL-17R<sup>+</sup> NL-SCs in #6507A (Sub5) tumors than in the bone marrow, spleen, and blood (Fig. 5F and Supplementary Fig. S5H). NL-SCs mainly comprised cells expressing IL-17 receptors (IL-17R and IL-21R) in the tumors (Fig. 5G and Supplementary Fig. S5I). Moreover, chemotaxis/migration assays showed that IL-17 strongly induced the recruitment of NL-SCs in vitro (Fig. 5H and I). Finally, we examined the immunosuppressive function of NL-SCs by evaluating T-cell proliferation using carboxyfluorescein succinimidyl ester (CFSE) assays in the presence of NL-SCs (Fig. 5J). #6507A (Sub5) tumors showed higher mean fluorescence intensities (MFI) of CFSE in CD8<sup>+</sup> T-cells with NL-SCs than in the control without NL-SCs (Fig. 5K and L). Moreover, among four major effector cytokines (Granzyme B, Perforin, IFN- $\gamma$ , and TNF- $\alpha$ ), expression of Granzyme B and Perforin was attenuated in CD8<sup>+</sup> cytotoxic T cells co-cultured with NL-SCs (Fig. 5M). These data indicate that T-cell proliferation was disrupted by NL-SCs, supporting that NL-SCs can contribute to suppression of T-cell activation in Sub5 and thus serve as a potential target for the treatment-resistant immune-cold Sub5.

## Discussion

We have performed comprehensive proteogenomic analysis for LACC. First, comparison of genomic alterations between early-stage UCC and LACC revealed potential factors (keratinization-TP53 regulation axis in SCC and O-glycosylation-TP53 regulation axis in ADC) that could develop aggressive phenotypes in LACC. Second, proteogenomic results revealed three treatment-resistant subtypes of LACC (Sub3, 5, and 6) and further identified the molecular and cellular characteristics for the treatment-resistant nature of Sub3 and 5, both of which belonged to the Keratin-low subtype in TCGA: activated stroma and ADC-like ECs for Sub3 and ECs with high EMT potential and CIC mutations/PVR/NL-SCs for Sub5. Therefore, our comprehensive proteogenomic analysis identified the subtypes of LACC and provided the more detailed molecular/cellular characterization of the treatment-resistant LACC subtypes compared to the previous omics analyses of early-stage UCC-enriched cohorts.

The proteogenomic characterization of the treatment-resistant LACC subtypes can facilitate therapeutic stratification of these subtypes. First, for the immune-hot Sub3, activated stroma (PDGFD and CXCL1<sup>high</sup> CAFs) and MUC5AC<sup>high</sup> ECs can be targeted simultaneously to inhibit the EC-driven tumor growth and the CAF-driven aggravated ECM activation at the same time. The activated stroma have been shown to cause chemoresistance [54] and radioresistance [55], which can decrease the efficacy of primary-CCRT in LACC. Targeting these CAFs



can thus improve the effectiveness of primary-CCRT using the following agents: 1) PDGFD neutralizing antibody, CR002 [56] and MAB1159 (R&D systems, Catalog #: MAB1159) or inhibitors (crenolanib and imatinib [57]) of the PDGFD receptor PDGFRB for PDGFD<sup>high</sup> CAFs; and 2) CXCL1 inhibitor, Reparixin [58], or CXCL1 neutralizing antibody [59] for CXCL1<sup>high</sup> CAFs. Moreover, MUC5AC inhibitors, such as acridinium [58] and curcumin [60], can be treated together to simultaneously target MUC5AC<sup>high</sup> ECs. Second, for the immune-cold squamous Sub5, aggressive ECs (PGK1 and CXCL10<sup>high</sup> ECs) can be targeted at the same time to suppress the elevated EMT using the following agents: PGK1 inhibitor, aryl/alkyl bisphosphonates [61], for PGK1<sup>high</sup> ECs and CXCL10 inhibitor, an food drug administration (FDA)-approved atorvastatin [62], for CXCL10<sup>high</sup> ECs. Moreover, the immunotherapy to inhibit PVR and NL-SCs (FCN1 and FCGR3B<sup>high</sup> macrophages) that suppress anti-tumor T-cell activation can be further used to release such suppression. TIGIT blocking antibody [63] can be employed to reduce the tumor-induced suppression of T-cells in Sub5 tumors. In addition, IL-17 and IL-17R blockers (e.g., FDA-approved secukinumab [64] and brodalumab [65]) could be used to reduce the recruitment of NL-SCs and thereby T-cell suppression mediated by NL-SCs. Combinations of these inhibitors and blockers should be explored to improve the therapeutic efficacy for Sub5. Finally, for the ADC Sub6, MUC5AC<sup>high</sup> ECs and aberrant O-glycosylated mucins can be targeted to inhibit aggressive tumor growth at the same time. Together with the aforementioned MUC5AC inhibitors targeting MUC5AC<sup>high</sup> ECs, Ac5GalNTGc can be employed to inhibit O-linked glycosylation of mucins [66]. All these data strongly suggest that combinatorial therapeutic means targeting multiple modalities (stromal, immune, and tumor cells) should be employed to treat the treatment-resistant LACCs. Of note, Sub4 showed one of treatment-sensitive LACC subtypes with better survival than the above treatment-resistant subtypes, but worse survival than the treatment-sensitive Sub1-2. Integrated analysis of DFS and LRFS revealed that the medium prognosis could be due to a higher rate of distant metastasis in Sub4 than Sub1-2, which suggest that the more potent systematic chemotherapy and/or immuno-oncology agents are needed to improve the effectiveness of primary CCRT.

To test the validity of the proposed treatment strategies, the organoid models or co-culture systems can be developed using cells derived from tumor tissues in the treatment-resistant subtypes. To this end, surface markers for 1) ECs for Sub3 and 6 (MUC5AC<sup>high</sup> ECs) and Sub5 (PGK1 and CXCL10<sup>high</sup> ECs), 2) CAFs for Sub3 (PDGFD and CXCL1<sup>high</sup> CAFs), and 3) NL-SCs for Sub5

(FCN1 and FCGR3B<sup>high</sup> macrophages) should be first identified from the marker genes of these subpopulations (scRNA-seq data), and the subpopulations can be then isolated from tumor tissues by FACS using their surface markers and/or the conventional markers for ECs (EPCAM), CAFs (FAP/ACTA2), and macrophages (CD68). After confirming upregulation of the marker genes used for labeling the subpopulations, the organoid models containing the subpopulations (e.g., MUC5AC<sup>high</sup> ECs + PDGFD/CXCL1<sup>high</sup> CAFs + immune cells for Sub3 and PGK1/CXCL10<sup>high</sup> ECs + FCN1/FCGR3B<sup>high</sup> macrophages + CD8<sup>+</sup> T-cells for Sub5) can be generated. However, the organoids may be difficult to generate or not represent the subtype characteristics, such as upregulation of the representative pathways (immune and ECM pathways for Sub3 and EMT-related signaling pathways for Sub5). Alternatively, the co-culture systems can be developed for the subpopulation combinations. Using the organoid models or co-culture systems, the efficacy of the proposed treatment strategies can be tested using the aforementioned inhibitors and/or blockers for the subtype-associated cell types (e.g., reparixin [67] or CXCL1 blocker [59] for CXCL1<sup>high</sup> CAFs). In this study, we demonstrated that the co-culture system of NL-SCs + CD8<sup>+</sup> T-cells isolated from the mouse model could be used to show that NL-SCs suppress activation of CD8<sup>+</sup> T-cells, which can be further used to test the efficacy of NL-SC inhibitors (e.g., IL-17 and IL-17R blockers).

For the above therapeutic stratification of LACC, determination of the LACC subtypes for patients should proceed. First, genomic screening for the genes having the mutations enriched in the treatment-resistant subtypes (Fig. 2G) can be developed to predict the treatment-resistant nature. However, our WES data showed that somatic mutations of these genes were detected only in a subpopulation of the patients in the subtypes, thereby leading to false negatives in prediction of the treatment-resistance. To improve the prediction accuracy, we can further utilize the representative molecules for the following 1) key pathways and 2) subpopulations of tumor, stromal, and immune cells enriched in each subtype identified from our proteogenomic and scRNA-seq analyses: 1) keratinization and apoptosis/TP53 regulation pathways for Sub1-2 enriched with KRT1/13, GADD45B, AKR1C2<sup>high</sup> ECs and additionally platelet/complement activation pathways for Sub2; 2) immune and ECM pathways for Sub3 enriched with the aforementioned CAFs and ECs; 3) conventional innate/adaptive immune pathways for Sub4 enriched with CD1C/LAMP3/CLEC9A<sup>high</sup> DC/pDC and C1QB<sup>high</sup> macrophages; 4) EMT and ECM remodeling pathways for Sub5 enriched with the aforementioned ECs and NL-SCs; and 5) O-line glycosylation of mucins for Sub6 enriched with MUC5AC<sup>high</sup>

ECs. Besides the genomic screening, in practice, a panel of the representative mRNAs for the key pathways can be first developed, and IHC analysis for the representative proteins for both the key pathways and the enriched cell subpopulations can then further augment the genomic screening and mRNA panel-based predictions.

Our *in vitro* experiments using #6507A tumor cells suggested that the levels of ETV4 and EMT markers could be increased in the Sub5 patients with the loss-of-function mutations of CIC. We thus compared both mRNA and protein levels of ETV4 and EMT-related markers (CDH2, VIM, BSG, VAV2, EGFR, and FOSL1) between Sub5 patients with and without CIC mutations. Unexpectedly, however, there were found to be no significant differences in mRNA and protein levels of these molecules. There could be several reasons: 1) there were only three patients who were found to carry the CIC mutations, and no reliable statistical decision could be drawn due to the low statistical power ( $n=3$ ); 2) the sequencing depth of our WES data were around 100x, and the CIC mutations could be failed to be detected even if small subpopulations of tumor cells actually carried the mutations, making the labels of patients with and without the CIC mutations uncertain; 3) although CIC mutations were detected, they could not lead to the loss-of-function of CIC (e.g., missense mutation causing single amino acid variation) because the amount of functional CIC proteins in the nucleus is important, which could not be distinguished from proteogenomic data; and 4) mRNA and protein levels of the molecules were measured from the entire cells in the tumor tissues, and the levels could not truly represent those of these molecules in the tumor cells where CIC mutations affected their levels. Consistent with this finding, the similar insignificant differences were observed in the IHC data of the EMT-related markers (BSG, VAV2, EGFR, and FOSL1), possibly for the same reasons mentioned above. These results suggest that further detailed functional and clinical studies would be needed for the effect of CIC mutations using organoids and *in vitro* experiments in a large clinical cohort and also its potential clinical application in treating Sub5 patients.

