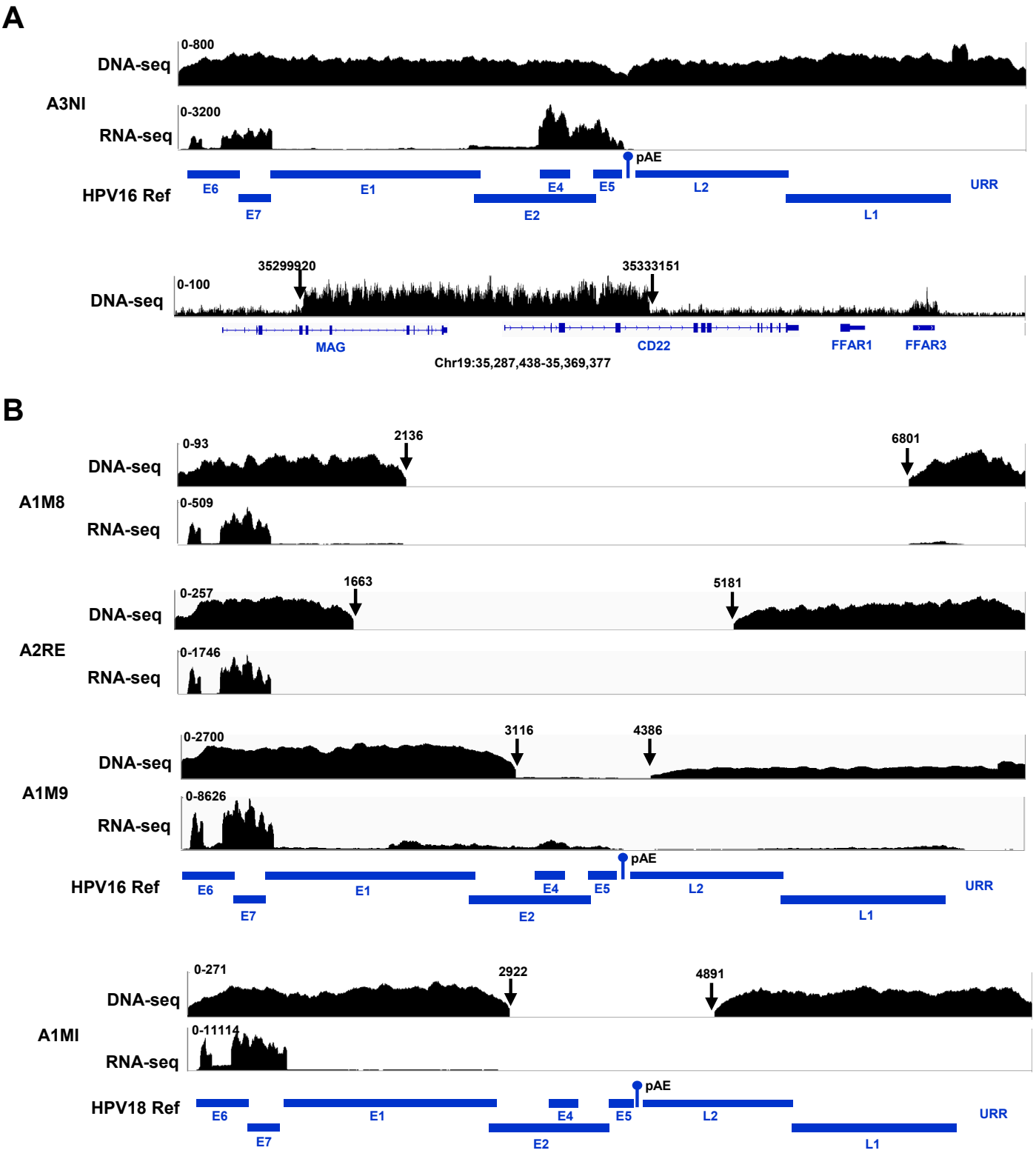
Fig. S1


A
DNA-seq
0-800
A3NI
RNA-seq
0-3200
pAE
HPV16 Ref
E6
E1
E4
E5
L2
URR
E7
E2
L1
35299920
35333151
DNA-seq
0-100
MAG
CD22
FFAR1
FFAR3
Chr19:35,287,438-35,369,377
B
DNA-seq
0-93
2136
6801
A1M8
RNA-seq
0-509
DNA-seq
0-257
1663
5181
A2RE
RNA-seq
0-1746
DNA-seq
0-2700
3116
4386
A1M9
RNA-seq
0-8626
pAE
HPV16 Ref
E6
E1
E4
E5
L2
URR
E7
E2
L1
DNA-seq
0-271
2922
4891
A1MI
RNA-seq
0-11114
pAE
HPV18 Ref
E6
E1
E4
E5
L2
URR
E7
E2
L1

C

A1M8 DNA-seq

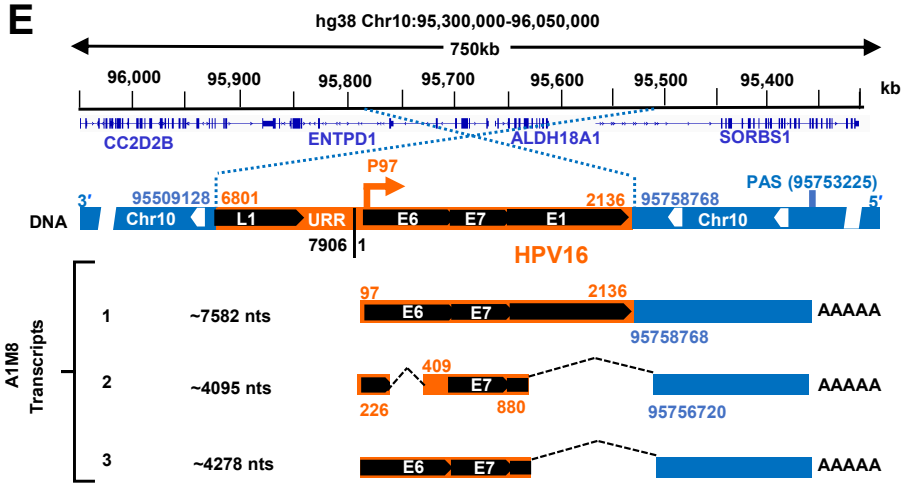
Rank	CJRs at HPV16 nt	Host Chr	Chr10 nt
1	2136	Chr10	95758768
2	6801	Chr10	95509128

D

A1M8 RNA-seq

No. of CJRs	CJRs at HPV16 nt	Host Chr	Chr10 nt
54	880	Chr10	95756720
3	880	Chr10	95756715

E



F

A1MI DNA-seq

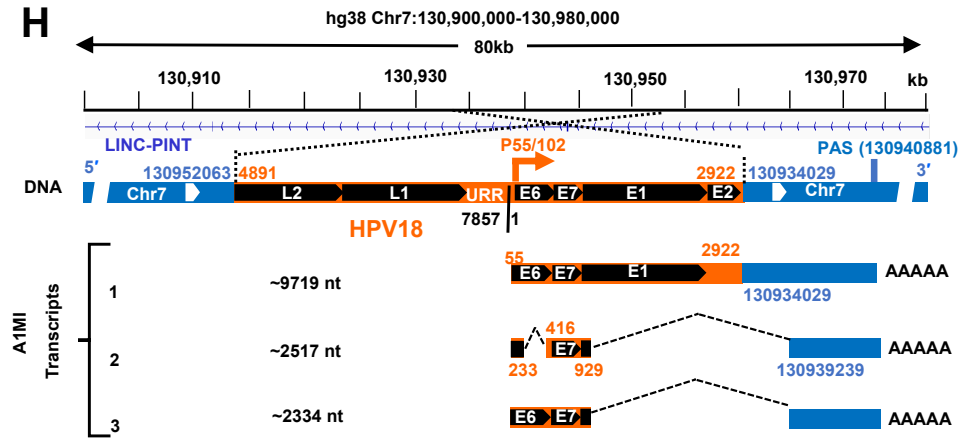
Rank	CJRs at HPV18 nt	Host Chr	Chr7 nt
1	2992	Chr7	130934029
2	4891	Chr7	130952063

G

A1MI RNA-seq

No. of CJRs	CJRs at HPV18 nt	Host Chr	Chr7 nt
1326	929	Chr7	130939239
8	929	Chr7	130939220
6	2992	Chr7	130934029

H



I

A2RE DNA-seq

Rank	DNA CJRs at HPV16 nt	Host Chr	Host genome nt
1	1663	Chr1	38021662
2	1199	Chr1	38022351
3	6985	Chr1	38022375
4	5288	Chr1	38146338
5	5517	Chr1	38147139
6	5181	Chr1	38147258
7	1589	Chr3	98040426
8	1261	Chr6	50856741
9	6904	Chr7	38577343
10	473	Chr11	121261348
11	1210	Chr12	80153683
12	5085	Chr14	81846144

J

A2RE RNA-seq

No. of CJRs	RNA CJRs at HPV16 nt	Host Chr	Host genome nt
102	880	Chr1	38022699
14	226	Chr1	98022699

K

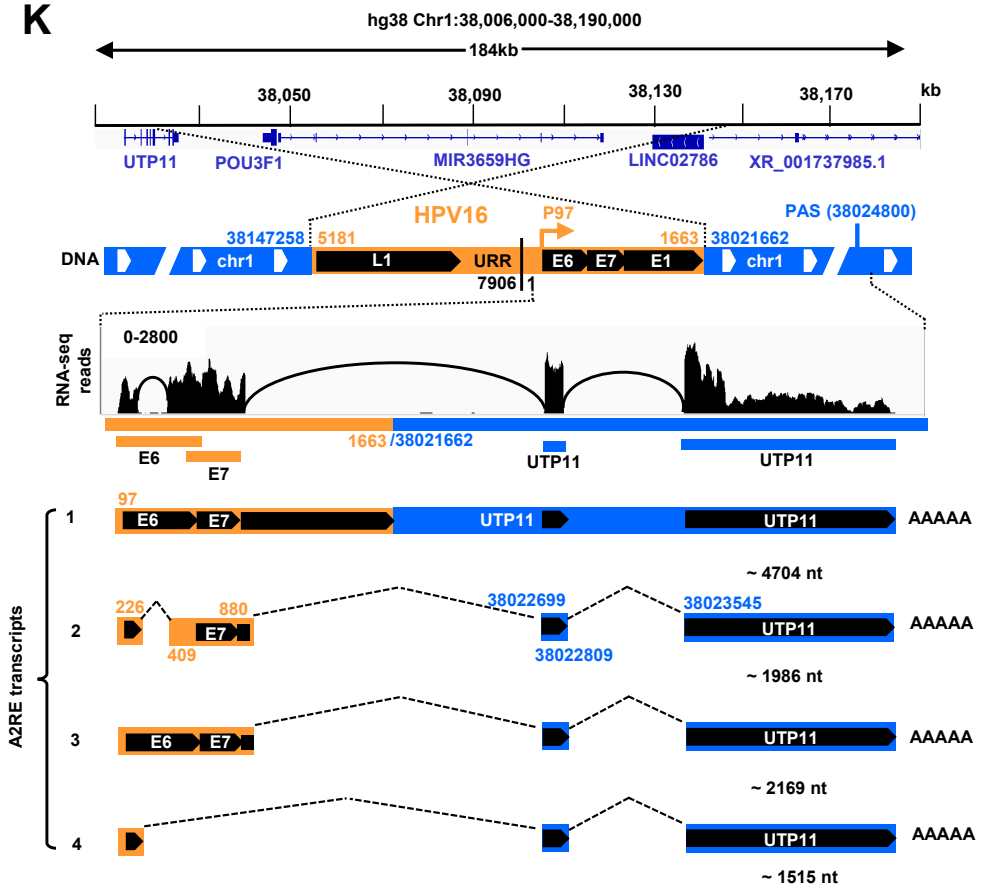


Fig. S1

L

A1M9 DNA-seq

Rank	CJRs at HPV16 nt	Host Chr	Chr nt
1	4386	Chr8	127981141
2	3116	Chr8	127981163
3	1749	Chr8	127983497
4	7547	Chr8	128000542
5	1785	Chr8	128001577
6	1684	Chr8	128005189
7	2068	Chr4	119703111
8	7614	Chr12	29998504
9	5495	Chr17	39718111
10	2448	Chr20	17390600

M

A1M9 RNA-seq

No. of CJRs	CJRs at HPV16 nt	Host Chr	Chr8 nt
717	880	Chr8	127989161
262	226	Chr8	127989161
127	880	Chr8	127982359
20	3116	Chr8	127981163
2	4386	Chr8	127981141

