Supplementary Methods

DNA isolation

DNA for 4 of the 8 samples used was identical to that sent to TCGA, while the remaining 4 samples consisted of DNA newly extracted from a section of tumor adjacent to that used for TCGA extraction.

HPV16-human hybrid capture, library preparation, and PacBio long-read sequencing

We subjected HPV-enriched tumor DNA from 8 samples to PacBio long-read sequencing. All 8 samples are represented in TCGA CESC cohort. Of note, TCGA Illumina data suggested integrations involving both HPV16 and HPV70 in 1 sample (TCGA-C5-A2LT). PacBio data analysis of TCGA-C5-A2LT confirmed 1 site involving HPV70 in the *E2* gene region, which shares ~70% homology to HPV16. The 8 samples were spread across 4 sequencing runs.

SMRTbell® libraries for sequencing on the PacBio Sequel® System were constructed as outlined in the PacBio Procedure & Checklist – Multiplexed Genomic DNA Target Capture Using IDT xGen® Lockdown® Probes (available at: https://www.pacb.com/documentation/procedure-checklist-multiplex-genomic-dna-target-capture-using-idt-xgen-lockdown-probes/). Briefly, gDNA (2 µg) was sheared to ~6-10 Kb fragments by g-TUBE (Covaris; USA) centrifugation. Sheared gDNA was concentrated using AMPure PB beads (Pacific Biosciences; USA), end-repaired and A-tailed, and ligated with barcoded adapters. After another AMPure bead purification, samples were PCR-amplified using the PacBio universal primer. PCR-amplified samples were size selected using the BluePippin system and pooled for hybridization with custom xGen® Lockdown® probes (Integrated DNA Technologies; USA) designed against the HPV16 genomic sequence. Hybridized DNA fragments were bead-captured and PCR-amplified, followed by DNA damage and end repair steps. Libraries were purified proceeding the primer annealing/polymerase binding steps. At time points specified in the PacBio protocol referenced above, sample/library size and concentration were assessed using the Fragment Analyzer (Agilent; USA).

Identification and prioritization of candidate IDGs

Survival association analysis. Using the maximally selected rank statistics, we first determined a cutoff point value for each candidate IDG corresponding to the most significant Logrank test statistics. We then converted the candidate IDG's expression to dichotomous variable, high vs. low, based on the cutoff point. This was done separately for each candidate IDG.

Functional testing of candidate IDGs

Western blotting. Cells (3.6 x 10^6) cells were washed with 1X PBS and lysed for 10 min on ice using Cell Extraction Buffer (Life Technologies) supplemented with 1 mM PMSF (Sigma; St. Louis, MO) and a Protease Inhibitor Cocktail (Sigma). Undigested cellular debris was pelleted by centrifugation at 10,000 RPM and protein was quantified by Bradford Assay (Sigma). Cell lysate (20 µg) was run on a Criterion Tris-HCl polyacrylamide gel (Bio-Rad), transferred to PVDF membranes, blocked with 1X TBS + 10% nonfat dry milk at room temperature for 1 h, and incubated with primary antibody overnight at 4°C. The following day, membranes were washed, incubated for 1 h at room temperature with HRP-conjugated goat anti-rabbit or mouse anti-rabbit IgG (Cell Signaling Technology; Danvers, MA), and developed using Pierce™ ECL western blotting substrate (Thermo Scientific).

	PacBio Subjects	MCW-ICC cohort for survival analysis
Total	8	145
Race		
White	6	112
Black	2	30
Asian	0	3
Age		
Mean age (y) ± SD	41.8 ± 9.9	47.8 ± 15.0
Stage		
I	7	91
II	1	24
III	0	19
IV	0	11
Histology		
Squamous cell	7	108
Adenocarcinoma	1	33
Neuroendocrine	0	2
Other/NA	0	2

Supplementary Table S1. Subject characteristics. Clinical information on 8 ICC PacBio patients and MCW-ICC patients used for survival analysis.

GeneSolution siRNA (Qiagen)		
Product Name	**Catalog Number**	**Target Sequence**
AllStars Negative Control	1027281	CAGGGTATCGACGATTACAAA
Hs_BNC1_6 (#1)	SI04281130	CCCTGCCATCGTGAATCAGTA
Hs_BNC1_7 (#2)	SI04302291	AAGCATAAGTGCACCATCGAA
Hs_RSBN1_6 (#1)	SI04229820	ATGGGCCTATAATGTGGGTAA
Hs_RSBN1_1 (#2)	SI00708519	CCCAGTGGATATTTCAATTAA
Hs_USP36_9 (#1)	SI05784954	CTGCTACGTGAAGGCAAGCAA
Hs_USP36_3 (#2)	SI00138082	CAAGAGCGTCTCGGACACCTA
Hs_TAOK3_4 (#1)	SI00115038	AAGAGCGGATCTCCAAACATA
Hs_TAOK3_5 (#2)	SI02224761	CAAACAGTATAAAGCACTCAA

qRT-PCR primers (IDT)		
Gene Name	**Primer**	**Sequence (5' to 3')**
BNC1	Forward	GCCTTGGACTGGACACTTC
	Reverse	GGTATGATGATGGATTGCTCTTC
RSBN1	Forward	ATGCCAAAGTCACCCTTCAA
	Reverse	CTTCAAAGAGAACTTCGCGG
USP36	Forward	CTGCACCAGGACCCACTTC
	Reverse	CGCTGTGTCTCCTGCTTTCT
TAOK3	Forward	GCCATTTTGGCAGAGCAGT
	Reverse	TCTTGAGCCTCATCTAGCCG
RPS18	Forward	TGTGGTGTTGAGGAAAGCA
	Reverse	CTTCAGTCGCTCCAGGTCTT

Supplementary Table S2. siRNA and qRT-PCR primer information. Sequence information for negative control and candidate IDG-specific siRNAs used in functional assays (top) and qRT-PCR primers (bottom) used for knockdown validation and survival analysis.

MCW PacBio HPVseq		
Sample	**Breakpoints**	**All Events**
A2LX	233	72
A1ML	14	5
A3HD	9	3
A0TN	2	2
A2LV	2	1
A8XH	2	1
A1M9	4	2
A2LT	1	1
Total (n=8)	**267**	**87**
Mean/Tumor	33.4	10.9
Median/Tumor	3	2

Supplementary Table S3. HPV integration breakpoints and events identified by PacBio long-read sequencing of HPV-enriched DNA from 8 ICC tumors. Summary of chimeric breakpoints and integration events identified by PacBio long-read sequencing of each ICC sample.

TCGA ID	Event #	HPV bp	HPV Gene	Human Chrom Band	Human bp	Human Gene/Locus	Gene Location	PacBio Read #	Validation	Clonality (IAF)
C5-A2LX	1	7271	URR	2p11.2	chr2:85,508,423	TCF7L1	Intron	52	TCGA (RNA)	*1.000
C5-A2LX	1	3077	E2		chr2:85,567,980	TGOLN2, RETSAT	NA	15		*0.875
C5-A2LX	2	1670	E1	2q36.3	chr2:227,065,855	LOC646736, MIR5702	NA	6	TCGA (WGS*)	*1.000
C5-A2LX	2	1844	E1		chr2:227,083,815	LOC646736, MIR5702	NA	72		*0.737
C5-A2LX	2	5211	L2		chr2:227,097,401	LOC646736, MIR5702	NA	18		*1.000
C5-A2LX	2	2679	E1		chr2:227,111,063	LOC646736, MIR5702	NA	14		*0.552
C5-A2LX	2	2249	E1		chr2:227,118,254	LOC646736, MIR5702	NA	87		*0.647
C5-A2LX	2	6137	L1		chr2:227,131,464	LOC646736, MIR5702	NA	165		*0.855
C5-A2LX	2	2056	E1		chr2:227,150,653	LOC646736, MIR5702	NA	142		*0.804
C5-A2LX	2	3883	E5		chr2:227,157,170	LOC646736, MIR5702	NA	25		*0.697
C5-A2LX	3	1950	E1	5p14.3	chr5:19,126,360	LINC02223, CDH18	NA	4	TCGA (WGS*)	
C5-A2LX	3	1130	E1		chr5:19,153,366	LINC02223, CDH18	NA	10		*0.748
C5-A2LX	4	3901	E5	8q24.3	chr8:142,417,944	PTP4A3	Intron	3	TCGA (WGS*)	
C5-A2LX	4	2946	E2		chr8:142,478,365	MROH5	Intron	2		*0.778
C5-A2LX	4	1173	E1		chr8:142,490,986	MROH5	Intron	65		*0.638
C5-A2LX	5	6424	L1	9p21.1	chr9:28,599,877	LINGO2	Intron	5	TCGA (WGS*)	
C5-A2LX	5	4631	L2		chr9:28,630,838	LINGO2	Intron	4		
C5-A2LX	5	3102	E2		chr9:28,636,652	LINGO2	Intron	14		
C5-A2LX	5	5234	L2		chr9:28,652,735	LINGO2	Intron	506		*0.781
C5-A2LX	5	482	E6		chr9:28,744,518	LINGO2	Intron	308		*0.813
C5-A2LX	6	2111	E1	12p13.31	chr12:6,401,803	CD9, PLEKHG6	NA	12	TCGA (WGS*)	*0.758
C5-A2LX	6	3463	E2		chr12:6,429,067	PLEKHG6	Intron	35		*0.789
C5-A2LX	6	477	E6		chr12:6,431,340	PLEKHG6	Intron	10		
C5-A2LX	6	6262	L1		chr12:6,434,810	PLEKHG6	Intron	9		*0.892
C5-A2LX	6	1151	E1		chr12:6,441,241	TNFRSF1A	Intron	15		*0.75; 0.76
C5-A2LX	6	6540	L1		chr12:6,441,633	TNFRSF1A	Intron	44		*0.75; 0.76
C5-A2LX	6	2721	E1		chr12:6,446,618	TNFRSF1A	Intron	8		*1.000
C5-A2LX	6	7677	URR		chr12:6,459,345	SCNN1A	Intron	125		*0.791
C5-A2LX	6	4882	L2		chr12:6,460,875	SCNN1A	Intron	72		*0.826
C5-A2LX	6	7315	URR		chr12:6,477,179	SCNN1A	Intron	13		
C5-A2LX	6	7567	URR		chr12:6,484,899	SCNN1A; LTBR	Intron	17		*0.778
C5-A2LX	6	4357	L2		chr12:6,521,227	LTBR, CD27-AS1	NA	20		
C5-A2LX	6	4829	L2		chr12:6,673,800	NOP2	Intron	19		

Supplementary Table S4. Summary of viral integrations identified by PacBio long-read sequencing of HPV-enriched ICC tumor DNA. Details of each unique, chimeric viral-human breakpoint identified by PacBio sequencing. For each breakpoint, we provide viral nucleotide and gene location, human chromosome band, human nucleotide and gene location, gene location (intron or exon), number of supporting PacBio reads, and clonality as quantified by integration allele fraction using sample matched TCGA data (IAF). *TCGA low-pass WGS used to calculate IAF. **TCGA WXS data used to calculate IAF. No value in IAF column = ViFi program did not detect viral integration within 500 bp of listed breakpoint coordinates.00