In this study, we proposed a hypothesis that Th17 cells release IL-17, which in turn activates NL-SCs and they then suppress activation of cytotoxic T-cells in Sub5 based on the results from our experiments and data analyses. First, T/NK cell subclustering analysis showed that the marker genes of CD4<sup>+</sup> T-cell subcluster had a significant overlap with mRNA signature of Sub5, relatively high signature scores in Sub5, and represented Th17 cell differentiation, in addition to the fact that the subcluster included a substantial proportion of Th17 cells expressing IL-17. Based on these results, we hypothesized that Th17

cells could be present in Sub5. Despite these indications, however, the presence of Th17 cells should be validated in Sub5 tumors by IHC analysis using RORC antibodies. Second, based on the previous findings [68], we hypothesized that Th17 cells can serve as a major source of IL-17. However, other cells also can produce IL-17 in inflammatory conditions, including group3 innate lymphoid cells (ILC3),  $\delta\gamma$ T cells, invariant NKT cells, and activated macrophages [68]. It is thus needed to validate that Th17 cells truly act as a source of IL-17 in Sub5 tumors. To this end, we can isolate and culture Th17 cells from tumor tissues in Sub5 patients and then measure the released IL-17 in the culture media. Third, we showed that IL-17 strongly increased the migration of NL-SCs isolated from tumors in our mouse models. These results demonstrate a crucial role of IL-17 in the recruitment and activation of NL-SCs, consistent with the previous finding [69]. It can be further confirmed that the released IL-17 in the culture media can activate the NL-SCs isolated from Sub5 tumors. Finally, our co-culture experiments of NL-SCs and CD8<sup>+</sup> T-cells isolated from our mouse models revealed that NL-SCs could suppress T-cell activation, consistent with the previous finding [70]. Despite the supporting data from our experiments and data analyses, the aforementioned experiments should be still carried out to validate our hypothesis on the regulatory axis of Th17-NL-SC-T-cell in Sub5.

Our study has several issues that previous omics studies also had. First, tumor samples can also include non-tumor cells, which dilutes tumor-associated molecular signatures. To reduce this purity issue, we tried to focus on molecular signatures having consistent alteration patterns across all samples in individual clusters. Second, given local heterogeneity, locally sampled tissues for proteogenomic analysis may not represent disease states of the whole tumor which can cause inconsistency with observed phenotypes (clinical subtypes, metastasis, and survival). For example, locally sampled ADC tissues may include squamous cancer cells, especially in tumors with the adenosquamous nature, and the proteogenomic data for these tissues may include the mixed signatures of ADC and SCC. To resolve this issue partly, we performed scRNA-seq and identified EC subpopulations associated with Sub1-6 and tried to focus on consistent cellular pathways in both bulk proteogenomic and scRNA-seq data. Third, for the integrative analysis, we combined RNA-seq data (merged dataset) from multiple cohorts that included patients with heterogeneous clinical parameters and sample collection/storage conditions. To resolve this issue, we attempted to remove batch effects and focus on statistically powerful conserved signatures across different datasets. However, how much all these problems have been corrected is unclear, and the

proposed signatures should be thus validated in larger clinical cohorts before clinical applications. Finally, our proteogenomic analyses are limited to providing correlative signatures, not causative associations.

## Methods

### Sample collection

UCC tissue and blood samples were collected from patients prior to radiotherapy at the department of radiotherapy, National Cancer Center (NCC) in Korea from July 2004 to March 2020. Most samples were obtained via biopsy, and 68 samples were obtained from the Bio Bank of NCC, Korea. Pathologic diagnosis was made by two gynecological pathologists. UCC tissues obtained from patients who had undergone radiotherapy as the primary treatment were collected at the outpatient clinic before radiotherapy was initiated along with the blood samples. The tissues collected via biopsy were immediately immersed in liquid nitrogen. Surgical samples were carried in liquid nitrogen as soon as the tissues were removed in the operating room and transferred to a -80 °C deep freezer within 30 min. Sample and clinical data collections were approved by the NCC Institutional Review Board (IRB No. NCC 2016–0019) and informed consent was obtained from all patients. Staging workup included bimanual physical examination, chest and abdomino-pelvic computed tomography (CT), pelvic magnetic resonance imaging (MRI), and positron emission tomography (PET) scans in all patients. Sigmoidoscopy and cystoscopy were performed for all patients. All patients were clinically staged as International Federation of Gynecology and Obstetrics (FIGO) 2014 staging system.

### Radiotherapy

Concomitant chemoradiotherapy (CCRT) consisted of whole pelvic external beam radiotherapy (EBRT) with chemotherapy and high-dose-rate (HDR) brachytherapy. For patients having poor physical performance, however, RT alone was administered with the curative aim. Patients with stage IVB patients were treated in diverse methods including CCRT, RT alone, post-RT combination chemotherapy, and CCRT+combination chemotherapy. Whole pelvic EBRT dose was 45 to 50.4 Gy. HDR brachytherapy consisted of 6 fractions of CT- (between 2004 and 2008) or MRI-based (since 2009) 3-dimensional image-guided intra-cavitary radiotherapy given twice a week with 5 Gy per fraction (total 30 Gy). The total biologically equivalent dose in 2-Gy fractions to Point A ranged from 72.3 to 102.2 Gy, with a median value of 87 Gy. The final primary tumor response was determined by physical examination, cervical cytology or biopsy if needed, and MRI at 3 months after radiotherapy. Local

recurrence was defined as the presence of residual disease not resolved at 3 months after radiotherapy, which was confirmed by biopsy, or as the relapse at the cervix, vaginal, and/or parametrium.

### WES and RNA sequencing analysis

Genomic DNA for WES was isolated from frozen biopsy tumor tissues and peripheral blood buffy coat of patients using the QIAamp DNA Mini kit (Qiagen GmbH, Hilden, Germany). Total RNA for RNA sequencing was extracted from the frozen biopsy tumor tissue using TRIzol (Invitrogen®, Carlsbad, CA, USA), followed by cleaning with the RNeasy Mini Kit (Qiagen GmbH®, Hilden, Germany). Using the isolated genomic DNA and RNA, we generated the sequencing libraries and then performed WES and RNA sequencing analyses following the Illumina's standard protocols, as described in details in Supplementary Methods. The reads resulted from WES and RNA sequencing were aligned to the GRCh38 reference genome using BWA MEM (version 0.7.17) [71] for WES data and STAR (version 2.4.0) [72] for RNA sequencing data. For the WES data, we identified somatic mutations using GATK3.8 and Strelka2 (version 2.9.10) [73], significantly mutated genes (SMGs) using MutSigCV (version 1.4.1) [74], and then copy number alterations (CNAs) using CNVkit (version 0.9.9) [75]. For the RNA sequencing data, the fragments per kilobase of transcript per million mapped reads (FPKM) at the gene level were calculated using RSEM (version 1.3.3) [76]. See Supplementary Methods for the details regarding the preprocessing and analysis of WES and RNA sequencing data.

### HPV integration analysis

To search for HPV integration sites, we attempted to identify the RNA-seq reads that contained the human genome component in one side and the viral genome component in the other side (called human-virus fusion reads) using the two previously reported approaches complementary to each other. In the first approach, we applied the Virus-Clip tool [77] to align RNA-seq reads onto viral genome and obtained potential human-virus fusion reads. To remove the false positives, we separated the fusion reads into human and virus parts and first selected the ones with both parts longer than 21 bp (the minimum length for blast). For each selected fusion read, we then aligned both human and virus parts onto human transcriptome sequences (NCBI refseq RNA) using blastn (version 2.15.0) [78] with default parameters. We further selected the fusion reads in which the human part was aligned (E-value < 1) onto human transcriptome while the virus part was not (E-value < 1) within 5 bp of the region (< 2-codon bp) where the human part was aligned. Based on these selected human-virus fusion reads, we next



identified HPV integration sites as the ones supported by two fusion reads and more. In the second approach, as previously described in the TCGA study [10], we first applied the PathSeq tool (GATK version 4.2.2.0) [79] to the RNA-seq data and identified non-human reads that were not aligned onto human genome (NCBI GRCh38). We then applied the CTAT-VIF tool (version 1.5.0) [80] to the non-human reads and identified the human-virus fusion transcripts and their associated HPV sites. Finally, we combined the HPV integration sites identified from both approaches.

### Estimation of tumor cellularity

Tumor cellularity was estimated based on histological images. For each sample, the percentage of tumor cells relative to all cells, including stromal, immune, and tumor cells, was estimated using the representative hematoxylin–eosin (HE)-stained sample slide. At least ten random high-power fields (HPFs) and up to 30 HPFs were microscopically evaluated, and the average tumor cell count of the total cell counts in each HE-stained slide was estimated as histological tumor cellularity.

### Global proteome and phosphoproteome analyses

We cryopulverized each tumor tissue individually into tissue powder, as described previously [81], transferred the power to a new tube containing lysis buffer, centrifuged the lysate, and then extracted proteins. The resulting protein was digested by trypsin (V5111, Promega) with a slightly modified filter-aided sample preparation method [82]. To generate the universal reference peptides, we pooled 60 µg of peptides from each tumor peptide sample. For each TMT set, the universal reference peptides (300 µg, 126 channel) and ten different tumor samples (300 µg each, 127N through 131C channels) were labeled with 11-plex TMT reagent, according to the manufacturer's instructions. After desalting, immobilized metal affinity chromatography (IMAC) phosphopeptide enrichment was performed on all TMT-labeled peptides. The flow-through non-phosphopeptide samples from the IMAC experiments were fractionated based on mid-pH reverse-phase liquid chromatography fractionation (mRP fractionation), as previously described [82]. The previously developed DO-NCFC-RP/RPLC system [83–85] was modified to produce up to 24 online NCFC fractions. This system was operated in one-dimensional RPLC mode (for global proteome) or two-dimensional RP/RPLC mode (for phosphoproteome). Both global proteome and phosphoproteome were analyzed using a quadrupole-orbitrap mass spectrometer (Q Exactive HF-X, Thermo Fisher Scientific) with an electric potential

of 2.4 kV and desolvation capillary temperature at 250 °C for electrospray ionization.

