The long noncoding RNA lnc-FANCI-2 restricts RAS signaling and phosphorylation of Akt and Erk in HPV16-infected cervical cancer cells

Haibin Liu[1a,b], Lulu Yu[1a], Vladimir Majerciak[1], Thomas Meyer[2], Ming Yi[3], Peter F. Johnson[4], Maggie Cam[2], Douglas R. Lowy[5], and Zhi-Ming Zheng[1]*

[1]Tumor Virus RNA Biology Section, HIV Dynamics and Replication Program, CCR/NCI/NIH

[2]Collaborative Bioinformatics Resource, CCR/NCI/NIH

[3]RAS Informatics Group, NCI/NIH

[4]Mouse Cancer Genetics Program, CCR/NCI/NIH

[5]Laboratory of Cellular Oncology, CCR/NCI/NIH

[a]*Co-first authors*

Running title: lnc-FANCI-2 regulates RAS signaling

[b]Present address: Center for Emerging Infectious Diseases, Wuhan Institute of Virology, and Center for Biosafety Mega-Science, Chinese Academy of Sciences, Wuhan, Hubei, 430207, China

*Corresponding: zhengt@exchange.nih.gov

ABSTRACT

We recently discovered increased expression of a cytoplasmic long noncoding RNA, lnc-FANCI-2, coinciding with cervical lesion progression from CIN1, CIN2-3 to cervical cancer. Viral E7 of high-risk HPVs and host transcription factor YY1 are two major factors promoting lnc-FANCI-2 expression. To explore possible roles of lnc-FANCI-2 in HPV-induced cervical carcinogenesis, we ablated the expression of *lnc-FANCI-2* in the HPV16-positive cervical cancer cell line, CaSki. Knock-out (KO) single cell clones expressed HPV16 oncogenes normally but displayed altered cell morphology and reduced proliferation and migration when compared with the parental cells. Proteomic profiling of cytosolic and secreted proteins from the parental and KO cells showed that lnc-FANCI-2 regulates expression of a subset of cell surface and adhesion-related proteins, including inhibition of MCAM, PODXL2 and ECM1 and increased levels of ADAM8 and TIMP2. RNA-seq analyses revealed that, relative to the parental cells, KO cells exhibited significantly increased RAS signaling but decreased IFN pathways. In KO cells, phosphorylated Akt and Erk1/2, two important RAS pathway effectors, were increased more than 3-fold, accompanied by increase of MCAM, VIM, and CCND2 (cyclin D2) and decrease of RAC3. Accordingly, high levels of lnc-FANCI-2 and lower levels of MCAM in cervical cancer patients are associated with improved survival. We found that lnc-FANCI-2 in CaSki cells interacts specifically with 32 host proteins, including H13, HNRH1, K1H1, MAP4K4, and RNPS1, and knockdown of MAP4K4 led to decreased phosphorylation of Akt and Erk1/2. In summary, a key function of lnc-FANCI-2 is to regulate RAS signaling, thereby affecting cervical cancer outcome.

INTRODUCTION

Cervical cancer is a leading cause of mortality in women worldwide. It is the second most diagnosed cancer among women, with an estimated 530,000 new cases worldwide each year, and the third most frequent cause of cancer-related death, accounting for 270,000 deaths annually[1]. Persistent infections in the cervical epithelium with the cancer-associated alpha human papillomaviruses (high-risk HPVs), in particular HPV16 and HPV18 with the sustained expression of viral oncoproteins, E6 and E7, contribute to squamous epithelial cell tumorigenesis [2-4]. Approximate 1.4% of high-grade neoplastic cervical lesions progress to invasive cervical cancer [5], while most lesions resolve spontaneously. The progression to cervical cancer typically takes 15 to 20 years after initial infection (https://www.who.int/news-room/fact-sheets/detail/cervical-cancer). These observations suggest that active host defense systems play important roles in preventing malignant transformation of neoplastic cervical lesions, and additional factors besides E6 and E7 are necessary for cervical cancer development. Indeed, one tumor suppressor network, the Fanconi anemia (FA) pathway [6], can be activated by HPV-induced DNA damage. FA proteins including FANCA, FANCD2 and FANCI are elevated upon HPV infection [7-9] and the activated FA pathway restricts HPV replication and guards host genome stability [10]. FANCD2 mutation stimulates HPV16 and HPV31 genome amplification and also promotes cervical and vaginal cancer development in HPV16 E7 transgenic mouse [11, 12].

Viral oncoprotein E7 plays a major role in the immortalization of primary epithelial cells mainly by inactivation of pRb family of proteins [13], whereas E6 mediates degradation of tumor suppressor p53 and activation of human telomerase reverse transcriptase (hTERT) transcription [14-16]. However, high-risk E6 and E7 are necessary but not sufficient to transform fully the immortalized cells into malignant cells [17-21] and additional oncogenic stress is needed. For

example, the HRAS^{G12V}, a constitutively activated RAS GTPase mutation, triggers malignant transformation of E6/E7-expressing keratinocytes[22-25]. Normal RAS regulates cell proliferation, cell differentiation and cell adhesion through cellular signal transduction from growth-factor receptors. Thus, overactive RAS signaling might facilitate transformation of E6/E7-expressing keratinocytes. In transgenic mice, HPV16 E6 and E7 can promote the development of spontaneous epithelial skin tumors, but not spontaneous tumors of the reproductive tract [26-29]. Prolonged estrogen treatment is required [30]. Estrogen receptor (ER) activation stimulates the mitogen-activated protein kinase (MAPK/Erk) and phosphoinositide 3-kinase (Pl3K/Akt) pathways, both of which are mediated by RAS [31]. However, the molecular mechanism underlying estrogen-induced cervical cancer in E6 and E7 transgene mice is not fully understood.

Long non-coding RNAs (lncRNAs) are RNAs over 200 nucleotides (nt) in length lacking a coding capacity for translation into functional proteins. Although several lncRNAs have been proposed to function in diverse biological processes, evidence is still lacking to support the functionality of the majority of lncRNAs [32]. Nevertheless, the identified lncRNA functions include regulating: (1) transcription of neighboring or distant genes by recruiting histone modifiers, as shown for Xist [33] and ANRIL [34], or by causing chromosome looping, as proposed for PVT1 [35] and CCAT1-L [36]. (2) RNA splicing by acting as decoy of splicing factors, as demonstrated for MALAT1 [37] and PNCTR [38]. (3) mRNA stability by directly base pairing as antisense transcripts [39], by acting as RNA binding protein (RBP) decoy to prevent mRNA-RBP interaction, such is the case for NORAD [40], or by functioning as miRNA sponges to prevent miRNA-induced degradation of mRNAs, as shown for PTEN [41] and PNUTS [42]. And (4) protein stability by directly binding, such as MaIL1 [43], or by acting as protein decoy to prevent ubiquitin-mediated proteasomal degradation, such as FAST [44]. Thus, lncRNA genomic position,

subcellular localization and their interactions with DNA, RNA and proteins are crucial determinants to the functionality of individual lncRNAs.

We recently reported an increased level of lnc-FANCI-2 expression in high-risk HPV-positive cervical intraepithelial neoplasia (CIN) and invasive cervical cancer (ICC) tissues [8]. At least 14 RNA isoforms of lnc-FANCI-2 are expressed by alternative promoter usage, alternative RNA splicing, and alternative selection of RNA polyadenylation sites from the genomic locus adjacent to the FANCI. Both lnc-FANCI-2 and FANCI are up-regulated simultaneously in neoplastic cervical lesions and cervical cancer by high-risk HPV infections. E6 and in particular, E7 are responsible for the enhanced expression of lnc-FANCI-2, the transcription of which is also regulated by YY1[8]. HPV infection increases YY1 levels but decreases the expression of p53-dependent miR-29a which targets the YY1 3' UTR. Viral E7 interacts with YY1 and facilitates YY1 transactivation of lnc-FANCI-2 promoter [8]. In situ hybridization has revealed that lnc-FANCI-2 is preferentially cytoplasmic, but its function in HPV infected cells has not been characterized. In this report, we demonstrate that knockout (KO) or knockdown (KD) of lnc-FANCI-2 promotes RAS signaling and phosphorylation of Akt and Erk.

RESULTS

lnc-FANCI-2 regulates proliferation of HPV-transformed cervical cancer cells.

lnc-FANCI-2 RNA is transcribed mainly from a proximal promoter TSS2, but also from an alternative minor, distal promoter TSS1. Two highly conserved YY1 binding motifs upstream of the TSS2 are essential for the TSS2 transcriptional activity [8] (Fig.1A). To elucidate the function of lnc-FANCI-2 in high-risk HPV-infected cervical cancer cells, we knocked out (KO) lnc-FANCI-2 expression in HPV16-positive CaSki cells using CRISPR/Cas9 deletion of either a

3-kb promoter region encompassing both TSS1 and TSS2 or an 86-bp region containing two YY1-binding motifs in the TSS2 promoter (Fig.1A). Two tested gRNAs with high KO efficiency were selected and cloned into a modified CRISPR/Cas9 expression vector to express both 5'-specific gRNA and 3'-specific gRNA simultaneously for efficient genome editing [45]. CaSki cells stably transfected with the dual gRNA expression vector were generated. Through a serial dilution, several single cell clones were isolated and verified for homozygous deletion by PCR screening. PCR genotyping indicated successful homozygous deletion of the promoter or YY1-binding motifs, respectively (Fig.1B).

Loss of lnc-FANCI-2 expression in the single-cell clones was examined by RT-qPCR (Fig.1C). When compared to all control clones transfected and selected from an empty vector containing no gRNA, all single cell clones with deletion of YY1-binding motifs (ΔYY1) showed 60-70% reduction of lnc-FANCI-2 expression, whereas the single cell clones with deletion of both TSS1 and TSS2 promoters (ΔPr) showed at least 90% reduction (Fig. 1C). Northern blot analysis confirmed the decrease in levels of the two major lnc-FANCI-2 isoforms (Fig. 1D), with an abundant 4-kb lnc-FANCI-2 RNA derived from a distal polyadenylation site PA2 and a less abundant 2-kb of lnc-FANCI-2 RNA derived from a proximal polyadenylation site PA1[8] (Fig. 1A and 1D). Reduced lnc-FANCI-2 RNA expression was also confirmed by RNAscope single molecule RNA in situ hybridization (RNA_ISH) (Fig. 1E and 1F). The parental (WT) CaSki cells expressed ~10 copies of lnc-FANCI-2 RNA per cell [8], whereas the copy number of lnc-FANCI-2 RNA in the ΔYY1-D5 cells dropped to ~5 copies and to ~2 copies per cell in the ΔPr-A9 cells (each averaged from 200 cells).

Although the WT CaSki cells grow as cell islands, the lnc-FANCI-2 KO cells displayed a dispersed cell growth pattern and an irregular or spindle-like cell morphology which was more

obvious in the ΔPr-A9 cells (Fig.S1A). The KO cells also grew significantly slower than the WT CaSki cells (Fig. 1G), formed fewer colonies with reduced size (Fig. 1H), and showed greatly decreased migration capacities (Fig. 1I), with ΔPr-A9 cells having a more severe phenotype than ΔYY1-D5 cells. The ΔPr-A9 cells were further examined for effects on HPV16 E6 and E7 expression and cell senescence. When compared with WT CaSki cells, the ΔPr-A9 cells showed little change in E6, E7, p53 (E6 downstream target) and E2F1 (E7 downstream target) protein levels (Fig. S1B) but a significant increase in cell senescence (Fig. S1C). Altogether, the data indicate that lnc-FANCI-2 RNA is important for cell growth, colony formation, and cell migration.

lnc-FANCI-2 regulates the expression and secretion of cell soluble receptors.

We considered that altered expression of cell membrane proteins and secreted factors in the lnc-FANCI-2 KO cells might contribute to the observed different cell morphology and growth properties. Therefore, we performed a Proteome Profiler Human sReceptor Array analysis to examine possible changes in expression of 105 well-characterized soluble protein receptors using total cell lysates and cell culture supernatants from ΔPr-A9 and WT CaSki cells.

We identified from the ΔPr-A9 cell lysate nine proteins with increased expression compared to the WT CaSki cells, including PODXL2, ECM1, NECTIN2, MCAM, ADAM9, ADAM10, CDH5, ITGA5 and NOTCH1, and six proteins with decreased expression including ITGB6, CDH13, LGALS3BP, TIMP2, ADMA8 and SCARF2 (Fig. 2A and 2B). We also found from the ΔPr-A9 cell culture supernatant five proteins with increased expression, including ADAM9, NECTIN2, ADAM10, CDH5 and ECM1, and the decreased expression of four proteins, including CRELD2, SDC1, SDC4 and TIMP2 (Fig. 2A and 2B). By immunoblot assays, we verified selectively the increase in PODXL2, MCAM, and ECM1 and the decrease in

ADAM8 and TIMP2 from the ΔPr-A9 cell lysates and the increase in ECM1 and the decrease in TIMP2 in the ΔPr-A9 cell culture supernatant (Fig. 2C). As ADAM8 is proteolytically processed into two protein isoforms [46], we confirmed by immunoblot the decreased expression of all three sizes of ADAM8 protein in the ΔPr-A9 cell lysate over the WT CaSki cells (Fig. 2C).

lnc-FANCI-2 regulates the expression of genes involved in RAS signaling.

We next conducted genome-wide RNA-seq analyses of WT and lnc-FANCI-2 KO cells (four samples/group) to determine the transcriptomic consequences of lnc-FANCI-2 deficiency. We obtained ~120 million mappable RNA reads to the human reference genome hg38 from each sample. Analysis of RNA-seq reads-coverage map by Integrative Genomics Viewer (IGV) showed ~90% reduction of lnc-FANCI-2 expression in ΔPr-A9 cells and ~72% decrease in ΔYY1-D5 cells compared to the WT CaSki cells (Fig. 3A), consistent with the results shown in Fig. 1C-F.

Hierarchical clustering analysis showed more global transcriptional similarity between ΔYY1-D5 and ΔPr-A9 cells (Fig. 3B). By applying a threshold of fold change (FC) ≥ 1.8 or FC ≤ -1.8 with a p value of FDR ≤ 0.01, we found 1230 genes in ΔPr-A9 and 797 genes in ΔYY1-D5 cells with significantly differential expression relative to the WT CaSki cells expressing ~15890 genes. The most significantly affected genes are shown in Volcano plots (Fig. 3C and Table S1). Among these, 211 were upregulated and 189 were downregulated in both ΔPr-A9 and Δ YY1-D5 cells (Fig. 3D and Table S2). By applying more stringent criteria with FPKM ≥7 in at least one of four RNA-seq samples as a cutoff, we profiled the genes with large expression differences in both ΔPr-A9 and ΔYY1-D5 cells. The results displayed in the heatmap of Fig. 3E included 51 upregulated with only one outlier, IGFBP3, and 47 downregulated genes excluding lnc-FANCI-2. Subsequently, we selectively verified by RT-PCR the decreased expression of

ITGB6, NLRP2, PLAC8, PSG4, PSMB9, and SERPINB1 and increased expression of CFH, CNTN5, EMP3, IGFBP3, PTGS2, and SERPINB2 in ΔPr-A9 cells (Fig. 3F), as well as the increased expression of IGFBP3 protein in ΔPr-A9 cells by immunoblot (Fig. 3G).

By performing the Gene Set Enrichment Analysis (GSEA) on the Hallmark gene sets, which provide more refined and concise inputs for GSEA [47], we found the most significantly upregulated pathways in both ΔPr-A9 and ΔYY1-D5 cells, when compared with the WT CaSki cells, were KRAS signaling and epithelial mesenchymal transition (EMT). The most significantly downregulated pathways in both ΔPr-A9 and ΔYY1-D5 cells over the WT CaSki cells were interferon gamma (IFN-γ) and interferon alpha (IFN-α) responses (Fig. 4A). GSEA plots for ΔPr-A9 shows 29 of 111 genes in RAS signaling and 39 of 130 genes in EMT (Fig. 4B and 4C, Table S3) were significantly upregulated, whereas 37 of 139 genes in IFN-γ response and 24 of 71 genes in IFN-α response were significantly downregulated (Fig. 4D and 4E, Table S3). Similar enriched gene sets with increased RAS signaling and EMT and decreased IFN-γ and IFN-α responses were observed in ΔYY1-D5 cells (Fig. S2, Table S3).

Since ΔPr-A9 cells exhibit higher KO efficiency of lnc-FANCI-2 and a higher number of differentially expressed genes (DEGs) and show more severe phenotypic changes in cell proliferation, migration, and colony formation than ΔYY1-D5 cells, the ΔPr-A9 cells were primarily used for further focused studies on RAS signaling in this report.

lnc-FANCI-2 regulates RAS GTPase activities to phosphorylate RAS signaling effectors Akt and Erk.

RAS activation triggers two major downstream signal transduction pathways, Raf/Mek/Erk and PI3K/Akt (Fig. S3), to transduce signals from extracellular stimuli to the cell

nucleus where specific effector genes are activated for their corresponding functions [48, 49]. To further confirm the GSEA data and verify that two major signaling pathways of RAS are activated, we first examined and compared the RAS GTPase activities in ΔPr-A9 cells and the WT CaSki cells. We found a significant increase in RAS GTPase activity in ΔPr-A9 cells (Fig. 5A, top bar graphs), indicating that the endogenous lnc-FANCI-2 RNA in CaSki cells suppresses RAS activation. Since ΔPr-A9 cells had undergone long term selection and adaption during single cell screening, the increased RAS GTPase activity might result from selection pressure for cell survival. However, by transient siRNA knockdown (KD) of lnc-FANCI-2 expression in the WT CaSki cells, we obtained a similar, albeit weaker, increase of RAS GTPase activity (Fig. 5A, lower bar graphs). This result suggests that the increased RAS GTPase activity in ΔPr-A9 cells was unrelated to the persistent selection pressure, but rather the deficiency of lnc-FANCI-2.

Next, we examined activation of the PI3K/Akt and Raf/Mek/Erk pathways [50-52]. We observed a 3-fold increase in phosphorylated Akt (p-Akt) and 2.5-fold increase of p-Erk1/2 (p44/p42 MAPK), after being normalized to total Akt and Erk, in ΔPr-A9 cells over the WT CaSki cells (Fig. 5B). These results were confirmed in ΔYY1-D5 cells (Fig. S4A). The increased p-Akt and p-Erk (mostly p-Erk2/p42) was accompanied by elevated expression of MCAM, VIM, and CCND2, and decreased expression of RAC3 (Fig. 5B). These randomly chosen potential RAS effector proteins facilitate membrane receptor function, cell proliferation, and EMT. Moreover, transient siRNA knockdown of lnc-FANCI-2 in the WT CaSki cells also led to the increased levels of p-Akt and p-Erk (Fig. 5C) and increased expression of MCAM and VIM at 48 h and 96 h post-transfection (Fig. 5C). These observations indicate that a transient loss of lnc-FANCI-2 in CaSki cells is sufficient to trigger RAS signaling. The same siRNA transfection in HeLa cells, a lnc-FANCI-2 negative cell line (Fig. S4B and S4C), exhibited no effect on p-Akt

or p-Erk1/2 levels (Fig. S4F), but in SiHa cells containing mainly nuclear lnc-FANCI-2 (Fig. S4D and S4E) enhanced the expression of p-Akt and p-Erk1/2 (Fig. S4F), verifying the specific effect of lnc-FANCI-2 depletion on RAS signaling pathways no matter where a nuclear or cytoplasmic lnc-FANCI-2 is predominant in the cancer cells. We found both cytoplasmic and nuclear locations of the increased lnc-FANCI-2 in HPV-infected human foreskin keratinocytes and cervical cancer tissues (Fig. S5). The lnc-FANCI-2 KO-mediated increase in p-Akt and p-Erk was sensitive to PI3K inhibitor LY294002 [53] and MEK1/2 inhibitor U0126 [54] (Fig. 5D), respectively. The inhibitors also decreased the expression of VIM and MCAM and cell proliferation, in particular by PI3K inhibitor LY294002 (Fig. 5D and 5E). These effects on cell proliferation were similar from ΔPr-A9 to the WT CaSki cells.

lnc-FANCI-2 inhibits MCAM expression to promote cervical cancer survival.