TCGA ID	Event #	HPV bp	HPV Gene	Human Chrom Band	Human bp	Human Gene/Locus	Gene Location	PacBio Read #	Validation	Clonality (IAF)
C5-A2LX	7	1308	E1	12q24.23	chr12:118,653,036	TAOK3	Intron	1	TCGA (RNA, WGS*)	*0.792
C5-A2LX	7	2624	E1		chr12:118,655,948	TAOK3	Intron	1		*0.828
C5-A2LX	8	807	E7	13q12.3	chr13:29,561,276	SLC46A3, MTUS2	NA	10	TCGA (WGS*)	*0.504
C5-A2LX	9	2615	E1	15q11.2	chr15:21,144,087	NF1P2, MIR5701	NA	208	TCGA (WGS*)	*0.763
C5-A2LX	9	7891	URR		chr15:21,147,233	LINC01193	Intron	62		
C5-A2LX	9	4166	E5α, L2		chr15:21,161,125	LINC01193	Intron	46		
C5-A2LX	10	2354	E1	17q12	chr17:37,862,099	ERBB2	Intron	4	TCGA (WGS*)	
C5-A2LX	10	5675	L1		chr17:37,886,244	MIEN1	Intron	4		
C5-A2LX	10	6430	L1		chr17:37,919,422	IKZF3	Exon	3		
C5-A2LX	10	2625	E1		chr17:37,924,795	IKZF3	Intron	420		*0.836
C5-A2LX	10	2455	E1		chr17:37,930,352	IKZF3	Intron	7		
C5-A2LX	10	7743	URR		chr17:37,935,663	IKZF3	Intron	80		
C5-A2LX	10	7367	URR		chr17:37,938,497	IKZF3	Intron	39		*1.000
C5-A2LX	10	4665	L2		chr17:37,939,932	IKZF3	Intron	31		
C5-A2LX	10	4665	L2		chr17:37,940,009	IKZF3	Intron	208		
C5-A2LX	10	3717	E2		chr17:37,944,545	IKZF3	Exon	138		*0.220
C5-A2LX	10	2290	E1		chr17:37,948,393	IKZF3	Intron	26		
C5-A2LX	10	1632	E1		chr17:37,962,079	IKZF3	Intron	107		*0.867
C5-A2LX	10	1729	E1		chr17:37,979,632	IKZF3	Intron	4551		*0.835
C5-A2LX	10	1404	E1		chr17:37,985,275	IKZF3	Intron	25		*0.273
C5-A2LX	10	5799	L1		chr17:37,991,088	IKZF3	Intron	1809		*0.820
C5-A2LX	10	4501	L2		chr17:37,991,745	IKZF3	Intron	825		*0.378
C5-A2LX	10	5421	L2		chr17:38,005,402	IKZF3	Intron	19		
C5-A2LX	10	7706	URR		chr17:38,007,730	IKZF3	Intron	7		*0.364
C5-A2LX	10	1681	E1		chr17:38,010,968	IKZF3	Intron	274		*0.833
C5-A2LX	10	4602	L2		chr17:38,045,349	ZPBP2, GSDMB	NA	27		*0.818
C5-A2LX	11	547	E7	17q21.31	chr17:42,893,220	GJC1	Intron	14	TCGA (WGS*)	*0.500
C5-A2LX	11	1918	E1		chr17:42,952,175	EFTUD2	Intron	17		*0.875
C5-A2LX	11	3159	E2		chr17:43,153,784	NMT1	Intron	10		*0.082
C5-A2LX	11	2615	E1		chr17:43,164,283	NMT1	Intron	12		*0.292
C5-A2LX	11	7058	L1		chr17:43,173,794	NMT1	Intron	59		*0.707
C5-A2LX	11	295	E6		chr17:43,182,811	NMT1	Intron	12		
C5-A2LX	11	5647	L1		chr17:43,202,273	PLCD3	Intron	69		
C5-A2LX	11	6014	L1		chr17:43,202,277	PLCD3	Intron	149		
C5-A2LX	11	6320	L1		chr17:43,206,277	PLCD3	Intron	33		*0.722
C5-A2LX	11	261	E6		chr17:43,230,898	HEXIM1, HEXIM2	NA	13		*0.828
C5-A2LX	11	7849	URR		chr17:43,412,179	MAP3K14, ARHGAP27	NA	106		*0.805

Supplementary Table S4 (continued). Summary of viral integrations identified by PacBio long-read sequencing of HPV-enriched ICC tumor DNA. Details of each unique, chimeric viral-human breakpoint identified by PacBio sequencing. For each breakpoint, we provide viral nucleotide and gene location, human chromosome band, human nucleotide and gene location, gene location (intron or exon), number of supporting PacBio reads, and clonality as quantified by integration allele fraction using sample matched TCGA data (IAF). *TCGA low-pass WGS used to calculate IAF. **TCGA WXS data used to calculate IAF. No value in IAF column = ViFi program did not detect viral integration within 500 bp of listed breakpoint coordinates.00

TCGA ID	Event #	HPV bp	HPV Gene	Human Chrom Band	Human bp	Human Gene/Locus	Gene Location	PacBio Read #	Validation	Clonality (IAF)
C5-A2LX	12	1185	E1	19p13.3	chr19:1,259,498	MIDN, CIRBP-AS1	NA	5	TCGA (RNA, WGS*)	*0.93; 0.73
C5-A2LX	12	2635	E1		chr19:1,261,480	MIDN, CIRBP-AS1	NA	5		*0.793
C5-A2LX	12	1911	E1		chr19:1,263,855	MIDN, CIRBP-AS1	NA	7		
C5-A2LX	13	1824	E1	22q13.33	chr22:50,894,323	SBF1	Intron	1	TCGA (RNA)	*0.222
C5-A2LX	14	831	E7	1p36.33	chr1:2,261,091	MORN1	Intron	6	PacBio Only	
C5-A2LX	14	7341	URR		chr1:2,297,772	MORN1	Intron	26		
C5-A2LX	14	1493	E1		chr1:2,298,615	MORN1	Intron	41		*1.000
C5-A2LX	14	1493	E1		chr1:2,298,698	MORN1	Intron	18		*1.000
C5-A2LX	15	3422	E2	1p31.2	chr1:69,303,092	DEPDC1-AS1, LINC01707	NA	16	PacBio Only	*0.750
C5-A2LX	15	1387	E1		chr1:69,303,105	DEPDC1-AS1, LINC01707	NA	12		*0.750
C5-A2LX	16	6004	L1	1q21.3	chr1:154,995,068	DCST2	Intron	5	PacBio Only	*0.667
C5-A2LX	16	2615	E1	1q21.3	chr1:155,067,370	EFNA3, EFNA1	NA	6	PacBio Only	
C5-A2LX	16	18	URR		chr1:155,084,125	EFNA3, EFNA1	NA	7		
C5-A2LX	16	6679	L1		chr1:155,084,503	EFNA3, EFNA1	NA	10		
C5-A2LX	17	4943	L2	1q25.1	chr1:173,911,868	RC3H1	Intron	13	PacBio Only	
C5-A2LX	17	2813	E2		chr1:173,921,675	RC3H1	Intron	14		*1.000
C5-A2LX	18	6709	L1	2p21	chr2:47,060,411	LINC01119	Intron	22	PacBio Only	
C5-A2LX	18	3547	E2		chr2:47,077,935	LINC01119	Intron	7		
C5-A2LX	19	218	E6	2p12	chr2:76,106,944	GCFC2, LRRTM4	NA	29	PacBio Only	
C5-A2LX	19	2044	E1		chr2:76,106,920	GCFC2, LRRTM4	NA	7		
C5-A2LX	20	2035	E1	2p12	chr2:78,323,777	LOC101927967; LOC101927948	Intron	9	PacBio Only	
C5-A2LX	21	472	E6	2q24.2	chr2:161,564,397	RBMS1, TANK	NA	13	PacBio Only	
C5-A2LX	22	3547	E2	2q33.3	chr2:205,183,112	ICOS, PARD3B	NA	4	PacBio Only	
C5-A2LX	23	1727	E1	4q12	chr4:55,051,331	GSX2, PDGFRA	NA	118	PacBio Only	
C5-A2LX	23	938	E1		chr4:55,119,291	PDGFRA	Intron	8		
C5-A2LX	23	7160	URR		chr4:55,171,192	PDGFRA, LINC02283	NA	5		
C5-A2LX	23	2776	E2		chr4:55,181,350	PDGFRA, LINC02283	NA	6		
C5-A2LX	24	2041	E1	4q13.2	chr4:68,892,428	TMPRSS11GP, TMPRSS11F	NA	3	PacBio Only	
C5-A2LX	24	4638	L2		chr4:68,901,804	TMPRSS11GP, TMPRSS11F	NA	7		
C5-A2LX	24	7077	L1		chr4:68,913,681	TMPRSS11GP, TMPRSS11F	NA	87		*0.934
C5-A2LX	24	4051	E5		chr4:68,967,947	TMPRSS11F	Intron	20		
C5-A2LX	24	214	E6		chr4:68,970,894	TMPRSS11F	Intron	8		*1.000
C5-A2LX	25	1233	E1	4q21.22	chr4:84,215,930	HPSE	Exon	35	PacBio Only	
C5-A2LX	25	6749	L1		chr4:84,215,942	HPSE	Exon	8		

Supplementary Table S4 (continued). Summary of viral integrations identified by PacBio long-read sequencing of HPV-enriched ICC tumor DNA. Details of each unique, chimeric viral-human breakpoint identified by PacBio sequencing. For each breakpoint, we provide viral nucleotide and gene location, human chromosome band, human nucleotide and gene location, gene location (intron or exon), number of supporting PacBio reads, and clonality as quantified by integration allele fraction using sample matched TCGA data (IAF). *TCGA low-pass WGS used to calculate IAF. **TCGA WXS data used to calculate IAF. No value in IAF column = ViFi program did not detect viral integration within 500 bp of listed breakpoint coordinates.00

TCGA ID	Event #	HPV bp	HPV Gene	Human Chrom Band	Human bp	Human Gene/Locus	Gene Location	PacBio Read #	Validation	Clonality (IAF)
C5-A2LX	26	7433	URR	4q35.2	chr4:187,516,526	FAT1	Intron	3	PacBio Only	
C5-A2LX	26	418	E6		chr4:187,538,209	FAT1	Exon	38		*0.185
C5-A2LX	27	7845	URR	5p14.3	chr5:21,162,007	LINC02241, GUSBP1	NA	4	PacBio Only	
C5-A2LX	28	7311	URR	5p14.1	chr5:26,464,862	LINC02211, CDH9	NA	5	PacBio Only	
C5-A2LX	29	6304	L1	5p13.2	chr5:37,843,222	GDNF-AS1	Intron	4	PacBio Only	
C5-A2LX	29	6618	L1		chr5:37,850,856	GDNF-AS1	Intron	9		*1.000
C5-A2LX	29	1173	E1		chr5:37,879,116	GDNF-AS1, LINC02110	NA	26		*0.909
C5-A2LX	29	1384	E1		chr5:37,879,119	GDNF-AS1, LINC02110	NA	54		*0.909
C5-A2LX	30	3744	E2	5q33.3	chr5:160,030,340	ATP10B	Intron	8	PacBio Only	
C5-A2LX	30	4383	L2		chr5:160,050,725	ATP10B	Intron	39		
C5-A2LX	30	3581	E2		chr5:160,058,398	ATP10B	Intron	9		
C5-A2LX	30	1201	E1		chr5:160,088,476	ATP10B	Intron	24		
C5-A2LX	30	2008	E1		chr5:160,088,556	ATP10B	Intron	66		
C5-A2LX	30	1201	E1		chr5:160,088,559	ATP10B	Intron	103		
C5-A2LX	30	5697	L1		chr5:160,146,649	ATP10B	Intron	78		
C5-A2LX	31	3227	E2	6q12	chr6:67,226,455	SLC25A51P1, LOC102723883	NA	5	PacBio Only	*0.866
C5-A2LX	31	372	E6		chr6:67,226,486	SLC25A51P1, LOC102723883	NA	9		*0.866
C5-A2LX	32	2043	E1	6q26	chr6:163,956,606	QKI	NA	8	PacBio Only	
C5-A2LX	32	4141	E5α, L2	6q26	chr6:163,956,674	QKI	NA	4		
C5-A2LX	33	711	E7	7p12.1	chr7:53,110,436	POM121L12, LINC01446	NA	15	PacBio Only	*0.867
C5-A2LX	34	6950	L1	7q11.23	chr7:72,214,534	TYW1B	Intron	9	PacBio Only	
C5-A2LX	34	3754	E2		chr7:72,231,432	TYW1B	Intron	5		
C5-A2LX	34	1865	E1		chr7:72,252,325	TYW1B	Intron	12		
C5-A2LX	34	3528	E2		chr7:72,254,760	TYW1B	Intron	34		
C5-A2LX	35	656	E7	8q22.2	chr8:100,334,082	VPS13B	Intron	4	PacBio Only	
C5-A2LX	36	5726	L1	8q24.21	chr8:128,703,146	CASC8, CASC11	NA	23	PacBio Only	
C5-A2LX	36	6305	L1		chr8:128,703,153	CASC8, CASC11	NA	4		
C5-A2LX	36	897	E1		chr8:128,888,384	PVT1	Intron	6		*1.000
C5-A2LX	37	4999	L2	9p23	chr9:9,166,889	PTPRD	Intron	7	PacBio Only	
C5-A2LX	38	737	E7	9q13	chr9:66,774,182	LOC728673, LOC101928381	NA	114	PacBio Only	
C5-A2LX	38	4109	E5α, L2		chr9:66,774,368	LOC728673, LOC101928381	NA	48		
C5-A2LX	38	843	E7		chr9:66,781,752	LOC728673, LOC101928381	NA	34		