A sample-specific customized database was constructed with a slight modification to a previously reported method [86]. MS/MS data for both global and phosphoproteome were processed using mPE-MMR for accurate precursor ion mass assignment. The refined MS/MS data were then subjected to MSGF+ (version 9949) [87] database search. For phosphoproteome datasets, the unidentified MS/MS data from the MSGF+ search were further subjected to a spectral library search [81]. Bipartite graph analysis using an in-house program was used to obtain protein groups from the identified peptides using a previously described process [81]. The 11-plex TMT labeling was used to quantify the protein abundances of the universal reference (channel 126) and ten tumor tissues (channels 127N–131C). After correcting the isotope impurity, the intensities of each TMT reporter ion (126–131C) were extracted with a mass tolerance of 0.005 Da from all MS/MS scans and then normalized using quantile normalization. See Supplementary Methods for the details regarding sample preparation, LC–MS/MS analysis, and database search for global proteome and phosphoproteome analyses.

### Identification of molecules correlated with genetic alterations

We used only expressed mRNAs, proteins, and phosphopeptides, respectively, that had FPKM > 1 (mRNAs) or were detected (proteins and phosphopeptides) in more than 50% of patients with protein and phosphorylation data available. For each SMG, we identified differentially expressed molecules between samples with and without alterations. To this end, for each molecule, we calculated a rank-sum statistic value and a log<sub>2</sub>-median-ratio. We then estimated the empirical null distributions of the rank-sum statistic value and log<sub>2</sub>-median-ratio via random permutation of all samples. Using the estimated empirical distributions, for each molecule, we computed adjusted p values for the observed rank-sum statistic value and log<sub>2</sub>-median-ratio, and then combined these p values using Stouffer's method [88]. Finally, we identified differentially expressed molecules as those with combined p values < 0.01 and absolute log<sub>2</sub>-median-ratios > a cutoff value, the mean of 2.5th and 97.5th percentiles of the empirical distribution for log<sub>2</sub>-median-ratios (e.g., log<sub>2</sub>-median-ratio = 0.45 for *KRT5/10/13/14/16*).

### Subtype identification

We first selected molecules (mRNAs, proteins, or phosphopeptides) with abundance data across all samples clustered to avoid the bias from missing values. For

the 1st stage clustering, we selected squamous tumors only and then top 10 (MAD10), 20 (MAD20), or 30% (MAD30) of molecules (mRNAs, proteins, or phosphopeptides) with the largest median absolute deviations (MADs). We performed a 1st orthogonal non-negative matrix factorization (ONMF [89]); clustering for the squamous tumors using MAD10, MAD20, or MAD30. Based on cophenetic correlations and consensus heat maps, we determined the number of clusters ( $hi\_k$ ) and cluster memberships, as previously described [90]. Before the 2nd ONMF clustering, we fixed the memberships of squamous tumors in an initial activation matrix, as described at the bottom of Supplementary Fig. S2A. Briefly, in the initial activation matrix, we set activation values for the squamous tumors to represent the memberships (e.g., [1 0 0 0] for a squamous tumor in s-Prot1). We also added zeros for the remaining activation values when the number of clusters ( $k$ ) was larger than  $hi\_k$ . For adenosquamous and adenocarcinoma, we assigned random activation values sampled from the uniform distribution. We performed the 2nd ONMF clustering with the initial activation matrix using the same MAD10, MAD20, or MAD30 molecules from squamous tumors with varying  $k$ . Based on cophenetic correlations, we finally determined  $k$  and cluster memberships.

#### Identification of molecular signatures defining the subtypes

To identify molecular signatures that defined the subtypes during clustering, we first defined ‘core samples’ for each subtype as the ones with positive silhouette width scores. To obtain the signature molecules defining each subtype, for each molecule, we compared  $\log_2$ -fold-changes in the core samples of the subtype with those of the other subtypes using the previously reported integrative statistical hypothesis testing method [91, 92] that computed an adjusted  $p$  value ( $p$ ) by combining  $p$  values obtained from two sample  $t$ -test and the median ratio test. For each comparison, the putative signature molecules were selected as those with  $p < 0.05$ . We then further filtered the selected molecules by choosing the ones with 1) a median value of patients in the subtype larger than zero, 2) a median value of the remaining patients less than zero, and 3) a median value of the patients in the subtype larger than that of the remaining patients.

#### Integrated clustering of subtypes identified from individual types of data

An integrated clustering was performed using all three types of data (mRNA, global proteome, and phosphoproteome data). Briefly, for each type of data, a cluster identified from the individual data clustering was first

converted to an indicator vector that included the ones for the samples in the subtype and zeros for the remaining samples. The indicator matrices for the three types of data were concatenated into the overall indicator matrix, which was then used as an input for  $k$ -means clustering. We selected the number of clusters ( $k$ , subtypes) as  $k=6$  (Sub1-6) after trying multiple  $k$  values and checking whether the subtypes resulted from the  $k$ -means clustering showed sufficient enrichment of clusters identified from the individual data. Cluster memberships of the samples were determined such that each sample was assigned to a subtype with the minimum distance to the subtype mean.

#### Survival analysis

Survival data were shown using Kaplan–Meier curves. Survival can be affected by treatment methods. Thus, to avoid biases in survival analysis, we excluded 36 patients who received RT alone, 15 patients who underwent surgery as a primary treatment, 8 patients who received palliative treatment or incomplete treatment, and 5 patients with treatment or survival information unavailable. In addition, we excluded 29 stage IVB patients that were treated with heterogeneous chemotherapeutic regimens during and after RT (Supplementary Table S1).

#### Pathway analysis

For integrated pathway analysis, we first identified molecular signatures for Sub1-6 identified from the integrated clustering based on the relationships between Sub1-6 and the clusters identified from the three types of data. For example, for Sub1, we identified the genes (S1-G) selected for RNA1 identified from mRNA data and the proteins (S1-P) selected for Prot1 and Phos1 identified from proteome and phosphoproteome data, respectively. To identify the pathways represented by the genes and proteins for Sub1-6, we performed an enrichment analysis of cellular pathways for the genes and proteins selected for each subtype (e.g., S1-G and S1-P for Sub1) using ConsensusPathDB [18]. The cellular pathways represented by the genes and proteins for each subtype were identified as those with  $p < 0.05$ , and the number of molecules involved in the pathway  $\geq 3$ .

#### Reconstruction of network models

We first obtained protein–protein interactions (PPIs) from ten interactome databases, including BioGRID [93], HuRI [94], IntAct [95], HitPredict [96], IID [97], MINT [98], DIP [99], HPRD [100], HTRIdb [101], and STRING [102]. For the list of molecules selected for network construction (molecules that have correlation with genetic alterations in Fig. 1G and 2J) and molecular signatures involved in activated stroma-related processes

associated with Sub3 in Fig. 3B, immune-related processes associated with Sub4 in Supplementary Fig. S3C, EMT-related processes associated with Sub5 in Fig. 3D, or the processes related to the production of proteoglycan and Tn antigen associated with Sub6 in Fig. 3F), we then extracted interactors of the selected molecules (i.e., the molecules and their 1st neighbors) as the nodes and the edges between them based on the PPIs. We visualized a network model to describe the extracted nodes and edges in Cytoscape (version 3.3.0) [103]. Among the 1st neighbors in the network model, we left only the key 1st neighbors without which the connection between the nodes for the selected molecules disappeared. We next added the activation and inhibition reactions between the nodes in the network model obtained from the relevant KEGG pathways (e.g., regulation of actin cytoskeleton in Fig. 1G) [104]. Finally, we arranged the nodes based on their localization obtained from the KEGG pathways.

#### scRNA-seq

Cellular suspensions were loaded onto a chromium controller (10×Genomics) to generate nanoliter-sized gel bead-in-emulsions (GEMs) containing single cells, reagents, and a single gel bead containing barcoded oligonucleotides. Barcoded sequencing libraries were prepared using Chromium Next GEM Single Cell 3' v3.1 Dual Index (10X Genomics) according to the manufacturer's protocol. The sequencing libraries were sequenced on NovaSeq 6000 (Illumina) with the following read lengths: 28 base pairs (bp) for Read 1 (16 bp 10×Barcode+12 bp UMI), 10 bp for Sample Index (dual), and 90 bp for Read 2. Next, we aligned the resulting reads from the sequencing to the GRCh38 reference genome and performed the unique molecular identifier (UMI) counting using Cell Ranger software (v.7.0.1) [105]. To improve statistical power for characterization of cellular heterogeneity in LACC, we also integrated scRNA-seq dataset generated from patients with SCC and ADC (GSE208653 [41], GSE236738 [42], E-MTAB-12305 [43], and S-BSST1035 [44]). Quality control, data normalization and integration, and cell clustering were performed using Seurat (v.4.0.4) [106]. See Supplementary Methods for the details regarding the preprocessing and analysis of scRNA-seq data.