As shown in Fig. 3E and Fig. S6A, the increased RNA expression of MCAM appears to be transcriptional or posttranscriptional in both ΔPr-A9 and ΔYY1-D5 cells. We confirmed by RT-PCR the lnc-FANCI-2 KO-mediated increase of MCAM RNA expression in ΔPr-A9 cells (Fig. 6A). A rescue experiment was further performed by transient expression of one major isoform of lnc-FANCI-2 RNA (a-PA2, GenBank: MT669800.1) [8] in ΔPr-A9 cells to determine both nuclear and cytoplasmic MCAM expression [55, 56] in individual lnc-FANCI-2 expressing cells. ΔPr-A9 cells ectopically expressing diffused cytoplasmic lnc-FANCI-2 RNA, as detected by RNAscope RNA-ISH, exhibited a marked reduction in nuclear MCAM protein (Fig. 6B-C) and cytoplasmic IGFBP3 RNA, another lnc-FANCI-2 effector gene (Fig. S6B-D). Quantitative analyses of the ΔPr-A9 cells with or without transient lnc-FANCI-2 RNA expression showed significant reduction of nuclear MCAM protein in the 35 cells with rescued lnc-FANCI-2 expression when compared to the 54 cells expressing no lnc-FANCI-2 (Fig. 6D). Due to

RNAscope procedures, the membrane-bound and cytoplasmic MCAM protein was mostly removed by protease III treatment when RNA-ISH was performed and thus only the nuclear signal of MCAM protein remained was detectable by IF staining. Using cell fractionation, we detected a cleaved MCAM of ~46 kDa mainly in the nuclear fraction by Western blot (Fig. 6E). The increased nuclear MCAM expression in ΔPr-A9 cells was proportional to the cytoplasmic and total MCAM levels when compared with the WT CaSki cells (Fig. 6E). The nuclear MCAM signal appeared as a proteolytic cleavage product presumably by the increased metalloprotease ADAM9 and ADAM10 proteins and decreased metalloprotease inhibitor TIMP2 (Fig. 2), as observed in other studies [56]. There was no secreted soluble MCAM detectable in the CaSki cell culture supernatant (Fig. 6F).

The significance of the inverse correlation of lnc-FANCI-2 and MCAM in CaSki cells was further investigated by analyzing their expression in 304 cervical cancer samples from the TCGA dataset. There was a significant negative correlation in lnc-FANCI-2 (LINC00925 being wrongly assigned by NCBI)[8] and MCAM RNA levels (Fig. 6G). The opposite effects of lnc-FANCI-2 and MCAM on cervical cancer survival (Fig. 6H and Fig. S7) show that the patients with a higher level of lnc-FANCI-2 [8] but a lower level of MCAM in the cervical tissue (Fig. S7 and Fig. 6H) exhibited a better survival prognosis for cervical cancer.

MAP4K4 association with lnc-FANCI-2 RNA in CaSki cells regulates RAS signaling and phosphorylation of Akt and Erk.

We speculated lnc-FANCI-2 restriction on RAS signaling through interactions with cellular proteins. Subsequently, comprehensive identification of lnc-FANCI-2-binding proteins in the WT CaSki cells was performed by an IRPCRP protocol (modified from published ChIRP [57]) in combination of mass spectrometry (Fig. 7A). In this protocol, the lnc-FANCI-2-binding

proteins in the WT CaSki cells were covalently crosslinked to lnc-FANCI-2 RNA via UV irradiation and pulled down by pooled 30 antisense oligos crossover the entire lnc-FANCI-2. The RNA extracted from the pulldowns were verified to be lnc-FANCI-2 specific by RT-PCR in the absence (-) or presence (+) of reverse transcriptase (RT) (Fig. 7B). The lnc-FANCI-2-binding proteins in the pulldowns were then subjected to LC-MS/MS analyses. We identified 32 specific lnc-FANCI-2-binding proteins, including H13, HNRH1, K1H1, MAP4K4, and RNPS1(Fig. 7C, Table S4).

MAP4K4, a serine/threonine protein kinase, stimulates cancer cell proliferation, invasion and migration, and is recently characterized as a novel MAPK/ERK pathway regulator [58-60] and regulates RAS signaling in the context of the early embryo [61]. Thus, whether MAP4K4 could upregulate PI3K/Akt signal transduction was investigated in the WT CaSki cells. We demonstrated that siRNA knockdown of MAP4K4 expression in the WT CaSki cells led to ~40% reduction of phosphorylation of both Akt and Erk1/2 (Fig. 7D). The data suggest that, through RNA-protein interactions, the increased lnc-FANCI-2 RNA in cells might titrate MAP4K4 and other cellular proteins to control RAS signaling and gene expression of its effectors (Fig. 7E).

DISCUSSION

HPV oncoproteins E6 and E7 are necessary but not sufficient for development of cervical cancer. Indeed, early studies indicated that other oncogenic stress, such as activated RAS mutant, is needed to trigger malignant transformation of E6 and E7 expressing cells [17, 22, 62, 63] and tumorigenesis [23, 24, 64-68]. Moreover, E7 has been also reported to repress phosphorylation of Akt

and Akt mediated signaling[69]. In this study, we show that lnc-FANCI-2, whose expression is highly dependent on E7 and YY1[8], restricts RAS GTPase activities and Akt and Erk phosphorylation presumably by interacting with one of cellular factors, MAP4K4, in HPV16-positive CaSki cells. Interestingly, early reports indicated that p-Akt attenuation by E7 could be abolished by introduction of H73E mutation in the E7 CR3 domain[69]. This E7 CR3 domain is also essential to interact with transcription factor YY1 to activate lnc-FANCI-2 transcription [8]. Thus, it would be presumable that the reported p-Akt attenuation by E7 might be mediated by the increased expression of lnc-FANCI-2. More importantly, these findings also suggest that additional oncogenic stress, such as activated RAS mutant, is required to overcome the lnc-FANCI-2 restriction on RAS signaling for the early observed malignant transformation [17, 22, 62, 63] and tumorigenesis [23, 24, 64-68] of E6 and E7 expressing cells. In CaSki cells, loss of lnc-FANCI-2 in this report was found not only to promote RAS signaling but also retard cell proliferation and induce cell senescence. This is consistent with previous observations showing that high levels of mutant $HRAS^{G12V}$ induce cell cycle arrest and cell death [70, 71].

Some lncRNAs function as RAS regulators [72, 73] by their ability to sequester RAS-targeting miRNAs, including MALAT1/ miR-217, RMRP/miR-206, and KRAS1P/miR-143/let-7 [73]. Conceivably, lncRNAs may also regulate RAS signaling through other mechanisms beyond miRNAs. In this report, loss of lnc-FANCI-2 in HPV16-positive CaSki and SiHa cells stimulates RAS signaling and expression of RAS signaling effectors (Fig. 7E). By profiling the proteins associated with lnc-FANCI-2 in CaSki cells, we identified MAP4K4, a serine/threonine protein kinase and a negative RAS signaling regulator in the context of the early embryo [59, 61], as one of major lnc-FANCI-2-binding proteins that may partially mediate the lnc-FANCI-2 restriction on RAS signaling. MAP4K4 is thought to regulate the c-Jun NH2-terminal kinase (JNK) signaling

[74, 75] and the MAPK/ERK pathway [59]. Indeed, silencing of MAP4K4 in CaSki cells led to reduction of both p-Akt and p-Erk1/2, rather than an expected enhancement of p-Akt and p-Erk1/2 expression according to the observations in the context of the early embryo[61].

Although all HPV16-positive cervical cancer and pre-cancer cell lines examined, such as CaSki, SiHa, W12 subclone cells, produce predominantly cytoplasmic or nuclear lnc-FANCI-2, unexpectedly, we found not all cervical cancer cell lines, including HPV18-positive HeLa and C4II cells, express lnc-FANCI-2, nor colorectal cancer cell line HCT116, adenovirus 5 DNA-transformed human embryonic kidney cell line HEK293, spontaneously transformed aneuploid immortal keratinocyte cell line HaCaT, and KSHV-infected B cell lymphoma cell line BCBL-1 cells which are all HPV-negative cell lines. It remains to know what negatively regulates lnc-FANCI-2 expression in those cell lines. However, we did find that HPV-negative cervical cancer cell line C33A cells expressing mutant p53 and mutant pRB produce a high level of lnc-FANCI-2 (Fig. S4B and S4C). In the lnc-FANCI-2 producing cells, we noticed that cellular location of lnc-FANCI-2 also varies from cell types, with lnc-FANCI-2 predominantly in the nucleus in SiHa cells, but in the cytoplasm in CaSki cells, although both contribute to regulate RAS signaling. This suggests that a variety of nuclear and cytoplasmic functions of lnc-FANCI-2 remains to be explored.

It is obvious from our RNA-seq and human soluble receptor array analysis that KO of lnc-FANCI-2 in CaSki cells affects the expression of a large set of genes responsible for RAS signaling, EMT pathway, and IFN responses. In addition to the notified activation of PI3K/Akt and Raf/Mek/Erk pathways, one of the major RAS signaling effectors in lnc-FANCI-2 KO or KD cells was MCAM. MCAM binds various cellular surface receptors or co-receptors to trigger signal transduction, cell proliferation and motility, tumor angiogenesis, and tumor cell metastasis

[76]. It is an integral, highly glycosylated membrane protein with three common protein isoforms [77-79]. Two membrane-anchored forms with a long (p113) or short (p76) cytoplasmic tail produced by alternative RNA splicing and a proteolytically cleaved soluble form without transmembrane and cytoplasmic regions [78-80]. In addition to its roles in cell signal transduction [76-79], the short isoform produced by an intramembrane cleavage may be directed towards the nucleus to regulate gene transcription [56]. We found that CaSki cells express mainly a cytoplasmic form of full-length MCAM (~113 kDa) and a small nuclear form (~46 kDa), but not the short ~78 kDa isoform nor the soluble product. Moreover, lnc-FANCI-2 KO cells exhibit high MCAM levels, which was most likely a result of the increased RAS signaling. Consistently, lnc-FANCI-2 and MCAM levels are negatively correlated in cervical cancer patients, and low levels of MCAM and high levels of lnc-FANCI-2 indicate better prognosis. This suggests that increased MCAM might promote cervical tumorigenesis in vivo.

In conclusion, this study provides the first evidence that one of lnc-FANCI-2 functions is to maintain epithelial functional integrity by repressing RAS signaling in preventing malignant transformation of neoplastic cervical lesions. In addition, more extensive studies are undergoing to elucidate the role of lnc-FANCI-2 in regulation of IFN responses, which are not covered in this report. Since transactivation of lnc-FANCI-2 expression is largely dependent on host transcription factor YY1 of which expression is regulated by HPV oncoproteins and host miR-29a, dissecting their transcription network on lnc-FANCI-2 and its effectors may further provide novel insights involved in host homeostasis against HPV infection-induced carcinogenesis. Our observations also highlight additional, unexpected players beyond viral E6 and E7 that drive cells toward malignant transformation.

MATERIALS AND METHODS

Cell cultures and treatment

CaSki, a HPV16-positive human cervical cancer cell line with wild-type p53, pRb, HRAS, KRAS, and NRAS (https://depmap.org/portal/cell_line/ACH-001336?tab=mutation) was obtained from the American Type Culture Collection (ATCC, Manassas, VA), maintained in Dulbecco's modified Eagle medium (DMEM) (Thermo Fisher Scientific) supplemented with 10% fetal bovine serum (FBS) and 1% penicillin-streptomycin, and grown in a cell culture incubator at 37°C with 5% CO_2.

SiHa (HPV16+), HeLa (HPV18+), C4II (HPV18+), C33A (no HPV), and other HPV-negative cell lines, HEK293 (Ad5 E1/E2-immortalized human kidney cells), and HaCaT (spontaneously immortalized human epidermal cell), were all obtained from ATCC and grown in Dulbecco's modified Eagle's medium (DMEM) with 10% FBS at 37°C and 5% CO_2. HCT116 cell line, a colorectal cancer cell line, was a gift from Dr. Bert Vogelstein of Johns Hopkins University and grown in McCoy's 5A medium with 10% FBS at 37°C and 5% CO_2. W12 subclones 20861 (HPV16+, integrated) and 20863 (HPV16+, episomal) were gift cells from Paul Lambert of University of Wisconsin and grown on pretreated NIH3T3 feeder cells in F12 medium at 37°C and 5% CO_2. BCBL-1 (body cavity B lymphoma cells with KSHV infection) was obtained from the AIDS Research and Reference Reagent Program, Division of AIDS, NIAID, NIH and cultured in RPMI 1640 containing 10% FBS at 37°C and 5% CO_2.

CaSki cells were treated with 20 µM phosphoinositide 3-kinase (PI3K) inhibitor, LY294002 (Cell Signaling Technology, Danvers, MA, #9901) or 10 µM mitogen-activated protein kinase kinase (MEK) inhibitor, U0126 (Cell Signaling Technology, Danvers, MA, #9903) and

examined for their proliferation and viability by CCK-8 assay and RAS effector expression by Western blot.

siRNAs, plasmid constructions and transfections

Synthetic double-stranded siRNA targeting the lnc-FANCI-2 transcript is listed in a supplementary Table S5. Human MAP4K4 siRNA (#M-003971) and non-targeting control siRNA(#D-001210-01) were purchased from Dharmacon, Inc (Lafayette, CO). Each siRNA transfection into CaSki or other type of cells was carried out by LipoJet in vitro transfection reagent (SignaGen, MD, USA). Plasmid pHBL03 has a short isoform of lnc-FANCI-2 cDNA (lnc-FANCI-2a-PA1, GenBank MT669801). Overlapping PCR products generated from 5' and 3' RACE PCR products with a primer pair of oHBL 17 and oHBL 18 were digested and inserted into pcDNA3 at Hind III and Xba I sites. The insertion was verified by sequencing. Plasmid pHBL05 has a long isoform of lnc-FANCI-2 (lnc-FANCI-2a-PA2, GenBank MT669800). PCR products generated from genomic DNA with oHBL 19 and oHBL 23 were digested and swapped into pHBL03 to replace lnc-FANCI-2a-PA1 at Bsmb I and Xba I sites.

CRISPR/Cas9 system was used to knockout lnc-FANCI-2 gene from CaSki cells. Two lnc-FANCI-2 specific guide RNAs (gRNAs) were designed by an on-line CRISPR design tool developed by the Zhang lab (http://crispr.mit.edu). A modified cloning strategy was used to clone the two gRNA into pSpCas9(BB)-2A-Puro vector (Addgene, Watertown, MA, #62988) and described in our previous publication [45]. This strategy allows all three components (two gRNAs and Cas9) express from the same plasmid to ensure the transfected cells expressing two gRNAs simultaneously knocking out the targeted gene. The sequences of primers to make guide RNAs were listed in Supplementary Table S5.

Knockout of lnc-FANCI-2 gene in CaSki cells, PCR screening and genotyping

Two gRNAs each under control by a U6 promoter in a pSpCas9(BB)-2A-Puro vector were constructed as described [45] and used for transfection of CaSki cells. Plasmid pHBL25 which contains gRNA 1 and 6 was used to delete the entire promoter region of lnc-FANCI-2 and plasmid pHBL27 which contains gRNA 3 and 4 was used to delete only two YY1-binding motifs spanning over an 87-bp region in the lnc-FANCI-2 promoter. CaSki cells transfected with the gRNA-expressing vectors were then under 2 weeks of puromycin selection (0.1 μg/ml). The puromycin-resistant CaSki cells were diluted to a final concentration of 0.5 cells per 100 μl, plated 100 μl of diluted cells into each well of a 96-well plate, and grew under purymycin selection. The colonies were inspected to ensure the clones were grown from one cell, and refed or re-plated as needed in 3 weeks of expansion.

A direct PCR was performed for number of single cell clones as described [45]. Briefly, several hundreds of cells expanded from a single cell clone were resuspended in 50 μl of phosphate-buffered saline (PBS) and then frozen-thawed on dry ice for three times. Cell lysates were treated with 4 μl of Qiagen protease (QIAGEN, Hilden, Germany, #1017782,) at 56°C for 10 min, and the protease was inactivated at 95°C for 5 mins. The resulting product was directly used for PCR screening. P1 + P5 were used for verification of promoter deletion and P3 + P4 for YY1 motifs deletion. To ensure the selected single cell clones displaying homozygous knockout of lnc-FANCI-2, total cell DNA was isolated using a QIAamp DNA Blood Minikit (QIAGEN, Hilden, Germany, #51106) and then the homozygous deletion was screened and the cells with homozygous KO were selected and confirmed by PCR. The primers used for genotyping are listed in supplementary Table S5. At least two single cell clones either with lnc-FANCI-2 promoter deletion or with YY1 motif deletion were selected. Three single cell clones from the

cells transfected with an empty vector, pSpCas9(BB)-2A-Puro, were also selected and served as controls for the genotyping screening.

Cell proliferation and viability assay

Cell proliferation and viability were determined by a CCK-8 assay (Dojindo Molecular Technologies, Rockville, MD). For proliferation assay, 500 µl of cell suspension (50,000 cells/well) were dispensed into individual wells in a 24-well plate. Cells were then treated with inhibitors (20 µM LY294002, or 10 µM U0126 respectively) after 24 h culture. At each time point of 24 h, 48 h and 72 h, 50 µl of CCK-8 solution was added in three wells of each group and incubated for 1 h in a cell culture incubator at 37°C. Cell proliferation was determined by absorbance at 450 nm viable cells, and normal medium was used to subtract background. Percentage of cell viability was calculated by the following formula: cell viability % = OD450 of inhibitors treated sample / OD450 of untreated sample × 100%.

Colony formation, cell migration, and senescence analysis

CaSki Cells were seeded at 1×10^4, 5×10^3 or 1×10^3 cells/well in 6-well plates. After 24 h of cell growth, the culture medium in each well was replaced by DMEM containing 1% methylcellulose and 10% FBS. The plates were fixed with 3.7% formaldehyde and stained with 1% crystal violet in 2 weeks. Transwell cell migration assays were performed as previously described [81]. Briefly, CaSki cells were re-suspended in serum free DMEM containing 0.1% BSA. 1×10^5 cells in 100 µl were then plated on top of the filter membrane in a transwell insert (SARSTEDT, Nümbrecht, Germany, # 83.3932.800) and incubated for 10 min at 37 °C to allow the cells to settle down, then 600 µl of complete media containing 10% FBS were added into the bottom of the lower chamber in a 24-well plate. After 24 h in culture, the cells on the top of the

insert membrane were removed, and cells on the bottom were fixed, stained by 0.2% crystal violet, and counted.

Cell senescence was determined by a Cellular Senescence Assay (MilliporeSigma, Burlington, MA, # QIA117-1KIT). Briefly, CaSki cells were plated at 2×10^5 cells/well in 6-well plates and cultured for 2 days to 50% confluence. Cells were washed once with 2 ml PBS and fixed with 0.5 ml of fixative solution at room temperature for 10 to 15 min. Then, cells were washed twice with 2 ml PBS, and 1 ml staining solution was added to each well. Cells were incubated at 37°C overnight without CO_2 and protected from the light. Blue stained cells were then counted under light microscopy.

Nuclear and cytoplasmic fractionation

CaSki and SiHa cells were fractionated by using Nuclei EZ Prep Kit (Sigma-Aldrich, #NUC-101) following the manufacturer's protocol. The details of the method could be found in our previous publication [82]. Western blot analysis for nuclear protein SRSF3 and cytoplasmic protein GAPDH was used to determine the fractionation efficiency. The fractionated cytoplasmic and nuclear RNAs were used for detection of lnc-FANCI-2 RNA with human GAPDH RNA serving as a control for RNA fractionation efficiency of the cytoplasmic RNAs.

RT-qPCR

Detection of lnc-FANCI-2 by RT-qPCR was performed as described [8]. Pre-designed primers for lnc-FANCI-2 are listed in Table S5. Briefly, 2 µg total RNA was converted to cDNA using Superscript First-stand Synthesis kit (Thermo Fisher Scientific, Waltham, MA, #11904018). qPCR was performed using TaqMan gene expression Master Mix (Applied Biosystems, Waltham, MA, #4369016) on a StepOne Plus Real-Time PCR system (Applied Biosystems).

TaqMan Gene Expression Assay (Applied Biosystems, #4331182) for GADPH (Hs02758991_g1) was served as an internal control. Data were plotted as fold change over the control group using the $2^{-\Delta\Delta Ct}$ method by which the data was normalized first to the values for GAPDH and then to the median value for control samples. Data are presented as a bar graph with mean ± SD for each group.