Supplementary Table S4 (continued). Summary of viral integrations identified by PacBio long-read sequencing of HPV-enriched ICC tumor DNA. Details of each unique, chimeric viral-human breakpoint identified by PacBio sequencing. For each breakpoint, we provide viral nucleotide and gene location, human chromosome band, human nucleotide and gene location, gene location (intron or exon), number of supporting PacBio reads, and clonality as quantified by integration allele fraction using sample matched TCGA data (IAF). *TCGA low-pass WGS used to calculate IAF. **TCGA WXS data used to calculate IAF. No value in IAF column = ViFi program did not detect viral integration within 500 bp of listed breakpoint coordinates.00

TCGA ID	Event #	HPV bp	HPV Gene	Human Chrom Band	Human bp	Human Gene/Locus	Gene Location	PacBio Read #	Validation	Clonality (IAF)
C5-A2LX	39	7833	URR	11q13.2	chr11:66,753,233	C11ORF86, SYT12	NA	119	PacBio Only	
C5-A2LX	39	2105	E1		chr11:66,758,636	C11ORF86, SYT12	NA	108		
C5-A2LX	39	7457	URR		chr11:66,763,423	C11ORF86, SYT12	NA	13		
C5-A2LX	39	1313	E1		chr11:66,766,216	C11ORF86, SYT12	NA	18		
C5-A2LX	39	1417	E1		chr11:66,772,159	C11ORF86, SYT12	NA	9		
C5-A2LX	39	834	E7		chr11:66,775,634	C11ORF86, SYT12	NA	14		
C5-A2LX	39	190	E6		chr11:66,777,054	C11ORF86, SYT12	NA	7		
C5-A2LX	39	2512	E1		chr11:66,850,202	RHOD, LOC107984341	NA	13		
C5-A2LX	39	2043	E1		chr11:66,854,210	RHOD, LOC107984341	NA	17		
C5-A2LX	40	950	E1	11q22.3	chr11:103,453,269	DYNC2H1, MIR4693	NA	6	PacBio Only	
C5-A2LX	40	5589	L2		chr11:103,521,980	DYNC2H1, MIR4693	NA	19		*1.000
C5-A2LX	40	1490	E1		chr11:103,526,657	DYNC2H1, MIR4693	NA	133		*0.915
C5-A2LX	40	1869	E1		chr11:103,535,176	DYNC2H1, MIR4693	NA	97		*0.861
C5-A2LX	41	2236	E1	12q12	chr12:45,211,481	NELL2	Intron	6	TCGA (RNA)	*1.000
C5-A2LX	41	7462	URR		chr12:45,229,727	NELL2	Intron	5		
C5-A2LX	42	597	E7	12q24.22	chr12:117,420,333	FBXW8; LOC100506551	Intron	9	PacBio Only	
C5-A2LX	43	6138	L1	13q21.31	chr13:63,365,463	LINC00448	Intron	4	PacBio Only	
C5-A2LX	44	390	E6	13q21.31	chr13:64,018,257	LINC00376, LINC00395	NA	6	PacBio Only	
C5-A2LX	44	6741	L1		chr13:64,021,811	LINC00376, LINC00395	NA	5		
C5-A2LX	45	2693	E1	13q21.31	chr13:75,110,900	LINC00381, LINC00347	NA	6	PacBio Only	
C5-A2LX	46	4039	E5	13q31.3	chr13:92,408,791	GPC5	Intron	4	PacBio Only	
C5-A2LX	47	7134	URR	14q23.2	chr14:64,556,051	SYNE2	Intron	5	PacBio Only	
C5-A2LX	48	607	E7	14q24.1	chr14:68,399,880	RAD51B	Intron	9	PacBio Only	
C5-A2LX	48	2648	E1		chr14:68,419,000	RAD51B	Intron	8		
C5-A2LX	48	4183	E5α, L2		chr14:68,419,136	RAD51B	Intron	9		
C5-A2LX	48	6567	L1		chr14:68,457,449	RAD51B	Intron	11		
C5-A2LX	48	1598	E1		chr14:68,793,696	RAD51B	Intron	4		
C5-A2LX	48	2357	E1		chr14:68,793,702	RAD51B	Intron	4		
C5-A2LX	49	1597	E1	14q32.11	chr14:91,473,610	RPS6KA5	Intron	2	PacBio Only	
C5-A2LX	49	5821	L1		chr14:91,532,990	DGLUCY	Intron	177		
C5-A2LX	49	5317	L2		chr14:91,536,079	DGLUCY	Intron	94		
C5-A2LX	49	7837	URR		chr14:91,554,259	DGLUCY	Intron	52		
C5-A2LX	49	462	E6		chr14:91,566,159	DGLUCY	Intron	170		
C5-A2LX	49	866	E1		chr14:91,615,337	DGLUCY	Intron	5		

Supplementary Table S4 (continued). Summary of viral integrations identified by PacBio long-read sequencing of HPV-enriched ICC tumor DNA. Details of each unique, chimeric viral-human breakpoint identified by PacBio sequencing. For each breakpoint, we provide viral nucleotide and gene location, human chromosome band, human nucleotide and gene location, gene location (intron or exon), number of supporting PacBio reads, and clonality as quantified by integration allele fraction using sample matched TCGA data (IAF). *TCGA low-pass WGS used to calculate IAF. **TCGA WXS data used to calculate IAF. No value in IAF column = ViFi program did not detect viral integration within 500 bp of listed breakpoint coordinates.00

TCGA ID	Event #	HPV bp	HPV Gene	Human Chrom Band	Human bp	Human Gene/Locus	Gene Location	PacBio Read #	Validation	Clonality (IAF)
C5-A2LX	50	1761	E1	14q32.32	chr14:103,474,848	CDC42BPB	Exon	2	PacBio Only	
C5-A2LX	51	7472	URR	15q14	chr15:37,604,504	MEIS2, TMCO5A	NA	136	PacBio Only	
C5-A2LX	52	1978	E1	15q15.1	chr15:41,522,299	EXD1	Intron	389	PacBio Only	
C5-A2LX	53	1746	E1	15q21.2	chr15:52,104,676	TMOD2	Exon	26	PacBio Only	
C5-A2LX	53	7808	URR		chr15:52,290,083	LEO1, MAPK6-DT	NA	7		
C5-A2LX	54	3447	E2	16p13.3	chr16:2,413,917	ABCA17P	Intron	34	PacBio Only	
C5-A2LX	54	787	E7		chr16:2,429,030	ABCA17P	Intron	15		
C5-A2LX	54	4064	E5		chr16:2,443,409	ABCA17P	Intron	46		
C5-A2LX	54	307	E6		chr16:2,586,534	MIR3178, PDPK1	NA	274		
C5-A2LX	54	5531	L2		chr16:2,589,133	PDPK1	Intron	10		
C5-A2LX	54	3665	E2		chr16:2,596,518	PDPK1	Intron	23		
C5-A2LX	54	629	E7		chr16:2,604,248	PDPK1	Intron	8		
C5-A2LX	55	1730	E1	16q21	chr16:58,385,562	PRSS54, GINS3	NA	13	PacBio Only	
C5-A2LX	55	996	E1		chr16:58,385,729	PRSS54, GINS3	NA	95		
C5-A2LX	55	3581	E2		chr16:58,387,057	PRSS54, GINS3	NA	52		
C5-A2LX	55	3596	E2		chr16:58,456,442	GINS3, NDRG4	NA	9		
C5-A2LX	55	3493	E2		chr16:58,476,739	GINS3, NDRG4	NA	4		
C5-A2LX	56	1935	E1	17q22	chr17:52,698,183	KIF2B, TOM1L1	NA	5	PacBio Only	
C5-A2LX	56	4912	L2		chr17:52,721,343	KIF2B, TOM1L1	NA	5		
C5-A2LX	56	1294	E1		chr17:52,721,272	KIF2B, TOM1L1	NA	2		
C5-A2LX	57	3631	E2	17q24.2	chr17:64,556,231	PRKCA	Intron	59	PacBio Only	
C5-A2LX	57	6798	L1		chr17:64,570,371	PRKCA	Intron	176		
C5-A2LX	57	1433	E1		chr17:64,576,064	PRKCA	Intron	184		
C5-A2LX	58	47	URR	17q24.3	chr17:70,815,972	SLC39A11	Intron	10	PacBio Only	
C5-A2LX	58	6314	L1		chr17:70,832,367	SLC39A11	Intron	3		
C5-A2LX	59	566	E7	17q25.2	chr17:75,035,488	LOC105371899, SNHG20	NA	37	PacBio Only	
C5-A2LX	59	637	E7		chr17:75,054,362	LOC105371899, SNHG20	NA	12		
C5-A2LX	59	6132	L1		chr17:75,061,005	LOC105371899, SNHG20	NA	112		
C5-A2LX	59	4640	L2		chr17:75,061,767	LOC105371899, SNHG20	NA	155		
C5-A2LX	59	1489	E1		chr17:75,070,448	LOC105371899, SNHG20	NA	6		
C5-A2LX	59	480	E6		chr17:75,075,293	LOC105371899, SNHG20	NA	10		
C5-A2LX	59	7031	L1		chr17:75,102,942	SEC14L1	Intron	7		
C5-A2LX	59	6287	L1		chr17:75,113,399	SEC14L1	Intron	5		
C5-A2LX	60	675	E7	18p11.32	chr18:1,070,771	LINC01904, LINC00470	NA	7	PacBio Only	
C5-A2LX	60	1862	E1		chr18:1,070,776	LINC01904, LINC00470	NA	9		

Supplementary Table S4 (continued). Summary of viral integrations identified by PacBio long-read sequencing of HPV-enriched ICC tumor DNA. Details of each unique, chimeric viral-human breakpoint identified by PacBio sequencing. For each breakpoint, we provide viral nucleotide and gene location, human chromosome band, human nucleotide and gene location, gene location (intron or exon), number of supporting PacBio reads, and clonality as quantified by integration allele fraction using sample matched TCGA data (IAF). *TCGA low-pass WGS used to calculate IAF. **TCGA WXS data used to calculate IAF. No value in IAF column = ViFi program did not detect viral integration within 500 bp of listed breakpoint coordinates.00