#### Identification of marker genes for the individual cell clusters

Using the final merged UMI matrix for each cell cluster (or subcluster), we computed the adjusted  $p$ -values, log-fold changes, proportion of cells in the cluster expressing the gene (pct.1) using the 'scanpy.tl.rank\_genes\_groups' function from scanpy (v.1.10.1) [107]. Adjusted  $p$ -values (Pw) were calculated for HVGs using the Wilcoxon

rank-sum test followed by the Benjamini–Hochberg method. Moreover, to focus on the genes predominantly expressed in the cluster compared to in the other clusters, we further computed adjusted  $p$ -values (Pt) from the previously reported empirical t-test for HVGs for the comparison of pct.1 values in the cluster vs. those in the other clusters. Finally, we identified marker genes for the cluster as the ones meeting the following criteria: 1) adjusted Pw < 0.01, 2) log-fold changes > 0.4 (~1.49-fold), and 3) adjusted Pt < 0.05, and 4) the highest pct.1 in the cluster compared to in the other clusters. To search for the subtype of LACC associated with the subtypes of the aforementioned four cell types, we first evaluated the significance of overlaps between 1) the bulk RNA and protein signatures for Sub1-6 (e.g., S1-G and S1-P for Sub1) and 2) marker genes from individual subclusters of the aforementioned four cell types (e.g., PGK1<sup>high</sup> EC subcluster) using the Fisher's exact test followed by the Benjamini–Hochberg method. Second, we calculated the signature score for each subcluster in individual samples of Sub1-6 as follows: 1) we selected expressed genes with FPKM > 1 in more than half of the samples in Sub1-6, and then applied quantile normalization to  $\log_2(\text{FPKM} + 1)$  for these expressed genes; and 2) after auto-scaling the normalized  $\log_2$ -FPKM values, for each sample, we calculated a signature score as the averaged auto-scaled value for the marker genes of the subcluster. Finally, we identified the pathways represented by the marker genes for each subcluster by performing the enrichment analysis of cellular pathways for the marker genes using ConsensusPathDB [18] and selecting the cellular pathways with  $p < 0.05$ , and the number of molecules involved in the pathways  $\geq 3$ .

#### Tissue microarrays and IHC

Tissue samples were composed of four or more pieces obtained by multiple punch biopsies, measuring approximately 3×3 mm each. All pieces from each tumor were formalin-fixed and paraffin-embedded into a single block, and 4 μm-thick sections were stained with hematoxylin and eosin (H&E). After evaluation of H&E tissue sections in each case, representative neoplastic areas were marked, and the corresponding paraffin block was retrieved. A tissue core of 2.0 mm in diameter was obtained from each selected block using Quick-Ray manual tissue microarrayer (UNITMA, Seoul, Korea). Tissue microarrays (TMAs) containing 2 normal cervical tissues and 93 patient tumor tissues were generated and used for IHC and multiplex fluorescence IHC (mIHC) stain. For this TMA slide generation, without any selection, we used all 93 tumor tissues having sufficient amounts available to mount on the TMA slides among the 146 tumor tissues after proteogenomic analysis- 17 of 33 tumors in



Sub1, 11 of 17 tumors in Sub2, 16 of 27 tumors in Sub3, 20 of 28 tumors in Sub4, 13 of 18 tumors in Sub5, and 16 of 23 tumors in Sub6.

IHC stain was carried out using Roche IHC/ISH system (BenchMark ULTRA) and OptiView DAB IHC Detection Kit (Ventana, 760–700). The following antibodies were used as the primary antibodies: anti-BSG (Abcam, Clone: 10E10), anti-EGFR (Ventana, Clone: 3C6), anti-FOSL1 (Santa Cruz, Clone: C-12), anti-VAV2 (Atlas, HPA003224), anti-ICOS (Abcam, Clone: EPR20560), anti-MMP2 (Invitrogen, Clone: 101), anti-GPC4 (Atlas, HPA030836). IHC slides were scanned using the Aperio AT2 DX System (Leica Biosystems) at 20× magnification and the images are displayed at 20× magnification using the Aperio ImageScope (Leica Biosystems). Quantification of IHC staining was performed using H-score, which was calculated by intensity and percentage of positively stained cells. The intensity was manually assessed in 4-grade (0, 1, 2, and 3) by two different pathologists as follows: If less than 10% of tumor cells were 1 or higher, the tumor was graded as 0 (negative); if more than 10% of tumor cells were 1 or higher but less than 10% of tumor cells were 2, the tumor was graded as 1 (weak positive); if more than 10% of tumor cells were 2 or higher but less than 10% of tumor cells were 3, the tumor was graded as 2 (moderate positive); and if more than 10% of tumor cells were 3, the tumor was graded as 3 (strong positive). The consensus grades between the two pathologists were determined. H-score was then obtained by multiplying the percentage and intensity:  $3 \times (\text{percentage of } 3+ \text{ cells}) + 2 \times (\text{percentage of } 2+ \text{ cells}) + 1 \times (\text{percentage of } 1+ \text{ cells})$ . Of note, for ICOS for which H-score was not reliably estimated, we used the density of the positive cells, instead of H-score: the number of positive cells (1+, 2+ or 3+) per mm<sup>2</sup>. This procedure was done for 1) the total area of the core regions mounted on the TMA slides and 2) subregions enriched with tumor cells. In addition, for MMP2 expressed in a significant portion of stromal cells across tumors, we defined subregions enriched with stromal cells in each tumor and estimated grades for these subregions because the percentage of positive cells could not be reliably determined. Of note, when selecting the subregions for tumor- or stroma-enriched subregions from the total image for each tumor, to focus on the representative subregions and avoid the selection bias, we tried to 1) select the subregions where the tumor or stroma contents were close to the average content in the total image, and also to 2) avoid the subregions that could contain the artifacts. In this manner, the number of the selected subregions varied from one to six, depending on the status of the tumor or stromal

distribution in the tumor, and grade/H-score was calculated for all the subregions at once.

### Multispectral imaging and analysis

mIHC stain was performed on prismCDX Co., Ltd (Gyeonggi-do, Korea) with a Leica Bond Rx<sup>TM</sup> Automated Stainer (Leica Biosystems). Briefly, the TMA slides were dewaxed with Leica Bond Dewax solution (Leica Biosystems, AR9222), followed by antigen retrieval with Bond Epitope Retrieval 2 (Leica Biosystems, AR9640). The staining process was performed four times in sequential rounds, including incubation with blocking buffer, primary antibody, and Mouse/Rabbit HRP secondary antibody (TheraNovis, C0105), and visualization of antigen with Astra-fluorophore (TheraNovis). Bond Epitope Retrieval 1 was treated to remove bound antibodies before proceeding to the next step. Finally, the nuclei were counterstained with DAPI (Thermo Scientific, 62,248). The slides were coverslipped using ProLong Gold antifade reagent (Invitrogen, P36930). The following antibodies were used: anti-PVR (Cell Signaling Technology, Clone: D8A5G), anti-TIGIT (Abcam, Clone: BLR047F), anti-CD66b (Novus, Clone: G10F5), anti-CK (Novus, Clone: AE-1/AE-3). The following fluorophores were used:

Astra-570 (TheraNovis, C0110), Astra-690 (TheraNovis, C0114), Astra-620 (TheraNovis, C0113), Astra-780 (TheraNovis, C0116).

mIHC slides were scanned using the PhenoImager<sup>TM</sup> HT (Akoya Biosciences) at 20× magnification. The representative images for training were selected in PhenoChart<sup>TM</sup> Whole Slide Viewer (1.1.0 version, Akoya Biosciences), and an algorithm was created in the inForm<sup>®</sup> Tissue Analysis software (2.6 version, Akoya Biosciences). Multispectral images were unmixed using the spectral library in the inForm software. Tissue was segmented based on cytokeratin (CK), an EC marker used to distinguish between parenchyma and stroma, and each single cell was segmented based on DAPI staining. The cells were phenotyped according to the expression compartment and intensity of each stained marker. After designating all tissue cores to be analyzed on the TMA slide, the created algorithm was applied for the batch analysis. The exported data were consolidated and analyzed in R studio (4.2.1 version) using the phenoptr (Akoya Biosciences) and phenoptrReport (Akoya Biosciences) packages. We calculated the distance between two cells located closest to each other using the 'find\_nearest\_distance' function and then determined the number of specified cells (cells expressing TIGIT) interacting with at least one reference cell (tumor cells or NL-SCs expressing PVR) at the proximity distance and the mean number of reference

cells within 30  $\mu\text{m}$  radius of one specified cell, as previously described [108].

#### Primary culture of cells derived from tumor tissues

Tumor samples were finely minced with scissors and dispersed into small aggregates by pipetting. Fine neoplastic tissue fragments were seeded into T-25 flasks. Tumor cells were initially cultured in Opti-MEMI (Thermo Fisher Scientific, MA, USA) with 5% fetal bovine serum (FBS). After primary culture, cells were sustained in RPMI 1640 (Thermo Fisher Scientific, MA, USA) with 10% fetal bovine serum and 1% (v/v) penicillin and streptomycin (10,000U/ml). Incubated flasks in humidified incubators at 37 °C in an atmosphere of 5%  $\text{CO}_2$  and 95% air [109].

#### Immunodetection

Western blot analysis and immunofluorescence stain were performed using 1) #6507A tumor cells derived from one tissue of a Sub5 patient carrying a frame shift deletion mutation in the C-terminal region of CIC and 2) #6595 tumor cells (control) derived from one tissue of a Sub3 patient as a control without CIC mutation. Each experiment was independently repeated twice. Cytoplasmic, membrane and nuclear protein extracts were prepared with a subcellular protein fraction Kit (Thermo Scientific, 78,840) and used for western blot analysis. Western blotting was carried out according to the standard procedures using the enhanced-chemiluminescence detection (Amersham™, RPN2232). The following antibodies were used: anti-CIC (Invitrogen, PA5-83,721), anti-ETV4 (Proteintech, 10,684-1-AP), anti-CDH1 (Cell Signaling Technology, Clone: 24E10), anti-CTNNA1 (BD Bioscience, Clone: 5), anti-CDH2 (BD Bioscience, Clone: 32), anti-VIM (BD Bioscience, Clone: RV202), anti- $\beta$ -Actin (Abcam, Clone: AC-15). Immunofluorescence stain was performed with patient-derived tumor cells. The cells were attached onto glass slides for O/N and fixed in 3.7% formaldehyde at 4°C cold room for 30 min. They were permeabilized with 0.5% Triton X-100 in PBS (2 mM  $\text{MgCl}_2$ ) at room temperature for 10 min and treated with 5% BSA in PBS (2 mM  $\text{MgCl}_2$ ) for 1 h to block nonspecific reaction. Sequential incubation of primary antibody and Alexa fluorescence-conjugated secondary antibody (Invitrogen, A11037 and A11029) was performed using DAPI (Invitrogen, D1306) as a counterstain. Immunofluorescence stain images were acquired at 40 $\times$  magnification using a Zeiss LSM 780 confocal microscope.