O

T1074 targeted DNA-seq

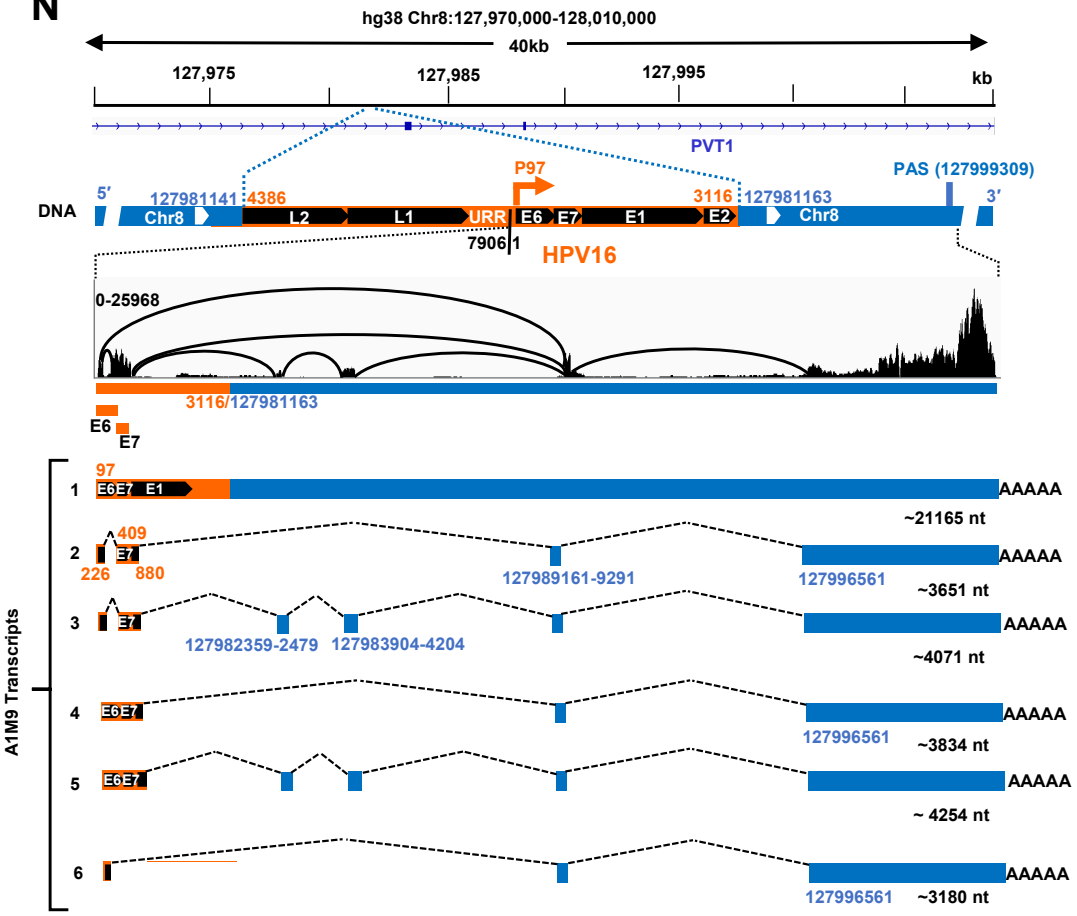
Rank	Host Chr	No. of mapped CJR locations
1	Chr3	22
2	Chr5	22
3	Chr2	16
4	Chr4	16
5	ChrX	13
6	Chr1	11
7	Chr7	11
8	Chr13	11
9	Chr18	11
10	Chr6	9
11	Chr8	9
12	Chr14	8
13	Chr10	7
14	Chr11	6
15	Chr15	6
16	Chr16	6
17	Chr9	5
18	Chr12	5
19	Chr17	5
20	Chr19	5
21	Chr20	3
22	Chr21	3
23	Chr22	3

P

T1074 RNA-seq

No. of CJRs	RNA CJRs at HPV16 nt	Host Chr	Host genome nt
86	880	Chr17	73338940
22	226	Chr17	73338940
9	5639	Chr17	73348599

N



Q

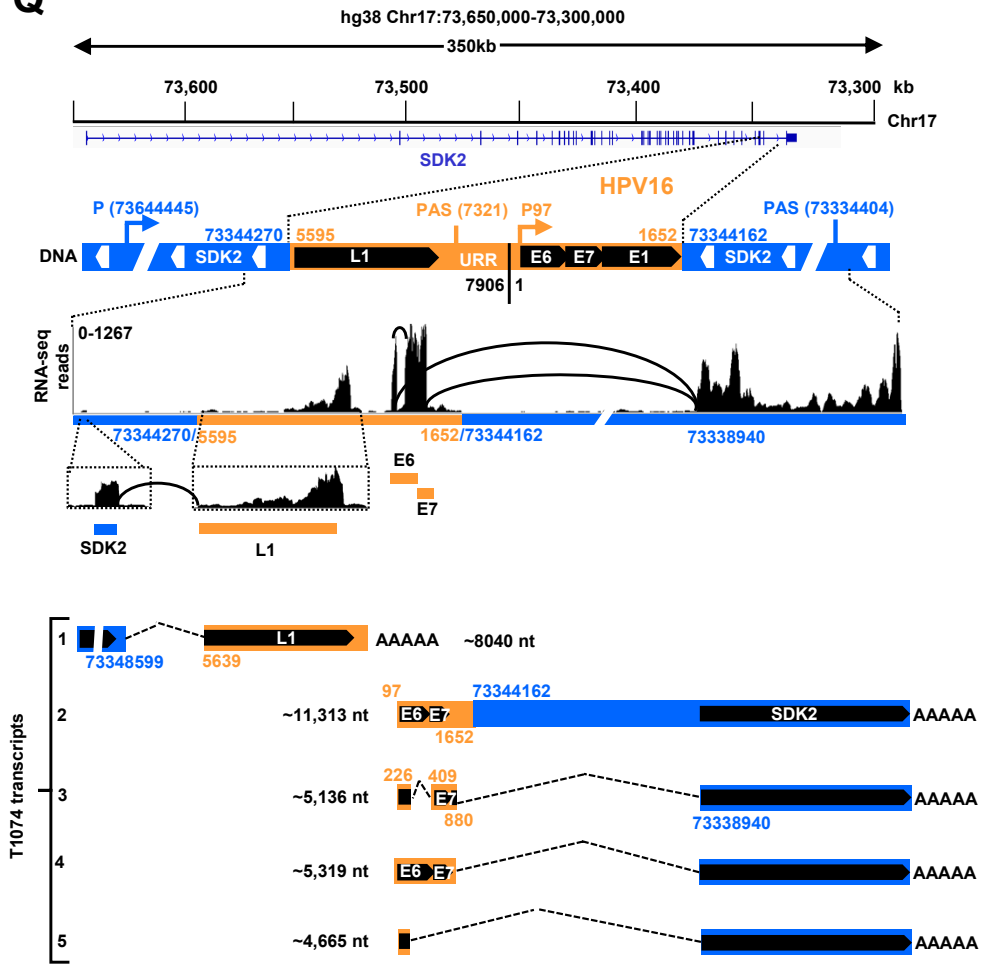
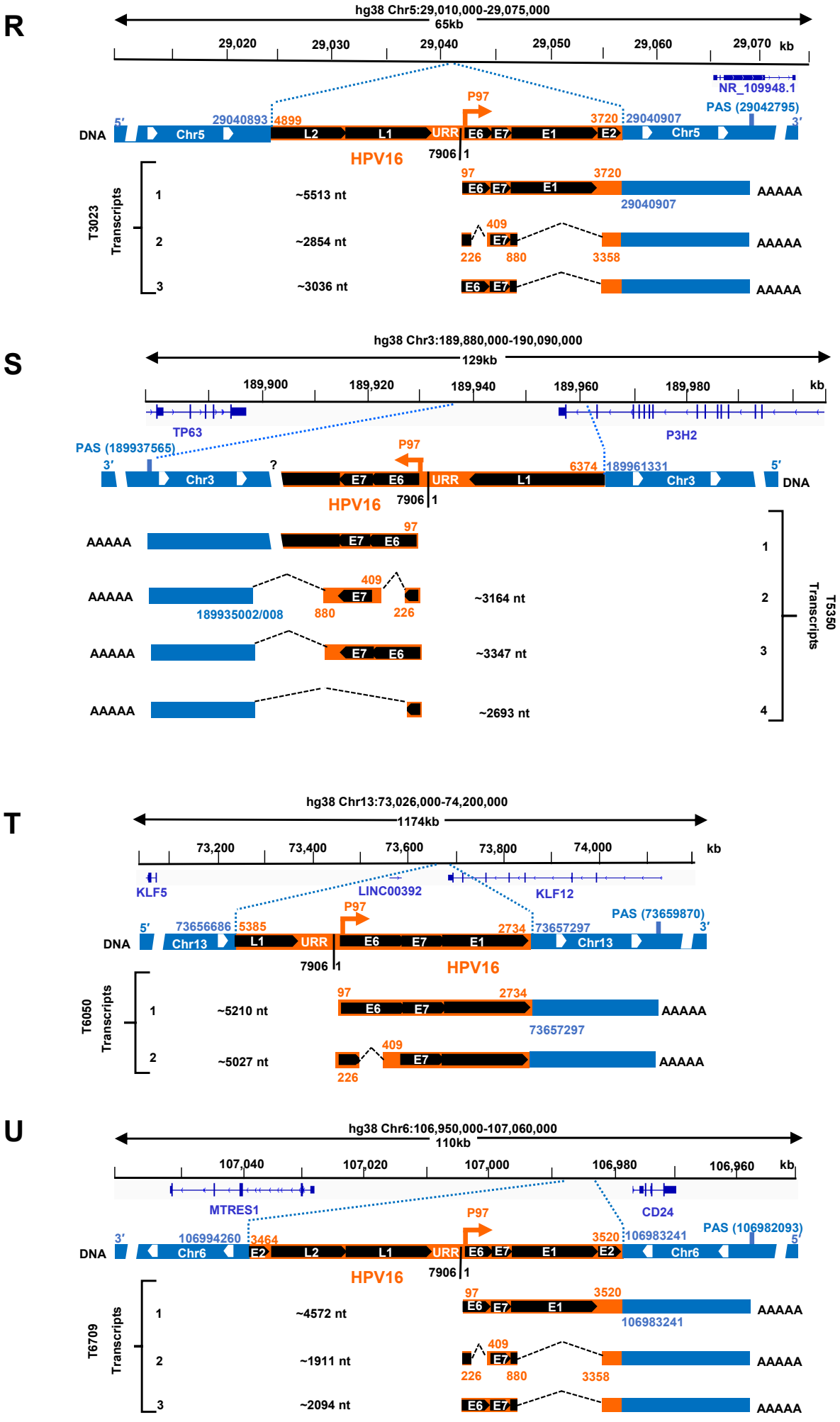


Fig. S1


R
hg38 Chr5:29,010,000-29,075,000
65kb
29,020
29,030
29,040
29,050
29,060
29,070
kb
NR_109948.1
PAS (29042795)
P97
DNA
5′
29040893
4899
Chr5
L2
L1
URR
E6
E7
E1
E2
3720
29040907
Chr5
3′
HPV16
7906
1
T3023 Transcripts
1
~5513 nt
97
3720
E6
E7
E1
AAAAA
29040907
2
~2854 nt
409
E7
AAAAA
226
880
3358
3
~3036 nt
E6
E7
AAAAA
S
hg38 Chr3:189,880,000-190,090,000
129kb
189,900
189,920
189,940
189,960
189,980
kb
TP63
P3H2
PAS (189937565)
P97
3′
Chr3
?
E7
E6
URR
L1
6374
189961331
Chr3
5′
DNA
HPV16
7906
1
AAAAA
97
E7
E6
1
AAAAA
409
E7
189935002/008
880
226
~3164 nt
2
AAAAA
E7
E6
~3347 nt
3
AAAAA
~2693 nt
4
T5350 Transcripts
T
hg38 Chr13:73,026,000-74,200,000
1174kb
73,200
73,400
73,600
73,800
74,000
kb
KLF5
LINC00392
KLF12
P97
PAS (73659870)
5′
73656686
5385
DNA
Chr13
L1
URR
E6
E7
E1
2734
73657297
Chr13
3′
7906
1
HPV16
T6050 Transcripts
1
~5210 nt
97
2734
E6
E7
AAAAA
73657297
2
~5027 nt
409
E7
AAAAA
226
U
hg38 Chr6:106,950,000-107,060,000
110kb
107,040
107,020
107,000
106,980
106,960
kb
MTRES1
CD24
P97
PAS (106982093)
3′
106994260
3464
DNA
Chr6
E2
L2
L1
URR
E6
E7
E1
E2
3520
106983241
Chr6
5′
HPV16
7906
1
T6709 Transcripts
1
~4572 nt
97
3520
E6
E7
E1
AAAAA
106983241
2
~1911 nt
409
E7
AAAAA
226
880
3358
3
~2094 nt
E6
E7
AAAAA

V

T6570 targeted DNA-seq

Order	DNA CJRs at		
	HPV18 nt	Host chr	Host chr nt
1	3631	Chr13	73350892
2	3455	Chr13	73658495
3	6264	Chr19	30548314
4	6864	Chr2	227416717

W

T6570 RNA-seq

No. of CJRs	RNA CJRs at		
	HPV18 nt	Host Chr	Host chr nt
67	3631	Chr13	73350892
50	929	Chr13	73648631
2	3284	Chr13	73655946
2	929	Chr13	73655946