TCGA ID	Event #	HPV bp	HPV Gene	Human Chrom Band	Human bp	Human Gene/Locus	Gene Location	PacBio Read #	Validation	Clonality (IAF)
C5-A2LX	61	6346	L1	18q11.2	chr18:23,664,478	SS18	Intron	3	PacBio Only	
C5-A2LX	61	1563	E1		chr18:23,721,584	PSMA8	Intron	3		
C5-A2LX	61	5780	L1		chr18:23,748,158	PSMA8	Intron	13		
C5-A2LX	61	1862	E1		chr18:23,748,391	PSMA8	Intron	20		
C5-A2LX	62	1860	E1	19p12	chr19:22,922,204	ZNF492, ZNF99	NA	43	PacBio Only	
C5-A2LX	62	3801	E2		chr19:22,922,111	ZNF492, ZNF99	NA	7		
C5-A2LX	63	1553	E1	19q13.12	chr19:37,204,836	ZNF567	Intron	4	PacBio Only	
C5-A2LX	64	6956	L1	20p13	chr20:640,969	SRXN1, SCRT2	NA	234	PacBio Only	
C5-A2LX	64	1884	E1		chr20:649,521	SCRT2	Intron	165		
C5-A2LX	65	2043	E1	20p12.2	chr20:11,443,787	LOC339593, LINC00687	NA	17	PacBio Only	
C5-A2LX	65	4524	L2		chr20:11,533,582	LOC339593, LINC00687	NA	11		
C5-A2LX	66	7778	URR	20q12	chr20:39,063,515	LINC01370, MAFB	NA	12	PacBio Only	
C5-A2LX	67	4340	L2	20q13.2	chr20:53,697,420	DOK5, LINC01441	NA	9	PacBio Only	
C5-A2LX	68	2324	E1	Xq11.2	chrX:64,439,258	ZC4H2, ZC3H12B	NA	24	PacBio Only	*0.667
C5-A2LX	69	153	E6	Xq21.32	chrX:92,285,160	MIR4454, NAP1L3	NA	4	PacBio Only	
C5-A2LX	70	5598	L2	Xq21.33	chrX:93,525,261	FAM133A, MIR548M	NA	9	PacBio Only	
C5-A2LX	70	2628	E1		chrX:93,525,234	FAM133A, MIR548M	NA	54		
C5-A2LX	70	472	E6		chrX:93,530,840	FAM133A, MIR548M	NA	4		
C5-A2LX	71	4086	E5	Xq25	chrX:126,550,794	PRR32, ACTRT1	NA	14	PacBio Only	*1.000
C5-A2LX	71	6402	L1		chrX:126,550,845	PRR32, ACTRT1	NA	7		*1.000
C5-A2LX	72	7894	URR	Xq28	chrX:151,980,117	MAGEA3, CETN2	NA	3	PacBio Only	
C5-A2LX	72	458	E6		chrX:151,981,250	MAGEA3, CETN2	NA	150		
C5-A2LX	72	2215	E1		chrX:152,002,205	NSDHL	Intron	7		
C5-A2LX	72	6800	L1		chrX:152,008,047	NSDHL	Intron	10		
C5-A2LX	72	2652	E1		chrX:152,013,670	NSDHL	Intron	9		
C5-A2LX	72	6616	L1		chrX:152,032,170	NSDHL	Intron	33		
C5-A2LX	72	936	E1		chrX:152,076,199	NSDHL, ZNF185	NA	3		
C5-A2LX	72	5947	L1		chrX:152,079,913	NSDHL, ZNF185	NA	3		
C5-A1ML	73	2175	E1	8p12	chr8:34,837,123	LINC01288, UNC5D	NA	3	TCGA (WGS)	0.069
C5-A1ML	73	7830	URR		chr8:34,837,123	LINC01288, UNC5D	NA	1		0.069
C5-A1ML	74	3081	E2	10q11.21	chr10:43,055,741	ZNF37BP, ZNF33B	NA	1	TCGA (WGS)	
C5-A1ML	74	371	E6		chr10:43,061,049	ZNF37BP, ZNF33B	NA	1		0.095; 0.046
C5-A1ML	75	4893	L2	Xq28	chrX:149,550,333	MAMLD1	Intron	152	TCGA (WGS); MCW Illumina	0.155
C5-A1ML	75	2851	E2		chrX:149,550,454	MAMLD1	Intron	79		0.155

Supplementary Table S4 (continued). Summary of viral integrations identified by PacBio long-read sequencing of HPV-enriched ICC tumor DNA. Details of each unique, chimeric viral-human breakpoint identified by PacBio sequencing. For each breakpoint, we provide viral nucleotide and gene location, human chromosome band, human nucleotide and gene location, gene location (intron or exon), number of supporting PacBio reads, and clonality as quantified by integration allele fraction using sample matched TCGA data (IAF). *TCGA low-pass WGS used to calculate IAF. **TCGA WXS data used to calculate IAF. No value in IAF column = ViFi program did not detect viral integration within 500 bp of listed breakpoint coordinates.00

TCGA ID	Event #	HPV bp	HPV Gene	Human Chrom Band	Human bp	Human Gene/Locus	Gene Location	PacBio Read #	Validation	Clonality (IAF)
C5-A1ML	76	2483	E1	7p22.2	chr7:4,200,858	SDK1	Intron	444	PacBio Only	
C5-A1ML	76	2689	E1		chr7:4,200,864	SDK1	Intron	119		
C5-A1ML	76	2556	E1		chr7:4,229,987	SDK1	Intron	81		
C5-A1ML	76	5687	L1		chr7:4,271,915	SDK1	Intron	15		
C5-A1ML	76	7399	URR		chr7:4,306,485	SDK1	Exon	28		
C5-A1ML	77	4764	L2	10q23.33	chr10:96,306,733	HELLS	Intron	11	PacBio Only	
C5-A1ML	77	3956	E5		chr10:96,312,993	HELLS	Intron	2		
C5-A1ML	77	7038	L1		chr10:96,319,211	HELLS	Intron	1		
C5-A3HD	78	1616	E1	1p13.2	chr1:114,306,721	RSBN1	Exon	33	TCGA (RNA, WGS*)	*0.103
C5-A3HD	78	2095	E1		chr1:114,340,522	RSBN1	Exon	29		*0.317
C5-A3HD	78	1112	E1		chr1:114,347,453	RSBN1	Intron	42		*0.167
C5-A3HD	78	3212	E2		chr1:114,364,269	AP4B1-AS1; PTPN22	Intron	6		*0.065
C5-A3HD	79	2815	E1	7p22.1	chr7:5,542,031	FBXL18	Intron	2	TCGA (WGS*)	*0.238
C5-A3HD	80	3426	E2	7q31.1	chr7:113,711,528	PPP1R3A, FOXP2	NA	91	TCGA (RNA, WGS*)	*0.179
C5-A3HD	80	331	E6		chr7:113,734,716	FOXP2	Intron	18		*0.037; 0.104
C5-A3HD	80	4161	E5α, L2		chr7:113,735,434	FOXP2	Intron	36		*0.104
C5-A3HD	80	2821	E2		chr7:113,735,913	FOXP2	Intron	84		*0.104
C5-A0TN	81	2593	E1	2q32.3	chr2:195,586,245	LINC01821, LINC01790	NA	7	TCGA (RNA, WGS)	0.233
C5-A0TN	82	6054	L1	3q21.3	chr3:126,849,193	PLXNA1, C3ORF56	NA	16	TCGA (WGS)	0.231
C5-A2LV	83	2467	E1	15q25.2	chr15:84,040,546	BNC1, SH3GL3	NA	132	TCGA (RNA, WGS)	0.141
C5-A2LV	83	5904	L1		chr15:84,040,546	BNC1, SH3GL3	NA	132		0.141
C5-A8XH	84	1292	E1	17q25.3	chr17:76,629,330	SCAT1, CYTH1	NA	7	MCW Illumina	**1.000
C5-A8XH	84	2960	E2		chr17:76,828,332	USP36	Intron	132		**0.878
C5-A1M9	85	3123	E2	8q24.21	chr8:128,993,409	PVT1	Intron	26	TCGA (RNA, WGS); MCW Illumina	
C5-A1M9	85	7611	URR		chr8:129,013,913	PVT1	Intron	5		0.129
C5-A1M9	85	*4857	*L2		*chr8:129,022,423	PVT1	Intron	*1		
C5-A1M9	86	*5489	*L2	17q12	*chr17:37,874,359	ERBB2	Intron	*4	TCGA (RNA, WGS)	0.173
C5-A2LT	87	*3138	*E2	13q12.11	*chr13:21,031,963	CRYL1	Intron	*1	TCGA (RNA, WGS)	0.099

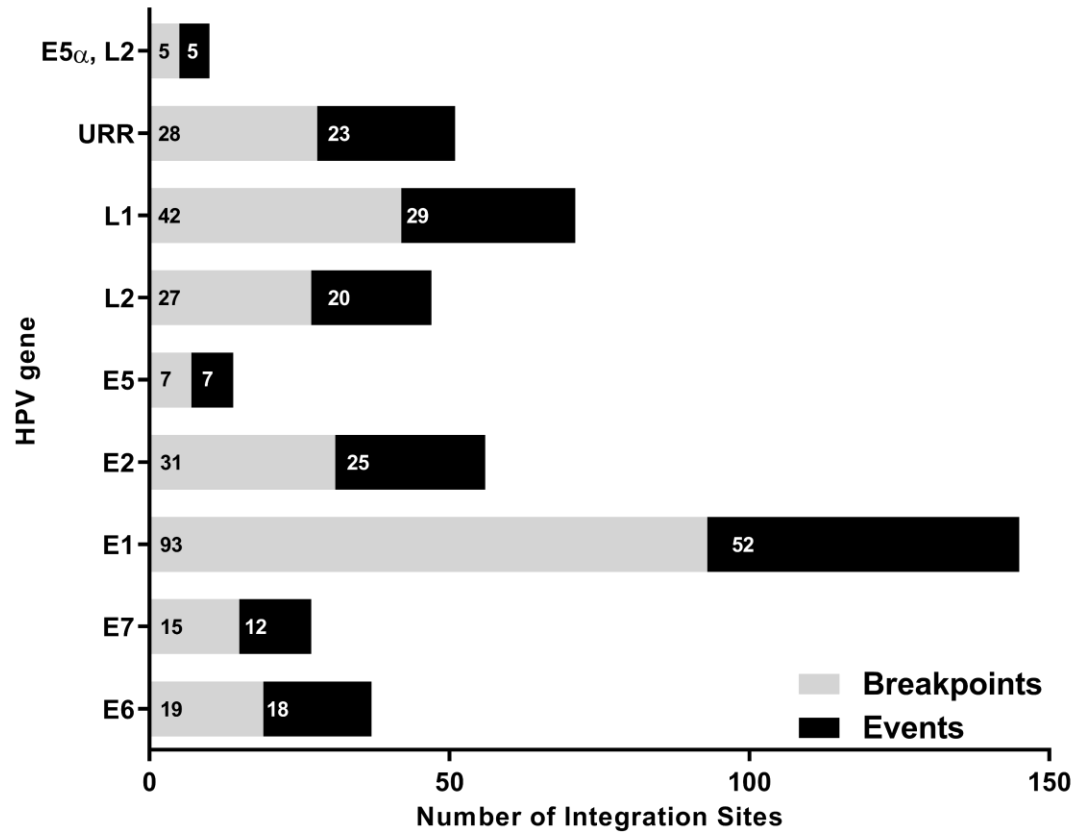
Supplementary Table S4 (continued). Summary of viral integrations identified by PacBio long-read sequencing of HPV-enriched ICC tumor DNA. Details of each unique, chimeric viral-human breakpoint identified by PacBio sequencing. For each breakpoint, we provide viral nucleotide and gene location, human chromosome band, human nucleotide and gene location, gene location (intron or exon), number of supporting PacBio reads, and clonality as quantified by integration allele fraction using sample matched TCGA data (IAF). *TCGA low-pass WGS used to calculate IAF. **TCGA WXS data used to calculate IAF. No value in IAF column = ViFi program did not detect viral integration within 500 bp of listed breakpoint coordinates.00