#### Intratongue cervical cancer model

We cultured #6507A (Sub5) cells derived from human UCC tissues in RPMI-1640 medium (Biowest, L0498) containing 10% fetal bovine serum (FBS). Subconfluent

cancer cells were harvested and washed with fresh RPMI-1640, followed by counting the number of cells. For the intratongue cervical cancer model,  $2 \times 10^6$  #6507A cancer cells were suspended in 40  $\mu\text{l}$  of RPMI-1640/Matrigel (Corning 354,262) mixture (1:1). Seven-ten-week-old Female BALB/c-nu mice were anesthetized with ketamine (100 mg/kg) and xylazine (10 mg/kg) via intraperitoneal injection. After anesthetization, mice were injected with #6507A cancer cells submucosally into the tongue using an insulin syringe with a 29-gauge needle (BD, 320,320). Tumor size was measured with a caliper every two days from 6 days after cancer cell inoculation until the end of the study. Tumor volume was calculated using the following formula: tumor volume =  $(\text{LD} \times \text{SD}^2)/2$ , where LD and SD indicate the long and short diameters, respectively. For processing of tumor tissue and blood, spleen, and bone marrow samples, see Supplementary Methods.

#### Flow cytometry

The single-cell suspensions ( $\sim 1.5 \times 10^6$  cells) from tumor, blood, spleen, and bone marrow samples were incubated with LIVE/DEAD™ Fixable Blue Dead Cell Stain Kit for UV excitation (Invitrogen, L23105) for 30 min at 4 °C and then stained with the following fluorescent monoclonal antibodies according to the manufacturer's protocol: Anti-CD45 (Invitrogen, Clone: 30-F11), anti-CD11b (Biolegend, Clone: M1/70), anti-Ly6C (Invitrogen, Clone: HK1.4), anti-Ly6G (Invitrogen, Clone: 1A8), anti-IL-17RB (Biolegend, Clone: 9B10), and anti-IL-21R (Biolegend, Clone: 4A9). To measure cytokine expression in  $\text{CD8}^+$  T cells, cells stained with surface monoclonal antibodies were fixed with intracellular fixation buffer (Invitrogen, 00-8222-49) for 25 min at room temperature, and then stained with the following fluorescent monoclonal antibodies according to the manufacturer's protocol: anti-TNF- $\alpha$  (Biolegend, Clone: MP6-XT22), anti-Granzyme B (Invitrogen, Clone: NGZB), anti-Perforin (Biolegend, Clone: S16009A), anti-IFN- $\gamma$  (Biolegend, Clone: XMG1.2). The stained samples were then analyzed using BD Symphony or BD LSRFortessa. Flowjo v10 was used for Flow cytometry data analysis.

#### CFSE T-cell proliferation assay

NL-SCs were sorted (FACS Aria III) from the spleen of #6507A tumor-bearing BALB/c-nu mice, and  $\text{CD4}^+$  T/ $\text{CD8}^+$  T-cells were sorted from the spleens of naïve BALB/c mice. The sorted T-cells were labeled with CFSE Cell Division Tracker Kit (BioLegend, 423,801) following the manufacturer's protocol. The final concentration of CFSE was 0.5  $\mu\text{M}$ . CFSE-labeled T-cells ( $1 \times 10^5$  cells) were co-cultured with NL-SCs ( $1 \times 10^5$  cells) for 4 days with or without NL-SCs in a 96-well plate pre-coated

with anti-CD3e (1 µg/mL, BioLegend, 145-2C11) and anti-CD28 antibodies (1 µg/mL, BioLegend, 37.51) overnight at 4 °C. CFSE levels in T-cells were analyzed by flow cytometry using BD LSRII. Flowjo v10 was used for Flow cytometry data analysis.

### Migration assay

NL-SCs ( $1 \times 10^5$ ) were seeded on a 3 mm pore PET membrane transwell insert (SPL 37124) in the upper chamber. The lower chamber included 10, 20, 50, or 100 ng/ml of recombinant mouse IL-17 (Peprotech 210–17), or triple combination of 20 ng/ml of IL-17, 20 ng/ml of IL-21 (Peprotech 210–21), and 20 ng/ml of IL-22 (Peprotech 210–22). After 6 h at 37 °C, 5% CO<sub>2</sub>, the medium in the upper chamber was removed and the membranes were fixed with 4% paraformaldehyde (PFA) for 20 min at room temperature (RT). After fixation, the membranes were dried for 10 min and stained with hematoxylin solution (Vitro Vivo Biotech) for 10 min at RT. After non-migrated cells were gently removed with a cotton tip, migrated cells were imaged using a bright-field microscope at 200X magnification. Image J was used to control the contrast images for accurate counting of migrated cells.

### Statistics

The measured values are presented as the mean ± SEM. Comparisons between multiple groups (> 2 groups) were made using ANOVA with Tukey's or Sidak's post hoc correction. Student's t-tests were used to compare data between two groups. Statistical significance was defined as  $p < 0.05$ .

### Abbreviations

LACC	Locally advanced cervical cancer
ECM	Extracellular matrix
EC	Epithelial cell
EMT	Epithelial-mesenchymal-transition
UCC	Uterine cervical cancer
CCRT	Concurrent cisplatin-based chemoradiotherapy
TCGA	The Cancer Genome Atlas
MS	Mass spectrometry
WES	Whole-exome sequencing
scRNA-seq	Single-cell RNA-seq
FIGO	International federation of gynecology and obstetrics
SMG	Significantly mutated gene
HPV	Human papillomavirus
CNAs	Copy number alterations
PAG-LACC	Predominantly altered genes in LACC
GSEA	Gene set enrichment analysis
ROS	Reactive oxygen species
LC-MS/MS	Liquid chromatography-tandem mass spectrometry
SCC	Squamous cell carcinoma
ADC	Adenocarcinoma
DFS	Disease-free survival
LRFS	Local recurrence-free survival
TME	Tumor microenvironment
IHC	Immunohistochemistry
DC	Dendritic cell
CAF	Cancer-associated fibroblast

SMC	Smooth muscle cell
NL-SC	Neutrophil-like suppressive cell
CFSE	Carboxyfluorescein succinimidyl ester
MFI	Mean fluorescence intensity
NCC	National Cancer Center
IRB	Institutional Review Board
CT	Computed tomography
MRI	Magnetic resonance imaging
PET	Positron emission tomography
EBRT	External beam radiotherapy
HDR	High-dose-rate
FPKM	Fragments per kilobase of transcript per million mapped reads
HE	Hematoxylin-eosin
HPF	High-power fields
ABC	Ammonium bicarbonate
TMT	Tandem mass tag
TEAB	Triethylammonium bicarbonate
IMAC	Immobilized metal affinity chromatography
DIW	Deionized water
ACN	Acetonitrile
NCFC	Noncontiguous fractionating and concatenating
AGC	Automatic gain control
HCD	Higher-energy collisional dissociation
TPM	Transcripts per million mapped reads
PSM	Peptide-spectrum match
UMC	Unique mass class
FDR	False discovery rate
NET	Normalized elution time
MAD	Median absolute deviation
ONMF	Orthogonal non-negative matrix factorization
PDAC	Pancreatic ductal adenocarcinoma
PPI	Protein-protein interaction
GEM	Gel bead-in-emulsion
GEO	Gene Expression Omnibus
UMI	Unique molecular identifier
HVG	Highly variable gene
PCA	Principal component analysis
PC	Principal component
NES	Normalized enrichment score
H&E	Hematoxylin and eosin
TMA	Tissue microarray
miHC	Multiplex fluorescence IHC
CK	Cytokeratin
FBS	Fetal bovine serum
PFA	Paraformaldehyde

### Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12943-025-02256-3>.

Supplementary Material 1  
Supplementary Material 2  
Supplementary Material 3  
Supplementary Material 4  
Supplementary Material 5  
Supplementary Material 6

### Acknowledgements

This work was supported by grants from the Proteogenomic Research Project of National Cancer Center, Korea (NCC-1810861, NCC-1810862, NCC-1711260, NCC-2210490), the Basic Science Research Program through the National Research Foundation (NRF) funded by the Ministry of Education, Korea (2018R1D1A1A09084250, 2021R1F1A1052407, RS-2023-00208004, RS-2023-00217571), and the NAVER Corporation (Grant No. 3720230110). This work was also supported by the Korea Basic Science Institute (KBSI) National Research Facilities & Equipment Center (NFEC) grant funded by the Korean government



(Ministry of Education) (2019R1A6C1010028) and by the NRF funded by the Korean government (Ministry of Science and ICT) (NRF-2022M3H9A2086450).

#### Authors' contributions

J-Y.K., S-W.L., D.H., S.S.K. and K.J. designed and directed the integrated proteogenomic analysis. H-J.S., M.C.L., C.W.Y., E-K.H., H.K., and J-Y.K. collected and characterized the tumor samples. D.Y.H., S.Y.C., Y-K.B., J.K.K., S-J.L., Y.S., D.H., and J-Y.K. performed genomic analysis of tumor samples and analyzed the genomic data. S-W.L. and D.H. designed the proteomic experiments. D.N., J-G.B., S-J.K., S.B., C.K., I.H.M., H.K., S.K., and S-W.L. performed global proteome and phosphoproteome profiling experiments and unified database searches and analyzed the proteomic data. D.Y.H., Y.S., and D.H. performed integrated analyses with genomic and proteomic data, as well as clustering and network analyses. E.J. and D.H. designed scRNA-seq analysis and performed integrated data analysis. H.J.B., S.S.K. performed functional experiments to test prognostic factors, and C.W.Y. and E-K.H. performed immunohistochemistry experiments. J.J., D.H.S., J-L.K., D.K.K., J.K., G.W.P., K.S.P. and K.J. developed primary cell and mouse models and performed functional experiments related to CAFs, NL-SCs, and T-cells. D.H., S-W.L., K.J., S.S.K., J-Y.K., D.Y.H., D.N., H-J.S., J.J., E.J., W.C., and S.Y.C. wrote the manuscript.