Northern blot

To validate lnc-FANCI-2 KO in CaSki cells, Northern blot was performed using poly(A)$^+$ mRNA isolated from 100 µg of total CaSki RNA with PolyATtract mRNA Isolation Systems (Promega, #Z5310) or 10 µg of total CaSki RNA as described previously [8]. Briefly, RNA samples were denatured in NorthernMax Formaldehyde loading dye (Thermo Fisher Scientific, #AM8552) at 75°C for 15 min, then separated in an 1% formaldehyde-containing agarose gel and transferred onto a GeneScreen Plus hybridization transfer membrane (Perkin Elmer, Waltham, MA, #NEF987001PK). RNAs on the membrane were crosslinked by exposing to UV light, and then prehybridized with PerfectHyb Plus hybridization buffer (Sigma-Aldrich, St. Louis, MO, #H7003) for 2 h at 42°C. Specific oligos against lnc-FANCI-2 or GAPDH listed in Table S5 were labeled with [γ-^{32}P]-ATP using T4 PNK (Thermo Fisher Scientific, Waltham, MA, #18004–010) and added into hybridization buffer for overnight incubation at 42°C. The membrane was then washed with 2× SSPE-0.5% SDS solution for 5 mins, followed by twice washes with 0.5× SSPE-0.5% SDS each for 15 min at 42°C and then exposed to a PhosphorImager screen.

RNA-seq and data analysis

Total RNA was from WT CaSki cells, ΔYY1-D5 and ΔPr-A9 cells were extracted using TRIzol reagent (Thermo Fisher Scientific, Waltham, MA, #15596018). Total ribo-minus RNA-seq libraries, four samples in each group, were prepared with TruSeq Stranded Total RNA Library Kit and then subjected to pair-end sequencing using Illumina-HiSeq3000/4000 platform. The samples had 123 to 140 million pass filter reads with more than 89% of bases above the quality score of Q30. Reads of the samples were trimmed from adapters and low-quality bases using Cutadapt before alignment with the reference genome (Human – hg38) and the annotated transcripts using STAR. The average mapping rate of all samples was 95%. Unique alignment was above 84%. There were 4.38 to 5.72% unmapped reads. The mapping statistics were calculated using Picard software. Percent coding bases were between 33-39%. Percent UTR bases were 25-31%, and mRNA bases were between 59-70% for all the samples. Library complexity was measured in terms of unique fragments in the mapped reads using Picard's MarkDuplicate utility. The samples had 71-76% non-duplicate reads. Raw data and the analyzed RNA-seq data supporting the findings in this study have been deposited in the NCBI GEO database (the accession#: GSE190904).

For RNA sequencing, sample reads were processed using the CCBR Pipeliner utility (https://github.com/CCBR/Pipeliner). Briefly, reads were trimmed from adapters and low-quality bases using Cutadapt (version 1.18) (https://bioweb.pasteur.fr/packages/pack@cutadapt@1.18) before alignment to the custom reference genome described below. The transcripts were aligned using STAR v2.5.2b in two-pass mode [83]. Expression levels were quantified using RSEM (version 1.3.0) [84] with a custom gene annotation described below.

The custom reference genome allowing quantification of both viral and host expression used in this alignment consisted of the human reference genome (hg38/Dec. 2013/GRCh38) FASTA with a HPV16 FASTA sequence added as an additional pseudochromosome. The custom gene annotation used for gene expression quantification consisted of a concatenation of the hg38 GENCODE annotation version 30 [85] and the HPV16 genome, with one other notable alteration. The hg38 v30 gene annotation is incorrect at the location of the lnc-FANCI-2 locus in the hg38 genome at chr15:89378104-89398487, which affects quantification of this gene. To correct this annotation and provide accurate quantification, therefore, we first performed BLAST against hg38 using the sequences of the 14 known isoform transcripts of lnc-FANCI-2 (MT669800 for lnc-FANCI-2a-PA2, MT669801 for lnc-FANCI-2a-PA1, MT669802 for lnc-FANCI-2b-PA2, MT669803 for lnc-FANCI-2b-PA1, MT669804 for lnc-FANCI-2c-PA2, MT669805 for lnc-FANCI-2c-PA1, MT669806 for lnc-FANCI-2d-PA2, MT669807 for lnc-FANCI-2d-PA1, MT669808 for lnc-FANCI-2e-PA2, MT669809 for lnc-FANCI-2e-PA1, MT669810 for lnc-FANCI-2f-PA2, MT669811 for lnc-FANCI-2f-PA1, MT669812 for lnc-FANCI-2g-PA2, and MT669813 for lnc-FANCI-2g-PA1). The results were used to determine the precise exon start and stop sites for each of the isoforms. These were then used to manually adjust the default hg38 v30 gene annotations at the locus indicated above to account for the correct structure of the lnc-FANCI-2 locus and its 14 isoforms.

Downstream analysis and visualization were performed within the NIH Integrated Analysis Portal (NIDAP) using R programs developed on the Foundry platform (Palantir Technologies). Briefly, raw counts data produced by RSEM were imported into the NIDAP platform, genes were filtered for low counts (<1 cpm) and the voom algorithm [86] from the Limma R package (version 3.40.6) [87] was used for quantile normalization and calculation of differentially expressed

genes. Pre-ranked GSEA was performed using the Molecular Signatures Database version 6.2 [88, 89] and the fgsea package [90].

Human Soluble Receptor Array

To detect the changes of soluble receptors expressed and released from parental WT CaSki cells and lnc-FANCI-2 KO cells, Human Soluble Receptor Array (R&D Systems, Minneapolis, MN, # ARY012) was performed according to a manufacturer protocol. Briefly, 7 x 10^6 WT CaSki cells and ΔPr-A9 cells were cultured in 15 ml of DMEM supplemented with 10% FBS in a T75 flask for 24 h. Cell culture supernatants were then collected and centrifuged at 4,000 x g for 15 min to remove cell debris. Cells were rinsed by 10 ml of PBS and solubilized in Lysis Buffer 17 at 4°C for 30 min. The cell lysis was then centrifuged at 14,000 x g for 5 min to remove cell debris. Protein concentration of cell lysis was measured by Micro BCA Protein Assay (PIERCE, Waltham, MA, #23235). Next, two sets of N and C membranes were blocking by Buffer 8/1 in 4-Well Multi-dish for 1 h on a rocking platform. 500 μl of culture supernatant or 100 μg cell lysis generated from the WT CaSki cells or ΔPr-A9 cells were diluted in 3 ml of Buffer 8/1, and then applied to each membrane after three washes with 1X Wash Buffer. The membranes were incubated overnight at 4 °C on a rocking platform and washed three times to remove unbound materials followed by incubation with their specific cocktail of biotinylated detection antibodies for 2 h at room temperature on a rocking platform. 2 ml of diluted Streptavidin-HRP was then applied to each membrane for 30 min at room temperature on a rocking platform after three times of wash. 1 ml of the prepared Chemi Reagent Mix was added onto the membranes after three times of wash. The signal was then captured by ChemiDoc™ Touch Imaging System (Bio-Rad). Quantification of the relative protein expression in ΔPr-A9 cells compared to the WT cells was determined in Image Lab software (Bio-Rad). The average signal of duplicate spots

subtracting background value from negative control spots represented the levels of each protein. The relative change in protein levels was then determined by comparing ΔPr-A9 cells to the WT cells.

Antibodies and Immunoblot

Immunoblot analysis was performed for individual proteins by using the following antibodies: Anti-MCAM (#17564-1-AP), anti-E2F1 (#12171-1-AP), anti-MAP4K4 (#55247-1-AP), and anti-p53 (#10442-1-AP) antibodies were from ProteinTech (Rosemont, IL). Anti-tubulin (#T5201) antibodies were from Sigma-Aldrich (St. Louis, MO). Anti-hnRNP C1/C2 (#Ab10294) and anti-RAC3 (#ab129062) antibodies were from Abcam (Cambridge, United Kingdom). Anti-CCND2 (Cyclin D2, #3741), anti-Akt (#9272), anti-phospho-Akt (#9271), anti-Erk1/2 (p44/42 MAPK, #4695), anti-phospho-Erk1/2 (#4370), and anti-IGFBP3 (D1U9C, #25864) were from Cell Signaling Technology (Danvers, MA). Anti-VIM (#MA5-11883) was from Thermo Fisher Scientific (Waltham, MA). Anti- PODXL2 (#AF1524), anti-ECM1 (#MAB39371), anti-TIMP2 (#AF971) and anti-ADAM8 (#AF1031) were from R&D Systems (Minneapolis, MN). Anti-SRSF3 (#NBP2-76892) was from Novus Biologicals (Centennial, CO). Rabbit polyclonal anti-HPV16 E7 (#GTX133411) was from GeneTex (Irvine, CA).

RAS GTPase Chemi ELISA

RAS GTPase activity in cell lysate of the WT CaSki or ΔPr-A9 cells was determined by RAS GTPase Chemi ELISA Kit (Active Motif, Carlsbad, CA, # 52097) according to the manufacturer protocol. Briefly, 2 x 10^7 cells were solubilized in 500 µl of Complete Lysis/Binding buffer at 4°C for 15 min. Protein concentration of cell extract was then measured by Micro BCA Protein

Assay (PIERCE, #23235) after cell debris was removed by centrifuged at 14,000 x g for 10 min. 2 µg of GST-Raf-RBD in 50 µl of Complete Lysis/Binding buffer was coated in each well of a 96-well plate by incubating for 1 h at 4°C with 100 rpm agitation. After three times of wash, 50 µg of extract in 50 µl Complete Lysis/Binding buffer was then added in each well. HeLa (EGF treated) extract was served as a positive control. The plate was incubated for 1 h at room temperature with 100 rpm agitation. The active RAS protein (GST-Raf-RBD binding) in each well was then detected by incubating with primary RAS antibody and three times of wash, and then, HRP secondary antibody and three times of wash. Chemiluminescence in each well was read in luminometer by adding 50 µl room-temperature Chemiluminescent Working Solution.

RNA *in situ* hybridization (RNA-ISH)

Endogenous lnc-FANCI-2 in CaSki cells was examined by RNAScope Multiplex Fluorescent V2 Assay (Advanced Cell Diagnostics, Minneapolis, MN, #323100) as described previously [8]. A custom-designed probe targeting to nt 359-1713 of GenBank MT669800.1 transcript for lnc-FANCI-2a-PA2 was utilized. Dual staining of lnc-FANCI-2 and IGFBP RNA was performed according to RNAscope Multplex Fluorescent v2 Manual part 2 (#323100), and the channel one probe for IGFBP3 (#310351) and channel three probe for lnc-FANCI-2 (#509061-C3) were applied.

For RNA-ISH combined IF staining of MCAM protein, 3 x 10^5 of CaSki cells were grown on a glass coverslip in a 6-well plate for 24 h before transfection. Plasmid pHBL05 containing a long isoform of lnc-FANCI-2a-PA2 (GenBank, MT669800.1) were transfected into cells with LipoD 293 DNA in vitro transfection reagent. Cultured Adherent Cell Sample Preparation for the RNAscope Multplex Fluorescent v2 was performed according to a manufacturer protocol (Advanced Cell Diagnostics, Minneapolis, MN, MK-50-010). Briefly, the cells were washed

with PBS and fixed by 10% neutral buffered formalin at room temperature for 30 min. The cells were washed with PBS three times and dehydrated by 50% ethanol for 5 min, 70% ethanol for 5 min, 100% ethanol for 5min, and 100% ethanol for 10 min. The cells were then rehydrated by 70% ethanol for 2 min, 50% ethanol for 2min, and PBS for 10 min. The cells were then applied to hydrogen peroxide at RT for 10 min and washed with distilled water for three times. The cells were then applied to Protease III (1:15 dilution with PBS) at RT for 10 min. The cells were ready for RNAscope Multplex Fluorescent v2 Manual part 2 (#323100). IF was performed after RNA-ISH but before counterstaining with DAPI. The cells were washed with PBS, and then were blocked with 2% bovine serum albumin (BSA) in PBS for 1 h at 37°C or overnight at 4°C. The cells were incubated with the MCAM primary antibody (ProteinTech, Rosemont, IL, #17564-1-AP, diluted 1:200 in 2% BSA blocking buffer) for 2 h at 37°C. An alexaFluor-conjugated secondary antibody (1: 500, ThermoFisher Scientific) was diluted in a blocking solution and incubated for 1 h at 37˚C. The slides were washed with PBS for three times. DAPI was used for nuclei counterstaining before mounting in a Prolong Gold Antifade mounting medium (Thermo Fisher Scientific, #P36934).

Confocal images were collected using a Zeiss LSM710 laser-scanning microscope equipped with a 63 × Plan-Apochromat (N.A. 1.4) objective lens. Three dimensional distributions of lnc-FANCI-2 were generated by Z stacks using ZEN 2.3 software (Zeiss).

Isolation of lnc-FANCI-2 RNA-protein complex by RNA Purification (IRPCRP) and LC-MS/MS analysis

To identify the lnc-FANCI-2 RNA-associated proteins, we pulled down lnc-FANCI-2 RNA from WT CaSki cells using IRPCRP, a modified ChIRP protocol from the published method [91], followed by mass spectrometry [57]. Briefly, a total of 30 lnc-FANCI-2 anti-sense oligo

probes (oLLY496-oLLY525), each with biotinTEG at 3' end, across entire lnc-FANCI-2 RNA (GenBank: MT669800.1) were designed using the online probe designer (singlemoleculefish.com). Two probe pools were used in IRPCRP assays: the probe pool 1 was the mixed oligos in even numbers (oLLY496 /498/500/502/504/506/508/510/512/514/516/518/520/522/524) and the probe pool 2 mixed with the oligos in odd numbers (oLLY497/499/501/503/505/507/509/511/513/515/517/519/521/523/525) (Table S5). CaSki cells were seeded and harvested at 24 h and the cell lysates from 200 million cells were used for each IRPCRP reaction. Instead of formaldehyde crosslink in the standard protocol [91], UV cross-linking (254 nm, energy at 480 mJ/cm^2) was applied in our study to minimize non-specific binding. After UV cross-linking, the cells were lysed on ice in 1× radioimmunoprecipitation assay (RIPA) buffer (Boston Bioproducts, Ashland, MA, #BP-115) (50 mM Tris base/Tris-HCl [pH 7.4], 150 mM NaCl, 0.5% sodium deoxycholate, 0.1% sodium dodecyl sulfate [SDS], 1% NP-40) supplemented with protease inhibitors (complete mini-EDTA-free protease inhibitor cocktail, Millipore Sigma, Burlington, MA, #469315900) and RNase inhibitor (Thermo Fisher Scientific, #AM2694) for 30 min, followed by a brief sonication. The obtained cell lysates were spin for 10 min of centrifugation at 20,000 × g at 4°C and the collected cell lysate supernatants were pre-absorbed with C-1 magnetic beads (Thermo Fisher Scientific, #65002) for 1 h at room temperature before proceeding to oligo probe hybridization. Two different hybridization protocols were used to pull down lnc-FANCI-2 RNA. One had the pre-absorbed cell lysate incubated with two separate probe pools in hybridization buffer (750 mM NaCl, 0.1% SDS, 50 mM Tris-Cl pH 7.0, 1 mM EDTA, 15% formamide) at room temperature overnight and the pre-washed C-1 magnetic beads was then added and incubated for additional 4 h; the other had two

separate oligo probe pools immobilized onto pre-washed C-1 magnetic beads first at room temperature for 1 h and then mixed with the pre-absorbed cell lysates in hybridization buffer overnight at room temperature, along with the negative control-beads only IRPCRP reaction without oligo probes serving as a control in the modified protocol. After hybridization, the beads were washed with 1x wash buffer (2x NaCl and sodium citrate, 0.1% SDS) for 5 times and resuspend in 1 ml lysis buffer. 100 µl beads from each IRPCRP were set aside for RNA isolation of lnc-FANCI-2 pulldown efficiency and the leftover beads were sent for mass spectrometry analysis. The input and IRPCRP RNA were isolated using a MIRNeasy mini kit (Qiagen, #217004) following the standard protocol after proteinase K (Millipore Sigma, #71049) digestion. For mass spectrometry analysis, the IRPCRP beads were processed for trypsin/LysC digestion before submitted to Thermo Scientific Orbitrap Exploris 240 Mass Spectrometer and a Thermo Dionex UltiMate 3000 RSLCnano System for Proteinomics. Peptides from trypsin digestion were loaded onto a peptide trap cartridge at a flow rate of 5 µl /min. The trapped peptides were eluted onto a reversed-phase Easy-Spray Column PepMap RSLC, C18, 2 µM, 100A, 75 µm x 250 mm (Thermo Scientific) using a linear gradient of acetonitrile (3-36%) in 0.1% formic acid. The elution duration was 110 min at a flow rate of 0.3 µl/min. Eluted peptides from the Easy-Spray column were ionized and sprayed into the mass spectrometer, using a Nano Easy-Spray Ion Source (Thermo Fisher Scientific) under the following settings: spray voltage, 1.6 kV, Capillary temperature, 275 °C. Other settings were empirically determined. Raw data files were searched against human protein sequences database using the Proteome Discoverer 2.4 software (Thermo, San Jose, CA) based on the SEQUEST algorithm. All the protein peptides identified in the IRPCRP pulldowns were summarized in Table S4.

ACKNOWLEDGEMENTS

We thank Louise T. Chow of the University of Alabama at Birmingham for her critical reading of the manuscript. We thank Craig Meyers of Penn State University Hershey Medical Center for providing HPV16- and HPV18- infected raft cultures, Xing Xie and Yang Li of Zhejiang University Women's Hospital of China for cervical tissue sections, and Johannes G. Schweizer from Arbor Vita corporation for anti-HPV16 E6-specific antibodies. This study was supported by the Intramural Research Program of the NIH, National Cancer Institute, Center for Cancer Research.

DATA AVAILABILITY

NCBI GEO database Accession# GSE190904 for CaSki cell RNA-seq.

REFERENCES

1. Torre, L.A., Bray, F., Siegel, R.L., Ferlay, J., Lortet-Tieulent, J. & Jemal, A. Global cancer statistics, 2012. *CA Cancer J Clin* **65**, 87-108 (2015).
2. Bosch, F.X., Lorincz, A., Munoz, N., Meijer, C.J. & Shah, K.V. The causal relation between human papillomavirus and cervical cancer. *J Clin Pathol* **55**, 244-265 (2002).
3. Munoz, N. *et al.* Epidemiologic classification of human papillomavirus types associated with cervical cancer. *N Engl J Med* **348**, 518-527 (2003).
4. Walboomers, J.M. *et al.* Human papillomavirus is a necessary cause of invasive cervical cancer worldwide. *J Pathol* **189**, 12-19 (1999).
5. McCredie, M.R. *et al.* Natural history of cervical neoplasia and risk of invasive cancer in women with cervical intraepithelial neoplasia 3: a retrospective cohort study. *Lancet Oncol* **9**, 425-434 (2008).
6. Niraj, J., Farkkila, A. & D'Andrea, A.D. The Fanconi Anemia Pathway in Cancer. *Annu Rev Cancer Biol* **3**, 457-478 (2019).
7. Hoskins, E.E. *et al.* Fanconi anemia deficiency stimulates HPV-associated hyperplastic growth in organotypic epithelial raft culture. *Oncogene* **28**, 674-685 (2009).
8. Liu, H. *et al.* Oncogenic HPV promotes the expression of the long noncoding RNA lnc-FANCI-2 through E7 and YY1. *Proc Natl Acad Sci U S A* **118** (2021).
9. Santegoets, L.A. *et al.* Different DNA damage and cell cycle checkpoint control in low- and high-risk human papillomavirus infections of the vulva. *Int J Cancer* **130**, 2874-2885 (2012).