Event	Gene symbol	molecule type (IPA)	Zscore	OS HR	OS *p*val	RFS HR	RFS *p*val
1	*MRPL35*	OTHER	-1.59	0.7	0.10	0.6	0.22
2	*C2orf83*	TRANSPORTER	4.02	1.6	0.05	0.6	0.27
2	*COL4A4*	OTHER	2.86	1.6	0.06	3.9	0.01
4	*MROH5*	OTHER	1.63	1.5	0.11	2.2	0.05
4	*GLI4*	TRANSCRIPTION REGULATOR	1.83	0.8	0.28	2.4	0.06
4	*DENND3*	OTHER	2.45	0.7	0.10	2.0	0.09
6	*SPSB2*	OTHER	1.77	1.3	0.36	2.0	0.07
6	*EMG1*	ENZYME	2.24	1.5	0.12	0.6	0.32
6	*USP5*	PEPTIDASE	1.75	1.8	0.06	1.9	0.18
6	*PHB2*	TRANSCRIPTION REGULATOR	1.76	1.6	0.05	1.6	0.36
7	*TAOK3*	KINASE	3.80	1.5	0.13	2.3	0.09
8	*POLR1D*	ENZYME	1.78	1.7	0.08	0.5	0.10
8	*POMP*	OTHER	4.83	1.4	0.15	2.8	0.02
9	*NF1P2*	OTHER	5.22	1.9	0.01	0.6	0.29
10	*C17orf98*	OTHER	2.65	1.4	0.17	1.7	0.17
10	*WIPF2*	OTHER	-1.76	0.7	0.09	1.9	0.09
10	*RPL23*	OTHER	1.74	1.5	0.12	0.5	0.15
11	*SPATA32*	OTHER	1.51	0.7	0.20	2.2	0.04
11	*MPP2*	KINASE	3.31	0.7	0.21	2.1	0.18
11	*HEXIM1*	TRANSCRIPTION REGULATOR	6.66	1.4	0.13	2.8	0.01
12	*GADD45B*	OTHER	1.96	0.6	0.05	2.4	0.03
12	*GNA11*	ENZYME	-1.76	0.7	0.25	0.6	0.19
13	*ZBED4*	TRANSCRIPTION REGULATOR	-1.64	1.8	0.07	0.5	0.11
13	*BRD1*	OTHER	-1.66	0.6	0.04	0.4	0.03
13	*TYMP*	GROWTH FACTOR	1.98	1.5	0.10	0.5	0.10
14	*LOC100129534*	OTHER	2.88	0.7	0.08	1.9	0.10
14	*MORN1*	OTHER	3.43	0.6	0.07	1.8	0.13
15	*GADD45A*	OTHER	3.49	1.7	0.05	0.6	0.15
16	*ISG20L2*	ENZYME	-1.62	1.6	0.12	0.5	0.06
16	*SCAMP3*	TRANSPORTER	1.55	2.3	0.01	2.4	0.09
16	*CHTOP*	OTHER	2.11	0.8	0.37	2.3	0.07
16	*LAMTOR2*	OTHER	1.55	1.4	0.12	0.6	0.25
16	*EFNA4*	KINASE	4.71	1.5	0.08	1.4	0.39
16	*EFNA3*	KINASE	5.55	1.8	0.02	3.0	0.06
16	*BGLAP*	OTHER	2.27	0.6	0.07	2.0	0.11
16	*RPS27*	OTHER	3.60	1.5	0.15	0.6	0.15
16	*BCAN*	OTHER	1.78	1.3	0.32	1.8	0.14
16	*ADAM15*	PEPTIDASE	2.46	1.2	0.46	1.8	0.14
16	*FDPS*	ENZYME	1.68	1.8	0.01	2.6	0.05
16	*CKS1B*	KINASE	2.57	1.5	0.14	1.6	0.26
17	*TNN*	OTHER	2.32	0.6	0.03	2.3	0.13
24	*CSN1S1*	OTHER	1.70	1.6	0.05	0.7	0.28

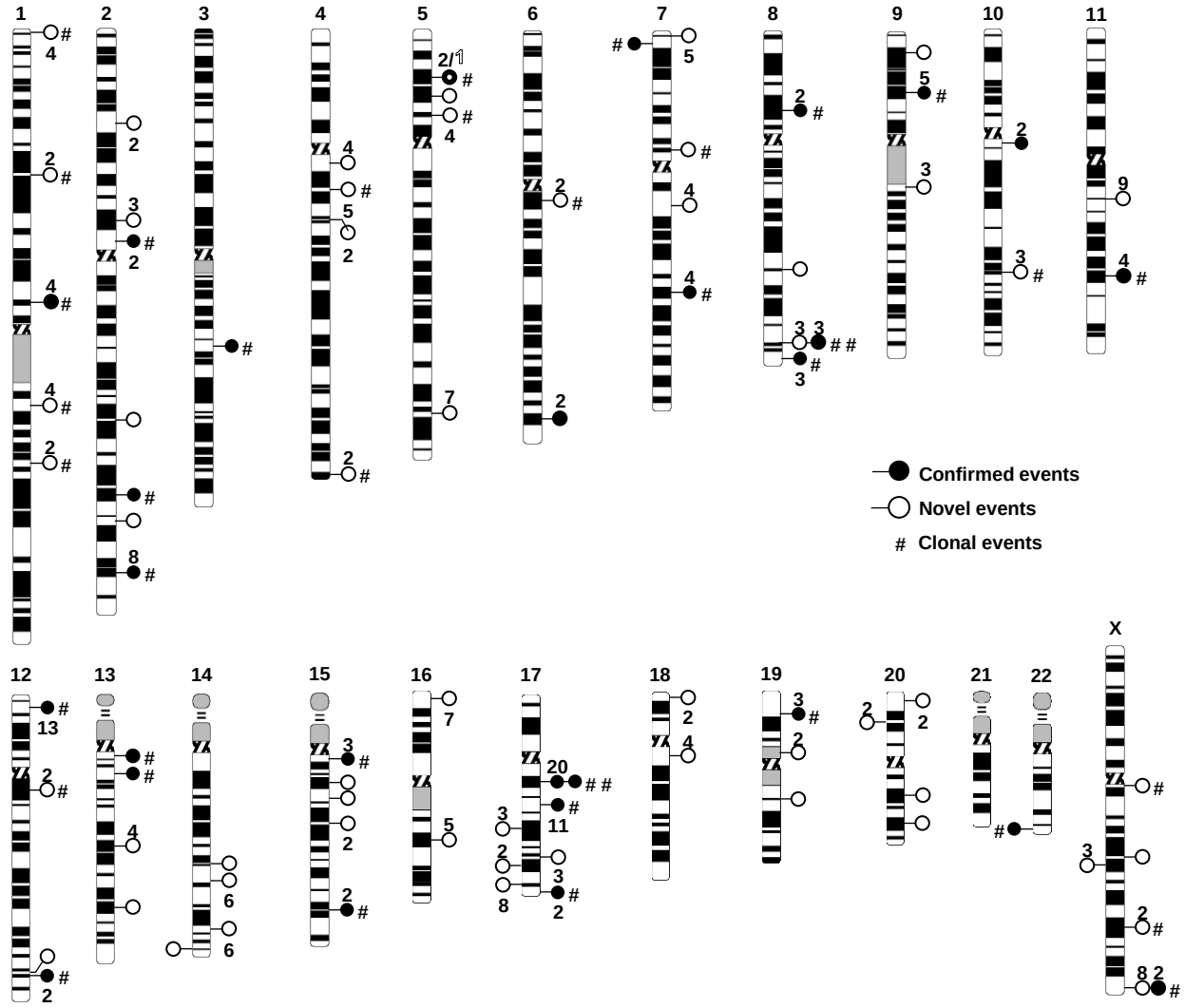
Supplementary Table S5. Final list of 84 candidate IDGs remaining after filtering based on integration event clonality, *Z*-score, and association with ICC survival. List of genes passing each of our criteria for qualification as a candidate IDG. Information included for each gene: Corresponding PacBio integration event number (Supplementary Table S4; column 2), gene symbol (NCBI), molecule type (IPA), *Z*-score, and overall survival (OS) and recurrence-free survival (RFS) hazard ratios (HR) and corresponding p-values (pval). IDGs highlighted in yellow are those further studied in functional assays.

Event	Gene symbol	molecule type (IPA)	Zscore	OS HR	OS *p*val	RFS HR	RFS *p*val
29	*LIFR*	TRANSMEMBRANE RECEPTOR	11.04	0.8	0.25	2.7	0.03
74	*HNRNPF*	OTHER	-1.93	0.6	0.06	2.1	0.07
75	*PRRG3*	OTHER	4.13	1.5	0.09	2.4	0.03
78	*PHTF1*	TRANSCRIPTION REGULATOR	4.77	0.8	0.33	2.6	0.01
78	*RSBN1*	OTHER	17.24	0.7	0.13	2.0	0.10
78	*TRIM33*	TRANSCRIPTION REGULATOR	1.93	0.6	0.06	2.2	0.10
78	*NRAS*	ENZYME	1.56	1.3	0.28	4.5	0.01
79	*RBAK-RBAKDN*	OTHER	2.26	1.4	0.13	2.0	0.20
79	*RBAK*	TRANSCRIPTION REGULATOR	2.26	1.7	0.04	0.7	0.45
80	*ZNF277*	TRANSCRIPTION REGULATOR	-2.35	1.5	0.14	0.6	0.14
81	*SLC39A10*	TRANSPORTER	1.93	2.0	0.00	4.0	0.00
82	*ZXDC*	TRANSCRIPTION REGULATOR	-1.95	0.5	0.02	0.6	0.16
82	*SNX4*	TRANSPORTER	-1.59	0.7	0.20	0.8	0.50
82	*ZNF148*	TRANSCRIPTION REGULATOR	-1.64	0.5	0.01	0.7	0.42
83	*ZNF592*	TRANSCRIPTION REGULATOR	-1.85	0.7	0.19	0.6	0.15
83	*WHAMM*	OTHER	-1.52	0.4	0.00	0.5	0.09
83	*SEC11A*	PEPTIDASE	1.67	1.4	0.19	2.0	0.08
83	*BNC1*	TRANSCRIPTION REGULATOR	1.85	1.5	0.11	0.4	0.02
84	*METTL23*	OTHER	1.93	0.7	0.12	2.0	0.19
84	*MFSD11*	OTHER	2.18	1.5	0.14	1.9	0.13
84	*SLC26A11*	TRANSPORTER	1.52	1.8	0.07	1.9	0.11
84	*SGSH*	ENZYME	3.97	1.7	0.03	2.1	0.18
84	*PGS1*	ENZYME	3.80	0.6	0.12	3.2	0.04
84	*CANT1*	ENZYME	2.50	1.7	0.03	3.2	0.00
84	*MGAT5B*	ENZYME	1.76	0.6	0.03	2.0	0.07
84	*USP36*	PEPTIDASE	15.68	1.4	0.19	2.4	0.07
84	*EIF4A3*	ENZYME	4.76	0.5	0.01	2.0	0.07
84	*RNF213*	ENZYME	2.50	0.7	0.16	2.0	0.18
84	*CBX2*	TRANSCRIPTION REGULATOR	2.04	1.4	0.15	1.5	0.32
84	*JMJD6*	TRANSMEMBRANE RECEPTOR	1.79	1.7	0.04	1.4	0.40
84	*LGALS3BP*	TRANSMEMBRANE RECEPTOR	4.81	1.6	0.04	5.8	0.00
84	*SEPTIN9*	ENZYME	2.23	1.6	0.05	2.7	0.01
84	*RPTOR*	OTHER	3.39	1.5	0.19	2.7	0.09
84	*GAA*	ENZYME	1.73	1.5	0.11	2.6	0.07
84	*SRSF2*	TRANSCRIPTION REGULATOR	1.60	0.4	0.00	2.7	0.06
84	*SOCS3*	PHOSPHATASE	1.77	1.4	0.27	2.6	0.08
85	*LRATD2*	OTHER	2.24	1.6	0.03	2.8	0.01
85	*PVT1*	OTHER	12.17	1.5	0.19	0.5	0.06
86	*MRPL45*	OTHER	2.10	0.5	0.04	2.2	0.04
86	*LASP1*	TRANSPORTER	2.12	1.7	0.02	2.7	0.01
86	*KRT19*	OTHER	1.64	1.4	0.18	2.0	0.08
86	*TOP2A*	ENZYME	1.68	0.7	0.09	2.2	0.05

Supplementary Table S5. Final list of 84 candidate IDGs remaining after filtering based on integration event clonality, *Z*-score, and association with ICC survival. List of genes passing each of our criteria for qualification as a candidate IDG. Information included for each gene: Corresponding PacBio integration event number (Supplementary Table S4; column 2), gene symbol (NCBI), molecule type (IPA), *Z*-score, and overall survival (OS) and recurrence-free survival (RFS) hazard ratios (HR) and corresponding p-values (pval). IDGs highlighted in yellow are those further studied in functional assays.