#### Funding

This work was supported by grants from the Proteogenomic Research Project of National Cancer Center, Korea (NCC-1810861, NCC-1810862, NCC-1711260, NCC-2210490), the Basic Science Research Program through the National Research Foundation (NRF) funded by the Ministry of Education, Korea (2018R1D1A1A09084250, 2021R1F1A1052407, RS-2023-00208004, RS-2023-00217571), and the NAVER Corporation (Grant No. 3720230110). This work was also supported by the Korea Basic Science Institute (KBSI) National Research Facilities & Equipment Center (NFEC) grant funded by the Korean government (Ministry of Education) (2019R1A6C1010028) and by the NRF funded by the Korean government (Ministry of Science and ICT) (NRF-2022M3H9A2086450).

#### Data availability

The exome and mRNA sequencing data generated in this study are available in dbGaP (<https://www.ncbi.nlm.nih.gov/gap/>; accession ID: phs002916.v1.p1). Mass-spectrometry-based global and phosphoproteomic data are available in the CPTAC data portal (<https://pdc.cancer.gov/pdc/>; accession ID: PDC000342 for global proteome and PDC000343 for phosphoproteome) and also in the Korea BioData Station (<https://kbds.re.kr>; Data ID: KPX10000219). Single-cell RNA-seq data are available in the GEO database (accession ID: GSE279998).

#### Declarations

##### Ethics approval and consent to participate

Biological samples from patients were obtained with the consent form approved by the IRB of NCC, Korea (NCC IRB No. 2016-0019). All animal experiments were approved by the Institutional Animal Care and Use Committee of Seoul National University (authorization no. SNU-201127-1-2).

##### Competing interests

The authors declare no competing interests.

##### Author details

<sup>1</sup>School of Biological Sciences, Seoul National University, Seoul 08826, Republic of Korea. <sup>2</sup>Department of Chemistry and Center for Proteogenome Research, Korea University, Seoul 02843, Republic of Korea. <sup>3</sup>Research Institute and Hospital, National Cancer Center, Goyang 10408, Republic of Korea. <sup>4</sup>Department of Anatomy and Cell Biology and Department of Biomedical Sciences, Seoul National University College of Medicine, Seoul 03080, Republic of Korea. <sup>5</sup>Korean Cell Line Bank, Laboratory of Cell Biology, Cancer Research Institute, Seoul National University College of Medicine, Seoul, Korea. <sup>6</sup>Institute of Allergy and Clinical Immunology, Seoul National University Medical Research Center, Seoul 03080, Republic of Korea.

Received: 7 November 2024 Accepted: 1 February 2025  
Published online: 14 March 2025

#### References

- Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2018;68:394–424.
- Monk BJ, Tan DSP, Hernandez Chagui JD, Takyar J, Paskow MJ, Nunes AT, Pujade-Lauraine E. Proportions and incidence of locally advanced cervical cancer: a global systematic literature review. *Int J Gynecol Cancer*. 2022;32:1531–9.
- Kim J, Lee D, Son KB, Bae S. The burden of cervical cancer in Korea: a population-based study. *Int J Environ Res Public Health*. 2020;17:6308.
- Kau YC, Liu FC, Kuo CF, Huang HJ, Li AH, Hsieh MY, Yu HP. Trend and survival outcome in Taiwan cervical cancer patients: a population-based study. *Medicine (Baltimore)*. 2019;98:e14848.
- Utada M, Chernyavskiy P, Lee WJ, Franceschi S, Sauvaget C, de Gonzalez AB, Withrow DR. Increasing risk of uterine cervical cancer among young Japanese women: Comparison of incidence trends in Japan, South Korea and Japanese-Americans between 1985 and 2012. *Int J Cancer*. 2019;144:2144–52.
- Cervical Cancer Treatment: Patient version. In: PDQ cancer information summaries [Internet]. Bethesda: National Cancer Institute (US); 2002. <https://pubmed.ncbi.nlm.nih.gov/26389422/>.
- Lorusso D, Xiang Y, Hasegawa K, Scambia G, Leiva M, Ramos-Elias P, Acevedo A, Sukhin V, Cloven N, Pereira de Santana Gomes AJ, et al. Pembrolizumab or placebo with chemoradiotherapy followed by pembrolizumab or placebo for newly diagnosed, high-risk, locally advanced cervical cancer (ENGOT-cx11/GOG-3047/KEYNOTE-A18): a randomised, double-blind, phase 3 clinical trial. *Lancet*. 2024;403:1341–50.
- Monk BJ, Toita T, Wu X, Vazquez Limon JC, Tarnawski R, Mandai M, Shapira-Frommer R, Mahantshetty U, Del Pilar E-D, Zhou Q, et al. Durvalumab versus placebo with chemoradiotherapy for locally advanced cervical cancer (CALLA): a randomised, double-blind, phase 3 trial. *Lancet Oncol*. 2023;24:1334–48.
- Marabelle A, Fakih M, Lopez J, Shah M, Shapira-Frommer R, Nakagawa K, Chung HC, Kindler HL, Lopez-Martin JA, Miller WH Jr, et al. Association of tumour mutational burden with outcomes in patients with advanced solid tumours treated with pembrolizumab: prospective biomarker analysis of the multicohort, open-label, phase 2 KEYNOTE-158 study. *Lancet Oncol*. 2020;21:1353–65.
- Cancer Genome Atlas Research N, Albert Einstein College of M, Analytical Biological S, Barretos Cancer H, Baylor College of M, Beckman Research Institute of City of H, Buck Institute for Research on A, Canada's Michael Smith Genome Sciences C, Harvard Medical S, Helen FGCC, et al. Integrated genomic and molecular characterization of cervical cancer. *Nature*. 2017;543:378–84.
- Espinosa AM, Alfaro A, Roman-Basaure E, Guardado-Estrada M, Palma I, Serralde C, Medina I, Juarez E, Bermudez M, Marquez E, et al. Mitosis is a source of potential markers for screening and survival and therapeutic targets in cervical cancer. *PLoS ONE*. 2013;8:e55975.
- Medina-Martinez I, Barron V, Roman-Bassaure E, Juarez-Torres E, Guardado-Estrada M, Espinosa AM, Bermudez M, Fernandez F, Venegas-Vega C, Orozco L, et al. Impact of gene dosage on gene expression, biological processes and survival in cervical cancer: a genome-wide follow-up study. *PLoS ONE*. 2014;9:e97842.
- Shin HJ, Joo J, Yoon JH, Yoo CW, Kim JY. Physical status of human papillomavirus integration in cervical cancer is associated with treatment outcome of the patients treated with radiotherapy. *PLoS ONE*. 2014;9:e78995.
- Manzo-Merino J, Contreras-Paredes A, Vazquez-Ulloa E, Rocha-Zavaleta L, Fuentes-Gonzalez AM, Lizano M. The role of signaling pathways in cervical cancer and molecular therapeutic targets. *Arch Med Res*. 2014;45:525–39.
- Green JA, Kirwan JM, Tierney JF, Symonds P, Fresco L, Collingwood M, Williams CJ. Survival and recurrence after concomitant chemotherapy and radiotherapy for cancer of the uterine cervix: a systematic review and meta-analysis. *Lancet*. 2001;358:781–6.
- Huang J, Qian Z, Gong Y, Wang Y, Guan Y, Han Y, Yi X, Huang W, Ji L, Xu J, et al. Comprehensive genomic variation profiling of cervical