X

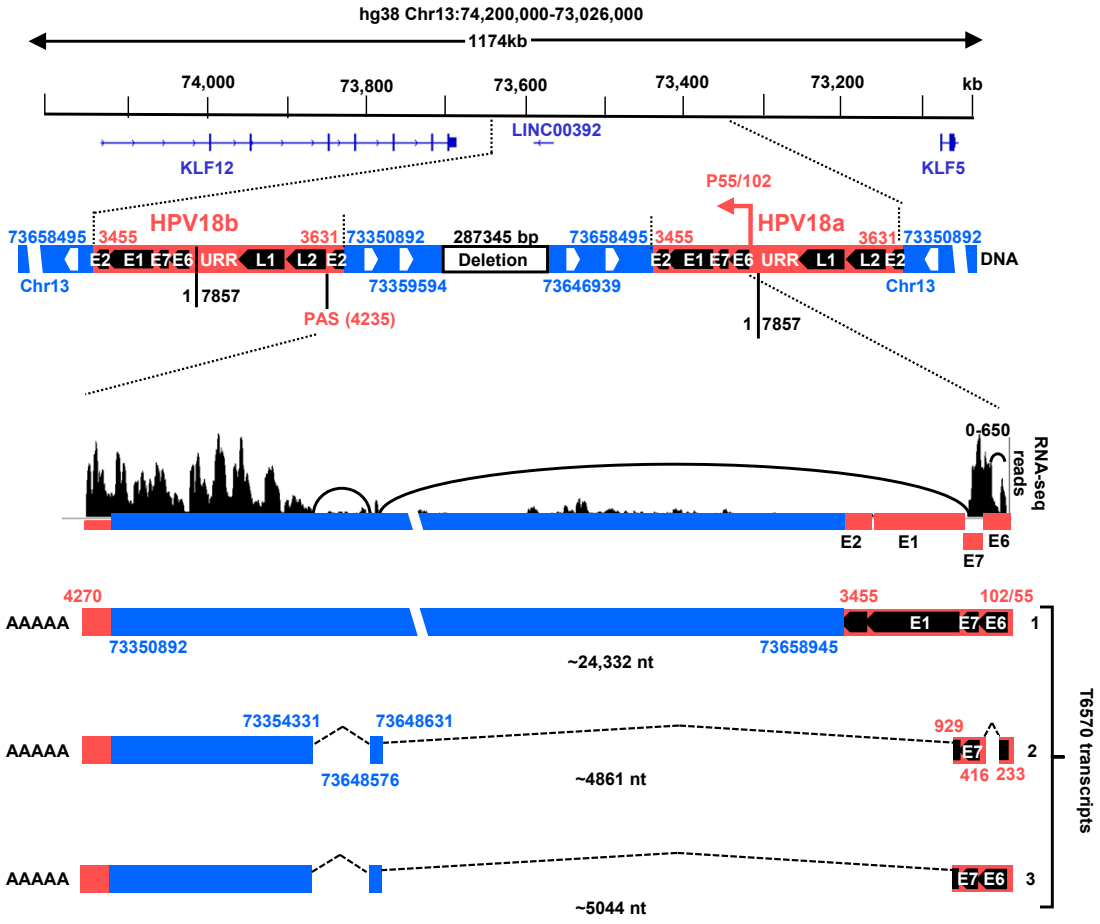
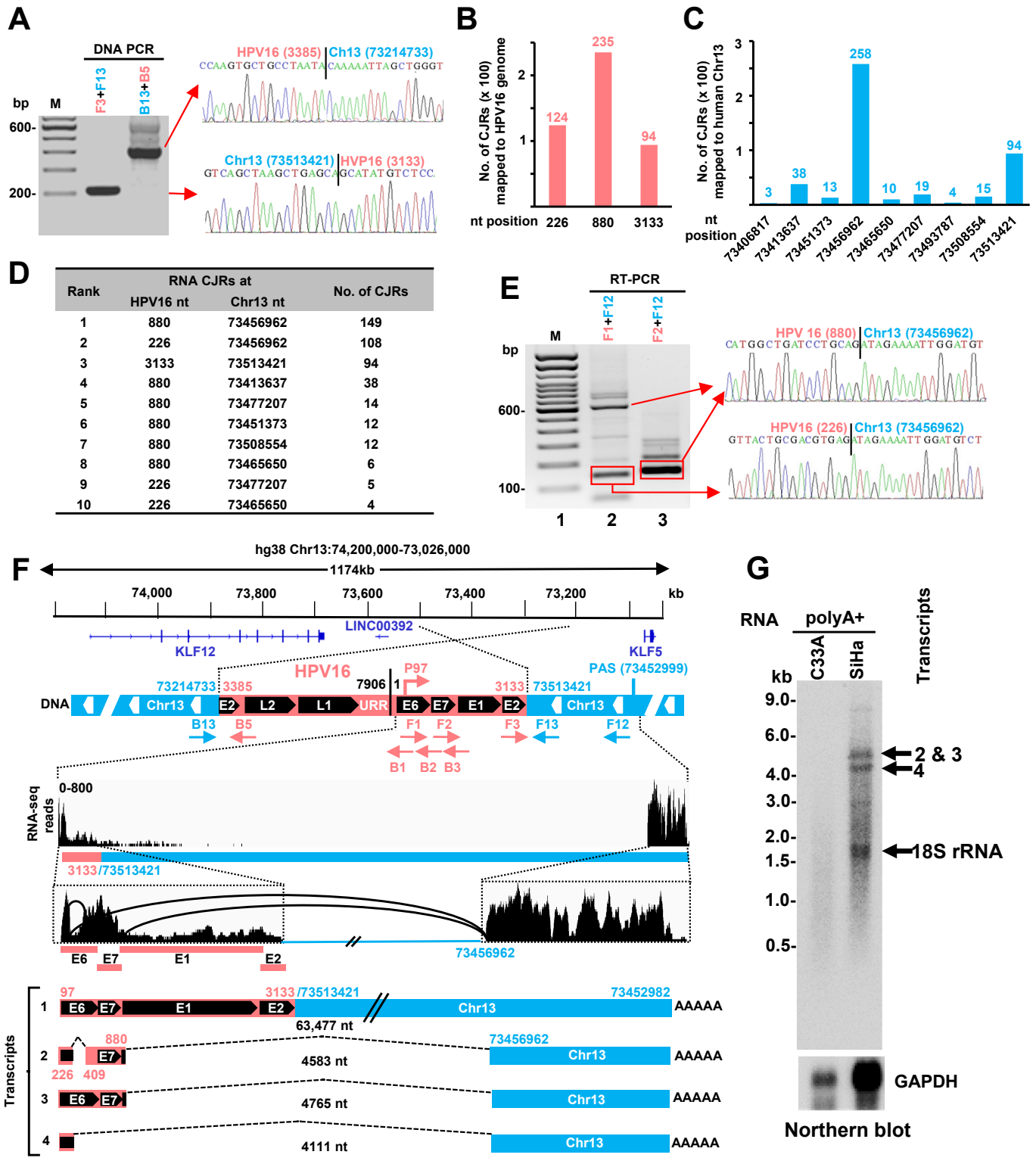


Figure S1. Identification of the mapped chromosomal HPV16 or HPV18 DNA integration sites expressing a bicistronic E6E7 RNA in cervical cancer tissues. (**A, B**) Coverage and distribution of HPV16- or HPV18-specific reads along with the viral genome in the indicated cervical cancer tissues. HPV16 DNA integration in the A3NI induced the genomic amplification of a Chr19 region covering MAG and CD22 genes. Arrows indicate the integration-mediated genome amplification or deletion. (**C-H**) E6E7 expression in the tissues A1M8 (HPV16, **C-E**) and A1MI (HPV18, **F-H**) bearing only a single integrated HPV DNA copy. The CJRs identified by both DNA-seq and RNA-seq were mapped to the HPV16 DNA integrated in the region between ENTPD1 and SORBS1 genes on Chr10 in the A1M8 (**E**) and the HPV18 DNA in the intron of LINC-PINT on Chr7 in the A1MI (**H**). The truncated viral DNA integrated in a reversed orientation expressed viral E6E7 RNA by using a viral early promoter P97 (HPV16) or P55/102 (HPV18) and a host polyadenylation signal (PAS) at Chr10 nt 95,753,225 in the A1M8 **(E**) and Chr7 nt 130,940,881 in the A1MI (**H**) to produce three virus-host chimeric transcripts from alternative RNA splicing in each tissue. (**I-U**) E6E7 expression selectively from a single integrated HPV16 DNA copy in cervical cancer tissues bearing integrated HPV16 DNAs on multiple chromosomes (Figure 1, Supplemental Tables S1-3). A bicistronic E6E7 RNA in each tissue was transcribed from a viral early promoter P97 of the integrated HPV16 DNA, alternatively spliced to produce multiple RNA isoforms, and polyadenylated by using a host PAS accessible downstream of the integrated viral DNA. A number below or on the right or left end of the isoform RNA indicates the estimated length of the RNA. URR, upstream regulatory region; nt 7906/1, tail-to-head junction. RNA-seq reads-coverage and Sashimi plots for splice junctions are visualized by IGV. In the A2RE tissue (**I-K**), DNA CJRs were mapped to multiple host chromosomes (**I**) but RNA CJRs only to a single HPV16 DNA integration site on Chr1 (**J**). The truncated HPV16 DNA was integrated in a reversed orientation in a region between UTP11 and AL139158.3 genes on Chr1 (**K**). The E6E7 RNA in the tissue expressed from the integrated HPV16 DNA was polyadenylated by using a host PAS at Chr1 nt 38,024,800 to produce four chimeric RNA transcripts. In the A1M9 tissue (**L-N**), DNA CJRs were mapped to multiple chromosomes (**L**) but the RNA CJRs only to a single chromosomal HPV DNA integration site on Chr8 (**M**). The integrated HPV16 DNA in the intron region of PVT1 on Chr8 expressed the E6E7 RNA by using a host PAS at nt 127,999,309 to produce six virus-host chimeric transcripts (**N**). In the T1074 tissue (**O-Q),** DNA CJRs were mapped to multiple host chromosomes **(O**) but RNA CJRs only to the integrated HPV16 DNA on Chr17(**P**). The truncated HPV16 DNA integrated in the last intron of the SDK2 gene on Chr17 (**Q**) expressed the E6E7 RNA by using a host PAS at Chr17 nt 73334404 to produce five chimeric virus-SDK2 RNA isoforms. In the tissues T3023 (**R**), T5350 (**S**), T6050 (**T**), and T6709 (**U**) with multiple chromosomal HPV16 DNA integration sites, E6E7 RNA in the T3023 (**R**) was expressed only from the integrated HPV16 DNA in the intron region of XR_001742622 gene on Chr5 and polyadenylated by using a host PAS at Chr5 nt 29,042,795 to produce three virus-host chimeric transcripts; in the tissue 5350 (**S**), expressed only from the integrated HPV16 DNA in an intergenic region between TP63 and P3H2 on Chr3 and terminated at a host PAS at Chr3 nt 189,937,565 to produce four virus-host chimeric transcripts; in the tissue T6050 (**T**), expressed from the integrated HPV16 DNA in an intergenic region between KLF5 and KLF12 on Chr13 and terminated at a host PAS at Chr3 nt 73,659,870 to produce two virus-host chimeric transcripts; and in the tissue T6709 (**U**), expressed from the integrated HPV16 DNA in an intergenic region between MTRES1 and CD24 on Chr6 and terminated at a host PAS at Chr6 nt 106,982,093 to produce three virus-host chimeric transcripts. (**V-X**) E6E7 expression selectively from a single integrated HPV18 DNA in the tissue T6570. DNA CJRs were mapped to multiple host chromosomes (**V**) but RNA CJRs only to a single HPV18 DNA integration site on Chr13 (**W**). (**X)** Two tandem copies of the truncated HPV18 DNA (HPV18a and HPV18b) with 176 bp deletion in the E2 region were integrated in a rearranged intergenic region between KLF12 and KLF5 on Chr13 but separated by a ~20-kb host DNA remanent from a large (~287-kb) deletion. Viral E6E7 RNA transcribed from the viral early promoter (P55/ P102) in the HPV18a and terminated at the viral early PAS at nt 4235 in the HPV18b to produce three virus-host chimeric transcripts.

Fig. S2


A
DNA PCR
F3+F13
B13+B5
bp
M
600-
200-
HPV16 (3385) | Ch13 (73214733)
CCAAGTGCTGCCTAATACAAAAATTAGCTGGGT
Chr13 (73513421) | HVP16 (3133)
GTCAGCTAAGCTGAGCAGCATATGTCTCC
B
No. of CJRs (x 100) mapped to HPV16 genome
124
235
94
nt position
226
880
3133
C
No. of CJRs (x 100) mapped to human Chr13
3
38
13
258
10
19
4
15
94
nt position
73406817
73413637
73451373
73456962
73465650
73477207
73493787
73508554
73513421
D
Rank
RNA CJRs at
HPV16 nt
Chr13 nt
No. of CJRs
1 880 73456962 149
2 226 73456962 108
3 3133 73513421 94
4 880 73413637 38
5 880 73477207 14
6 880 73451373 12
7 880 73508554 12
8 880 73465650 6
9 226 73477207 5
10 226 73465650 4
E
RT-PCR
F1+F12
F2+F12
M
bp
600-
100-
1
2
3
HPV 16 (880) | Chr13 (73456962)
CATGGCTGATCCTGCAGATAGAAAATTGGATGT
HPV16 (226) | Chr13 (73456962)
GTTACTGCGACGTGAGATAGAAAATTGGATGTCT
F
hg38 Chr13:74,200,000-73,026,000
1174kb
74,000
73,800
73,600
73,400
73,200
kb
LINC00392
KLF12
KLF5
PAS (73452999)
HPV16
P97
7906
1
73214733
3385
3133
73513421
DNA
Chr13
E2
L2
L1
URR
E6
E7
E1
E2
Chr13
B13
B5
F1
F2
F3
F13
F12
B1
B2
B3
RNA-seq reads
0-800
3133/73513421
E6
E7
E1
E2
73456962
Transcripts
1
97
3133/73513421
73452982
E6
E7
E1
E2
Chr13
AAAAA
63,477 nt
2
880
73456962
E7
Chr13
AAAAA
226
409
4583 nt
3
E6
E7
Chr13
AAAAA
4765 nt
4
Chr13
AAAAA
4111 nt
G
RNA
polyA+
C33A
SiHa
Transcripts
kb
9.0-
6.0-
4.0-
3.0-
2.0-
1.5-
1.0-
0.5-
2 & 3
4
18S rRNA
GAPDH
Northern blot