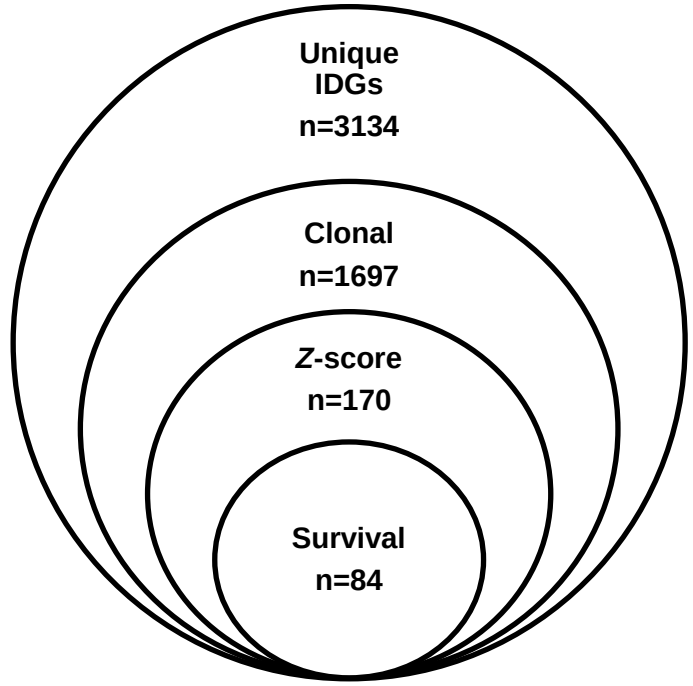


Supplementary Figure S1. Frequency of viral breakpoints and integration events in HPV gene regions revealed by PacBio long-read sequencing. The number of chimeric HPV-human breakpoints (gray) or integration events (black) found in each viral gene is plotted for all 8 samples. A breakpoint is defined as the unique nucleotide position of each human-HPV chimeric junction, while an integration event is defined as a collection of one or more chimeric HPV-human junctions occurring within 1.5Mb of each other.



Supplementary Figure S2. Chromosomal map of HPV integration sites detected by PacBio long-read sequencing of HPV-enriched ICC tumor DNA. Chromosomal ideograms marked with each circle representing an integration event and number above or below representing the number of breakpoints comprising that event. No number indicates only one breakpoint identified in that event. Black-filled circles = sites identified in other datasets (TCGA). Open circles = sites found by PacBio sequencing and not originally reported by TCGA. # denotes a clonal event (IAF $\geq$ 10%, meaning at least 10% of reads covering the area of the integration had to harbor a human-viral chimeric breakpoint). ## denotes 2 unique clonal events

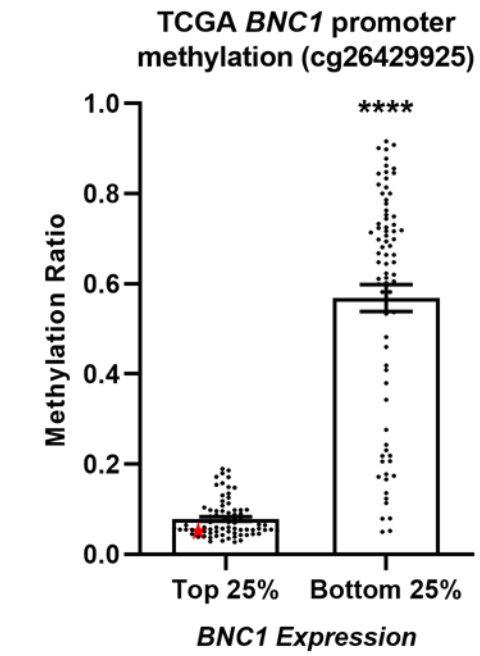
A

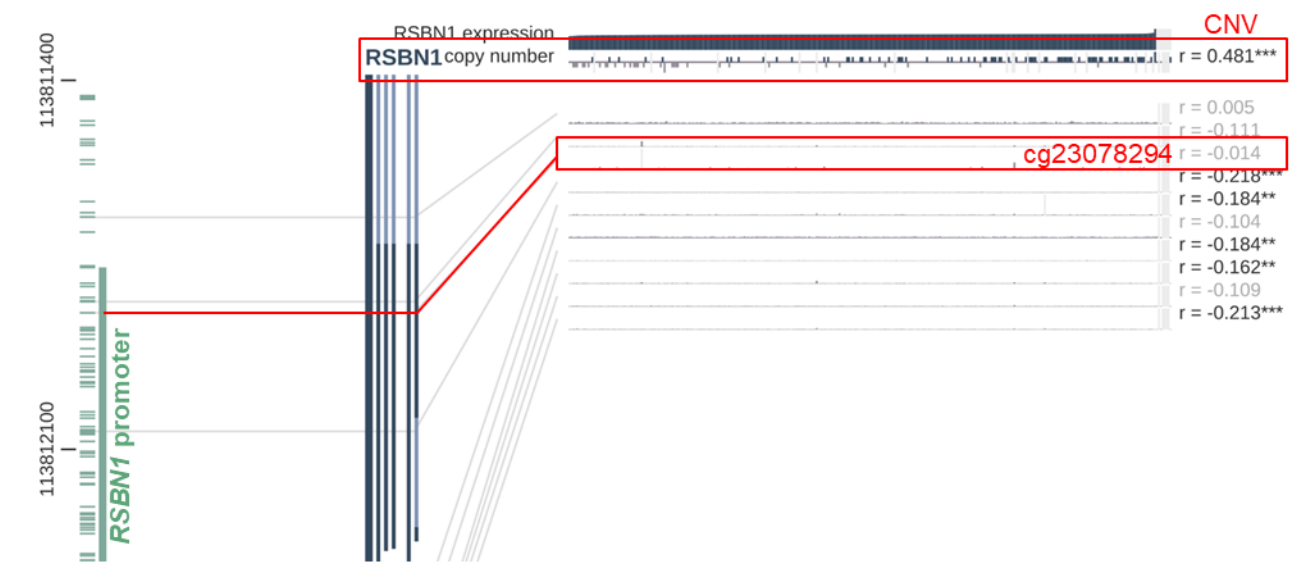


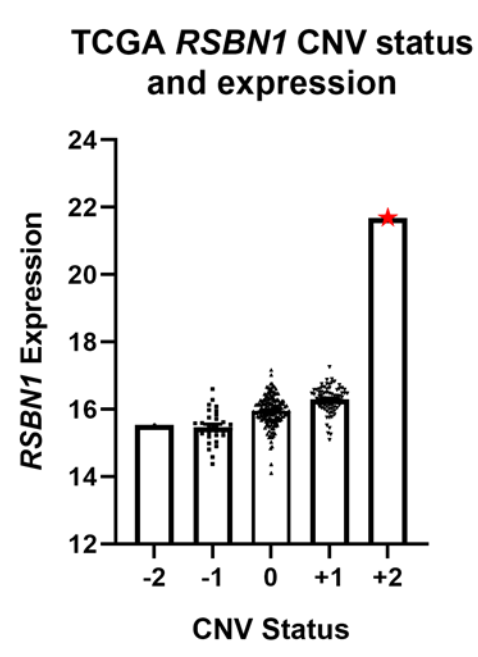
B

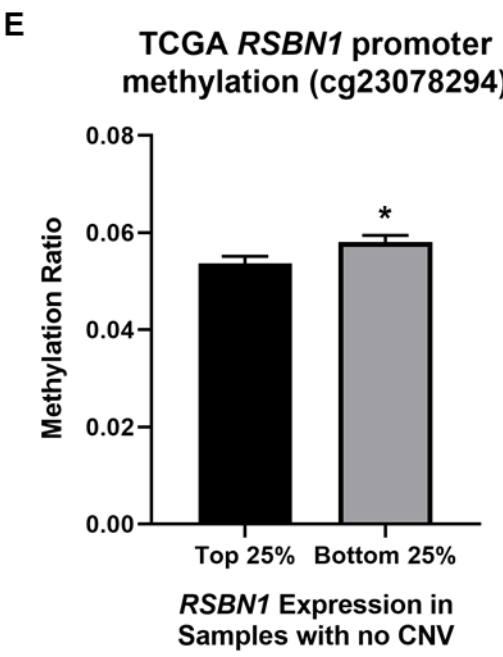
Gene	Publication number	
Known Oncogenes	**Pan-cancer**	**ICC-specific**
ERBB2	>29,000	224
RAD51B	2,593	36
PVT1	481	16
Candidate IDGs		
BNC1	35	0
RSBN1	3	0
USP36	14	0
TAOK3	8	0

Supplementary Figure S3. Filtering of candidate IDGs for functional prioritization. Filtering of candidate IDGs (A) began with genes within 2Mb on either side of each integration event for each of the 8 samples, providing a total of 3399 genes (3134 unique). IDGs were first filtered based on clonal representation of their associated integration site, leaving 1697 potential IDGs. Next, genes were removed if they did not meet our tumor-specific expression criteria (*Z*-score cutoff of 1.5/-1.5), leaving 170 potential IDGs. The last filter was association with survival in TCGA ICC cohort (HR, *p*-value cutoff of 0.2), leaving a total of 84 candidate IDGs. PubMed literature search results of the 4 candidate IDGs suggest their unknown function in cervical cancer (B).