- intraepithelial neoplasia and cervical cancer identifies potential targets for cervical cancer early warning. *J Med Genet.* 2019;56:186–94.
17. Ojesina AI, Lichtenstein L, Freeman SS, Pedamallu CS, Imaz-Rosshandler I, Pugh TJ, Cherniack AD, Ambrogio L, Cibulskis K, Bertelsen B, et al. Landscape of genomic alterations in cervical carcinomas. *Nature.* 2014;506:371–5.
  18. Kamburov A, Wierling C, Lehrach H, Herwig R. ConsensusPathDB—a database for integrating human functional interaction networks. *Nucleic Acids Res.* 2009;37:D623–628.
  19. Matsuo K, Mandelbaum RS, Machida H, Purushotham S, Grubbs BH, Roman LD, Wright JD. Association of tumor differentiation grade and survival of women with squamous cell carcinoma of the uterine cervix. *J Gynecol Oncol.* 2018;29:e91.
  20. Cai BH, Hsu PC, Hsin IL, Chao CF, Lu MH, Lin HC, Chiou SH, Tao PL, Chen JY. p53 acts as a co-repressor to regulate keratin 14 expression during epidermal cell differentiation. *PLoS ONE.* 2012;7:e41742.
  21. Deng Z, Matsuda K, Tanikawa C, Lin J, Furukawa Y, Hamamoto R, Nakamura Y. Late Cornified Envelope Group I, a novel target of p53, regulates PRMT5 activity. *Neoplasia.* 2014;16:656–64.
  22. Aljawad MF, Faisal A, Alqanbar MF, Wilmarth PA, Hassan BQ. Tandem mass tag-based quantitative proteomic analysis of cervical cancer. *Proteomics Clin Appl.* 2023;17:e2100105.
  23. Ramirez-Torres A, Gil J, Contreras S, Ramirez G, Valencia-Gonzalez HA, Salazar-Bustamante E, Gomez-Caudillo L, Garcia-Carranca A, Encarnacion-Guevara S. Quantitative proteomic analysis of cervical cancer tissues identifies proteins associated with cancer progression. *Cancer Genom Proteom.* 2022;19:241–58.
  24. Wattanawongdon W, Simawaranon Bartpho T, Tongtawee T. Expression of CD44 and MDM2 in cholangiocarcinoma is correlated with poor clinicopathologic characteristics. *Int J Clin Exp Pathol.* 2019;12:3961–7.
  25. Tian Y, Denda-Nagai K, Tsukui T, Ishii-Schrade KB, Okada K, Nishizono Y, Matsuzaki K, Hafley M, Bresalier RS, Irimura T. Mucin 21 confers resistance to apoptosis in an O-glycosylation-dependent manner. *Cell Death Discov.* 2022;8:194.
  26. Zhang Y, Sun L, Lei C, Li W, Han J, Zhang J, Zhang Y. A sweet warning: mucin-type O-Glycans in cancer. *Cells.* 2022;11:3666.
  27. Egan K, Cooke N, Kenny D. Living in shear: platelets protect cancer cells from shear induced damage. *Clin Exp Metastasis.* 2014;31:697–704.
  28. Krug K, Jaehnig EJ, Satpathy S, Blumenberg L, Karpova A, Anurag M, Miles G, Mertins P, Geffen Y, Tang LC, et al. Proteogenomic landscape of breast cancer tumorigenesis and targeted therapy. *Cell.* 2020;183:1436–1456 e1431.
  29. Vasaiikar S, Huang C, Wang X, Petyuk VA, Savage SR, Wen B, Dou Y, Zhang Y, Shi Z, Arshad OA, et al. Proteogenomic analysis of human colon cancer reveals new therapeutic opportunities. *Cell.* 2019;177:1035–1049 e1019.
  30. Xu JY, Zhang C, Wang X, Zhai L, Ma Y, Mao Y, Qian K, Sun C, Liu Z, Jiang S, et al. Integrative proteomic characterization of human lung adenocarcinoma. *Cell.* 2020;182:245–261 e217.
  31. Azevedo Martins JM, Rabelo-Santos SH, do Amaral Westin MC, Zeferino LC. Tumoral and stromal expression of MMP-2, MMP-9, MMP-14, TIMP-1, TIMP-2, and VEGF-A in cervical cancer patient survival: a competing risk analysis. *BMC Cancer.* 2020;20:660.
  32. Xiao Z, Mayer AT, Nobashi TW, Gambhir SS. ICOS is an indicator of T-cell-mediated response to cancer immunotherapy. *Cancer Res.* 2020;80:3023–32.
  33. Wang C, Zhang J, Fok KL, Tsang LL, Ye M, Liu J, Li F, Zhao AZ, Chan HC, Chen H. CD147 induces epithelial-to-mesenchymal transition by disassembling cellular apoptosis susceptibility protein/E-Cadherin/beta-catenin complex in human endometriosis. *Am J Pathol.* 2018;188:1597–607.
  34. Usman S, Waseem NH, Nguyen TKN, Mohsin S, Jamal A, Teh MT, Waseem A. Vimentin is at the heart of Epithelial Mesenchymal Transition (EMT) mediated metastasis. *Cancers (Basel).* 2021;13:4985.
  35. Takeda T, Tsubaki M, Matsuda T, Kimura A, Jinushi M, Obana T, Takegami M, Nishida S. EGFR inhibition reverses epithelial-mesenchymal transition, and decreases tamoxifen resistance via Snail and Twist downregulation in breast cancer cells. *Oncol Rep.* 2022;47:109.
  36. Casalino L, Talotta F, Matino I, Verde P. FRA-1 as a regulator of EMT and metastasis in breast cancer. *Int J Mol Sci.* 2023;24:8307.
  37. Hirao T, Sakamoto Y, Kamada M, Hamada S, Aono T. Tn antigen, a marker of potential for metastasis of uterine cervix cancer cells. *Cancer.* 1993;72:154–9.
  38. O'Connell F, Cibas ES. Cytologic features of ciliated adenocarcinoma of the cervix: a case report. *Acta Cytol.* 2005;49:187–90.
  39. Jiang W, Xu Y, Chen X, Pan S, Zhu X. E26 transformation-specific variant 4 as a tumor promoter in human cancers through specific molecular mechanisms. *Mol Ther Oncolytics.* 2021;22:518–27.
  40. Lee Y. Regulation and function of capicua in mammals. *Exp Mol Med.* 2020;52:531–7.
  41. Guo C, Qu X, Tang X, Song Y, Wang J, Hua K, Qiu J. Spatiotemporally deciphering the mysterious mechanism of persistent HPV-induced malignant transition and immune remodelling from HPV-infected normal cervix, precancer to cervical cancer: Integrating single-cell RNA-sequencing and spatial transcriptome. *Clin Transl Med.* 2023;13:e1219.
  42. Dai D, Pei Y, Zhu B, Wang D, Pei S, Huang H, Zhu Q, Deng X, Ye J, Xu J, et al. Chemoradiotherapy-induced ACKR2(+) tumor cells drive CD8(+) T cell senescence and cervical cancer recurrence. *Cell Rep Med.* 2024;5:101550.
  43. Li C, Liu D, Yang S, Hua K. Integrated single-cell transcriptome analysis of the tumor ecosystems underlying cervical cancer metastasis. *Front Immunol.* 2022;13:966291.
  44. Li C, Liu D, Zhao Y, Ding Y, Hua K. Diverse intratumoral heterogeneity and immune microenvironment of two HPV-related cervical cancer types revealed by single-cell RNA sequencing. *J Med Virol.* 2023;95:e28857.
  45. Yan ZX, Dong Y, Qiao N, Zhang YL, Wu W, Zhu Y, Wang L, Cheng S, Xu PP, Zhou ZS, et al. Cholesterol efflux from C1QB-expressing macrophages is associated with resistance to chimeric antigen receptor T cell therapy in primary refractory diffuse large B cell lymphoma. *Nat Commun.* 2024;15:5183.
  46. Heger L, Hofer TP, Bigley V, de Vries IJM, Dalod M, Dudziak D, Ziegler-Heitbrock L. Subsets of CD11c(+) DCs: Dendritic cell versus monocyte lineage. *Front Immunol.* 2020;11:559166.
  47. Nishimura J, Tanaka H, Yamakoshi Y, Hiramatsu S, Tamura T, Toyokawa T, Muguruma K, Maeda K, Hirakawa K, Ohira M. Impact of tumor-infiltrating LAMP-3 dendritic cells on the prognosis of esophageal squamous cell carcinoma. *Esophagus.* 2019;16:333–44.
  48. Zhang JG, Czabotar PE, Policheni AN, Caminschi I, Wan SS, Kitsoulis S, Tullett KM, Robin AY, Brammananth R, van Delft MF, et al. The dendritic cell receptor Clec9A binds damaged cells via exposed actin filaments. *Immunity.* 2012;36:646–57.
  49. Monu NR, Frey AB. Myeloid-derived suppressor cells and anti-tumor T cells: a complex relationship. *Immunol Invest.* 2012;41:595–613.
  50. Hyeon DY, Nam D, Han Y, Kim DK, Kim G, Kim D, Bae J, Back S, Mun DG, Madar IH, et al. Proteogenomic landscape of human pancreatic ductal adenocarcinoma in an Asian population reveals tumor cell-enriched and immune-rich subtypes. *Nat Cancer.* 2023;4:290–307.
  51. Masood R, Hochstim C, Cervenka B, Zu S, Baniwal SK, Patel V, Kobiak A, Sinha UK. A novel orthotopic mouse model of head and neck cancer and lymph node metastasis. *Oncogenesis.* 2013;2:e68.
  52. Myers JN, Holsinger FC, Jasser SA, Bekele BN, Fidler IJ. An orthotopic nude mouse model of oral tongue squamous cell carcinoma. *Clin Cancer Res.* 2002;8:293–8.
  53. Nair S, Pillai MR. Human papillomavirus and disease mechanisms: relevance to oral and cervical cancers. *Oral Dis.* 2005;11:350–9.
  54. Plava J, Cihova M, Burikova M, Matuskova M, Kuceroval L, Miklikova S. Recent advances in understanding tumor stroma-mediated chemoresistance in breast cancer. *Mol Cancer.* 2019;18:67.
  55. Krisnawan VE, Stanley JA, Schwarz JK, DeNardo DG. Tumor microenvironment as a regulator of radiation therapy: new insights into stromal-mediated radioresistance. *Cancers (Basel).* 2020;12:2916.
  56. Ostendorf T, van Roeyen CR, Peterson JD, Kunter U, Eitner F, Hamad AJ, Chan G, Jia XC, Macaluso J, Gazit-Bornstein G, et al. A fully human monoclonal antibody (CR002) identifies PDGF-D as a novel mediator of mesangioproliferative glomerulonephritis. *J Am Soc Nephrol.* 2003;14:2237–47.
  57. Melaiu O, Catalano C, De Santi C, Cipollini M, Figlioli G, Pelle L, Barone E, Evangelista M, Guazzelli A, Boldrini L, et al. Inhibition of the platelet-derived growth factor receptor beta (PDGFRB) using gene silencing,