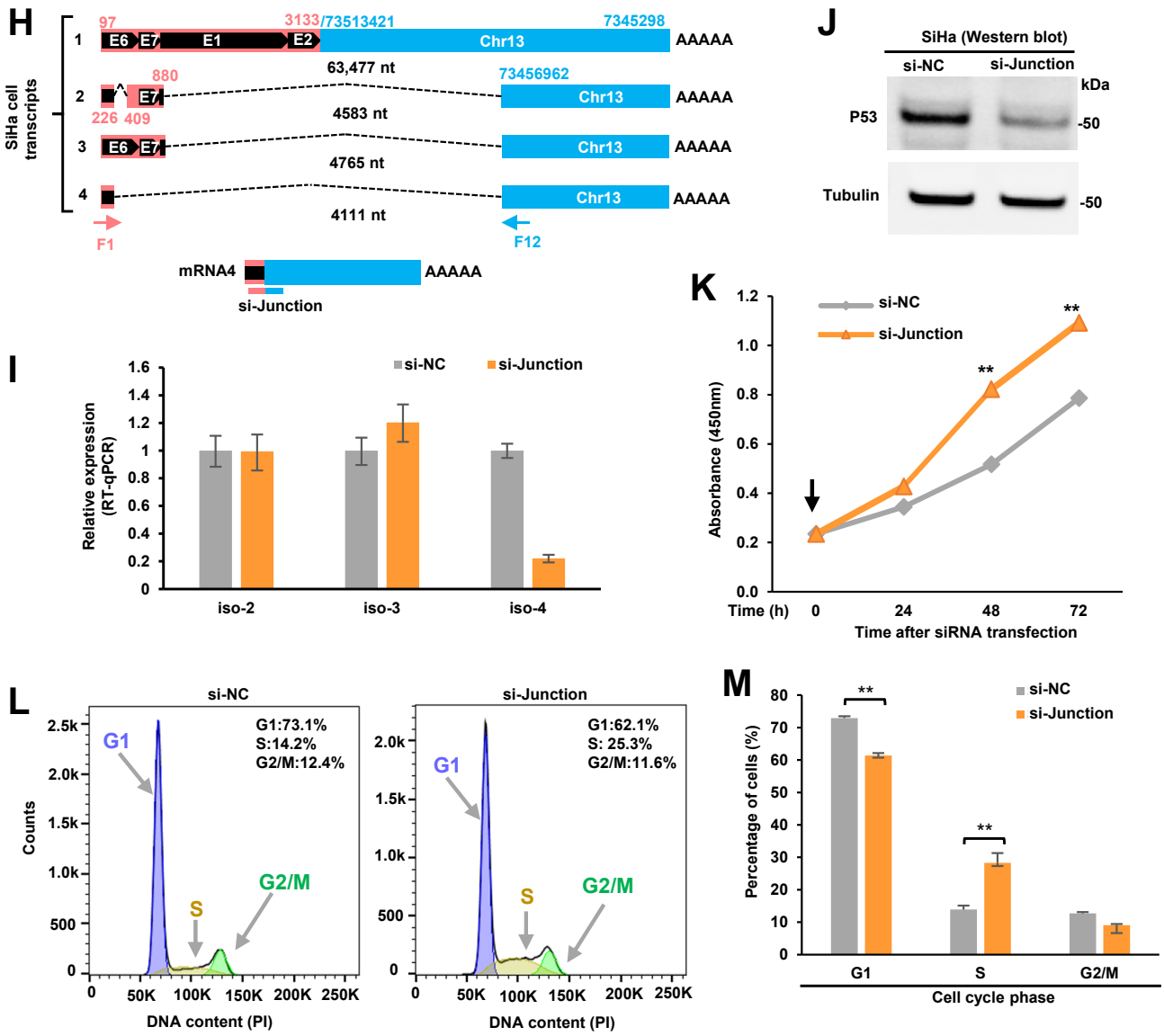

H
SiHa cell transcripts
97
3133/73513421
7345298
E6
E7
E1
E2
Chr13
AAAAA
880
63,477 nt
73456962
226 409
4583 nt
4765 nt
4111 nt
F1
F12
mRNA4
si-Junction
J
SiHa (Western blot)
si-NC
si-Junction
kDa
P53
Tubulin
-50
I
si-NC
si-Junction
Relative expression (RT-qPCR)
iso-2
iso-3
iso-4
K
Absorbance (450nm)
Time (h)
Time after siRNA transfection
L
si-NC
si-Junction
G1:73.1%
S:14.2%
G2/M:12.4%
G1:62.1%
S: 25.3%
G2/M:11.6%
G1
S
G2/M
Counts
DNA content (PI)
M
Percentage of cells (%)
Cell cycle phase

Figure S2. Bicistronic E6E7 RNA expression from a single integrated viral DNA copy in HPV16-positive SiHa cells and potential function of the virus-host chimeric transcript-4. (**A**) Two integration junctions of the HPV16 DNA integration site were verified by PCR (see two primer sets F3+F13 and B13+B5 in **F**) and by Sanger-sequencing, with the sequenced integration junction on the right. (**B-C**) Virus-host RNA CJRs were mapped to the HPV16 genome (**B**) and to host Chr13 (**C**). **(D**) Distribution of the top 10 RNA CJRs mapped to HPV16 and host Chr13. (**E**) Validation of two RNA splice junctions by RT-PCR using two primer sets F1+F12 and F2+F12 in **F** and Sanger-sequencing, with the sequenced splice junctions on the right. (**F**) The illustration shows the truncated HPV16 DNA integrated in a rearranged region (dashed lines) on Chr13 between nt 73,214,733 and nt 73,513,421 (top), with the primers used for validation of the integrated viral DNA. Total RNA-seq reads-coverage along with the chimeric host-virus genome region and Sashimi plots for splice junctions are visualized by IGV (middle). The integrated viral DNA expresses the E6E7 RNA by using the viral early promoter P97 and a host PAS at Chr13 nt 73,452,999 to produce four virus-host chimeric RNA isoforms (bottom). The number in nt below each isoform RNA indicates the estimated length of the RNA without a pA tail. (**G**) Identification of the major viral RNA transcripts in SiHa cells by Northern blot analysis using polyA+ RNA selected from 100 μg of total cell RNA. A mixture of three viral probes (B1, B2 and B3) located in the HPV16 E6 and E7 regions (**F**) were used for Northern blot analysis. RNA from HPV-negative C33A cells served as a negative control. GAPDH RNA served as a sample loading control. (**H-M**) Reduced expression of a viral-host fusion transcript (transcript-4) promotes the proliferation of SiHa cells by increasing cell cycle G1 to S phase transition. **(H**) Major viral-host fusion transcripts in SiHa cells and detection strategies by RT-PCR with the indicated primers (arrows). A synthetic siRNA was designed to specifically target transcript-4 (si-Junction). (**I-K**), Reduced expression of the transcript-4 level in SiHa cells led to a reduction of p53 protein and enhancement of cell proliferation. SiHa cells were harvested at 48 h after siRNA transfection. The knockdown efficiency of the transcript-4 was confirmed by RT-qPCR using a virus-host specific junction TaqMan probe (**I**). The effects of the knockdown of transcript-4 on p53 expression was measured by Western blot with tubulin serving as a sample loading control (**J**). (**K**) SiHa cell proliferation was examined at the indicated times (h) after transfection of 40 nM of the indicated siRNAs (arrow) by counting cell number (mean ± SE from three independent experimental repeats). ** indicates $p<0.01$ by unpaired, two-tailed Student's *t* test. (**L** and **M**) Reduction of transcript-4 expression by the junction-specific siRNA promotes cell entry into S phase. Flow cytometry analysis of the fraction of cells in different phases of the cell cycle was performed after indicated siRNA knockdown. SiHa cells 24 h after plating were transfected with 40 nM of the indicated siRNAs and harvested for flow cytometry 48 h after siRNA transfection. Data are from a representative of three separate experiments, each in triplicate (**L**). Percentage of cells in the indicated cell cycle after siRNA knockdown of the transcript-4 expression are shown as bar graphs (**M**). Data are mean ± SD; $n=3$. **, $p<0.05$ by unpaired, two-tailed Student's *t* test.

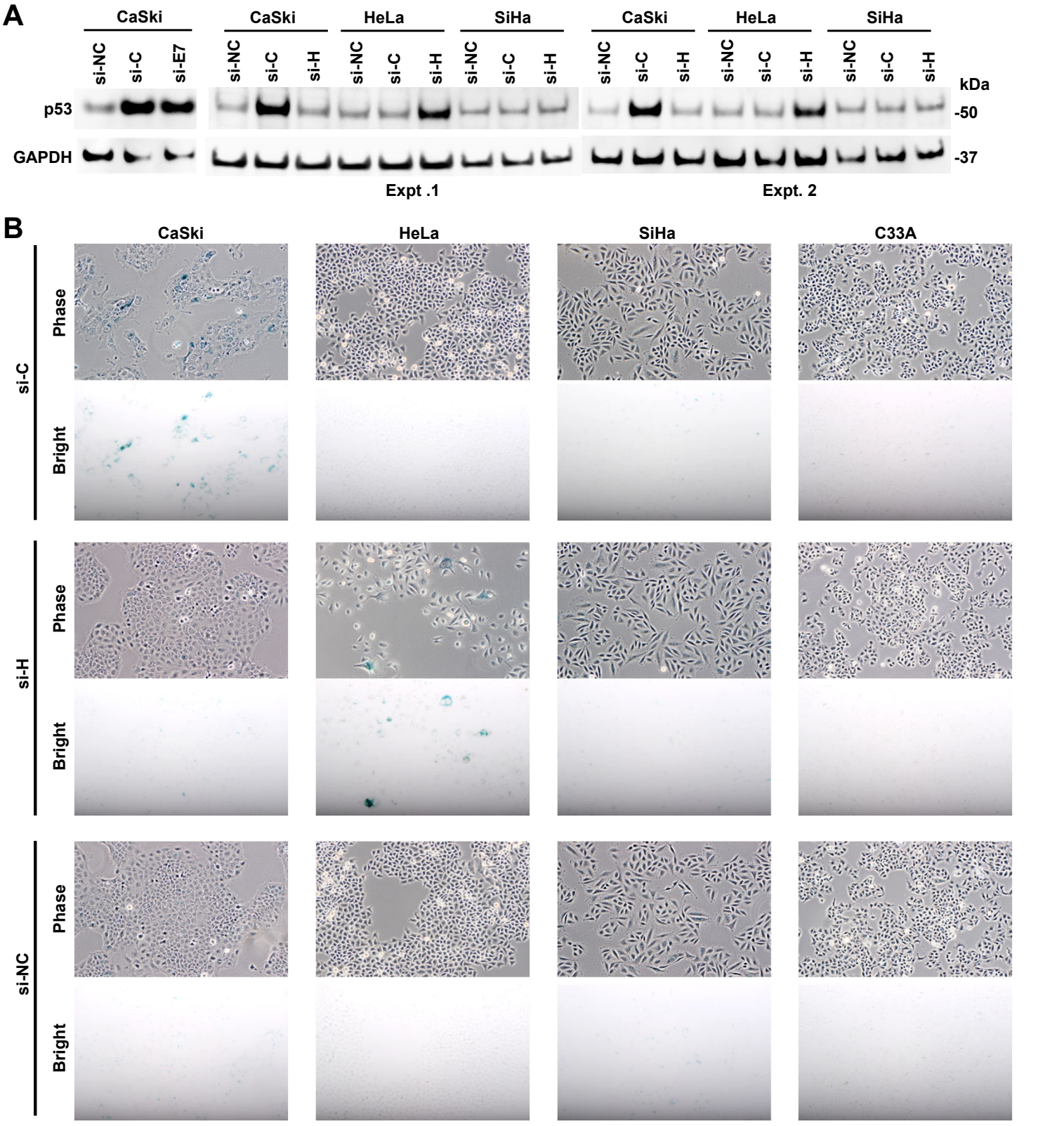
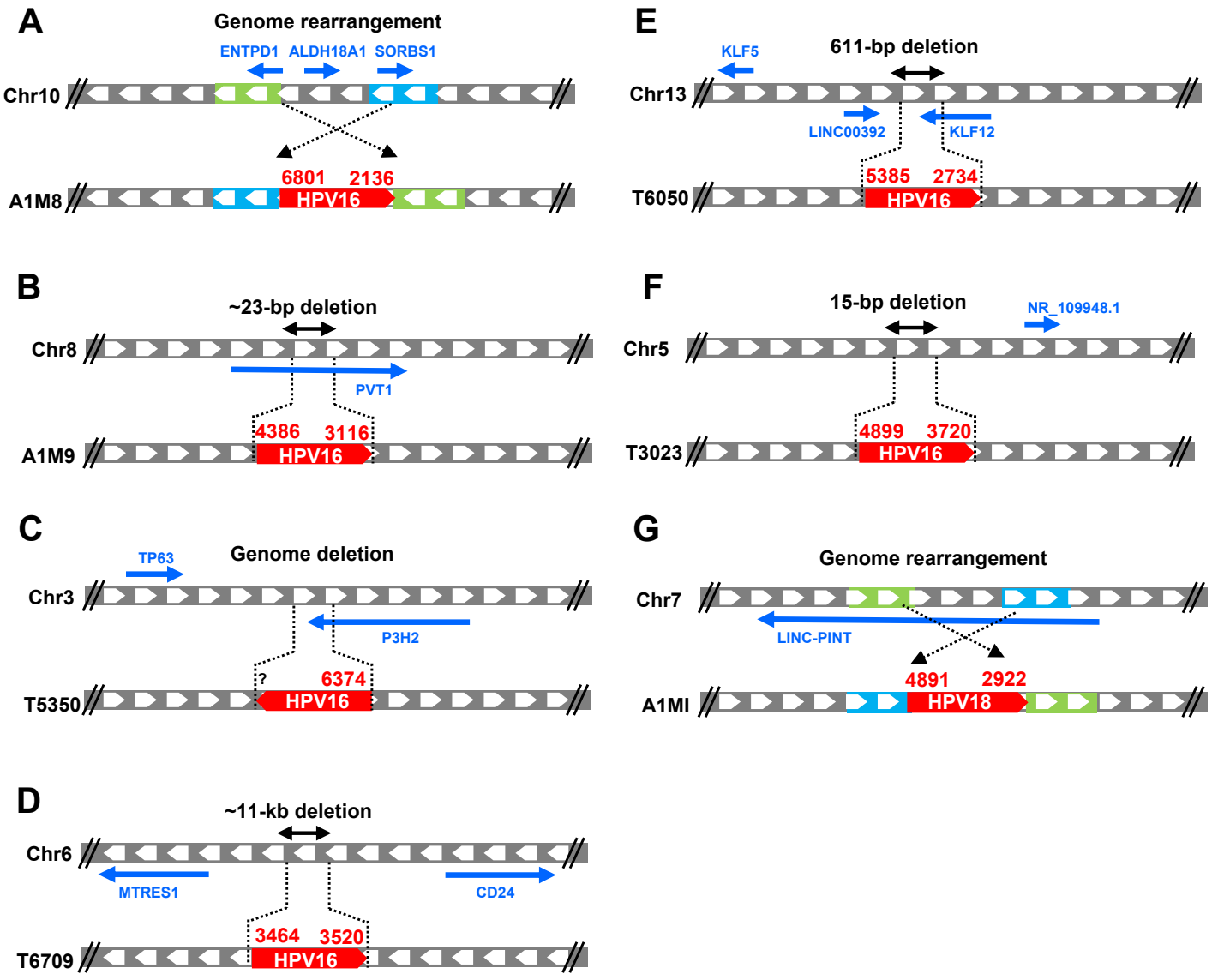
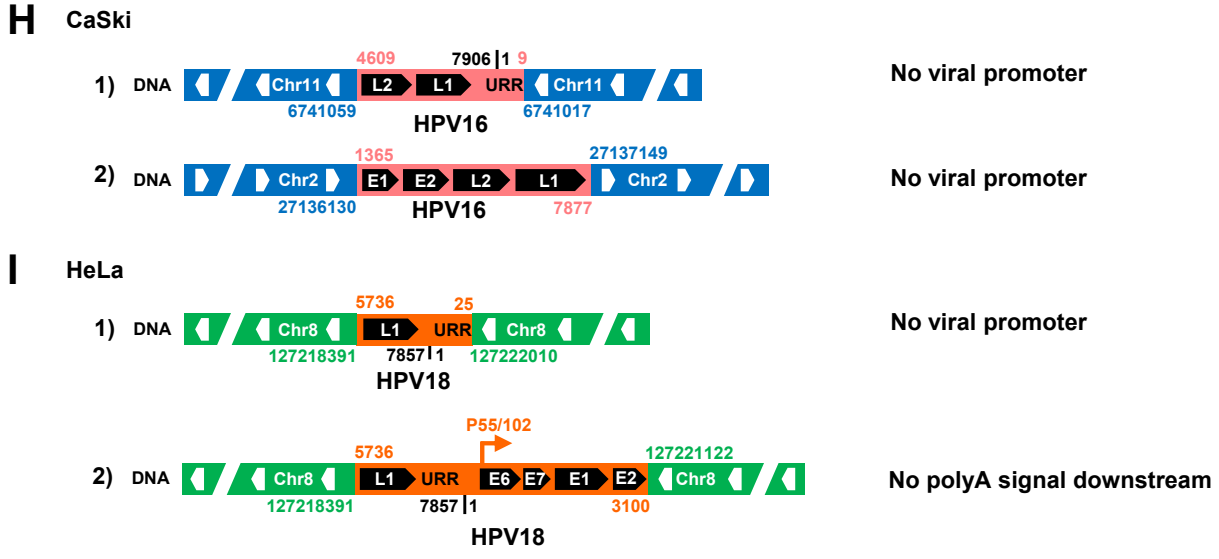