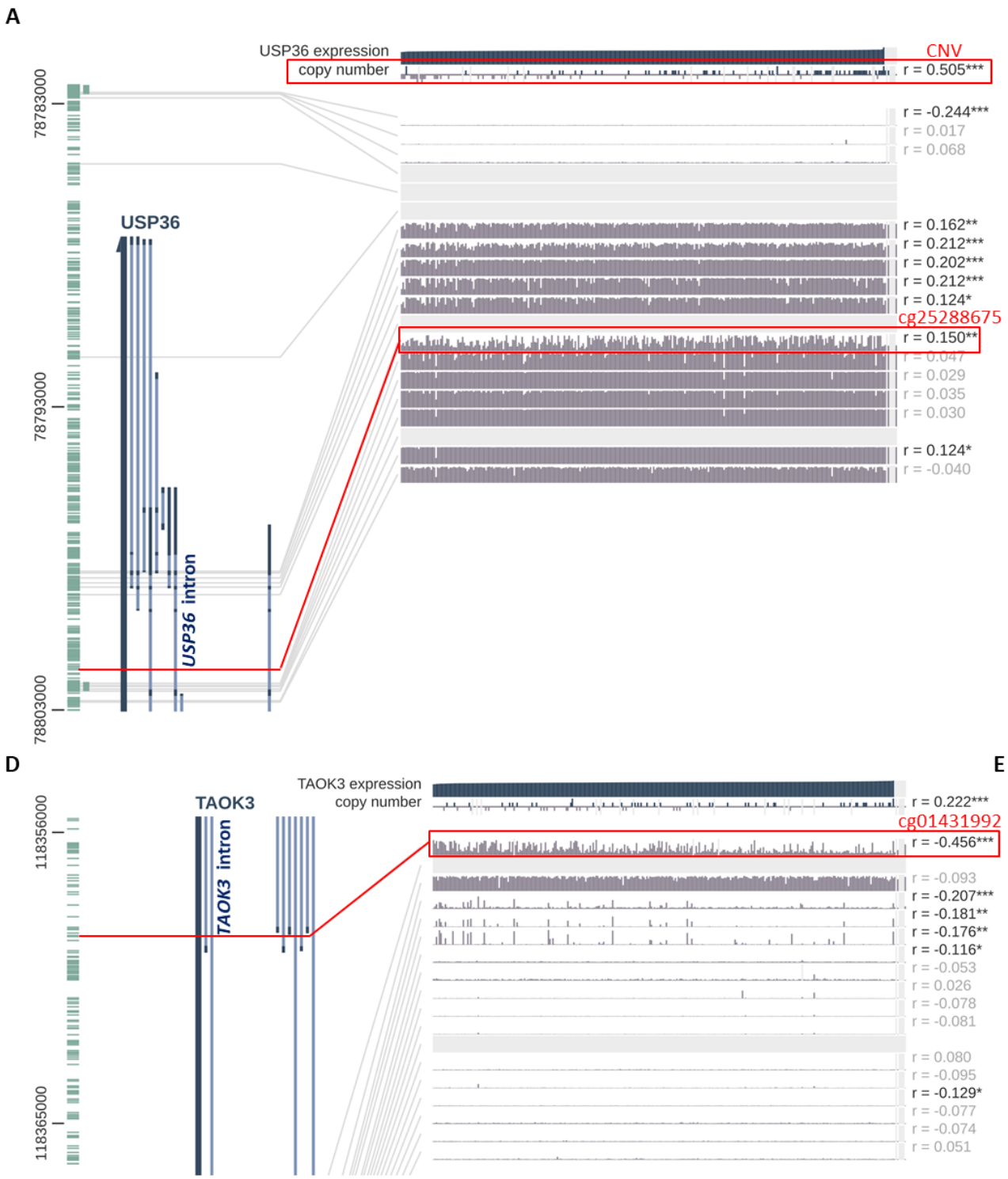




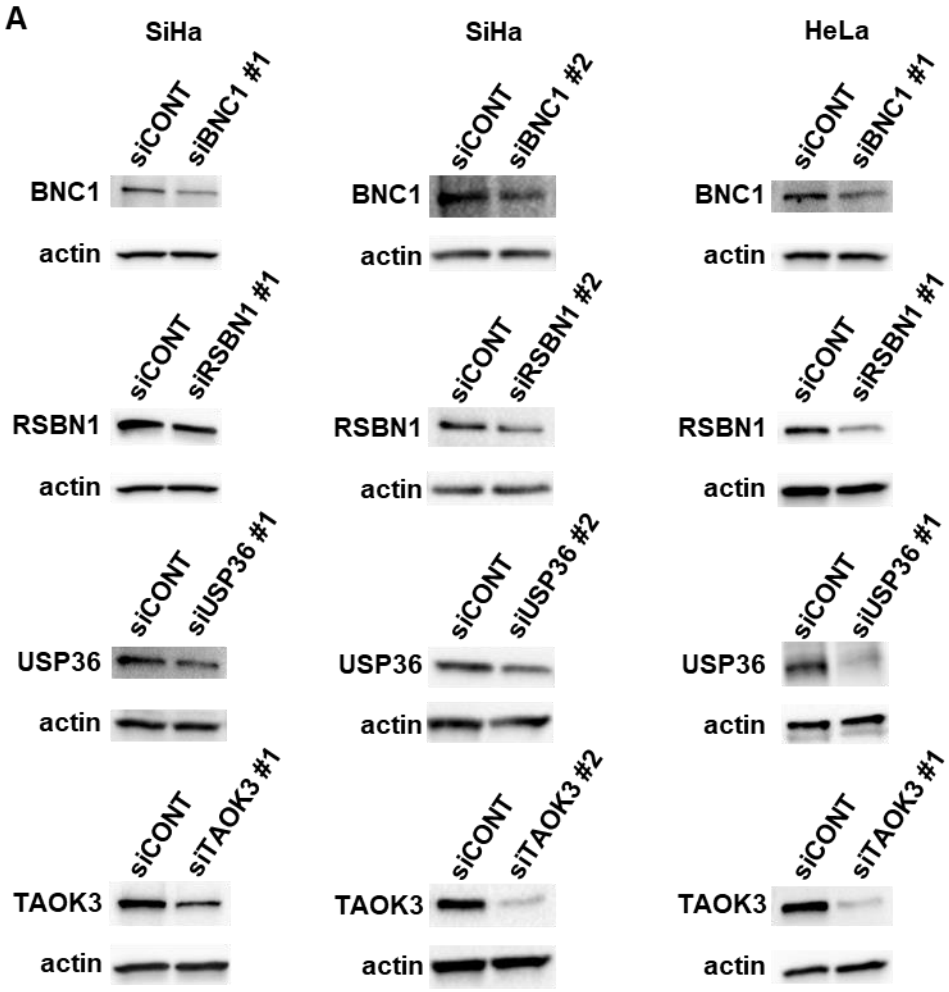


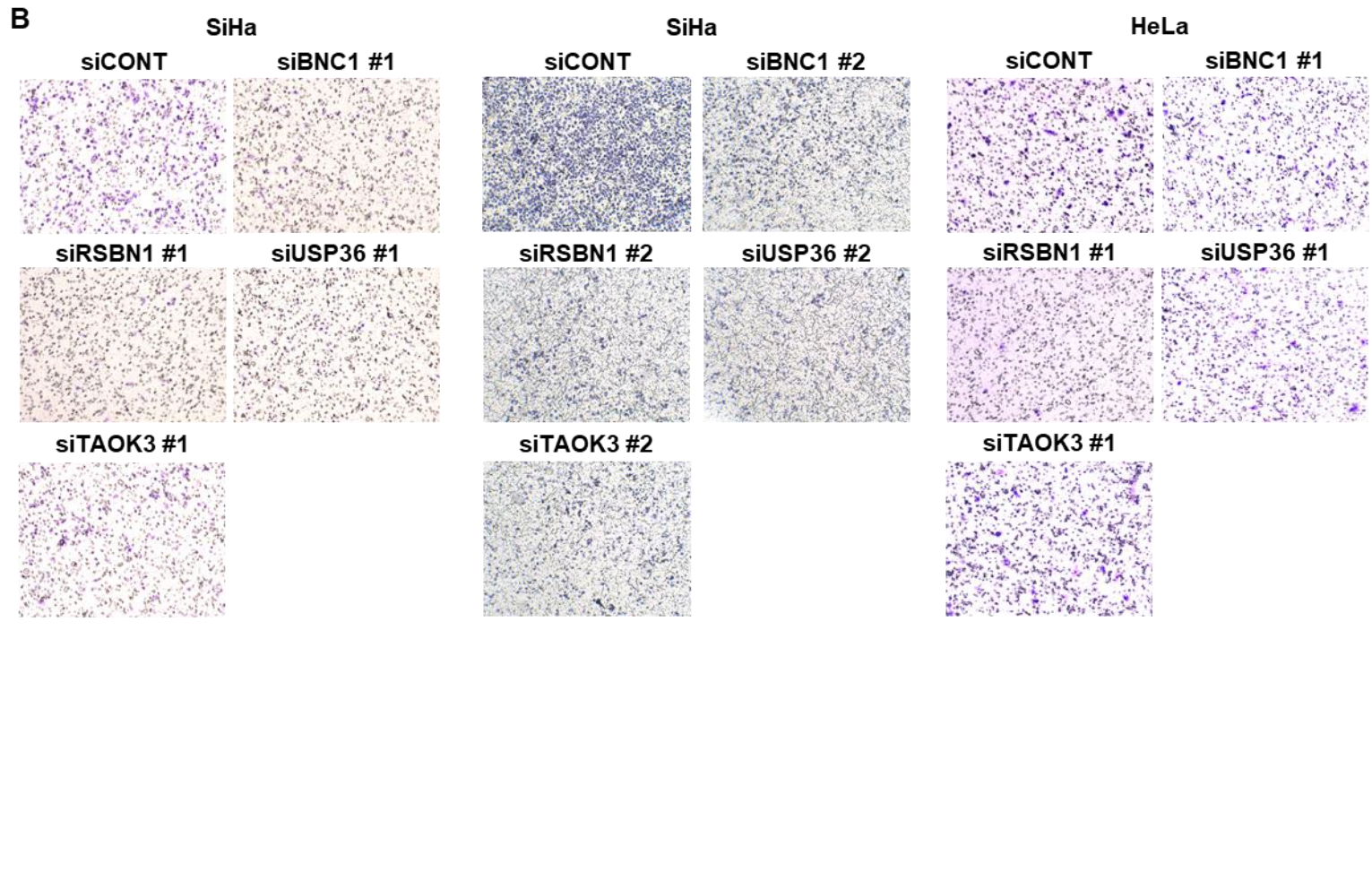


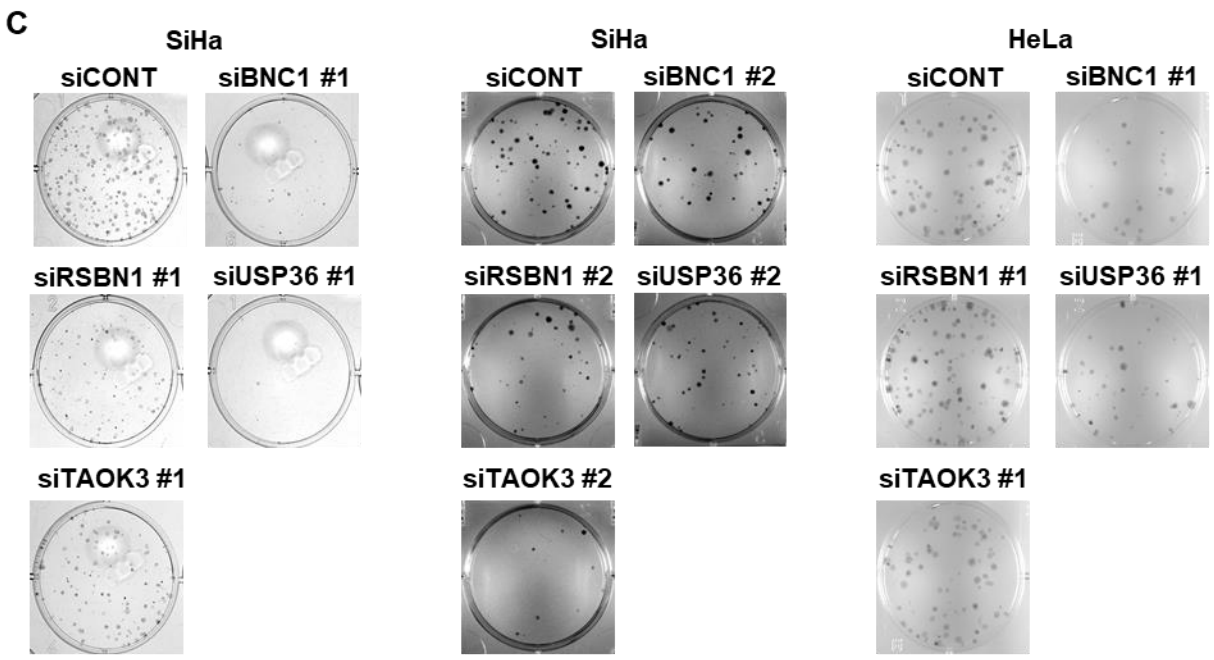
Supplementary Figure S4. Potential regulation of *BNC1* and *RSBN1* expression in non-integrated ICC samples. (A) MEXPRESS data depicting correlation of *BNC1* CpG methylation and expression in TCGA ICC cohort (n=304). *BNC1* promoter hypomethylation displayed the most significant correlation with high *BNC1* expression (cg26429925; ***p < 0.001). (B) Analysis of cg26429925 methylation in TCGA ICC cohort revealed significantly higher methylation of this site in samples comprising the bottom 25% of *BNC1* expressors versus those comprising the top 25% (****p < 0.0001; integrated sample TCGA-C5-A2LV marked with red star). (C) High *RSBN1* expression was significantly correlated with copy number variation in the TCGA ICC cohort (MEXPRESS CNV; ***p < 0.001). (D) Plot of *RSBN1* expression as a function of CNV status using data from TCGA ICC cohort with the integrated sample (TCGA-C5-A3HD; red star). In TCGA ICC samples with no reported CNV in this area, methylation of a CpG site in the *RSBN1* promoter (cg23078294) was most negatively correlated with its expression. (E) Methylation of this site was significantly higher in samples comprising the bottom 25% of *RSBN1* expressors versus those comprising the top 25% (*p < 0.05). Plotted data is presented as mean ± standard error of the mean.