708 10. Hoskins, E.E. *et al.* The fanconi anemia pathway limits human papillomavirus replication. *J Virol*
709 **86**, 8131-8138 (2012).
710 11. Park, J.W. *et al.* Deficiencies in the Fanconi anemia DNA damage response pathway increase
711 sensitivity to HPV-associated head and neck cancer. *Cancer Res* **70**, 9959-9968 (2010).
712 12. Park, J.W., Shin, M.K. & Lambert, P.F. High incidence of female reproductive tract cancers in FA-
713 deficient HPV16-transgenic mice correlates with E7's induction of DNA damage response, an
714 activity mediated by E7's inactivation of pocket proteins. *Oncogene* **33**, 3383-3391 (2014).
715 13. Roman, A. & Munger, K. The papillomavirus E7 proteins. *Virology* **445**, 138-168 (2013).
716 14. Vande Pol, S.B. & Klingelhutz, A.J. Papillomavirus E6 oncoproteins. *Virology* **445**, 115-137 (2013).
717 15. Liu, X., Dakic, A., Zhang, Y., Dai, Y., Chen, R. & Schlegel, R. HPV E6 protein interacts physically and
718 functionally with the cellular telomerase complex. *Proc Natl Acad Sci U S A* **106**, 18780-18785
719 (2009).
720 16. Klingelhutz, A.J., Foster, S.A. & McDougall, J.K. Telomerase activation by the E6 gene product of
721 human papillomavirus type 16. *Nature* **380**, 79-82 (1996).
722 17. Storey, A., Pim, D., Murray, A., Osborn, K., Banks, L. & Crawford, L. Comparison of the in vitro
723 transforming activities of human papillomavirus types. *EMBO J* **7**, 1815-1820 (1988).
724 18. Barbosa, M.S. & Schlegel, R. The E6 and E7 genes of HPV-18 are sufficient for inducing two-stage
725 in vitro transformation of human keratinocytes. *Oncogene* **4**, 1529-1532 (1989).
726 19. Hawley-Nelson, P., Vousden, K.H., Hubbert, N.L., Lowy, D.R. & Schiller, J.T. HPV16 E6 and E7
727 proteins cooperate to immortalize human foreskin keratinocytes. *EMBO J* **8**, 3905-3910 (1989).
728 20. Munger, K., Phelps, W.C., Bubb, V., Howley, P.M. & Schlegel, R. The E6 and E7 genes of the
729 human papillomavirus type 16 together are necessary and sufficient for transformation of
730 primary human keratinocytes. *J Virol* **63**, 4417-4421 (1989).
731 21. Halbert, C.L., Demers, G.W. & Galloway, D.A. The E7 gene of human papillomavirus type 16 is
732 sufficient for immortalization of human epithelial cells. *J Virol* **65**, 473-478 (1991).
733 22. Phelps, W.C., Yee, C.L., Munger, K. & Howley, P.M. The human papillomavirus type 16 E7 gene
734 encodes transactivation and transformation functions similar to those of adenovirus E1A. *Cell*
735 **53**, 539-547 (1988).
736 23. DiPaolo, J.A., Woodworth, C.D., Popescu, N.C., Notario, V. & Doniger, J. Induction of human
737 cervical squamous cell carcinoma by sequential transfection with human papillomavirus 16 DNA
738 and viral Harvey ras. *Oncogene* **4**, 395-399 (1989).
739 24. Schreiber, K. *et al.* Strong synergy between mutant ras and HPV16 E6/E7 in the development of
740 primary tumors. *Oncogene* **23**, 3972-3979 (2004).
741 25. Narisawa-Saito, M. *et al.* An in vitro multistep carcinogenesis model for human cervical cancer.
742 *Cancer Res* **68**, 5699-5705 (2008).
743 26. Song, S., Pitot, H.C. & Lambert, P.F. The human papillomavirus type 16 E6 gene alone is
744 sufficient to induce carcinomas in transgenic animals. *J Virol* **73**, 5887-5893 (1999).
745 27. Herber, R., Liem, A., Pitot, H. & Lambert, P.F. Squamous epithelial hyperplasia and carcinoma in
746 mice transgenic for the human papillomavirus type 16 E7 oncogene. *J Virol* **70**, 1873-1881
747 (1996).
748 28. Song, S., Liem, A., Miller, J.A. & Lambert, P.F. Human papillomavirus types 16 E6 and E7
749 contribute differently to carcinogenesis. *Virology* **267**, 141-150 (2000).
750 29. Riley, R.R., Duensing, S., Brake, T., Munger, K., Lambert, P.F. & Arbeit, J.M. Dissection of human
751 papillomavirus E6 and E7 function in transgenic mouse models of cervical carcinogenesis. *Cancer*
752 *Res* **63**, 4862-4871 (2003).
753 30. Brake, T. & Lambert, P.F. Estrogen contributes to the onset, persistence, and malignant
754 progression of cervical cancer in a human papillomavirus-transgenic mouse model. *Proc Natl*
755 *Acad Sci U S A* **102**, 2490-2495 (2005).

31. Kato, S. *et al.* Activation of the estrogen receptor through phosphorylation by mitogen-activated protein kinase. *Science* **270**, 1491-1494 (1995).
32. Statello, L., Guo, C.J., Chen, L.L. & Huarte, M. Gene regulation by long non-coding RNAs and its biological functions. *Nat Rev Mol Cell Biol* **22**, 96-118 (2021).
33. Zhao, J., Sun, B.K., Erwin, J.A., Song, J.J. & Lee, J.T. Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. *Science* **322**, 750-756 (2008).
34. Yap, K.L. *et al.* Molecular interplay of the noncoding RNA ANRIL and methylated histone H3 lysine 27 by polycomb CBX7 in transcriptional silencing of INK4a. *Mol Cell* **38**, 662-674 (2010).
35. Cho, S.W. *et al.* Promoter of lncRNA Gene PVT1 Is a Tumor-Suppressor DNA Boundary Element. *Cell* **173**, 1398-1412 e1322 (2018).
36. Xiang, J.F. *et al.* Human colorectal cancer-specific CCAT1-L lncRNA regulates long-range chromatin interactions at the MYC locus. *Cell Res* **24**, 513-531 (2014).
37. Tripathi, V. *et al.* The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation. *Mol Cell* **39**, 925-938 (2010).
38. Yap, K., Mukhina, S., Zhang, G., Tan, J.S.C., Ong, H.S. & Makeyev, E.V. A Short Tandem Repeat-Enriched RNA Assembles a Nuclear Compartment to Control Alternative Splicing and Promote Cell Survival. *Mol Cell* **72**, 525-540 e513 (2018).
39. Nishizawa, M., Ikeya, Y., Okumura, T. & Kimura, T. Post-transcriptional inducible gene regulation by natural antisense RNA. *Front Biosci (Landmark Ed)* **20**, 1-36 (2015).
40. Lee, S. *et al.* Noncoding RNA NORAD Regulates Genomic Stability by Sequestering PUMILIO Proteins. *Cell* **164**, 69-80 (2016).
41. Poliseno, L., Salmena, L., Zhang, J., Carver, B., Haveman, W.J. & Pandolfi, P.P. A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. *Nature* **465**, 1033-1038 (2010).
42. Grelet, S. *et al.* A regulated PNUTS mRNA to lncRNA splice switch mediates EMT and tumour progression. *Nat Cell Biol* **19**, 1105-1115 (2017).
43. Aznaourova, M. *et al.* Noncoding RNA MaIL1 is an integral component of the TLR4-TRIF pathway. *Proc Natl Acad Sci U S A* **117**, 9042-9053 (2020).
44. Guo, C.J. *et al.* Distinct Processing of lncRNAs Contributes to Non-conserved Functions in Stem Cells. *Cell* **181**, 621-636 e622 (2020).
45. BeltCappellino, A., Majerciak, V., Lobanov, A., Lack, J., Cam, M. & Zheng, Z.M. CRISPR/Cas9-Mediated Knockout and In Situ Inversion of the ORF57 Gene from All Copies of the Kaposi's Sarcoma-Associated Herpesvirus Genome in BCBL-1 Cells. *J Virol* **93** (2019).
46. Schlomann, U. *et al.* The metalloprotease disintegrin ADAM8. Processing by autocatalysis is required for proteolytic activity and cell adhesion. *J Biol Chem* **277**, 48210-48219 (2002).
47. Liberzon, A., Birger, C., Thorvaldsdottir, H., Ghandi, M., Mesirov, J.P. & Tamayo, P. The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst* **1**, 417-425 (2015).
48. Downward, J. Targeting RAS signalling pathways in cancer therapy. *Nat Rev Cancer* **3**, 11-22 (2003).
49. Simanshu, D.K., Nissley, D.V. & McCormick, F. RAS Proteins and Their Regulators in Human Disease. *Cell* **170**, 17-33 (2017).
50. Castellano, E. & Downward, J. RAS Interaction with PI3K: More Than Just Another Effector Pathway. *Genes Cancer* **2**, 261-274 (2011).
51. Asati, V., Mahapatra, D.K. & Bharti, S.K. PI3K/Akt/mTOR and Ras/Raf/MEK/ERK signaling pathways inhibitors as anticancer agents: Structural and pharmacological perspectives. *Eur J Med Chem* **109**, 314-341 (2016).

803 52. Cuesta, C., Arevalo-Alameda, C. & Castellano, E. The Importance of Being PI3K in the RAS
804 Signaling Network. *Genes (Basel)* **12** (2021).
805 53. Vanhaesebroeck, B., Perry, M.W.D., Brown, J.R., Andre, F. & Okkenhaug, K. PI3K inhibitors are
806 finally coming of age. *Nat Rev Drug Discov* **20**, 741-769 (2021).
807 54. Favata, M.F. *et al.* Identification of a novel inhibitor of mitogen-activated protein kinase kinase. *J*
808 *Biol Chem* **273**, 18623-18632 (1998).
809 55. Kebir, A. *et al.* CD146 short isoform increases the proangiogenic potential of endothelial
810 progenitor cells in vitro and in vivo. *Circ Res* **107**, 66-75 (2010).
811 56. Stalin, J. *et al.* Soluble CD146 boosts therapeutic effect of endothelial progenitors through
812 proteolytic processing of short CD146 isoform. *Cardiovasc Res* **111**, 240-251 (2016).
813 57. Chu, C. & Chang, H.Y. ChIRP-MS: RNA-Directed Proteomic Discovery. *Methods Mol Biol* **1861**, 37-
814 45 (2018).
815 58. Gao, X., Gao, C., Liu, G. & Hu, J. MAP4K4: an emerging therapeutic target in cancer. *Cell Biosci* **6**,
816 56 (2016).
817 59. Gao, X. *et al.* MAP4K4 is a novel MAPK/ERK pathway regulator required for lung
818 adenocarcinoma maintenance. *Mol Oncol* **11**, 628-639 (2017).
819 60. Gonzalez-Montero, J., Rojas, C.I. & Burotto, M. MAP4K4 and cancer: ready for the main stage?
820 *Front Oncol* **13**, 1162835 (2023).
821 61. Patterson, V. *et al.* Abrogation of MAP4K4 protein function causes congenital anomalies in
822 humans and zebrafish. *Sci Adv* **9**, eade0631 (2023).
823 62. Crook, T., Storey, A., Almond, N., Osborn, K. & Crawford, L. Human papillomavirus type 16
824 cooperates with activated ras and fos oncogenes in the hormone-dependent transformation of
825 primary mouse cells. *Proc Natl Acad Sci U S A* **85**, 8820-8824 (1988).
826 63. Matlashewski, G., Schneider, J., Banks, L., Jones, N., Murray, A. & Crawford, L. Human
827 papillomavirus type 16 DNA cooperates with activated ras in transforming primary cells. *EMBO J*
828 **6**, 1741-1746 (1987).
829 64. Greenhalgh, D.A. *et al.* Transgenic mice expressing targeted HPV-18 E6 and E7 oncogenes in the
830 epidermis develop verrucous lesions and spontaneous, rasHa-activated papillomas. *Cell Growth*
831 *Differ* **5**, 667-675 (1994).
832 65. Riou, G., Barrois, M., Sheng, Z.M., Duvillard, P. & Lhomme, C. Somatic deletions and mutations
833 of c-Ha-ras gene in human cervical cancers. *Oncogene* **3**, 329-333 (1988).
834 66. Eiben, G.L. *et al.* Establishment of an HLA-A*0201 human papillomavirus type 16 tumor model
835 to determine the efficacy of vaccination strategies in HLA-A*0201 transgenic mice. *Cancer Res*
836 **62**, 5792-5799 (2002).
837 67. Golijow, C.D., Mouron, S.A., Gomez, M.A. & Dulout, F.N. Differences in K-ras codon 12 mutation
838 frequency between "high-risk" and "low-risk" HPV-infected samples. *Gynecol Oncol* **75**, 108-112
839 (1999).
840 68. Prokopakis, P., Sourvinos, G., Koumantaki, Y., Koumantakis, E. & Spandidos, D.A. K-ras mutations
841 and HPV infection in cervicitis and intraepithelial neoplasias of the cervix. *Oncol Rep* **9**, 129-133
842 (2002).
843 69. Strickland, S.W. & Vande Pol, S. The Human Papillomavirus 16 E7 Oncoprotein Attenuates AKT
844 Signaling To Promote Internal Ribosome Entry Site-Dependent Translation and Expression of c-
845 MYC. *J Virol* **90**, 5611-5621 (2016).
846 70. Cararo-Lopes, E. *et al.* Autophagy buffers Ras-induced genotoxic stress enabling malignant
847 transformation in keratinocytes primed by human papillomavirus. *Cell Death Dis* **12**, 194 (2021).
848 71. Singh, A. *et al.* A gene expression signature associated with "K-Ras addiction" reveals regulators
849 of EMT and tumor cell survival. *Cancer Cell* **15**, 489-500 (2009).

850 72. Yang, J., Nie, J., Ma, X., Wei, Y., Peng, Y. & Wei, X. Targeting PI3K in cancer: mechanisms and
851 advances in clinical trials. *Mol Cancer* **18**, 26 (2019).
852 73. Saliani, M., Mirzaiebadizi, A., Javadmanesh, A., Siavoshi, A. & Ahmadian, M.R. KRAS-related long
853 noncoding RNAs in human cancers. *Cancer Gene Ther* (2021).
854 74. Su, Y.C., Han, J., Xu, S., Cobb, M. & Skolnik, E.Y. NIK is a new Ste20-related kinase that binds NCK
855 and MEKK1 and activates the SAPK/JNK cascade via a conserved regulatory domain. *EMBO J* **16**,
856 1279-1290 (1997).
857 75. Collins, C.S. *et al.* A small interfering RNA screen for modulators of tumor cell motility identifies
858 MAP4K4 as a promigratory kinase. *Proc Natl Acad Sci U S A* **103**, 3775-3780 (2006).
859 76. Joshkon, A. *et al.* Role of CD146 (MCAM) in Physiological and Pathological Angiogenesis-
860 Contribution of New Antibodies for Therapy. *Biomedicines* **8** (2020).
861 77. Anfosso, F. *et al.* Activation of human endothelial cells via S-endo-1 antigen (CD146) stimulates
862 the tyrosine phosphorylation of focal adhesion kinase p125(FAK). *J Biol Chem* **273**, 26852-26856
863 (1998).
864 78. Dye, D.E. *et al.* hShroom1 links a membrane bound protein to the actin cytoskeleton. *Cell Mol*
865 *Life Sci* **66**, 681-696 (2009).
866 79. Wang, Z., Xu, Q., Zhang, N., Du, X., Xu, G. & Yan, X. CD146, from a melanoma cell adhesion
867 molecule to a signaling receptor. *Signal Transduct Target Ther* **5**, 148 (2020).
868 80. Bardin, N., Frances, V., Combes, V., Sampol, J. & Dignat-George, F. CD146: biosynthesis and
869 production of a soluble form in human cultured endothelial cells. *FEBS Lett* **421**, 12-14 (1998).
870 81. Justus, C.R., Leffler, N., Ruiz-Echevarria, M. & Yang, L.V. In vitro cell migration and invasion
871 assays. *J Vis Exp* (2014).
872 82. Yu, L. & Zheng, Z.M. Human Papillomavirus Type 16 Circular RNA Is Barely Detectable for the
873 Claimed Biological Activity. *mBio* **13**, e0359421 (2022).
874 83. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21 (2013).
875 84. Li, B. & Dewey, C.N. RSEM: accurate transcript quantification from RNA-Seq data with or without
876 a reference genome. *BMC Bioinformatics* **12**, 323 (2011).
877 85. Harrow, J. *et al.* GENCODE: the reference human genome annotation for The ENCODE Project.
878 *Genome Res* **22**, 1760-1774 (2012).
879 86. Law, C.W., Chen, Y., Shi, W. & Smyth, G.K. voom: Precision weights unlock linear model analysis
880 tools for RNA-seq read counts. *Genome Biol* **15**, R29 (2014).
881 87. Smyth, G.K. Linear models and empirical bayes methods for assessing differential expression in
882 microarray experiments. *Stat Appl Genet Mol Biol* **3**, Article3 (2004).
883 88. Liberzon, A., Subramanian, A., Pinchback, R., Thorvaldsdottir, H., Tamayo, P. & Mesirov, J.P.
884 Molecular signatures database (MSigDB) 3.0. *Bioinformatics* **27**, 1739-1740 (2011).
885 89. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for
886 interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* **102**, 15545-15550
887 (2005).
888 90. Korotkevich, G., Sukhov, V., Budin, N., Shpak, B., Artyomov, M.N. & Sergushichev, A. Fast gene
889 set enrichment analysis. *bioRxiv* (2021).
890 91. Chu, C., Quinn, J. & Chang, H.Y. Chromatin isolation by RNA purification (ChIRP). *J Vis Exp* (2012).

891

892

893

894

Figure legends

Fig. 1. Knockout (KO) of lnc-FANCI-2 in CaSki cells affects cell proliferation, colony formation and migration. (A) Diagram and KO strategies of lnc-FANCI-2 gene. On the figure top is the lnc-FANCI-2 gene structure and its alternative transcription start sites (TSS) and polyadenylation sites (pA). TSS2 or pA2 with a heavier arrow are predominately used for lnc-FANCI-2 expression. The lower figure part shows KO strategies. Red slashes represent gRNA-targeted sites to create genomic DNA deletions by CRISPR-Cas9 technology. Deletion of YY1-binding motifs (ΔYY1 motifs or ΔYY1) led to a deletion of 86-bp DNA fragment containing two YY1-binding motifs. Deletion of lnc-FANCI-2 promoter (ΔPromoter or ΔPr) led to delete a ~3.3-kb promoter region. Primers (P1-P4) used for deletion screening were shown as black arrows. (B) PCR screening of single cell clones with homozygous lnc-FANCI-2 KO, with indicated primer sets diagramed (A). Single cell clones selected from the cells transfected with an empty vector served as control cells. (C) Evaluation of lnc-FANCI-2 KO efficiency in the selected single cell clones (B) by RT-qPCR. (D) Northern blot validation of lnc-FANCI-2 KO efficiency from the individual single cell clones on polyA$^+$ RNA (enriched from 100 μg of total cell RNA, Lanes 1-4) or 10 μg of total cell RNA (lanes 5-9). Total cell RNA (2 μg) of HEK293T cells with ectopic expression of two isoforms (short, S or long, L) of lnc-FANCI-2 cDNA served as a control. Oligo probe P5 (A) labeled with ^{32}P was used for hybridization. GAPDH RNA served as a loading control and hybridized with a GAPDH-specific oligo probe. (E-F) RNA in situ hybridization (RNA-ISH) validation of lnc-FANCI-2 KO efficiency by RNAscope technology. Two single cell clones were examined with red color for lnc-FANCI-2 and blue color for the nucleus (E). Bar graphs show the copy number of lnc-FANCI-2 per cell in the parental (WT), ΔYY1-D5 or ΔPr-A9 cells (each averaged from 200 cells) (F). (G-I) lnc-FANCI-2 expression is

required for cell proliferation, colony formation, and cell invasion of CaSki cells. Representative pictures are shown from three separate experiments of cell proliferation by CCK8 assay (G), cell colony formation under 1% methylcellulose solution (H), and cell migration by transwell assay (I) of WT, ΔYY1-D5 or ΔPr-A9 cells. Cell colonies were stained with crystal violet as described in the procedure. Numbers of the invaded cells crossed the transwell membrane were quantified and shown as bar graphs (I). *, P< 0.05; **, P<0.01; ***, P<0.001 by two tailed Student *t* test.

Fig. 2. KO of lnc-FANCI-2 in CaSki cells affects expression of cellular soluble receptors. (A) Dot blots shows differentially expressed soluble receptors in ΔPr-A9 cells when compared to those in the WT CaSki cells. A Proteome Profiler Human sReceptor Array was used to examine 105 cellular soluble receptors. Label in red, upregulated soluble receptors; label in blue, downregulated soluble receptors. (B) Quantification of differentially expressed and/or released soluble receptors (A) in ΔPr-A9 cells when compared to those in the WT CaSki cells. Error bar represents standard deviation of replicates in each proteome array. The dash line indicates the threshold of fold change-1 with a score above or below the threshold being determined as differentially expressed and/or released soluble receptors. (C) Validation of several differentially expressed soluble receptors by immunoblot analysis using specific antibodies. Tubulin served as an internal loading control. * non-specific band in MCAM immunoblot.

Fig. 3. Transcriptomic effect of lnc-FANCI-2 KO in CaSki cells by RNA-seq analysis. (A) RNA-seq reads-coverage maps by IGV showing the expression levels of lnc-FANCI-2 from two KO cell clones, ΔPr-A9 and ΔYY1-D5, to the WT CaSki cells. one representative coverage profile of four in each type of cells is shown by IGV. Red slashes represent the deleted genomic region, with validated fourteen lnc-FANCI-2 RNA isoforms shown below [8]. (B) Similarity in heatmap comparison among ΔPr-A9, ΔYY1-D5 and WT cells was generated using Limma-

normalized counts from each sample by Pearson complete linkage method. (C) Volcano plot visualization of differentially expressed genes (DEGs) in ΔPr-A9 and ΔYY1-D5. The genes with the most significant change in the increased (red) or decreased (blue) expression are indicated. FC, fold change. (D) Venn diagram depicting the overlapped DEGs between ΔPr-A9 and ΔYY1-D5 cells over the WT CaSki cells. (E) Heatmap shows overlapped DEGs with FPKM equal or more than 7 in ΔPr-A9 and ΔYY1-D5 cells when compared with the WT CaSki cells. (F) Validation of selective 12 upregulated or downregulated DEGs shown in the heatmap by RT-PCR with gene-specific primers in the presence (+) or absence (-) of reverse transcriptase (RT). GAPDH served as an internal RNA control. Relative expression level of each gene was calculated based on band density after normalizing to GAPDH RNA band, with the expression level in the WT CaSki cells setting as 1. (G) Validation of the upregulated expression of IGFBP3 in ΔPr-A9 cells over the WT CaSki cells by immunoblot analysis.