Figure S3. A siRNA targeting the host portion of the virus-host chimeric E6E7 RNA stabilizes p53 and promotes cell senescence in CaSki and HeLa cells, but not in SiHa or C33A cells. (A) Immunoblot of p53 protein expression in CaSki, HeLa and SiHa cells upon indicated siRNAs: si-NC, non-targeting siRNA control; si-C, CaSki-specific; si-E7, HPV16 E7-specific siRNA 198 (6); si-H, HeLa-specific. **(B)** Representative images of β-galactosidase (β-gal) staining. Cells were transfected twice with 40 nM of the indicated siRNAs before β-gal staining. The cells were plated on day 1, transfected with the siRNAs on day 2 and passed on day 4, and transfected again with the siRNAs on day 5. The β-gal staining was performed on day 8. The stained cells were examined on day 9 by both phase field and bright field microscopy.

Viral E6E7 expressible integration sites

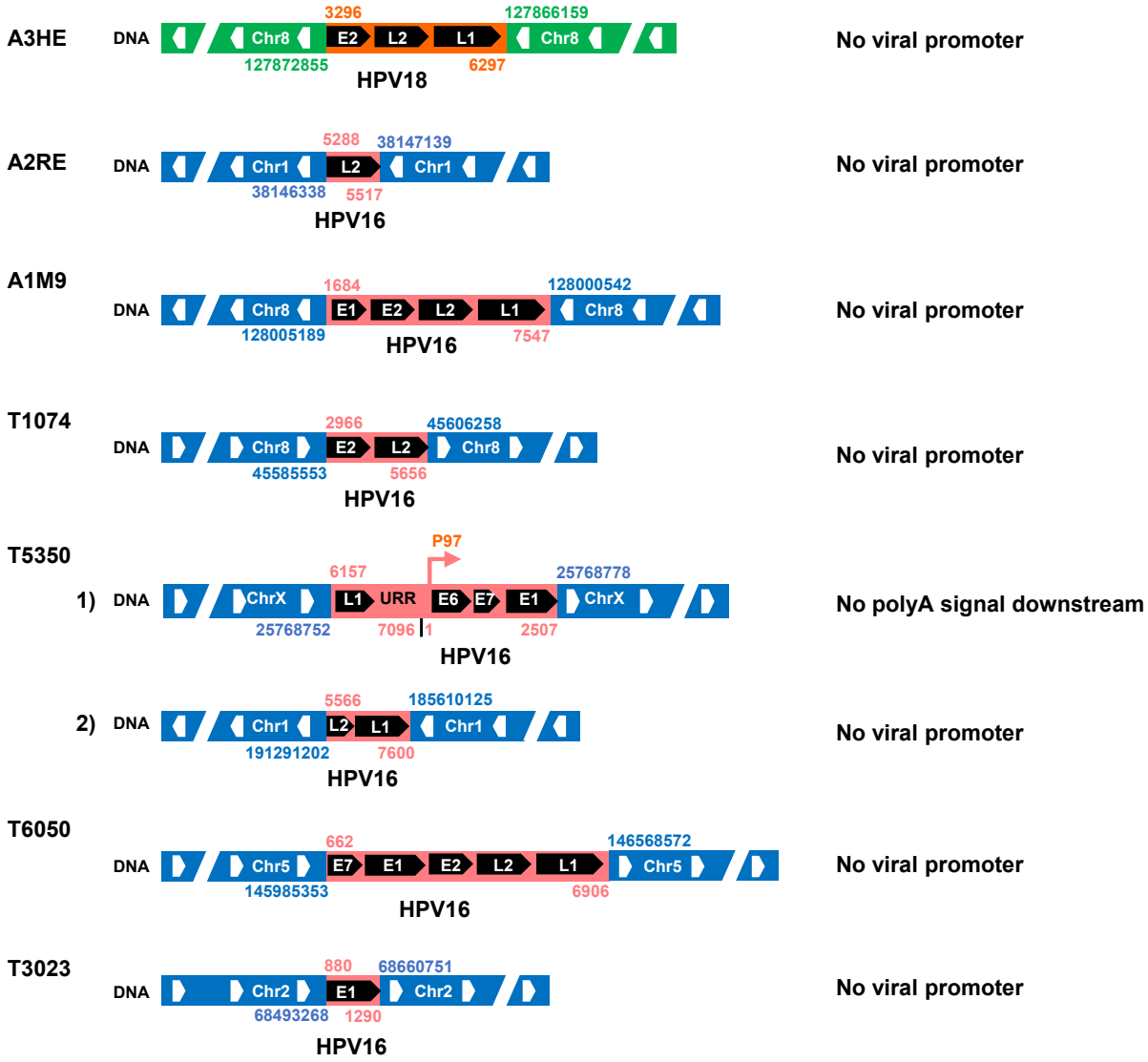


Viral E6E7 non-expressible integration sites



Viral E6E7 non-expressible integration sites

J Cervical cancer tissues



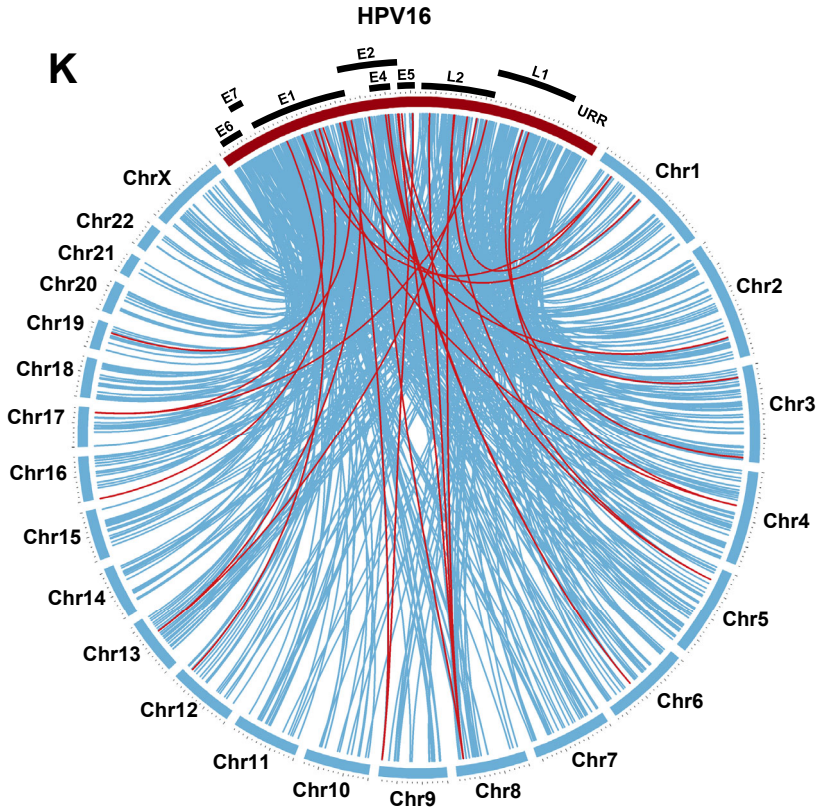


Figure S4. Simplified structures of viral E6E7 expressible or non-expressible HPV16 and HPV18 integration sites in cervical cancer tissues and derived cell lines. (A-G) E6E7 expressible HPV16 (**A-F**) and HPV18 (**G**) DNA integration sites**. (H-J)** Examples of E6E7 non-expressible HPV16 and HPV18 integration sites in CaSki (**H**), HeLa (**I**) and cervical cancer tissues (**J**). No viral promoter means no detectable viral early promoter P97 from the integrated HPV16 or P55/102 from HPV18 DNA. No PAS for the integrated viral DNA means an optimal poly A signal (UUAUUU or its equivalents) not available 10 kb downstream of the integrated viral DNA by the online Poly(A) Finder tool (http://dnafsminer.bic.nus.edu.sg/PolyA.html). An optimal poly(A) signal within the searched region must be determined by a matrix score ≥0.5 and the presence of additional UGUA motifs upstream and G/U- or U-rich downstream elements (DSE) of the PAS. (**K**) Genome-wide distributions of HPV16 integration events in 17 HPV16-positive cervical cancer tissues. A total of 525 expressible and non-expressible viral DNA integration junctions were mapped from the HPV16 genome (red) to individual human chromosomes (blue). Relative viral ORF positions are shown above the linearized HPV16 genome in a scale of 100,000x. URR, upstream regulatory region. The size of human chromosomes (Chr1-22 + ChrX) is in proportion to their length. Each “tick” in an internal circle indicates the genome positions of a detected virus-host DNA junction (red, expressible; blue non-expressible) from the virus to host genome. The circular map was created using Circos package (doi:10.1101/gr.092759.109).