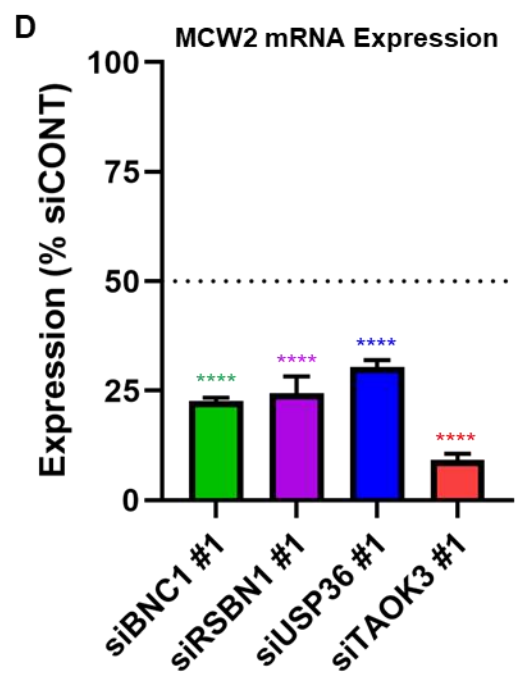


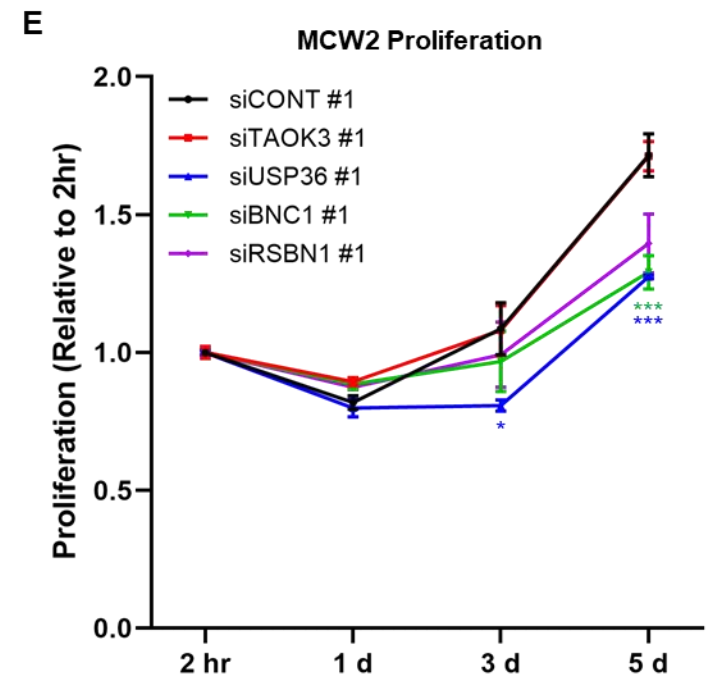
Supplementary Figure S5. Potential regulation of *USP36* and *TAOK3* expression in non-integrated ICC samples. (A) MEXPRESS showed *USP36* expression correlated most with copy number variation in TCGA ICC cohort (CNV; ***p < 0.001). (B) Plot of *USP36* expression as a function of CNV status using data from TCGA ICC cohort with the integrated sample (TCGA-C5-A8XH; red star). In TCGA ICC samples with no reported CNV in this area, methylation of a CpG site within intron 17 of *USP36* (cg25288675) was most positively correlated with its expression. (C) Methylation of cg25288675 was significantly lower in samples comprising the bottom 25% of *USP36* expressors versus those comprising the top 25% (****p < 0.0001). (D) MEXPRESS showed *TAOK3* expression most correlated with methylation of a site within intron 1 (cg01431992; ***p < 0.001). (E) Analysis of cg01431992 methylation in TCGA ICC cohort revealed significantly higher methylation of this site in samples comprising the bottom 25% of *TAOK3* expressors versus those comprising the top 25% (****p < 0.0001; red star denotes methylation ratio of integrated sample, TCGA-C5-A2LX). Plotted data is presented as mean ± standard error of the mean.



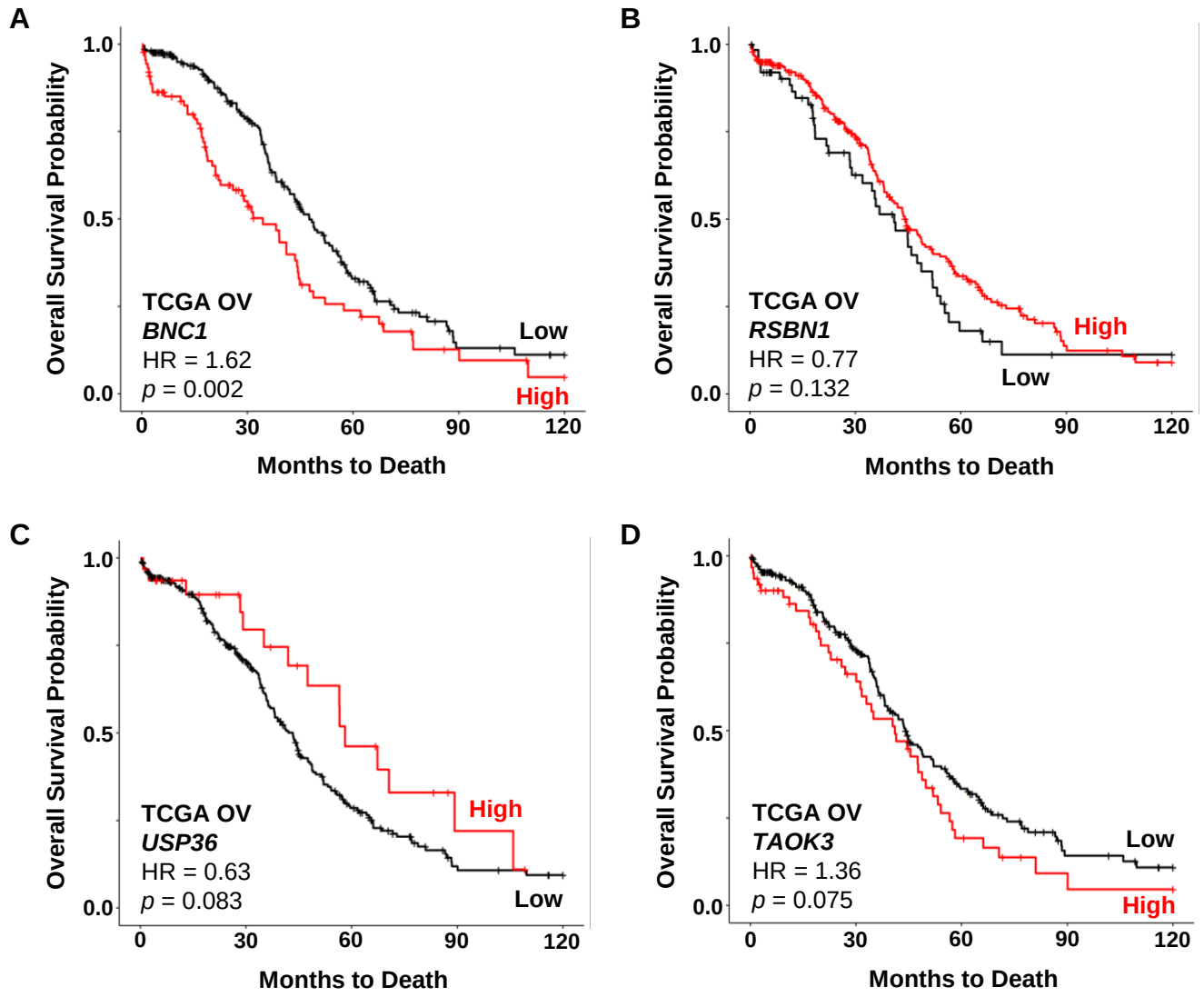




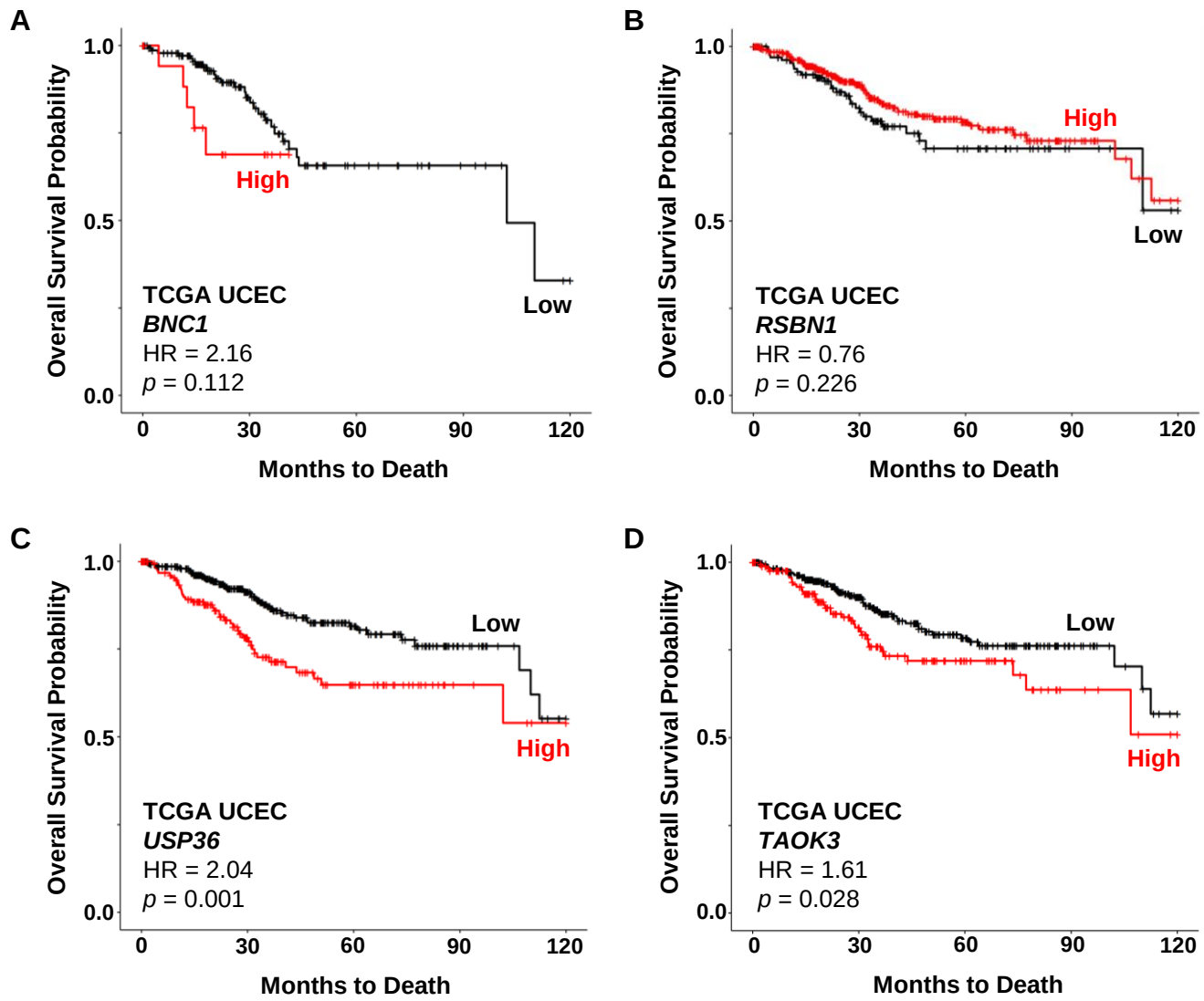




Supplementary Figure S6. siRNA-mediated knockdown of candidate IDGs in 2 cervical cancer cell lines. Representative western blot images (A) confirming decreased protein levels of candidate IDGs in SiHa (2 unique siRNA for each IDG; #1 and #2) and HeLa knockdown cells compared to control (siCONT). Representative images used to quantify SiHa and HeLa cell migration (B) and colony formation (C) presented in Fig. 6. IDG knockdown in the HPV18+ patient-derived cervical cancer cell line, MCW2 (D; ****$p < 0.0001$), revealed significantly decreased MCW2 proliferation upon knockdown of *USP36* and *BNC1* (E; *$p < 0.05$; ***$p < 0.001$).



Supplementary Figure S7. IDG expression correlation with overall survival in TCGA ovarian cancer cohort. Kaplan-Meier plots were generated using RNAseq data from TCGA ovarian cancer cohort (TCGA OV; n=299). Similar to results from the TCGA ICC cohort, high expression of *BNC1* (A) and *TAOK3* (D; not significant) was associated with poorer overall survival in ovarian cancer patients. Conversely, low *RSBN1* (B) and *USP36* (C) expression demonstrated a trend toward associating with poorer overall survival in these patients. HR = Hazard ratio.



Supplementary Figure S8. IDG expression correlation with overall survival in TCGA endometrial cancer cohort. Kaplan-Meier plots were generated using RNAseq data from TCGA endometrial cancer cohort (TCGA UCEC). Similar to results from the TCGA ICC cohort, high expression of *BNC1* (A), *USP36* (C), and *TAOK3* (D) tended to associated with poorer overall survival in endometrial cancer patients, although the results did not reach significance. Conversely, low *RSBN1* (B) expression tended to predicted poorer overall survival of these patient, but also did not reach statistical significance. HR = Hazard ratio. *BNC1* survival analysis from the UCEC cohort was limited to n=172 samples (versus n=526 for *USP36*, *TAOK3*, and *RSBN1*) due to residual batch effects exhibited by some genes after normalization of TCGA pan-cancer RNAseq data.