Fig. 4. Pathway analyses of DEGs identified by RNA-seq. (A) Top 3 upregulated and top 4 downregulated pathways in ΔPr-A9 and ΔYY1-D5 cells over the WT CaSki cells. Data were generated by Gene set enrichment analysis (GSEA) performed with Hallmark gene sets. NES stands for normalized enrichment score. (B-E) Each GSEA Enrichment plot shows the enrichment score and gene hits enriched in ΔPr-A9 cells using Hallmark gene sets, KRAS_SIGNALING_UP (B), EPITHELIAL_MESENCHYMAL_TRANSITION (C), INTERFERON_GAMMA_ RESPONSE (D) or INTERFERON_ALPHA_RESPONSE (E). padj, adjusted p value; Zero crossed 7342, the middle number of total genes in the GSEA at ranking. The heatmaps below the enrichment plots (B and C) visualize the genes enriched in respective pathways of KRAS_SIGNALING_UP (B) and

EPITHELIAL_MESENCHYMAL_TRANSITION (C) in ΔPr-A9 cells when compared to the WT CaSki cells.

Fig. 5. CaSki cells with lnc-FANCI-2 KO exhibit activation of RAS signaling pathway. (A) CaSki cells with lnc-FANCI-2 KO (top, ΔPr-A9 cells) or knockdown (KD, lower) by lnc-FANCI-2-specific siRNAs display increased RAS GTPase activity when compared to the WT CaSki cells (top) or the WT cells treated with a nonspecific control siRNA (siNS). The relative RAS GTPase activity was measured by a RAS GTPase Chemi ELISA assay. *, $P< 0.05$; ***, $P<0.001$ by two tailed Student *t* test. (B and C) Selective validation of increased expression of RAS signaling-related downstream genes in lnc-FANCI-2 KO ΔPr-A9 cells (B) or lnc-FANCI-2 KD CaSki cells (C) over the WT CaSki cells. Relative expression of the indicated genes in ΔPr-A9 cells in comparison to the WT cells (B) or in the WT CaSki cells treated for 48 h and 96 h with a nonspecific control siRNA (siNS) or a siRNA specifically targeting the lnc-FANCI-2 exon 3 (C) were examined by immunoblot analyses using corresponding antibodies as indicated. Tubulin served as an internal control for each blot. The level of p-Akt or p-Erk was calculated by normalizing to total Akt or Erk protein and the other proteins by normalizing to tubulin. The protein level in the WT cells or WT cells treated with siNS set as 1. The KD efficiency of lnc-FANCI-2 RNA (C, top bar graphs) was examined by RT-qPCR. (D) The effect of blocking RAS signaling on the expression of MCAM and VIM using PI3K inhibitor LY294002 (20 μM) and MEK inhibitor U0126 (10 μM). The protein at each time point was examined by immunoblot analysis with a corresponding antibody. The level of MCAM or VIM was calculated after normalizing to tubulin. The protein level in ΔPr-A9 cells without the inhibitor was set as 100%. *, nonspecific protein band. (E) The time dependent cell viability of the WT CaSki and ΔPr-A9 cells in the presence of 20 μM LY294002 or 10 μM U0126. Data was obtained in each time point

after normalizing to the cells treated with DMSO. The mean ± SD at each data point was calculated from 6 samples combined from two independent experiments.

Fig. 6. Negative correlation of lnc-FANCI-2 and MCAM predicts cervical cancer patient prognosis. (A) Validation of differential expression of MCAM RNA in Δpr-A9 cells by RT-PCR in the presence (+) or absence (-) of reverse transcriptase (RT). One pair of primers with one primer at the exons 11 and the other at exon 13 of MCAM RNA (NM_006500) were used. GAPDH served as a loading control. (B) The major isofrom lnc-FANCI-2a-PA2 (GenBank, MT669800.1) repressed the expression of MCAM. Δpr-A9 cells were transfected with lnc-FANCI-2a-PA2 cDNA plasmid. lnc-FANCI-2 RNA in red was detected by RNAscope ISH and MCAM protein in green was detected by IF with an anti-MCAM antibody. (C) Expression levels of lnc-FANCI-2 RNA and MCAM protein in three neighboring cells were measured by signal intensity of a line crossing over the stained cells (B, white line arrow). (D) Calculation of the expression levels of MCAM in lnc-FANCI-2 positive cells (n=35) and lnc-FANCI-2 negative cells (n=54) by fluorescent intensity. (E) Subcellular MCAM distributions in the nucleus (N) and cytoplasm (C) of WT CaSki cells and Δpr-A9 by immunoblot analysis. Fractionation efficiency and sample loading were controlled by cytoplasmic (represented by tubulin) and nuclear (represented by hnRNP C1/C2) proteins. (F). No detectable MCAM in the culture supernatant by immunoblot. (G-H) Correlation and survival analysis of lnc-FANCI-2 and MCAM expression with cervical squamous cell carcinoma (CESC) cases from the cancer genome atlas (TCGA) datasets by GEPIA web server (http://gepia.cancer-pku.cn/). The negative correlation at r=-0.321 (with p-value=1.02e-08) of Inc-FANCI-2 with MCAM in cervical cancer patients was obtained from the RNA-seq data from the TCGA for CESC tumor type downloaded from the TCGA data portal (https://portal.gdc.cancer.gov/). Only primary solid tumor samples (n=304 after exclusion

of 2 metastatic samples and 3 normal samples) were subjected to analysis, with the data showing as a Scatter plot (G). Kaplan-Meier plot with log rank test p-value at 5.74e-05 (H) shows MCAM a biomarker for poor prognosis of cervical cancer survival with a lower quartile group cutoff. RNA-seq and survival data are derived from the TCGA CESC cancer patients.

Fig. 7. lnc-FANCI-2 interacts with host factors to regulate RAS signaling pathway. (A) The lnc-FANCI-2-associated proteins in the WT CaSki cells were identified by isolation of RNA-protein complexes using RNA purification (IRPCRP)-Mass spectrometry technology. (B) lnc-FANCI-2 RNA in the IRPCRP-1 and IRPCRP-2 pulldowns had the pooled antisense biotinylated oligos (pool 1 with oligos in even numbers and pool 2 with oligos in odd numbers) immobilized to avidin-beads first before mixed with cell lysates, while the IRPCRP-3 and IRPCRP-4 pulldowns had the oligos pool 1 and 2 separately mixed with cell lysates first before addition to the avidin-beads for the RNA pull-downs. RT-PCR in the absence (-) or presence (+) of reverse transcriptase (RT) was carried out using the RNA isolated from the individual IRPCRP experiments using a primer pair of oHBL5 and oHBL12 (Table S5) specific for lnc-FANCI-2 RNA detection. Beads only (no oligos) IRPCRP experiments served as a negative control. Total RNA from the WT CaSki cells after sonication was used as an input control. Arrow indicates the detected lnc-FANCI-2 RNA. (C) Proteins associated with lnc-FANCI-2 RNA identified from lnc-FANCI-2 IRPCRP pulldowns. A Venn diagram circle in orange shows the proteins pulled down specifically from lnc-FANCI-2 IRPCRP reactions 1-4 (PSM>=2 from two separate reactions), and in blue the proteins non-specifically pulled down by beads only (oligo-free) in the IRPCRP reactions (PSM >=1). Top 10 proteins binding lnc-FANCI-2 are shown on the left in the order by the number of identified PSM. (D) Expression of p-Akt and p-Erk from CaSki WT cells 48 h after siRNA-mediated MAP4K4 knockdown (KD) was immunoblotted by the

corresponding antibodies. GAPDH served as a sample loading control. Fold change of the indicated proteins in the cells with MAP4K4 KD over the cells treated by non-targeting siRNA (siRNA-NS) was calculated after normalizing to GAPDH. (E) A proposed model illustrates how lnc-FANCI-2-protein complexes regulate the RAS signaling pathway to control Akt/ErK phosphorylation and expression of host genes.

Fig. 1

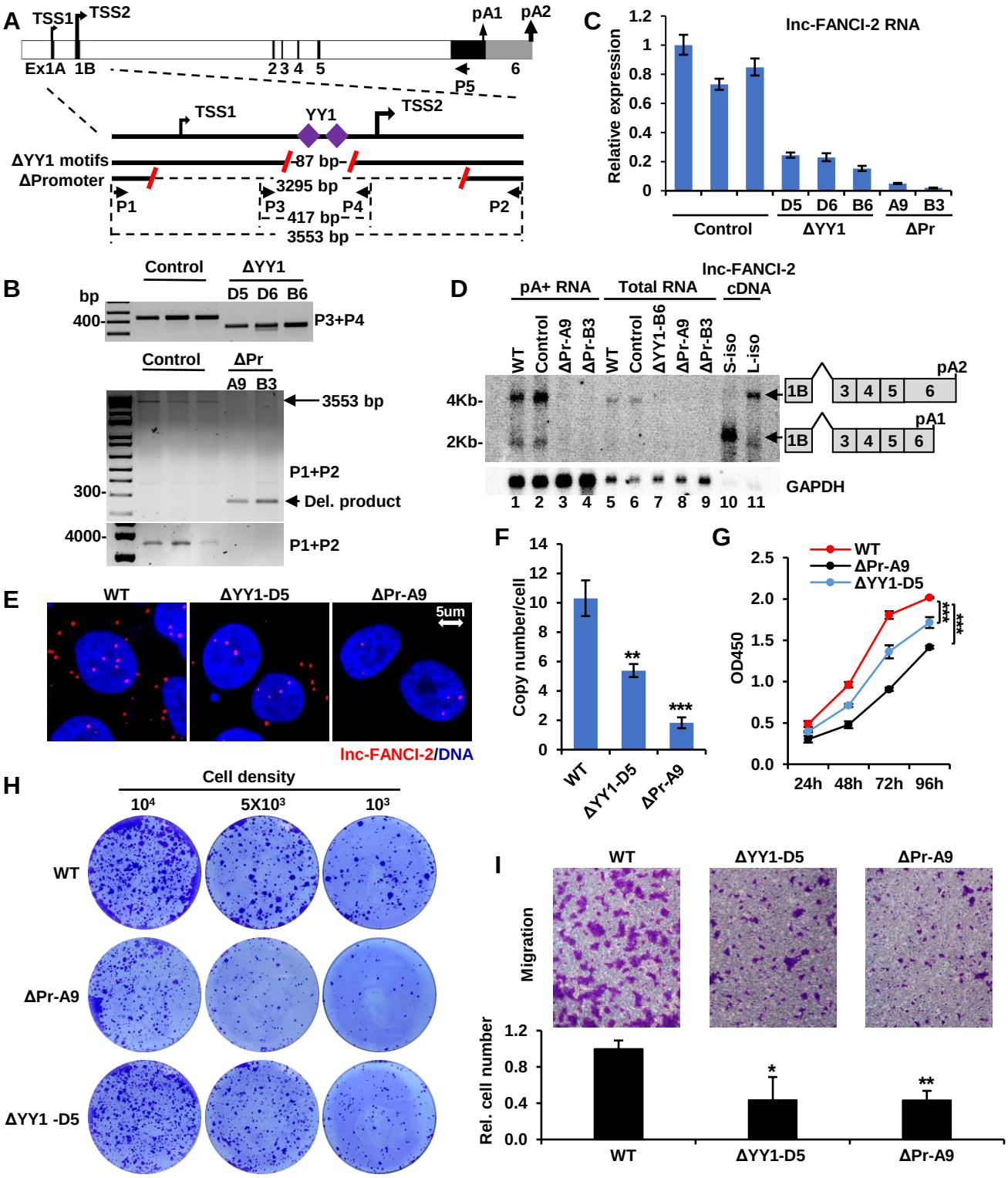

A
TSS1
TSS2
pA1
pA2
Ex1A 1B
2 3 4 5
6
P5
TSS1
YY1
TSS2
ΔYY1 motifs
87 bp
ΔPromoter
3295 bp
P1
P3
P4
P2
417 bp
3553 bp
B
Control
ΔYY1
bp
D5 D6 B6
400-
P3+P4
Control
ΔPr
A9 B3
3553 bp
P1+P2
300-
Del. product
4000-
P1+P2
C
lnc-FANCI-2 RNA
Relative expression
1.2
1
0.8
0.6
0.4
0.2
0
D5 D6 B6 A9 B3
Control
ΔYY1
ΔPr
D
pA+ RNA
Total RNA
lnc-FANCI-2 cDNA
WT
Control
ΔPr-A9
ΔPr-B3
WT
Control
ΔYY1-B6
ΔPr-A9
ΔPr-B3
S-iso
L-iso
pA2
4Kb-
1B 3 4 5 6
pA1
2Kb-
1B 3 4 5 6
GAPDH
1 2 3 4 5 6 7 8 9 10 11
E
WT
ΔYY1-D5
ΔPr-A9
5um
lnc-FANCI-2/DNA
F
Copy number/cell
14
12
10
8
6
4
2
0
**
***
WT
ΔYY1-D5
ΔPr-A9
G
WT
ΔPr-A9
ΔYY1-D5
OD450
2.5
2.0
1.5
1.0
0.5
0.0
***
***
24h 48h 72h 96h
H
Cell density
10⁴
5X10³
10³
WT
ΔPr-A9
ΔYY1 -D5
I
WT
ΔYY1-D5
ΔPr-A9
Migration
Rel. cell number
1.2
0.8
0.4
0.0
*
**
WT
ΔYY1-D5
ΔPr-A9

Fig. 2

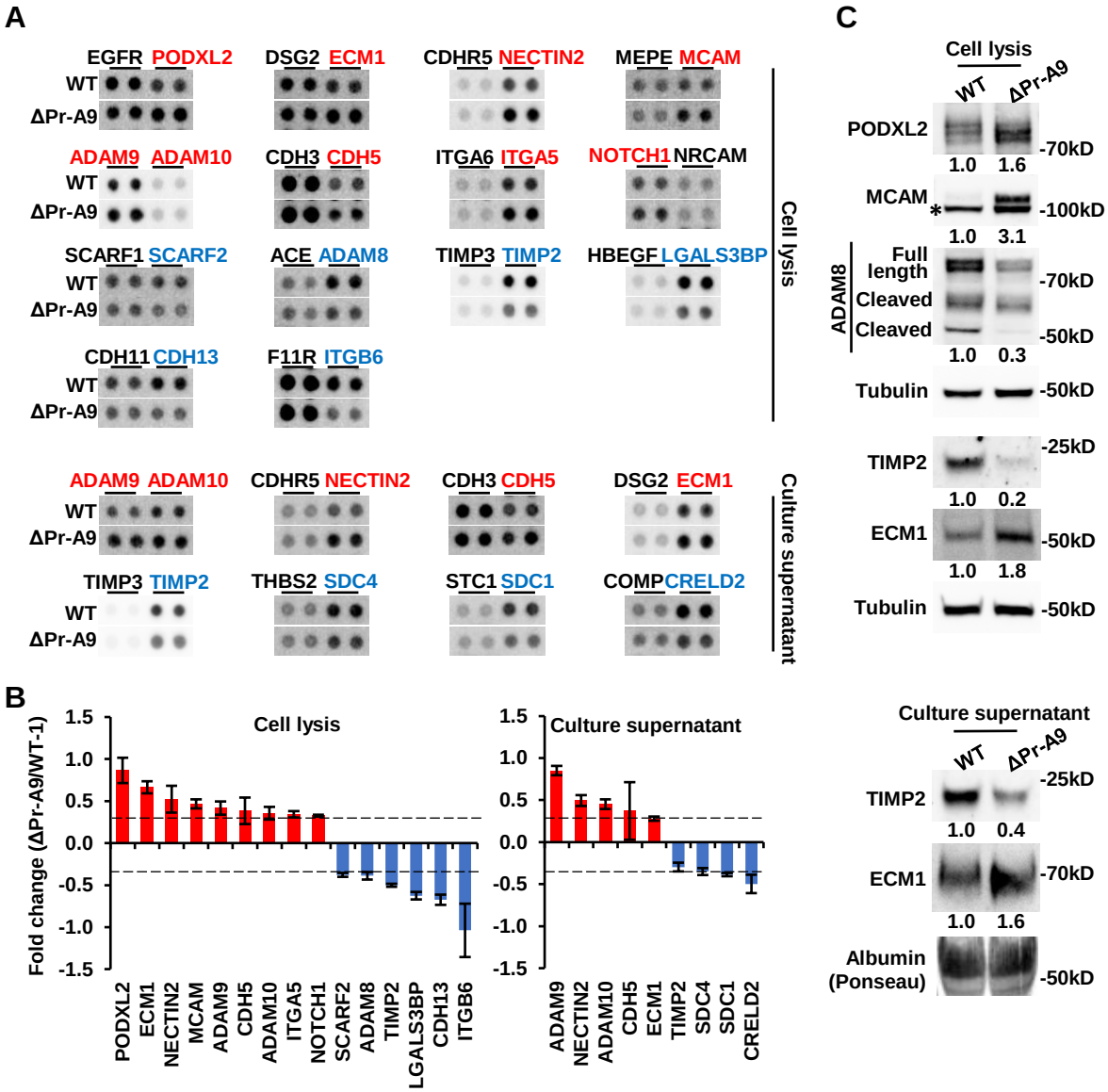
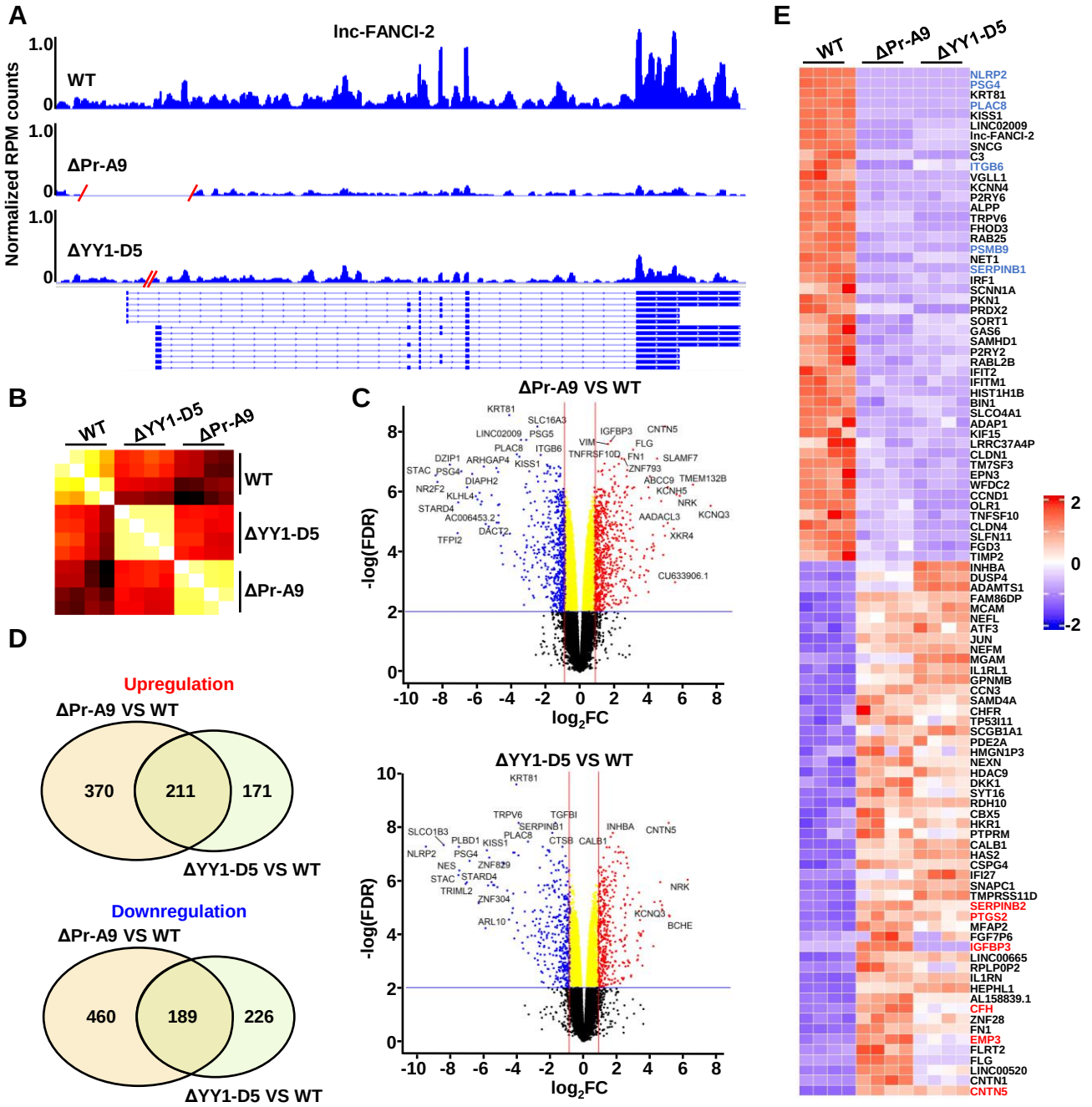