- crenolanib besylate, or imatinib mesylate hampers the malignant phenotype of mesothelioma cell lines. *Genes Cancer*. 2017;8:438–52.
58. Cortijo J, Mata M, Milara J, Donet E, Gavalda A, Miralpeix M, Morcillo EJ. Acridinium inhibits cholinergic and tobacco smoke-induced MUC5AC in human airways. *Eur Respir J*. 2011;37:244–54.
  59. Kunisada M, Hosaka C, Takemori C, Nakano E, Nishigori C. CXCL1 inhibition regulates UVB-induced skin inflammation and tumorigenesis in Xpa-deficient mice. *J Invest Dermatol*. 2017;137:1975–83.
  60. Tang L, Chen Q, Sun L, Zhu L, Liu J, Meng Z, Ni Z, Wang X. Curcumin suppresses MUC5AC production via interfering with the EGFR signaling pathway. *Int J Mol Med*. 2018;42:497–504.
  61. Zhang K, Sun L, Kang Y. Regulation of phosphoglycerate kinase 1 and its critical role in cancer. *Cell Commun Signal*. 2023;21:240.
  62. Wilson NO, Solomon W, Anderson L, Patrickson J, Pitts S, Bond V, Liu M, Stiles JK. Pharmacologic inhibition of CXCL10 in combination with anti-malarial therapy eliminates mortality associated with murine model of cerebral malaria. *PLoS ONE*. 2013;8:e60898.
  63. Liu L, Wang A, Liu X, Han S, Sun Y, Zhang J, Guo L, Zhang Y. Blocking TIGIT/CD155 signalling reverses CD8(+) T cell exhaustion and enhances the antitumor activity in cervical cancer. *J Transl Med*. 2022;20:280.
  64. Berry SPD, Dossou C, Kashif A, Sharifinejad N, Azizi G, Hamedifar H, Sabzvari A, Zian Z. The role of IL-17 and anti-IL-17 agents in the immunopathogenesis and management of autoimmune and inflammatory diseases. *Int Immunopharmacol*. 2022;102:108402.
  65. Liu N, Su D, Liu K, Liu B, Wang S, Zhang X. The effects of IL-17/IL-17R inhibitors on atherosclerosis in psoriasis and psoriatic arthritis: A protocol for systematic review and meta analysis. *Medicine (Baltimore)*. 2021;100:e24549.
  66. Wang SS, Solar VD, Yu X, Antonopoulos A, Friedman AE, Agarwal K, Garg M, Ahmed SM, Addhya A, Nasirikenari M, et al. Efficient inhibition of O-glycan biosynthesis using the hexosamine analog Ac(5)GalNTGc. *Cell Chem Biol*. 2021;28:699–710 e695.
  67. Verachi P, Gobbo F, Martelli F, Martinelli A, Sarli G, Dunbar A, Levine RL, Hoffman R, Massucci MT, Brandolini L, et al. The CXCR1/CXCR2 inhibitor reparixin alters the development of myelofibrosis in the Gata1 (low) mice. *Front Oncol*. 2022;12:853484.
  68. Song Y, Yang JM. Role of interleukin (IL)-17 and T-helper (Th)17 cells in cancer. *Biochem Biophys Res Commun*. 2017;493:1–8.
  69. He D, Li H, Yusuf N, Elmetts CA, Li J, Mountz JD, Xu H. IL-17 promotes tumor development through the induction of tumor promoting micro-environments at tumor sites and myeloid-derived suppressor cells. *J Immunol*. 2010;184:2281–8.
  70. Hyeon DY, Nam D, Han Y, Kim DK, Kim G, Kim D, Bae J, Back S, Mun DG, Madar IH, et al. Proteogenomic landscape of human pancreatic ductal adenocarcinoma in an Asian population reveals tumor cell-enriched and immune-rich subtypes. *Nat Cancer*. 2023;4:290–307.
  71. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2010;26:589–95.
  72. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013;29:15–21.
  73. Kim S, Scheffler K, Halpern AL, Bekritsky MA, Noh E, Kallberg M, Chen X, Kim Y, Beyter D, Krusche P, Saunders CT. Strelka2: fast and accurate calling of germline and somatic variants. *Nat Methods*. 2018;15:591–4.
  74. Lawrence MS, Stojanov P, Mermel CH, Robinson JT, Garraway LA, Golub TR, Meyerson M, Gabriel SB, Lander ES, Getz G. Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature*. 2014;505:495–501.
  75. Talevich E, Shain AH, Botton T, Bastian BC. CNVkit: genome-wide copy number detection and visualization from targeted DNA sequencing. *PLoS Comput Biol*. 2016;12:e1004873.
  76. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinform*. 2011;12:323.
  77. Ho DW, Sze KM, Ng IO. Virus-Clip: a fast and memory-efficient viral integration site detection tool at single-base resolution with annotation capability. *Oncotarget*. 2015;6:20959–63.
  78. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. BLAST+: architecture and applications. *BMC Bioinform*. 2009;10:421.
  79. Kostic AD, Ojesina AI, Pedamallu CS, Jung J, Verhaak RG, Getz G, Meyerson M. PathSeq: software to identify or discover microbes by deep sequencing of human tissue. *Nat Biotechnol*. 2011;29:393–6.
  80. Haas BJ, Dobin A, Li B, Stransky N, Pochet N, Regev A. Accuracy assessment of fusion transcript detection via read-mapping and de novo fusion transcript assembly-based methods. *Genome Biol*. 2019;20:213.
  81. Mun DG, Bhin J, Kim S, Kim H, Jung JH, Jung Y, Jang YE, Park JM, Kim H, Jung Y, et al. Proteogenomic characterization of human early-onset gastric cancer. *Cancer Cell*. 2019;35:111–124 e110.
  82. Park JM, Park JH, Mun DG, Bae J, Jung JH, Back S, Lee H, Kim H, Jung HJ, Kim HK, et al. Integrated analysis of global proteome, phosphoproteome, and glycoproteome enables complementary interpretation of disease-related protein networks. *Sci Rep*. 2015;5:18189.
  83. Chae SY, Nam D, Hyeon DY, Hong A, Lee TD, Kim S, Im D, Hong J, Kang C, Lee JW, et al. DNA repair and cholesterol-mediated drug efflux induce dose-dependent chemoresistance in nutrient-deprived neuroblastoma cells. *iScience*. 2021;24:102325.
  84. Lee H, Mun DG, Bae J, Kim H, Oh SY, Park YS, Lee JH, Lee SW. A simple dual online ultra-high pressure liquid chromatography system (sDO-UHPLC) for high throughput proteomic analysis. *Analyst*. 2015;140:5700–6.
  85. Lee H, Mun DG, So JE, Bae J, Kim H, Masselon C, Lee SW. Efficient exploitation of separation space in two-dimensional liquid chromatography system for comprehensive and efficient proteomic analyses. *Anal Chem*. 2016;88:11734–41.
  86. Park H, Bae J, Kim H, Kim S, Kim H, Mun DG, Joh Y, Lee W, Chae S, Lee S, et al. Compact variant-rich customized sequence database and a fast and sensitive database search for efficient proteogenomic analyses. *Proteomics*. 2014;14:2742–9.
  87. Kim S, Pevzner PA. MS-GF+ makes progress towards a universal database search tool for proteomics. *Nat Commun*. 2014;5:5277.
  88. Hwang D, Rust AG, Ramsey S, Smith JJ, Leslie DM, Weston AD, de Atauri P, Aitchison JD, Hood L, Siegel AF, Bolouri H. A data integration methodology for systems biology. *Proc Natl Acad Sci U S A*. 2005;102:17296–301.
  89. Kim Y, Kim TK, Kim Y, Yoo J, You S, Lee I, Carlson G, Hood L, Choi S, Hwang D. Principal network analysis: identification of subnetworks representing major dynamics using gene expression data. *Bioinformatics*. 2011;27:391–8.
  90. Brunet JP, Tamayo P, Golub TR, Mesirov JP. Metagenes and molecular pattern discovery using matrix factorization. *Proc Natl Acad Sci U S A*. 2004;101:4164–9.
  91. Bhin J, Jeong HS, Kim JS, Shin JO, Hong KS, Jung HS, Kim C, Hwang D, Kim KS. PGC-enriched miRNAs control germ cell development. *Mol Cells*. 2015;38:895–903.
  92. Hwang D, Smith JJ, Leslie DM, Weston AD, Rust AG, Ramsey S, de Atauri P, Siegel AF, Bolouri H, Aitchison JD, Hood L. A data integration methodology for systems biology: experimental verification. *Proc Natl Acad Sci U S A*. 2005;102:17302–7.
  93. Stark C, Breitkreutz BJ, Reguly T, Boucher L, Breitkreutz A, Tyers M. BioGRID: a general repository for interaction datasets. *Nucleic Acids Res*. 2006;34:D535–539.
  94. Luck K, Kim DK, Lambourne L, Spirohn K, Begg BE, Bian W, Brignall R, Cafarelli T, Campos-Laborie FJ, Charlotteaux B, et al. A reference map of the human binary protein interactome. *Nature*. 2020;580:402–8.
  95. Hermjakob H, Montecchi-Palazzi L, Lewington C, Mudali S, Kerrien S, Orchard S, Vingron M, Roechert B, Roepstorff P, Valencia A, et al. IntAct: an open source molecular interaction database. *Nucleic Acids Res*. 2004;32:D452–455.
  96. Patil A, Nakai K, Nakamura H. HitPredict: a database of quality assessed protein-protein interactions in nine species. *Nucleic Acids Res*. 2011;39:D744–749.
  97. Kotlyar M, Pastrello C, Malik Z, Jurisica I. IID 2018 update: context-specific physical protein-protein interactions in human, model organisms and domesticated species. *Nucleic Acids Res*. 2019;47:D581–9.
  98. Chatr-aryamontri A, Ceol A, Palazzi LM, Nardelli G, Schneider MV, Castagnoli L, Cesareni G. MINT: the Molecular INTERaction database. *Nucleic Acids Res*. 2007;35:D572–574.



99. Xenarios I, Salwinski L, Duan XJ, Higney P, Kim SM, Eisenberg D. DIP, the Database of Interacting Proteins: a research tool for studying cellular networks of protein interactions. *Nucleic Acids Res.* 2002;30:303–5.
100. Keshava Prasad TS, Goel R, Kandasamy K, Keerthikumar S, Kumar S, Mathivanan S, Telikicherla D, Raju R, Shafreen B, Venugopal A, et al. Human protein reference database–2009 update. *Nucleic Acids Res.* 2009;37:D767–772.
101. Bovolenta LA, Acencio ML, Lemke N. HTRIdb: an open-access database for experimentally verified human transcriptional regulation interactions. *BMC Genomics.* 2012;13:405.
102. Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, Simonovic M, Roth A, Santos A, Tsafou KP, et al. STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.* 2015;43:D447–452.
103. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 2003;13:2498–504.
104. Kanehisa M, Goto S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 2000;28:27–30.
105. Zheng GX, Terry JM, Belgrader P, Ryvkin P, Bent ZW, Wilson R, Ziraldo SB, Wheeler TD, McDermott GP, Zhu J, et al. Massively parallel digital transcriptional profiling of single cells. *Nat Commun.* 2017;8:14049.
106. Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM 3rd, Hao Y, Stoeckius M, Smibert P, Satija R. Comprehensive integration of single-cell data. *Cell.* 2019;177:1888–1902 e1821.
107. Wolf FA, Angerer P, Theis FJ. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* 2018;19:15.
108. Tosi A, Cappellesso R, Dei Tos AP, Rossi V, Aliberti C, Pigozzo J, Fabozzi A, Sbaraglia M, Blandamura S, Del Bianco P, et al. The immune cell landscape of metastatic uveal melanoma correlates with overall survival. *J Exp Clin Cancer Res.* 2021;40:154.
109. Ku JL, Kim WH, Park HS, Kang SB, Park JG. Establishment and characterization of 12 uterine cervical-carcinoma cell lines: common sequence variation in the E7 gene of HPV-16-positive cell lines. *Int J Cancer.* 1997;72:313–20.

## Publisher's Note

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