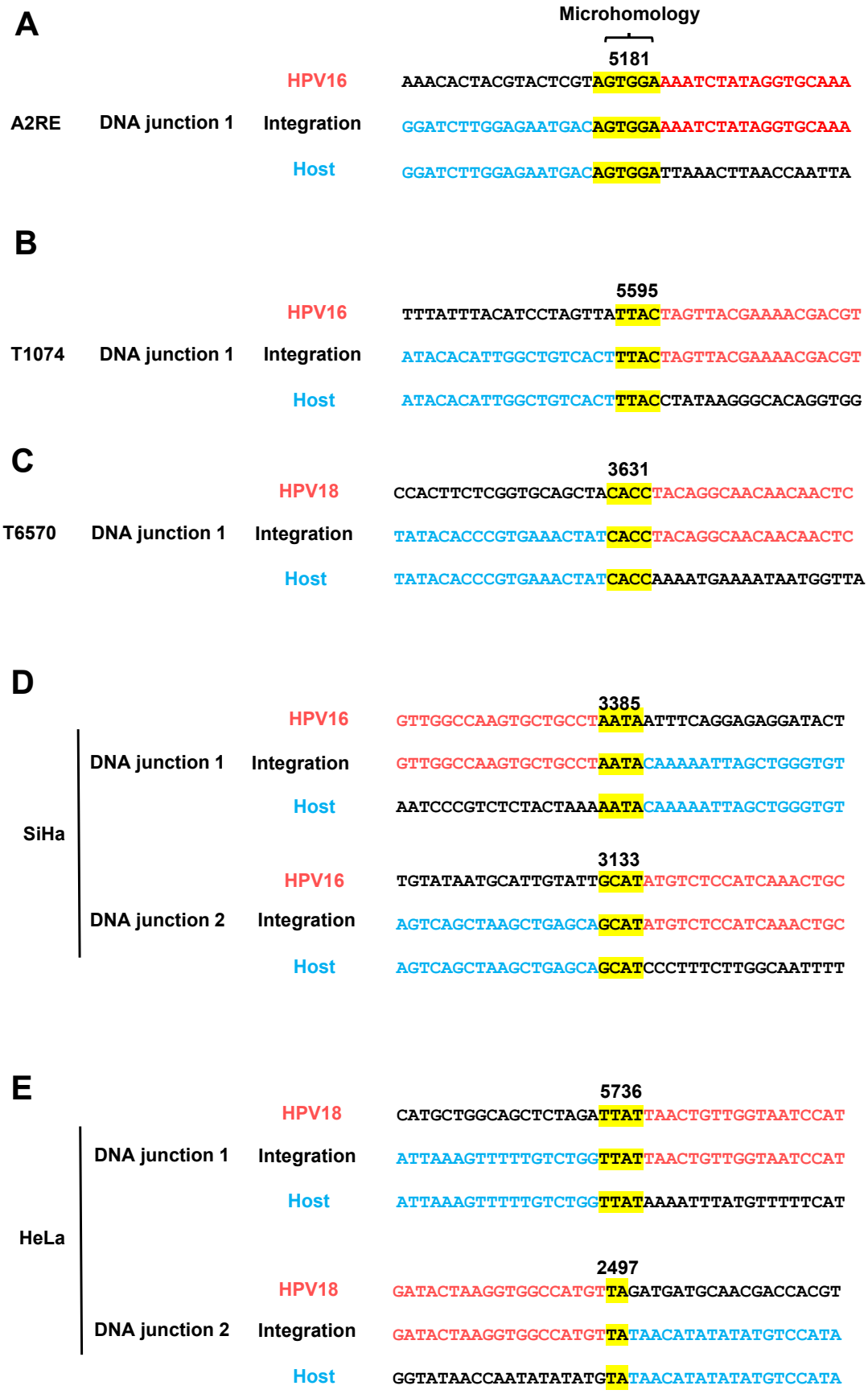
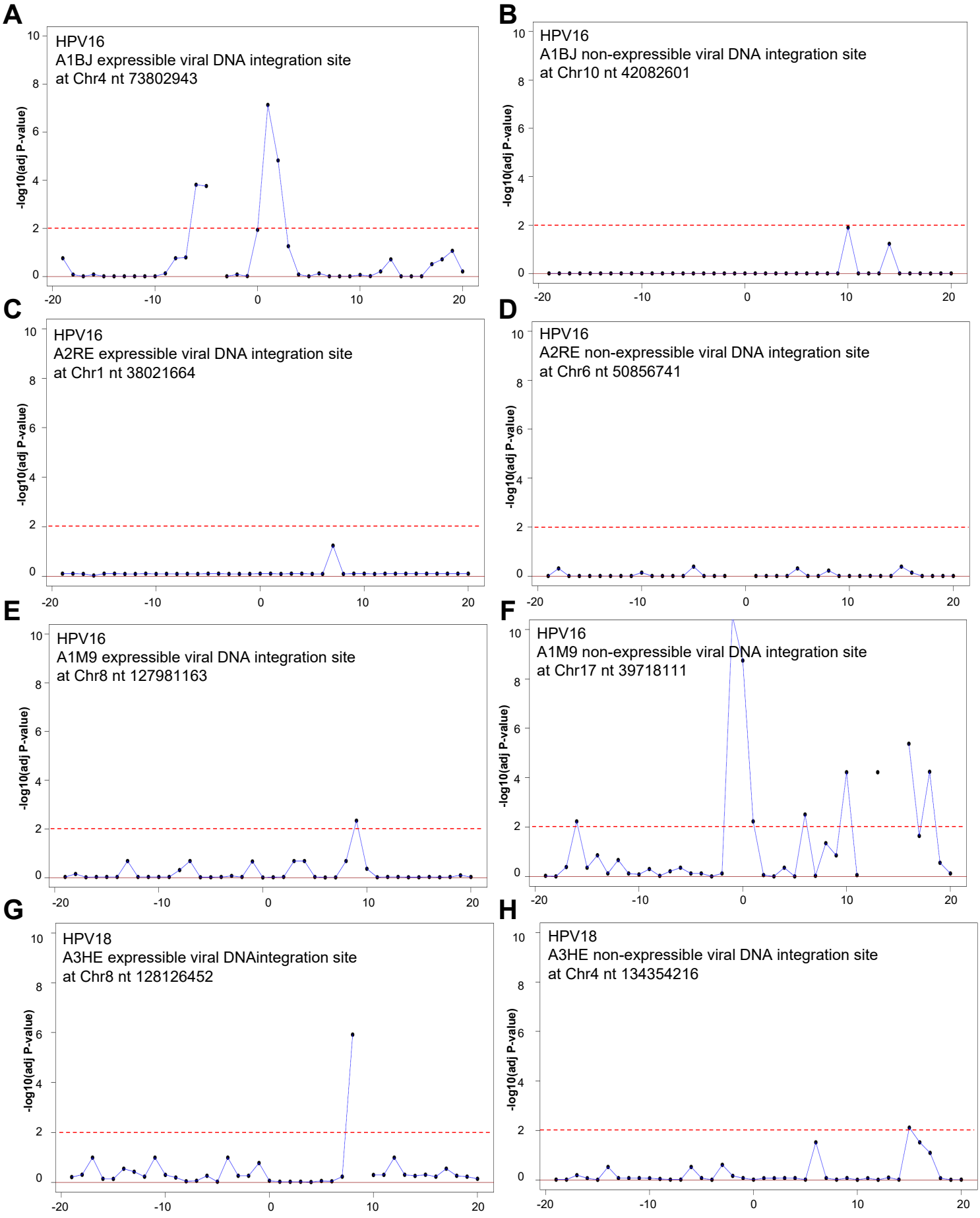
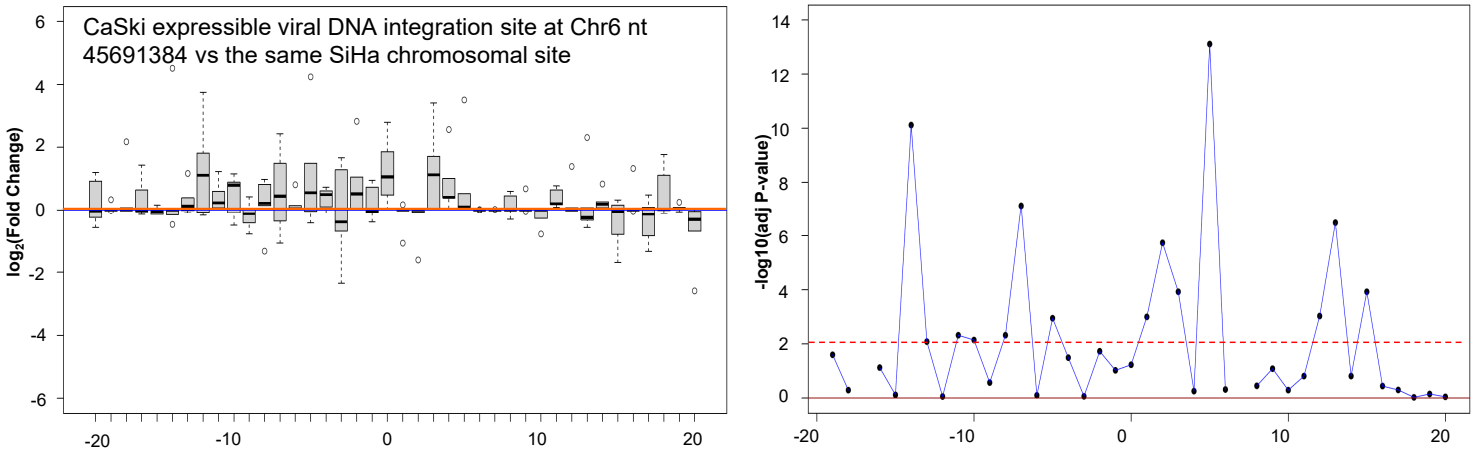


Figure S5. Regions of the microhomology sequence (MHs) were identified at virus-host integration junctions in the cervical cancer tissues A2RE (A), T1074 (B) and T6570 (C) and in SiHa (D) and HeLa (E) cells. Yellow color shows the MHs at the integration junction between HPV16/18 and human (hg38) reference genomes. Number above the yellow color indicates the last nt position in the virus genome at the integration junction.

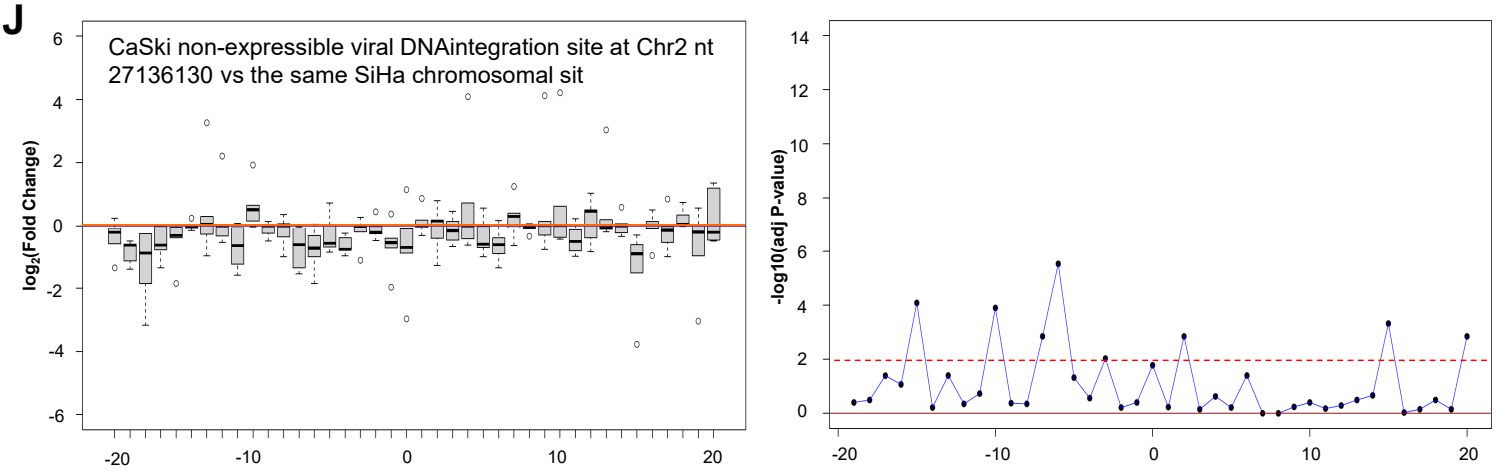
Fig. S6


A
HPV16
A1BJ expressible viral DNA integration site
at Chr4 nt 73802943
B
HPV16
A1BJ non-expressible viral DNA integration site
at Chr10 nt 42082601
C
HPV16
A2RE expressible viral DNA integration site
at Chr1 nt 38021664
D
HPV16
A2RE non-expressible viral DNA integration site
at Chr6 nt 50856741
E
HPV16
A1M9 expressible viral DNA integration site
at Chr8 nt 127981163
F
HPV16
A1M9 non-expressible viral DNA integration site
at Chr17 nt 39718111
G
HPV18
A3HE expressible viral DNAintegration site
at Chr8 nt 128126452
H
HPV18
A3HE non-expressible viral DNA integration site
at Chr4 nt 134354216
-log10(adj P-value)

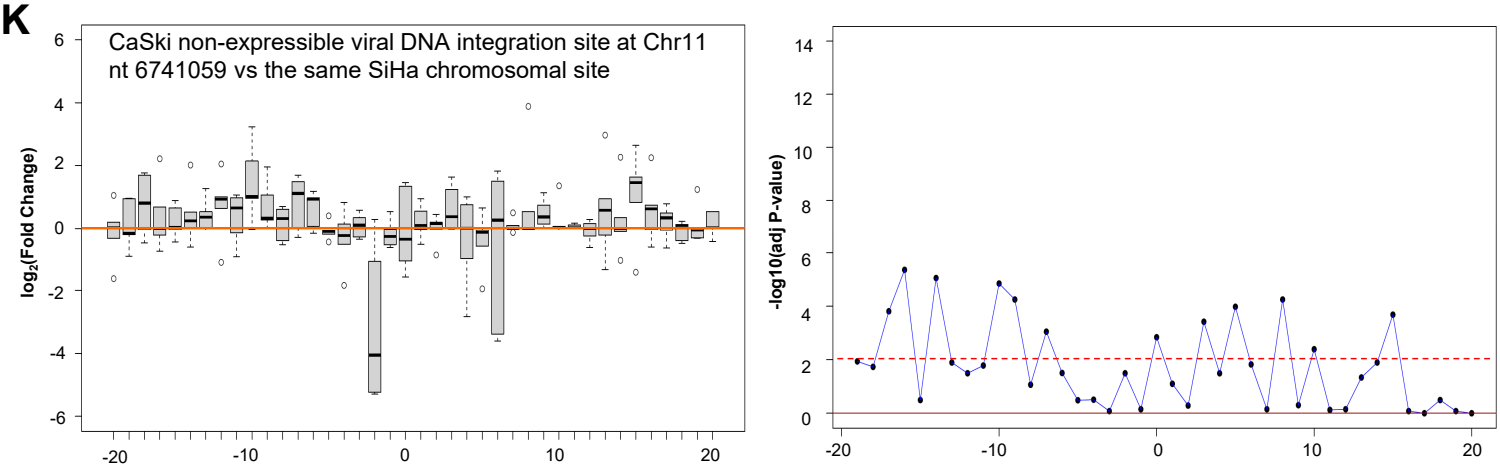
I


CaSki expressible viral DNA integration site at Chr6 nt 45691384 vs the same SiHa chromosomal site
log2(Fold Change)
-log10(adj P-value)