A
EGFR PODXL2
DSG2 ECM1
CDHR5 NECTIN2
MEPE MCAM
ADAM9 ADAM10
CDH3 CDH5
ITGA6 ITGA5
NOTCH1 NRCAM
SCARF1 SCARF2
ACE ADAM8
TIMP3 TIMP2
HBEGF LGALS3BP
CDH11 CDH13
F11R ITGB6
WT
ΔPr-A9
Cell lysis
ADAM9 ADAM10
CDHR5 NECTIN2
CDH3 CDH5
DSG2 ECM1
TIMP3 TIMP2
THBS2 SDC4
STC1 SDC1
COMP CRELD2
Culture supernatant
B
Fold change (ΔPr-A9/WT-1)
Cell lysis
Culture supernatant
1.5
1.0
0.5
0.0
-0.5
-1.0
-1.5
PODXL2 ECM1 NECTIN2 MCAM ADAM9 CDH5 ADAM10 ITGA5 NOTCH1 SCARF2 ADAM8 TIMP2 LGALS3BP CDH13 ITGB6
ADAM9 NECTIN2 ADAM10 CDH5 ECM1 TIMP2 SDC4 SDC1 CRELD2
C
Cell lysis
WT
ΔPr-A9
PODXL2
1.0 1.6
-70kD
MCAM
*
1.0 3.1
-100kD
ADAM8
Full length
Cleaved
Cleaved
-70kD
-50kD
1.0 0.3
Tubulin
-50kD
TIMP2
-25kD
1.0 0.2
ECM1
-50kD
1.0 1.8
Tubulin
-50kD
Culture supernatant
WT
ΔPr-A9
TIMP2
-25kD
1.0 0.4
ECM1
-70kD
1.0 1.6
Albumin (Ponseau)
-50kD

Fig. 3


A
lnc-FANCI-2
Normalized RPM counts
WT
ΔPr-A9
ΔYY1-D5
B
WT
ΔYY1-D5
ΔPr-A9
C
ΔPr-A9 VS WT
-log(FDR)
log2FC
ΔYY1-D5 VS WT
D
Upregulation
ΔPr-A9 VS WT
370
211
171
ΔYY1-D5 VS WT
Downregulation
ΔPr-A9 VS WT
460
189
226
ΔYY1-D5 VS WT
E
WT
ΔPr-A9
ΔYY1-D5

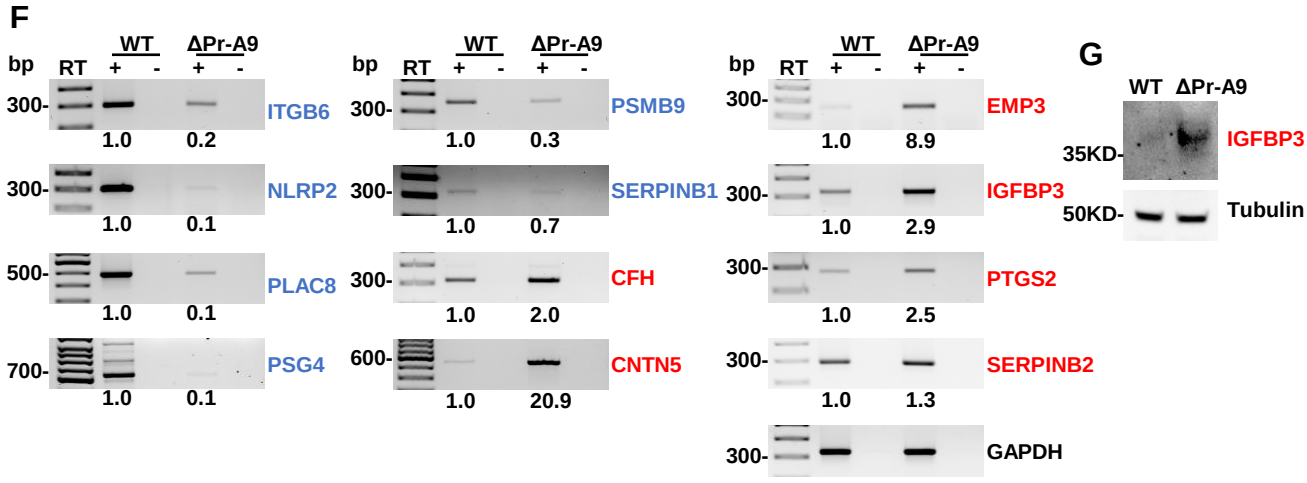

F
bp
RT
WT
ΔPr-A9
+
-
ITGB6
1.0
0.2
NLRP2
1.0
0.1
PLAC8
1.0
0.1
PSG4
1.0
0.1
PSMB9
1.0
0.3
SERPINB1
1.0
0.7
CFH
1.0
2.0
CNTN5
1.0
20.9
EMP3
1.0
8.9
IGFBP3
1.0
2.9
PTGS2
1.0
2.5
SERPINB2
1.0
1.3
GAPDH
G
WT
ΔPr-A9
35KD-
IGFBP3
50KD-
Tubulin

Fig. 4

A

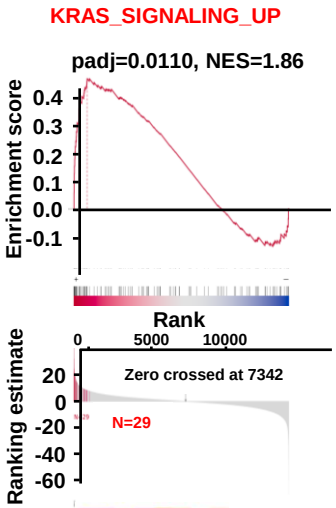
ΔPr-A9 VS WT

pathway	NES
KRAS_SIGNALING_UP	1.858008
EPITHELIAL_MESENCHYMAL_TRANSITION	1.771138
INFLAMMATORY_RESPONSE	1.490227
INTERFERON_GAMMA_RESPONSE	-1.47303
ADIPOGENESIS	-1.50073
INTERFERON_ALPHA_RESPONSE	-1.62585
BILE_ACID_METABOLISM	-1.64176

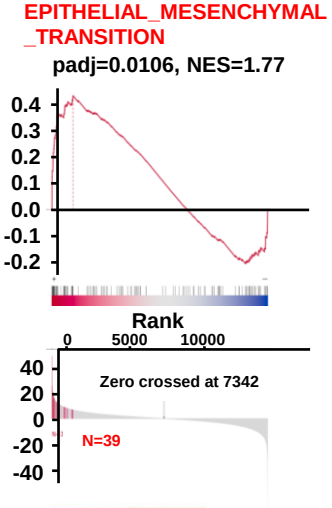
ΔYY1-D5 VS WT

pathway	NES
TNFA_SIGNALING_VIA_NFKB	1.911117
KRAS_SIGNALING_UP	1.746758
EPITHELIAL_MESENCHYMAL_TRANSITION	1.711204
ESTROGEN_RESPONSE_LATE	-1.63355
INTERFERON_GAMMA_RESPONSE	-1.69812
ESTROGEN_RESPONSE_EARLY	-1.71013
INTERFERON_ALPHA_RESPONSE	-2.22382

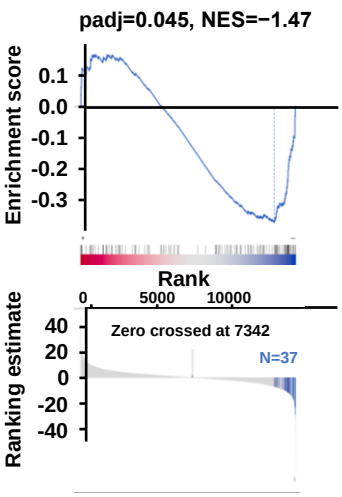
B



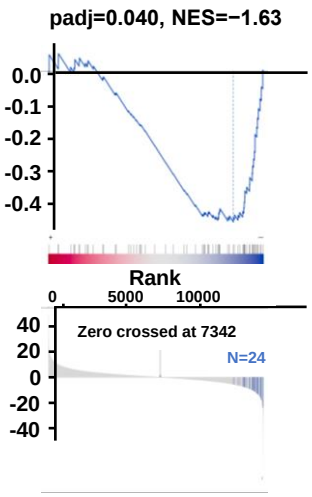
C

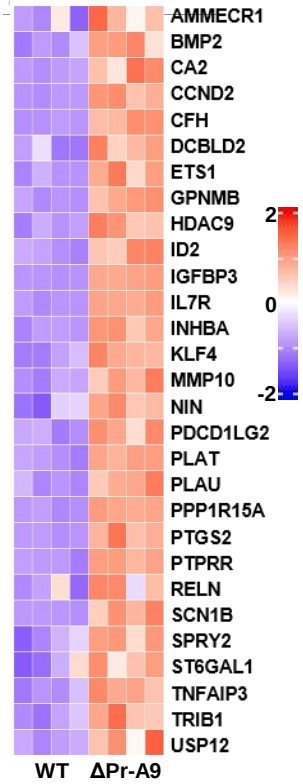


D



E





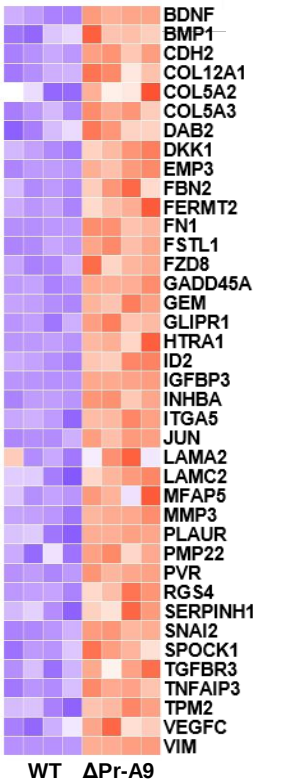


Fig. 5

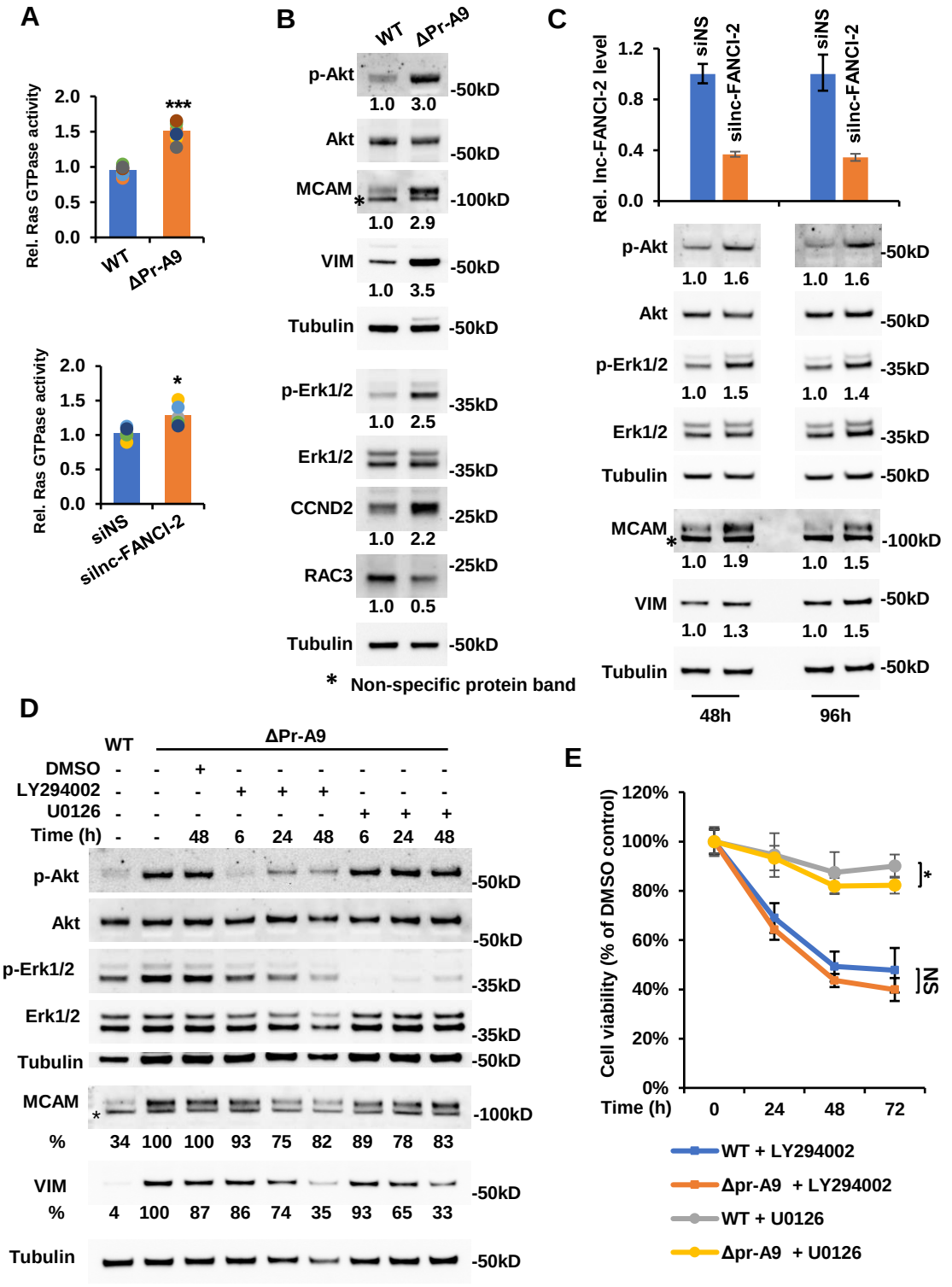

A
Rel. Ras GTPase activity
2.0
1.5
1.0
0.5
0.0
***
WT
ΔPr-A9
Rel. Ras GTPase activity
2.0
1.5
1.0
0.5
0.0
*
siNS
siInc-FANCI-2
B
WT
ΔPr-A9
p-Akt
1.0 3.0
-50kD
Akt
-50kD
MCAM
*
1.0 2.9
-100kD
VIM
1.0 3.5
-50kD
Tubulin
-50kD
p-Erk1/2
1.0 2.5
-35kD
Erk1/2
-35kD
CCND2
1.0 2.2
-25kD
RAC3
1.0 0.5
-25kD
Tubulin
-50kD
* Non-specific protein band
C
Rel. lnc-FANCI-2 level
1.2
0.8
0.4
0.0
siNS
siInc-FANCI-2
siNS
siInc-FANCI-2
p-Akt
1.0 1.6
1.0 1.6
-50kD
Akt
-50kD
p-Erk1/2
1.0 1.5
1.0 1.4
-35kD
Erk1/2
-35kD
Tubulin
-50kD
MCAM
*
1.0 1.9
1.0 1.5
-100kD
VIM
1.0 1.3
1.0 1.5
-50kD
Tubulin
-50kD
48h
96h
D
WT
ΔPr-A9
DMSO - - + - - - - - -
LY294002 - - - + + + - - -
U0126 - - - - - - + + +
Time (h) - - 48 6 24 48 6 24 48
p-Akt
-50kD
Akt
-50kD
p-Erk1/2
-35kD
Erk1/2
-35kD
Tubulin
-50kD
MCAM
*
-100kD
% 34 100 100 93 75 82 89 78 83
VIM
-50kD
% 4 100 87 86 74 35 93 65 33
Tubulin
-50kD
E
Cell viability (% of DMSO control)
120%
100%
80%
60%
40%
20%
0%
Time (h)
0
24
48
72
*
NS
WT + LY294002
Δpr-A9 + LY294002
WT + U0126
Δpr-A9 + U0126

Fig. 6

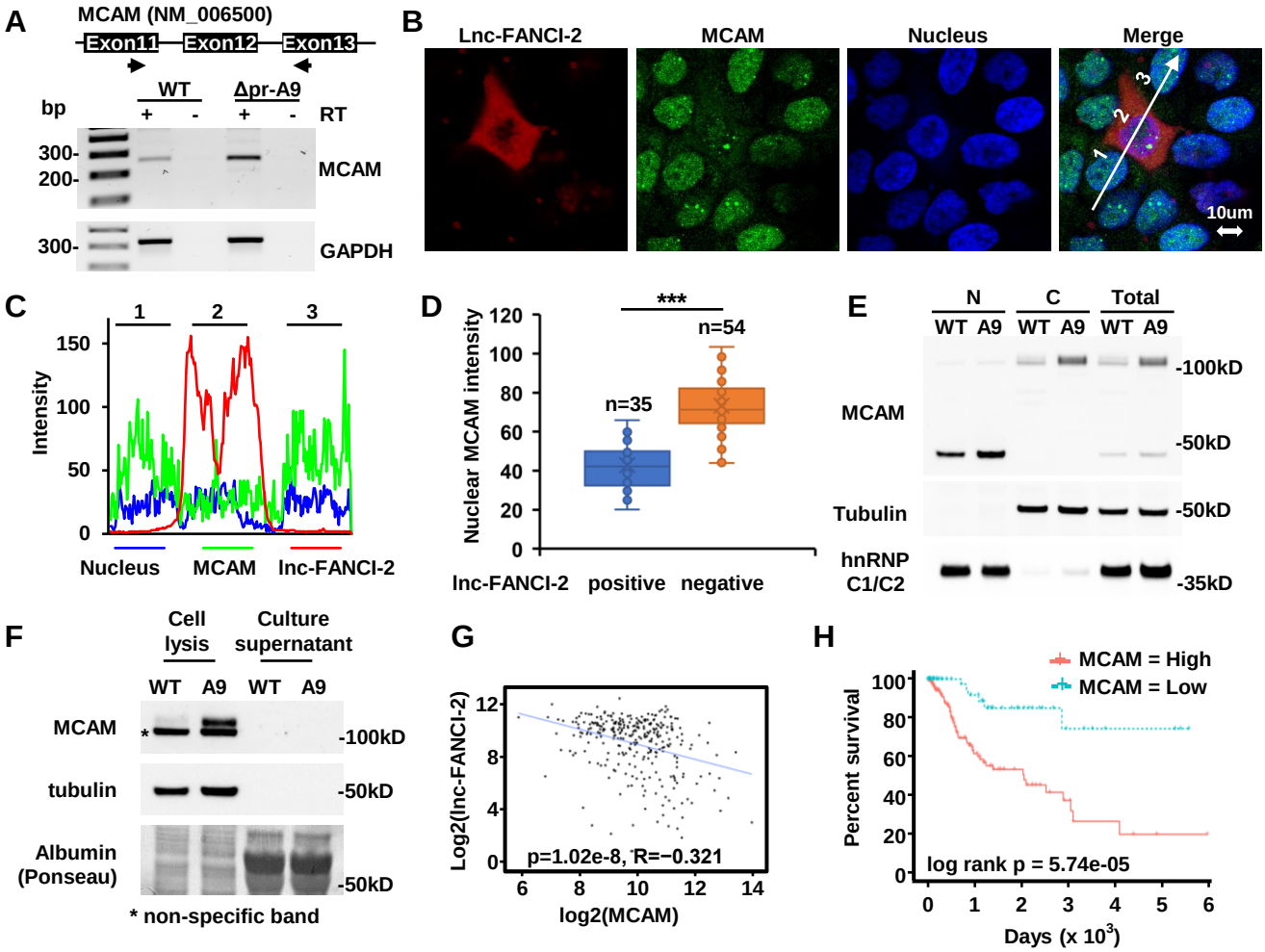
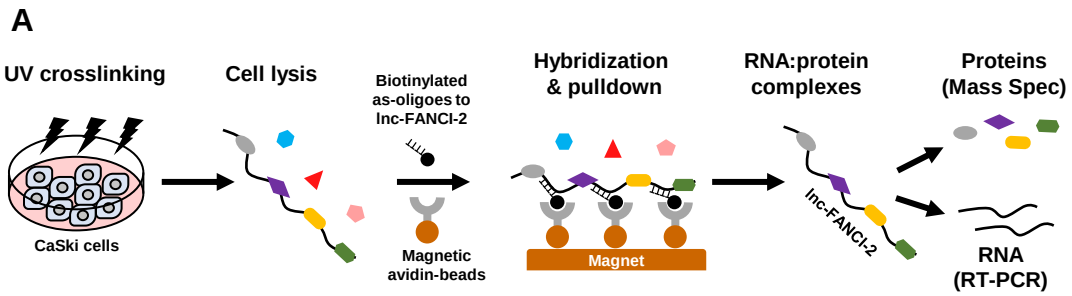
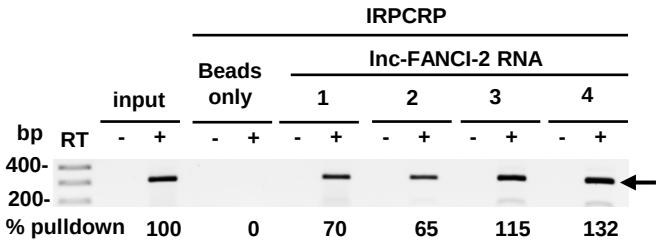
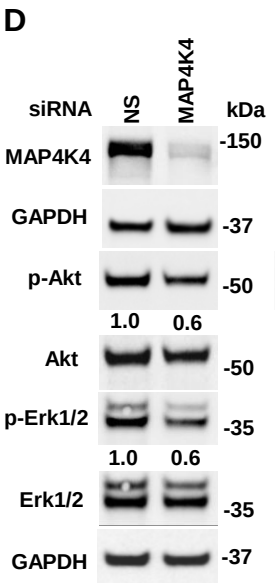
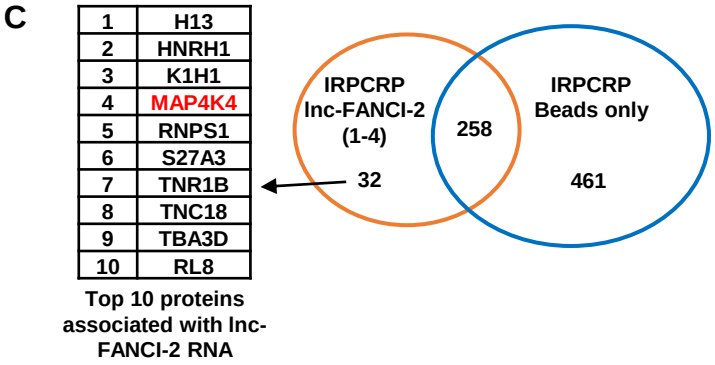

A
MCAM (NM_006500)
Exon11
Exon12
Exon13
WT
Δpr-A9
bp
+ - + - RT
300-
200-
MCAM
300-
GAPDH
B
Lnc-FANCI-2
MCAM
Nucleus
Merge
1
2
3
10um
C
1
2
3
Intensity
150
100
50
0
Nucleus
MCAM
lnc-FANCI-2
D
***
n=54
n=35
Nuclear MCAM intensity
120
100
80
60
40
20
0
lnc-FANCI-2
positive
negative
E
N
C
Total
WT A9 WT A9 WT A9
MCAM
-100kD
-50kD
Tubulin
-50kD
hnRNP C1/C2
-35kD
F
Cell lysis
Culture supernatant
WT A9 WT A9
MCAM
*
-100kD
tubulin
-50kD
Albumin (Ponseau)
-50kD
* non-specific band
G
Log2(lnc-FANCI-2)
12
8
4
0
p=1.02e-8, R=−0.321
6
8
10
12
14
log2(MCAM)
H
MCAM = High
MCAM = Low
Percent survival
100
80
60
40
20
0
log rank p = 5.74e-05
0
1
2
3
4
5
6
Days (x 10^3)

Fig. 7

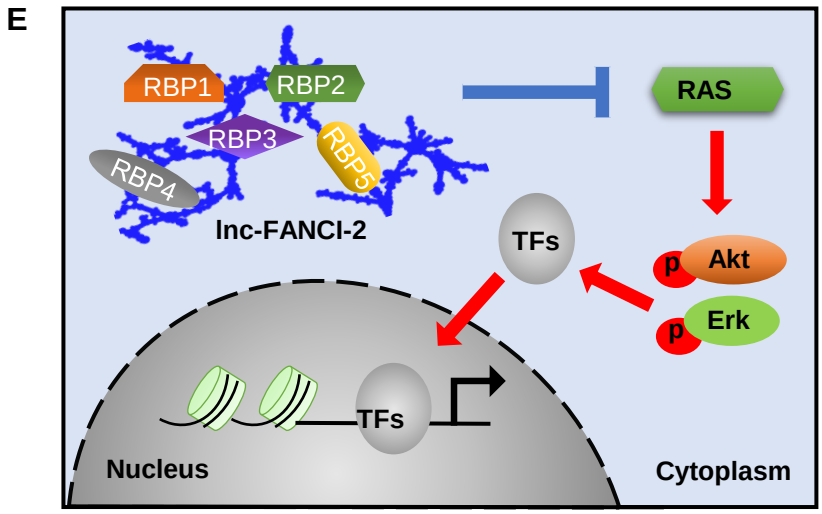








1	H13
2	HNRH1
3	K1H1
4	MAP4K4
5	RNPS1
6	S27A3
7	TNR1B
8	TNC18
9	TBA3D
10	RL8



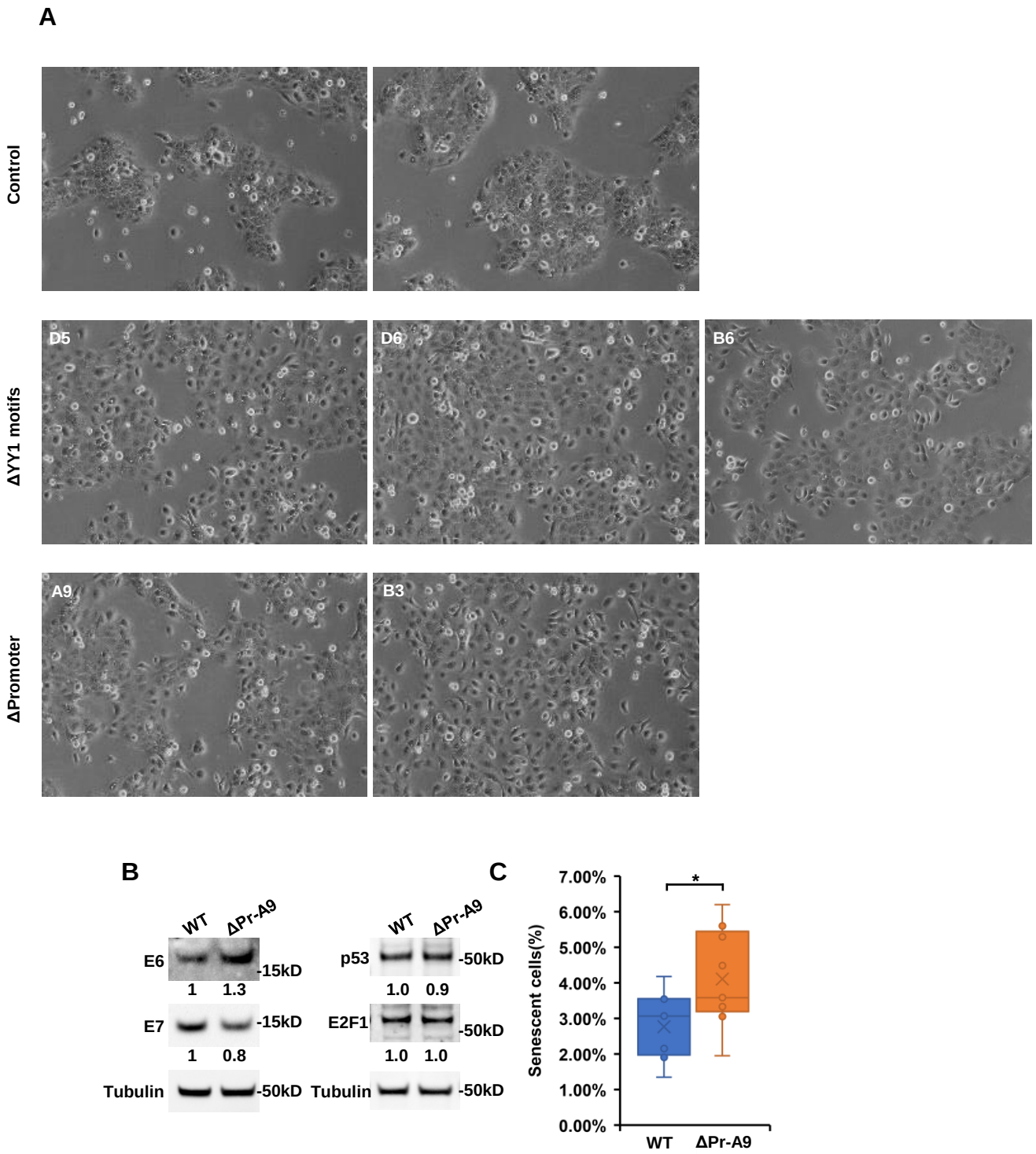


Fig. S1. Characterization of lnc-FANCI-2 KO on CaSki cell growth, viral oncoprotein expression, and cell senescence. (A) The characteristic cell growth behavior and morphology was imaged at 24 h after spreading. (B) Effect of lnc-FANCI-2 KO in CaSki cells on the expression of HPV16 E6 and E7 and their downstream targets. Total cell extracts from parental CaSki cells and ΔPr-A9 cells were examined by immunoblotting with corresponding antibodies. Tubulin served as a protein loading control. The relative protein levels of E6, E7, p53, and E2F1 were calculated after normalizing to tubulin. (C) KO of lnc-FANCI-2 expression enhances cell senescence in b-gal analysis.

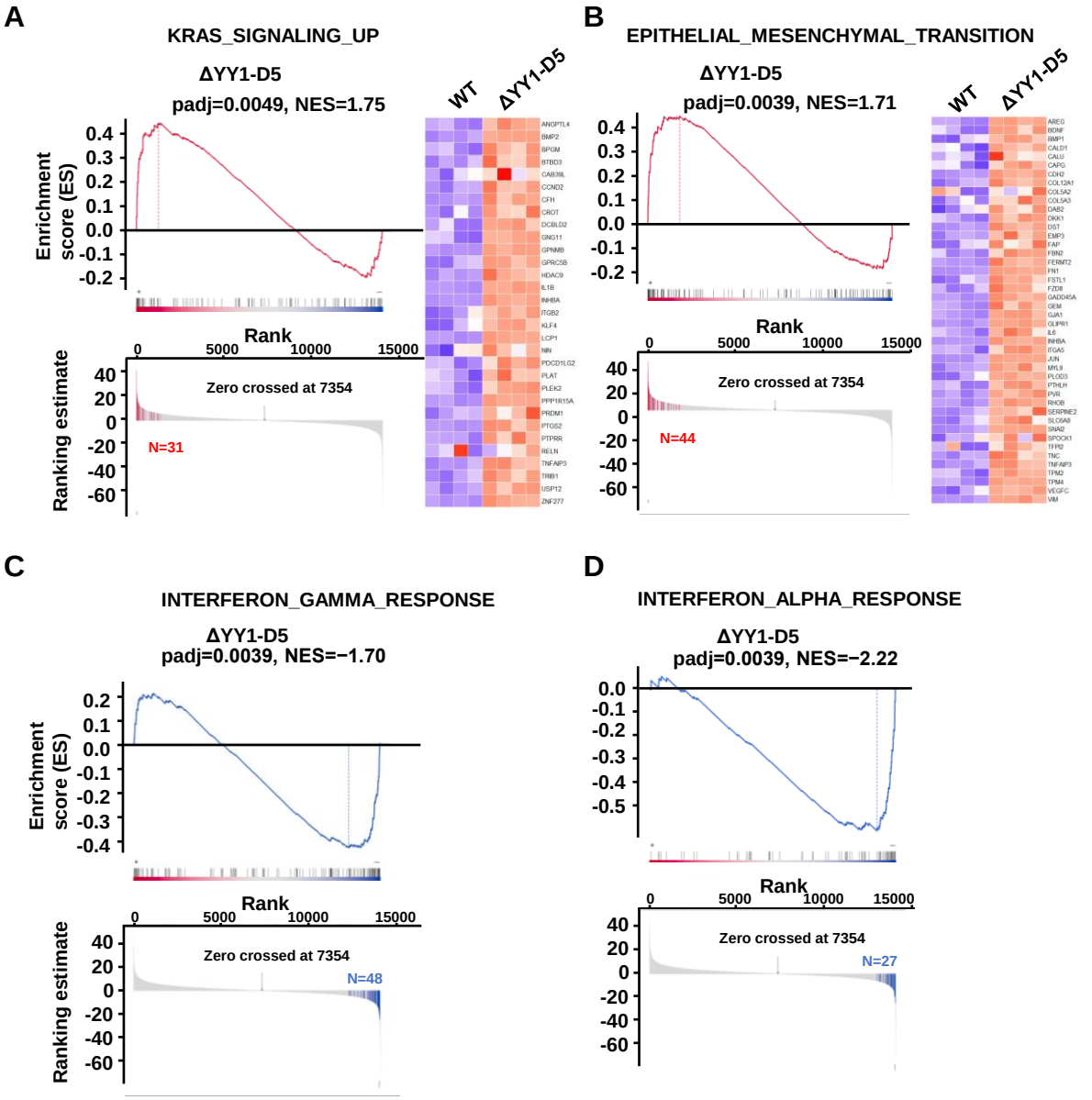


Fig S2. GSEA Enrichment plot shows enrichment score and gene hits enriched in ΔYY1-D5 cells using Hallmark gene sets, (A) KRAS_SIGNALING_UP. (B) EPITHELIAL_MESENCHYMAL_TRANSITION. (C) INTERFERON_ALPHA_RESPONSE. (D) INTERFERON_GAMMA_RESPONSE. The heatmap on the right side of Enrichment plot in each panel visualize the genes with differentiated expression enriched in each pathway in ΔYY1-D5 cells.

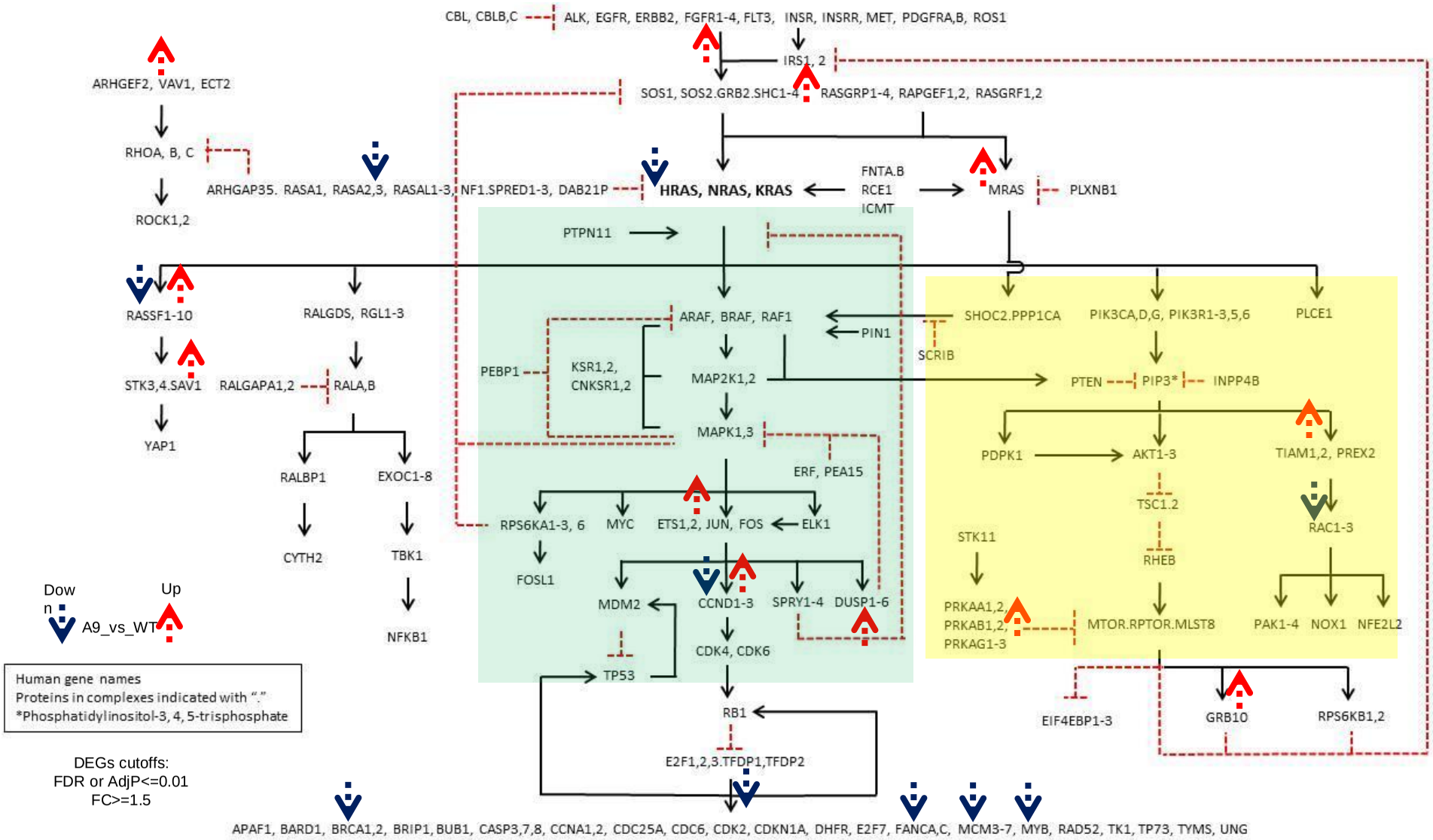

CBL, CBLB,C
ALK, EGFR, ERBB2, FGFR1-4, FLT3, INSR, INSRR, MET, PDGFRA,B, ROS1
IRS1, 2
ARHGEF2, VAV1, ECT2
SOS1, SOS2.GRB2.SHC1-4
RASGRP1-4, RAPGEF1,2, RASGRF1,2
RHOA, B, C
ARHGAP35. RASA1, RASA2,3, RASAL1-3, NF1.SPRED1-3, DAB21P
HRAS, NRAS, KRAS
FNTA.B
RCE1
ICMT
MRAS
PLXNB1
ROCK1,2
PTPN11
RASSF1-10
RALGDS, RGL1-3
ARAF, BRAF, RAF1
PIN1
SHOC2.PPP1CA
PIK3CA,D,G, PIK3R1-3,5,6
PLCE1
SCRIB
STK3,4.SAV1
RALGAPA1,2
RALA,B
PEBP1
KSR1,2,
CNKSR1,2
MAP2K1,2
PTEN
PIP3*
INPP4B
YAP1
MAPK1,3
PDPK1
AKT1-3
TIAM1,2, PREX2
RALBP1
EXOC1-8
ERF, PEA15
TSC1.2
RAC1-3
RPS6KA1-3, 6
MYC
ETS1,2, JUN, FOS
ELK1
STK11
CYTH2
TBK1
RHEB
FOSL1
MDM2
CCND1-3
SPRY1-4
DUSP1-6
PRKAA1,2,
PRKAB1,2,
PRKAG1-3
MTOR.RPTOR.MLST8
PAK1-4
NOX1
NFE2L2
NFKB1
CDK4, CDK6
TP53
EIF4EBP1-3
GRB10
RPS6KB1,2
RB1
E2F1,2,3.TFDP1,TFDP2
Down
Up
A9_vs_WT
Human gene names
Proteins in complexes indicated with "."
*Phosphatidylinositol-3, 4, 5-trisphosphate
DEGs cutoffs:
FDR or AdjP<=0.01
FC>=1.5
APAF1, BARD1, BRCA1,2, BRIP1, BUB1, CASP3,7,8, CCNA1,2, CDC25A, CDC6, CDK2, CDKN1A, DHFR, E2F7, FANCA,C, MCM3-7, MYB, RAD52, TK1, TP73, TYMS, UNG

Fig. S3. Pathway map of RAS Initiative with highlighted differentially expressed genes (DEGs) between A9 cells vs wild type cells. DEGs are mainly distributed and clustered in two main branches of RAS pathway: MAPK signaling branch and PI3K/AKT branches. DEGs were derived from RNA-seq data at cutoff of adjusted p-value <=0.01 and fold change>1.5 for both up- and down-regulated genes (indicated by red and blue arrows respectively) between A9 cells vs WT CaSki cells using three common analysis methods (DESeq2, edgeR, and limma-voom) and in-house BRB analysis pipeline from NCI. Highlighted genes in the pathway map are derived as DEGs in at least three of the analysis methods, although in most cases, these DEGs behaved consistently across all 4 methods. MAPK signaling branch is highlighted in transparent green box and PI3K/AKT branch is highlighted in transparent yellow box within the pathway map of RAS signaling pathway, which was collectively collated from community inputs that were organized and stimulated by RAS Initiative, and maintained as a common knowledge basis at URL below: https://www.cancer.gov/research/key-initiatives/ras/ras-central/blog/2015/ras-pathway-v2, which was also described in recent review (Figure 1 in Nissley and McCormick, 2022).