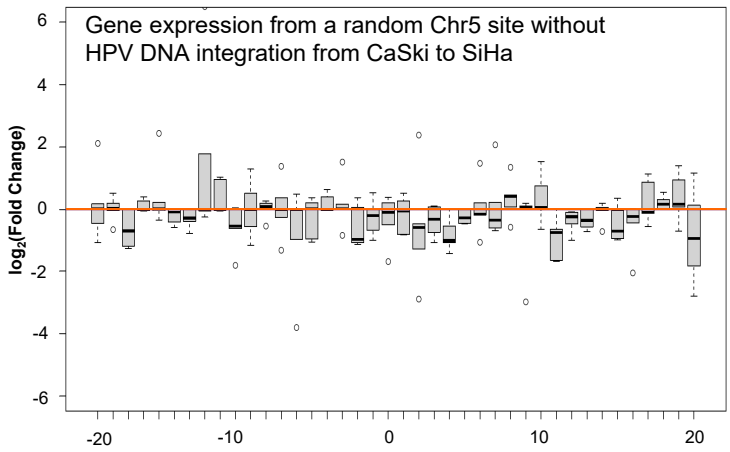
J

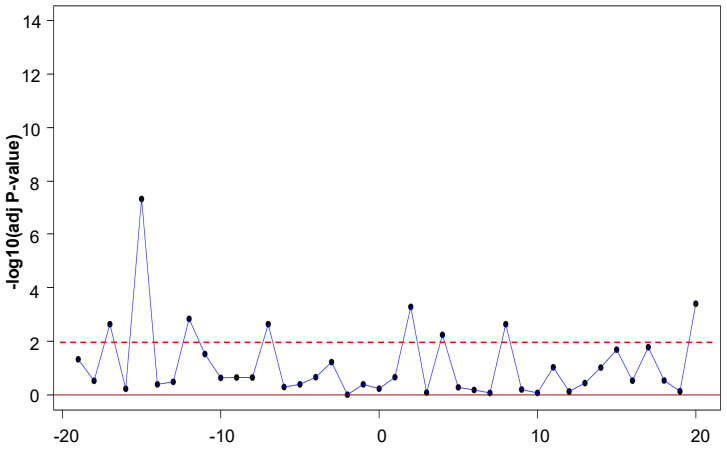

CaSki non-expressible viral DNAintegration site at Chr2 nt 27136130 vs the same SiHa chromosomal sit
log2(Fold Change)
-log10(adj P-value)

K

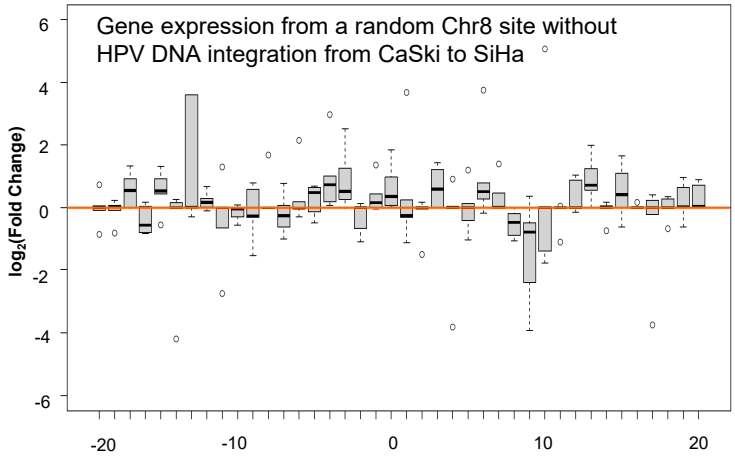

CaSki non-expressible viral DNA integration site at Chr11 nt 6741059 vs the same SiHa chromosomal site
log2(Fold Change)
-log10(adj P-value)

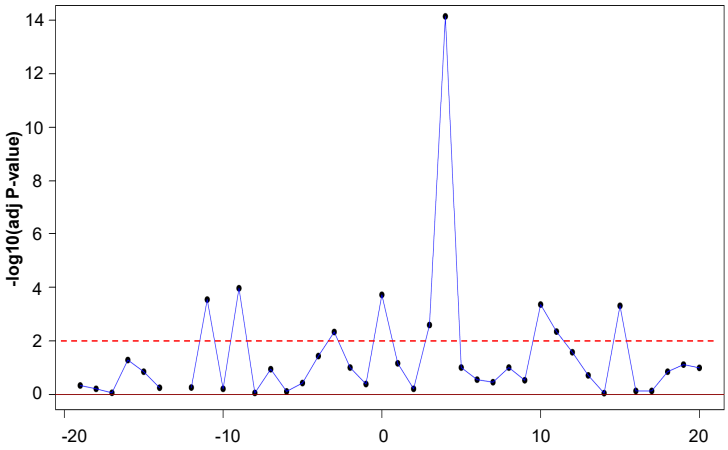
L


Gene expression from a random Chr5 site without
HPV DNA integration from CaSki to SiHa
log2(Fold Change)

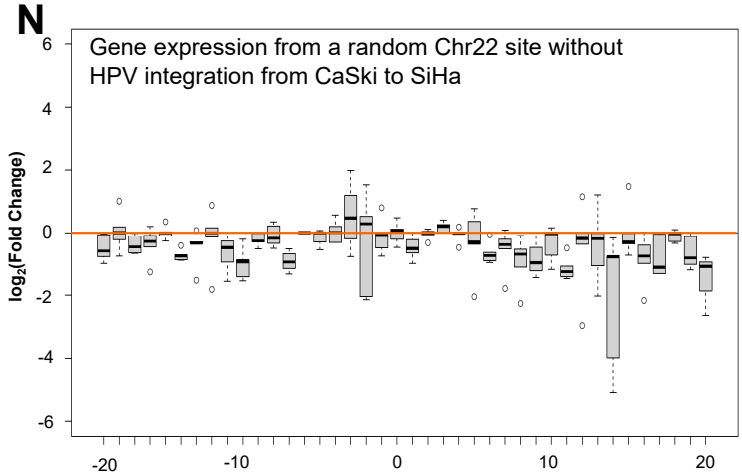

-log10(adj P-value)

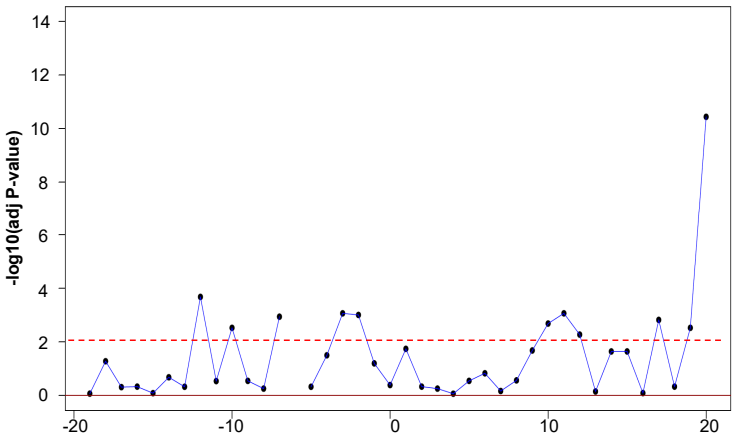
M


Gene expression from a random Chr8 site without
HPV DNA integration from CaSki to SiHa
log2(Fold Change)

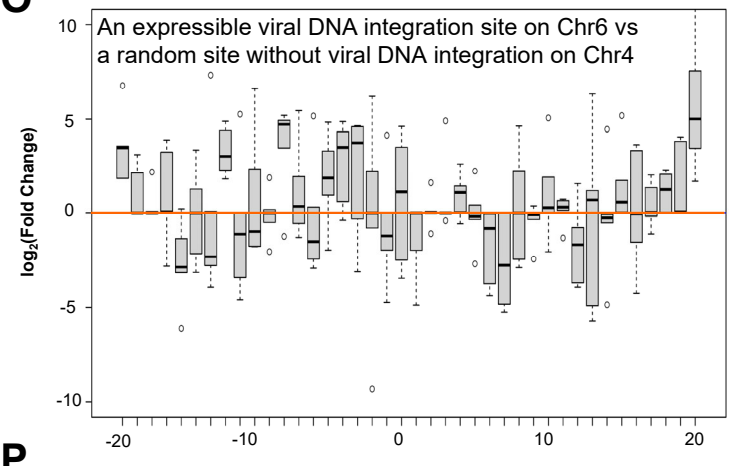

-log10(adj P-value)

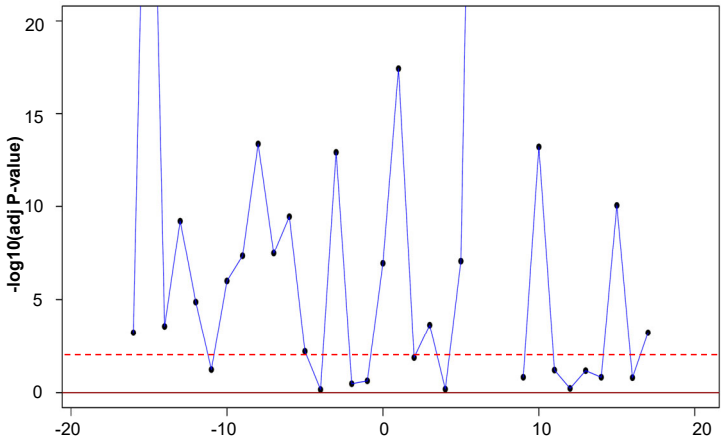
N


Gene expression from a random Chr22 site without
HPV integration from CaSki to SiHa
log2(Fold Change)

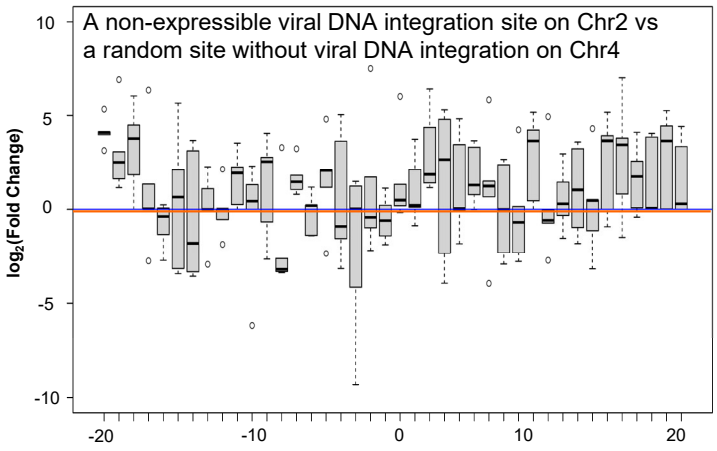

-log10(adj P-value)

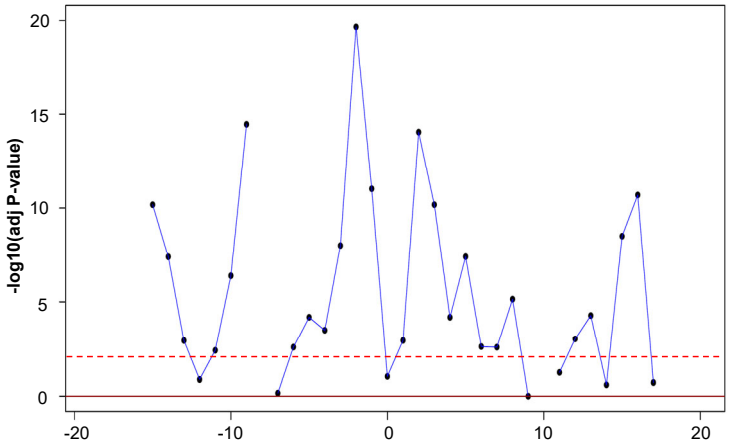
O


An expressible viral DNA integration site on Chr6 vs
a random site without viral DNA integration on Chr4
log2(Fold Change)

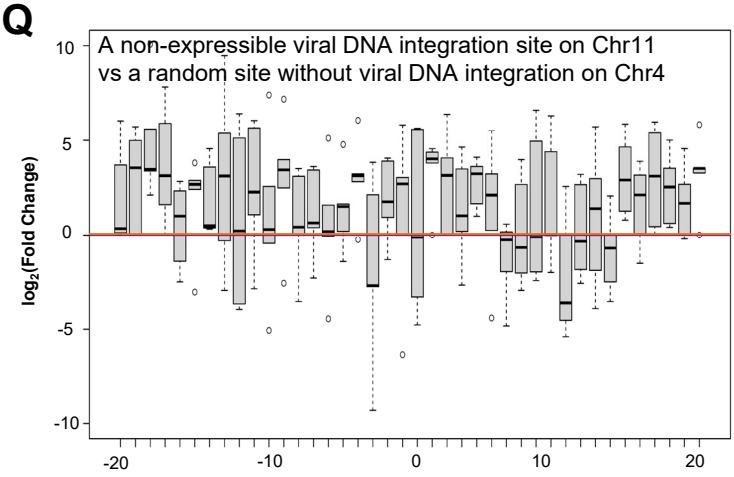

-log10(adj P-value)

P


A non-expressible viral DNA integration site on Chr2 vs
a random site without viral DNA integration on Chr4
log2(Fold Change)


-log10(adj P-value)

Q


A non-expressible viral DNA integration site on Chr11
vs a random site without viral DNA integration on Chr4
log2(Fold Change)