Reference (where the Ras pathway map has been described):

Nissley, D. V. and McCormick, F. (2022). RAS at 40: update from the RAS Initiative. *Cancer Discov.* https://doi.org/10.1158/2159-8290.CD-21-1554

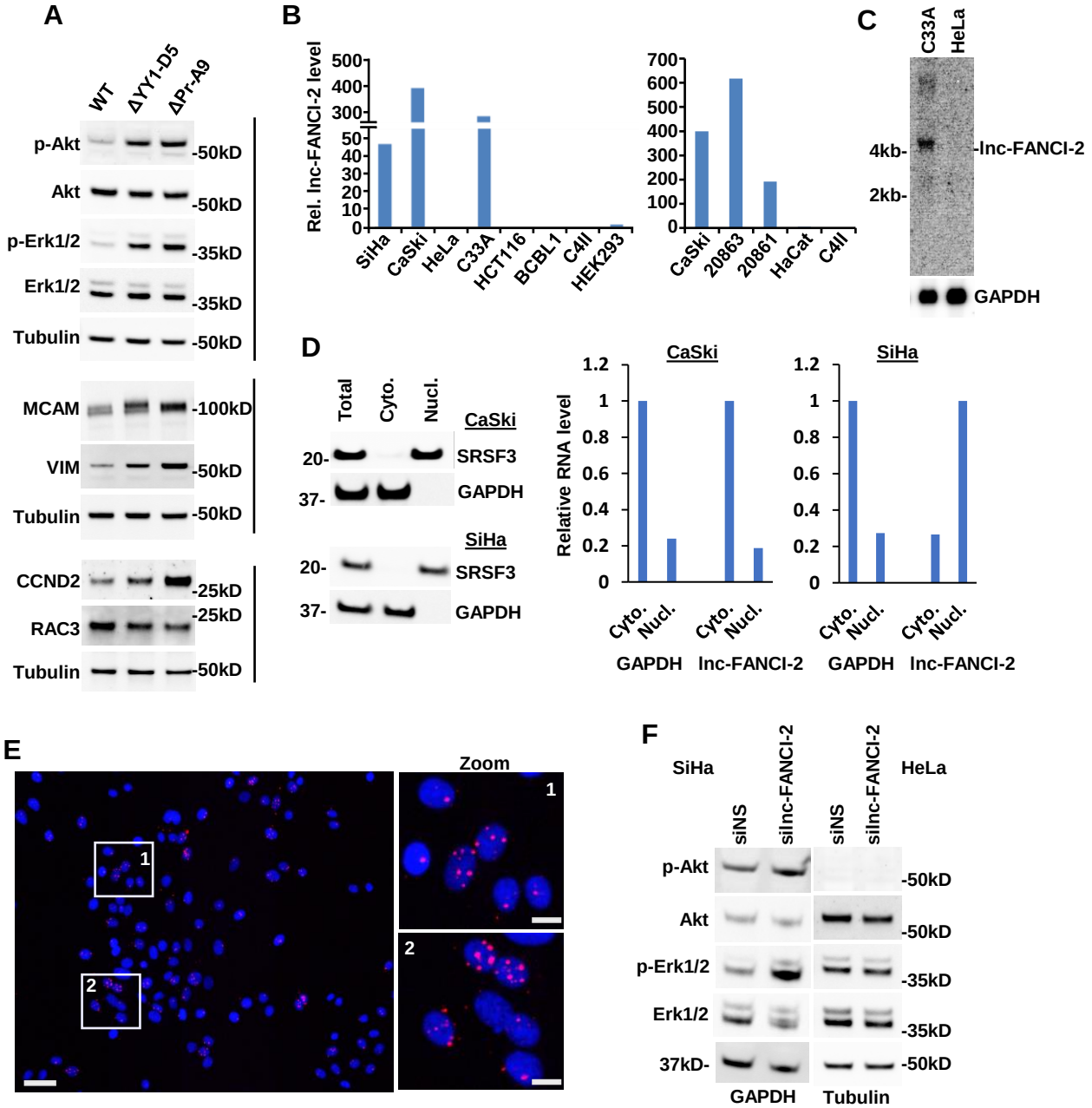


Fig S4. (A) Selective validation of increased expression of RAS signaling-related downstream genes in ΔPr-A9 and ΔYY1-D5 cells. (B) Selective expression of lnc-FANCI-2 RNA in a subset of cervical cancer cell lines as determined by RT-qPCR. Cervical cancer cell lines: SiHa (HPV16+), CaSki (HPV16+), HeLa (HPV18+), C33A (no HPV), C4II (HPV18+), and W12 subclones 20861 (HPV16+, integrated) and 20863 (HPV16+, episomal); other HPV negative cell lines: HCT116 (colorectal cancer cells), BCBL-1 (body cavity B lymphoma cells), HEK293 (Ad5 E1/E2-immortalized human kidney cells), HaCaT (spontaneously immortalized human epidermal cells). (C) HeLa cells express no lnc-FANCI-2 when compared with C33A cells as determined by Northern blot (C). (D) Subcellular determination of lnc-FANCI-2 as being predominantly cytoplasmic in CaSki but nuclear in SiHa cells. Cytoplasmic and nuclear fractionation efficiency was determined by Western blot analyses of nuclear protein SRSF3 (serine- and arginine-rich splicing factor 3) and cytoplasmic GAPDH protein. Total fractionated cytoplasmic and nuclear RNAs from CaSki and SiHa cells were quantified for lnc-FANCI-2 by RT-qPCR, with cytoplasmic GAPDH RNA serving as internal controls for RNA fractionation efficiency. (E) Subcellular lnc-FANCI-2 (red) localization in SiHa cells was examined by RNAscope RNA ISH analysis. Nuclei were stained with DAPI (blue). Scale bars: 25 μm in the left and 10 μm in the zoom. (F) KD of lnc-FANCI-2 expression in SiHa cells enhanced phosphorylation of Akt and Erk1/2 but did not in HeLa cells expressing no lnc-FANCI-2 (B, C).

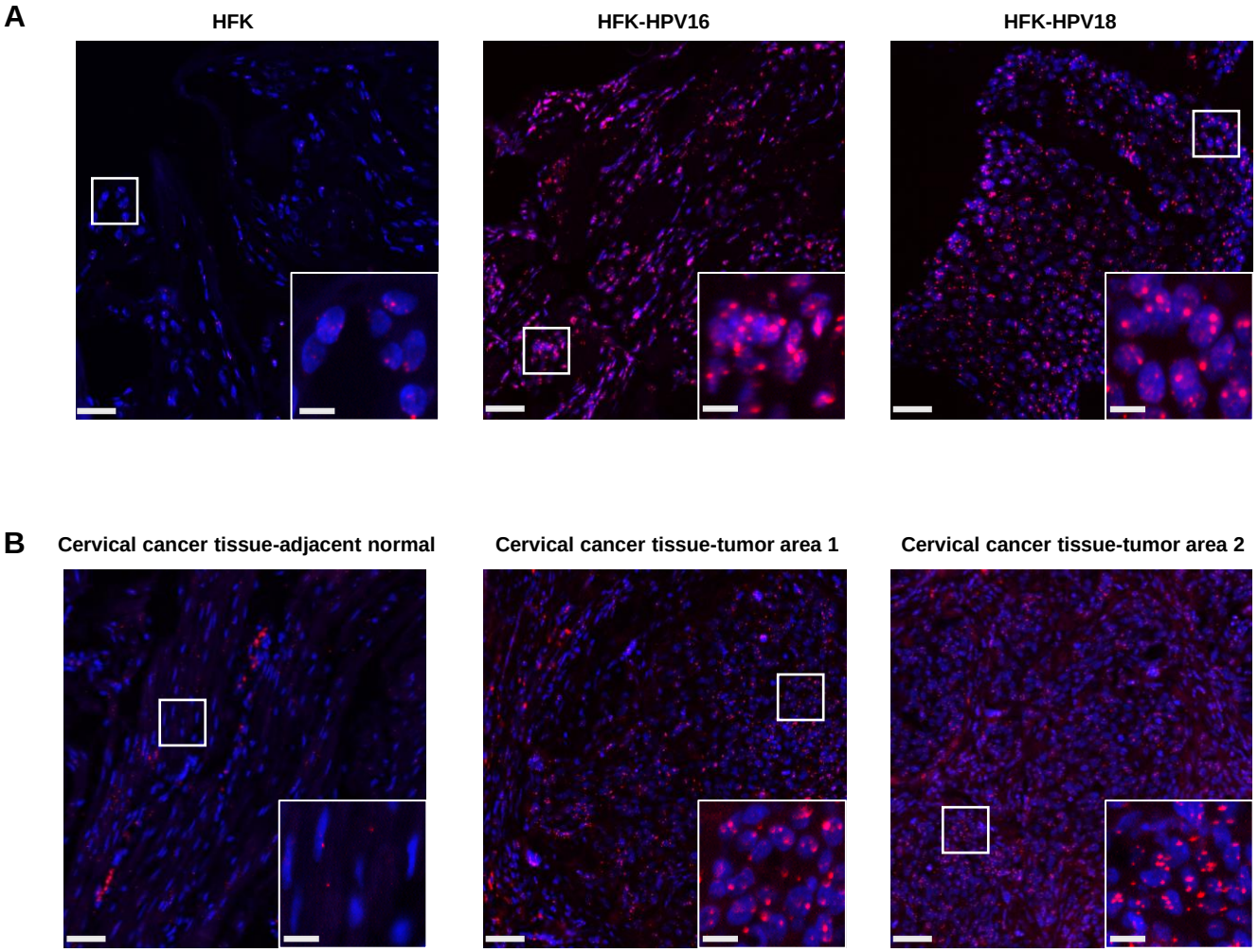


Fig. S5. Increased expression of lnc-FANCI-2 in HPV 16- and HPV18-infected raft culture tissues and cervical cancer tissues. (A) HFK-derived raft cultures without (HFK) or with HPV16 or HPV18 infections (HFK-HPV16 or HFK-HPV18) was examined at day 10 for lnc-FANCI-2 RNA by RNAscope RNA-ISH analysis. (B) lnc-FANCI-2 expression in HPV+ cervical cancer tissues was examined by RNAscope RNA-ISH analysis. Nuclei were stained with DAPI (blue). Scale bars: 25 μm in the original figures and 10 μm in the zoomed insets.

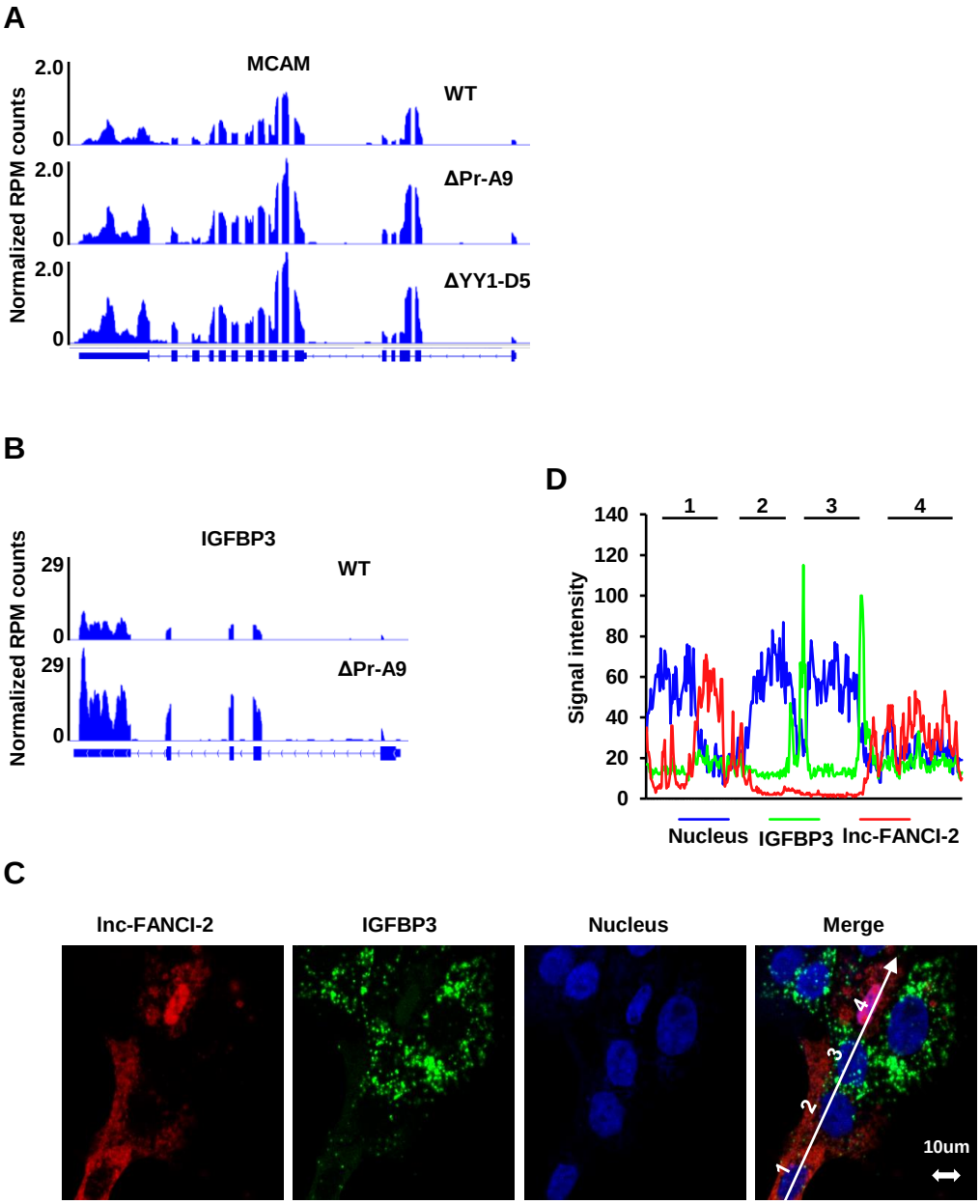


Fig. S6. MCAM expression and lnc-FANCI-2. (A-B) RNA-seq reads-coverage of MCAM and IGFBP3 by IGV illustrates the increased expression of MCAM and IGFBP3 in lnc-FANCI-2 KO cells ΔPr-A9 and/or ΔYY1-D5 when compared to the parental WT CaSki cells. (C-D) Transient rescue expression of lnc-FANCI2 in Δpr-A9 cells inhibits expression of IGFBP3. Δpr-A9 cells were transfected with a lnc-FANCI-2a-PA2 cDNA plasmid. lnc-FANCI-2 in red and IGFBP3 RNA in green were detected by RNA-ISH at 24h post transfection using each specific antisense RNA probe and imaged by confocal microscopy (C). Expression levels of lnc-FANCI-2 RNA and IGFBP3 RNA in four neighboring cells (D) were measured by signal intensity of a line crossing over the stained cells in C (white line arrow).

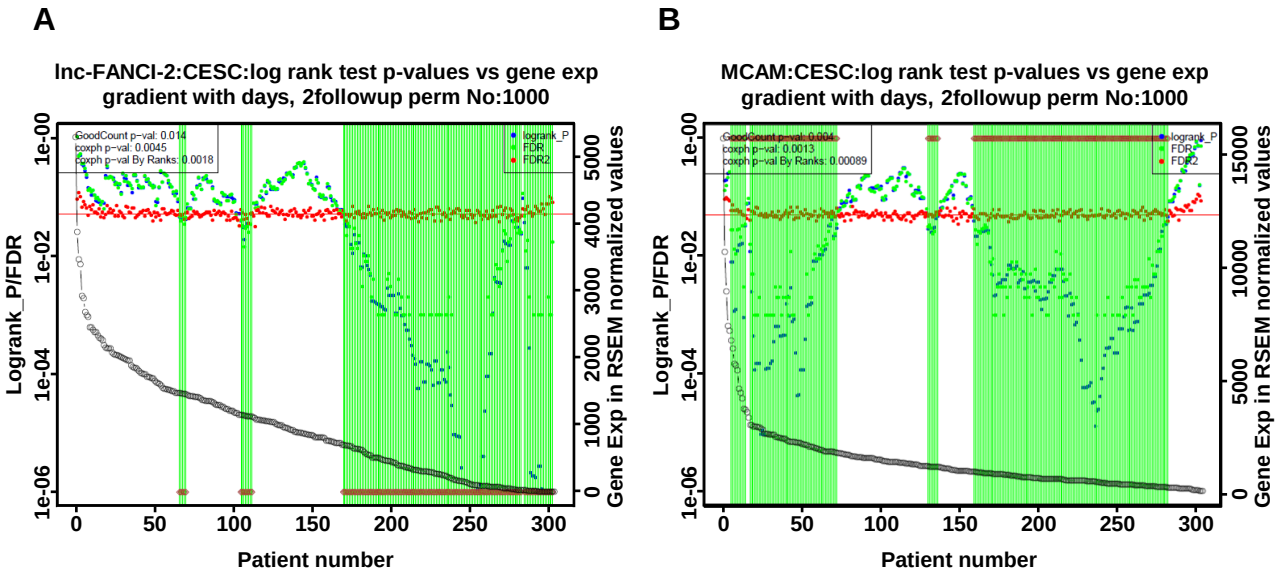


Fig. S7. Expression levels of lnc-FANCI-2 and MCAM are significantly associated with TCGA CESC cancer patients' survival outcome. The expression of lnc-FANCI-2 (A) and of MCAM (B) along with survival of the same group of 304 cervical cancer patients were analyzed by our in-house GradientScanSurv pipeline as described (Yi et al 2018, PLoS One 13(12):e0207590). The **open grey circles** in each plot are ranked expression levels from high (left) to low (right) of lnc-FANCI-2 (A) and MCAM (B) from all patient samples as indicated by y-axis at the right side of the plot (in RSEM normalized value of RNAseq data). At each breaking point denoted by patient number (x-axis), the log rank test (y-axis at the left side of the plot) was performed between the higher expression group at the left side of the cut-point vs the lower expression group at the right side of the cut-point and the indicated log rank p-values (**blue dots**) are shown as at each cut-point at x-axis. The **horizontal red line** is the p=0.05 cutoff line for the log rank test p-values. If a log rank p-value (**blue dot**) at a cut-point is below the cutoff (p<0.05), there will be a **vertical green line** shown at the corresponding cut-point, indicative of a significant log rank test p-value, and also a **brown** diamond plotted either at the bottom part of the **vertical green line** (A) if higher expression of lnc-FANCI-2 led to less severe outcome (dying slower than the lower expressors), or at the top part of the **vertical green line** (B) if higher expression of MCAM led to more severe outcome (dying faster than the lower expressors).

The original GoodCount was defined as the number of significant log rank tests (the default significance level was set as 0.05) across all possible cut-points for the original dataset. We employed the bootstrap approach to permutate the data for 1000 times in this case, and for each time, a similar procedure was performed like the original dataset, and a permutated GoodCount was obtained for each permutated dataset. The GoodCount p-value was derived as the proportion of the times that the GoodCounts of permutated datasets were no less than that of the original dataset (see Yi et al 2018, PLoS One 13(12):e0207590) and shown at the left top corner of each plot, 0.014 for lnc-FANCI-2 in panel A and 0.004 for MCAM in panel B. A significant GoodCount p-value (**blue dots** in both panels A and B) is indicative of significant association of the expression levels with the corresponding survival outcomes of the analyzed patient samples. Derived p-values from univariant expression-based cox regression models (coxph p-value: derived by original gene expression values; or coxph p-value By Rank: derived by corresponding ranks of gene expression values) were shown at the top left corner of each plot. The cut-point-specific FDR (**green dots**) and FDR2 (**red dots**) were shown at the top right corner of each plot. FDR for each cut-point was defined as the portion of cases of permutated datasets that have log rank test p-values no higher than that of original dataset. FDR2 for each cut-point was defined as the portion of cases of permutated datasets that have log rank test p-values no higher than default setting of significance as 0.05.