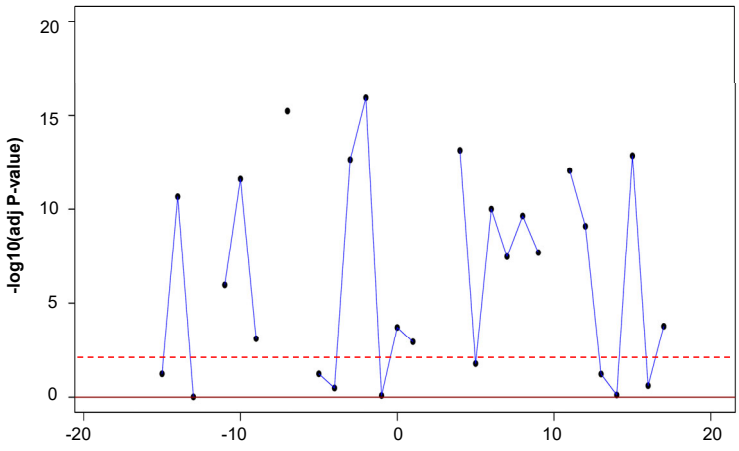

-log10(adj P-value)

Fig. S6

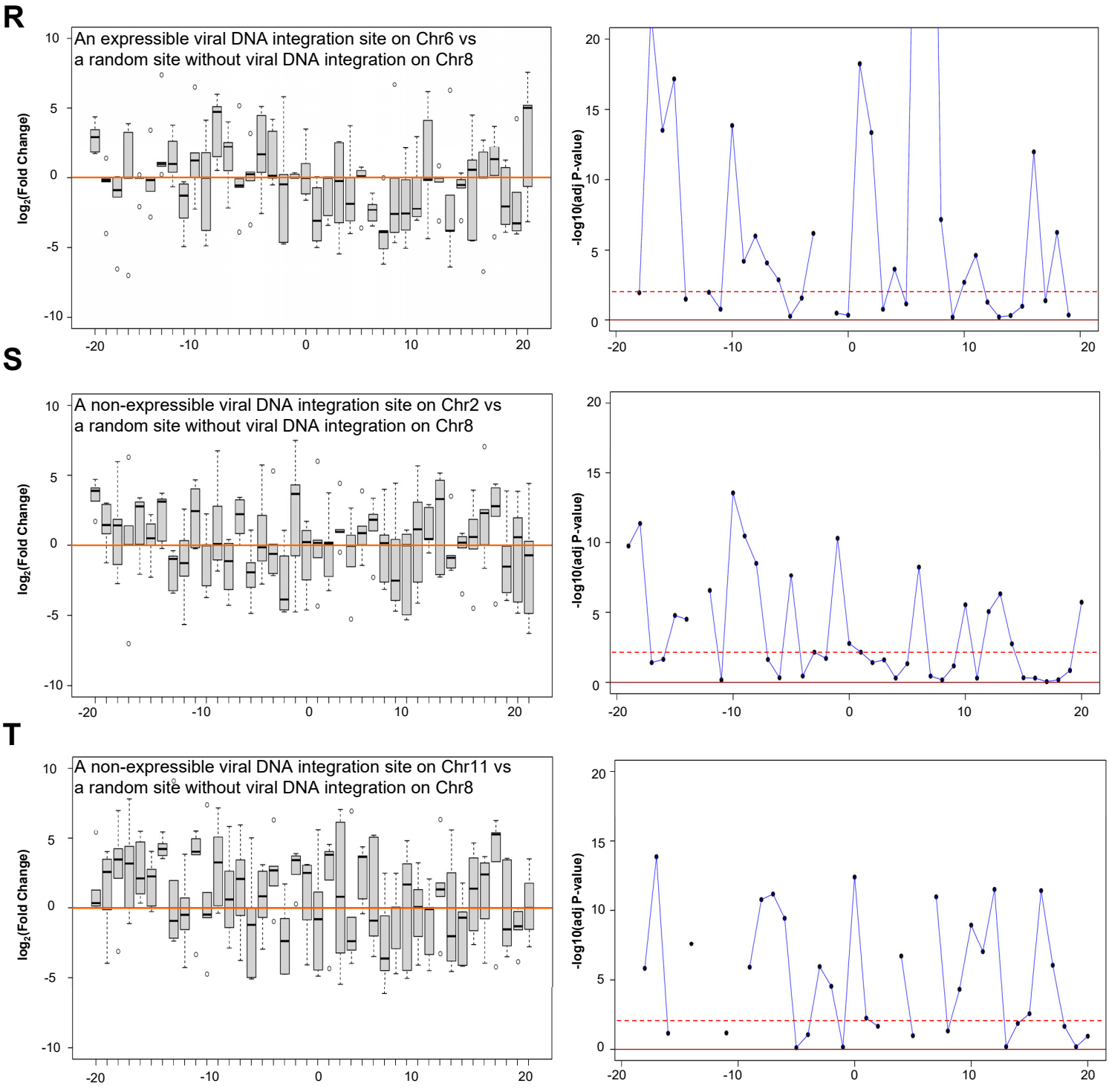
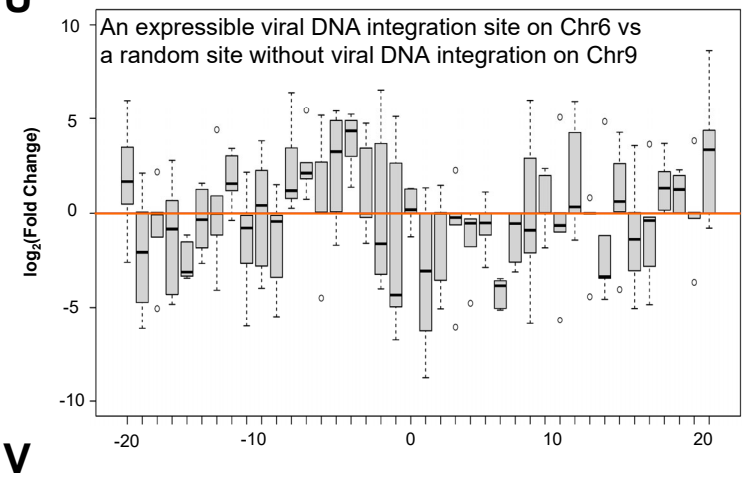
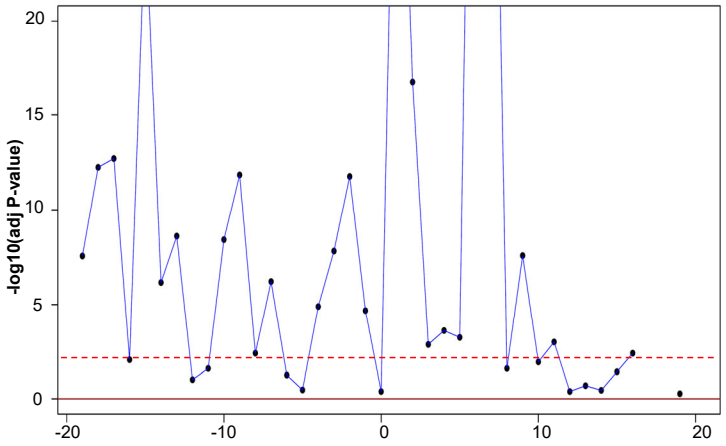

R
An expressible viral DNA integration site on Chr6 vs
a random site without viral DNA integration on Chr8
log2(Fold Change)
-log10(adj P-value)
S
A non-expressible viral DNA integration site on Chr2 vs
a random site without viral DNA integration on Chr8
log2(Fold Change)
-log10(adj P-value)
T
A non-expressible viral DNA integration site on Chr11 vs
a random site without viral DNA integration on Chr8
log2(Fold Change)
-log10(adj P-value)

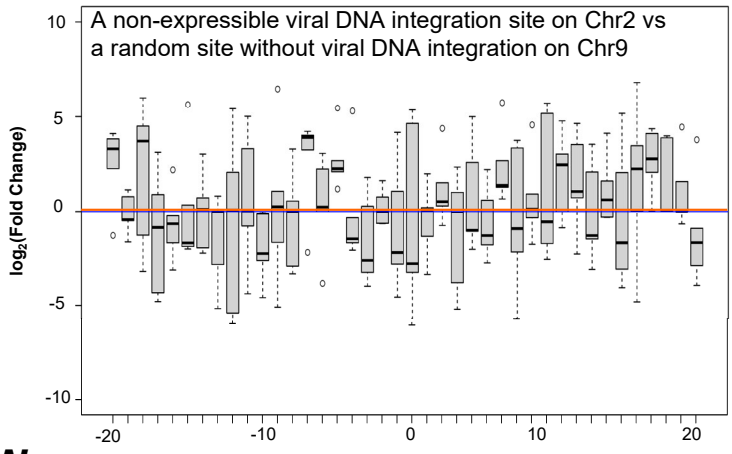
Fig. S6

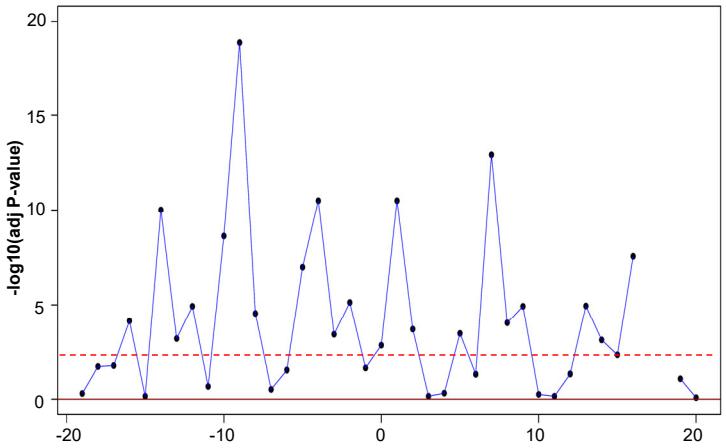
U


An expressible viral DNA integration site on Chr6 vs
a random site without viral DNA integration on Chr9
log2(Fold Change)

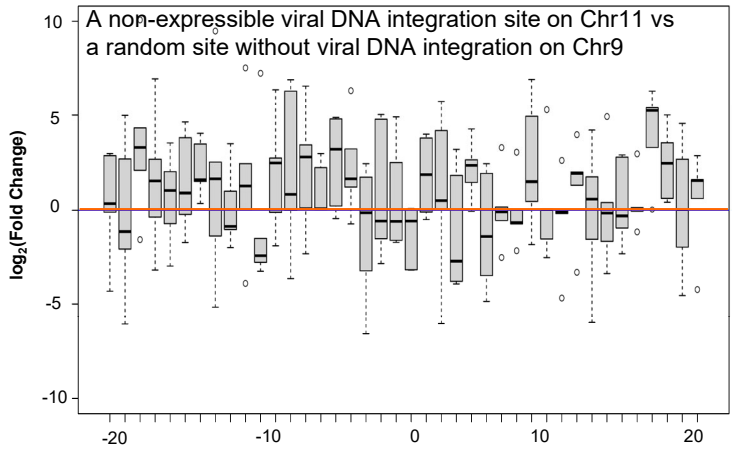

-log10(adj P-value)

V


A non-expressible viral DNA integration site on Chr2 vs
a random site without viral DNA integration on Chr9
log2(Fold Change)


-log10(adj P-value)

W


A non-expressible viral DNA integration site on Chr11 vs
a random site without viral DNA integration on Chr9
log2(Fold Change)

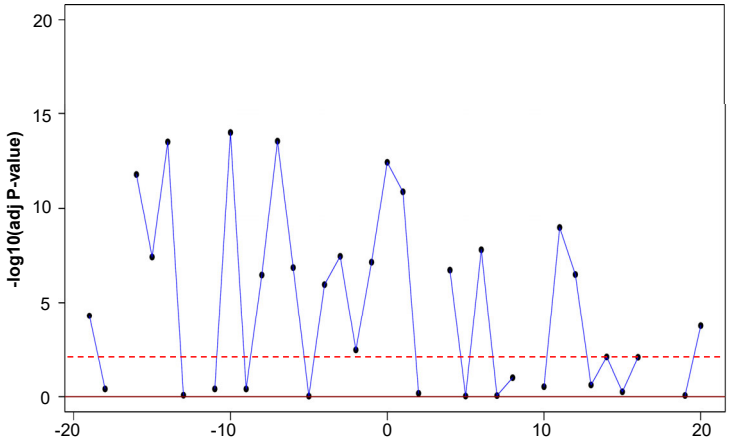

-log10(adj P-value)

Figure S6. HPV DNA integration may or may not impact the expression of the nearby host genes in cervical cancer tissues. (**A-H**) Significance of the variance by F-test with Bonferroni correction within each bin shown in Figure 7 for individual expressible vs non-expressible HPV DNAs integrated in the mapped tissue chromosomal regions**.** Each point represents a five-gene bin, corresponding to those in Figure 7. The missing point in each bin indicates the P-value close to zero. The points above the dashed red line indicate P< 0.01. (**I-N**) Comparison of the host gene expression nearby an expressible or non-expressible HPV integration site in CaSki cells over the same chromosomal position in SiHa cells (**I-K**) or from the same randomly selected chromosomal site without HPV integration from CaSki to SiHa cells (**L-N**). See Figure 7 and above for other description details. The genome range for each panel is shown as below: Chr6 nt 37,482,938-75,250,323 (**I**), Chr2 nt 885,290-43,767,987 (**J**), Chr11 nt 3,671,083-18,452,063 (**K**), Chr5 nt 39,105,252-93,594,611 (**L**), Chr8 nt 22,138,020- 66,178,464 (**M**), Chr22 nt 18,149,899-35,394,207 (**N**). (**O-W**) Comparison of host gene expression nearby the expressible or non-expressible HPV DNA integration site to a randomly selected chromosomal site without HPV DNA integration in CaSki cells. See Figure 7 and above for other description details. The genome ranges for comparison in each panel are shown below: Chr6 nt 37,482,938-75,250,323 vs Chr4 nt 9,224,896-70,058,845 (**O**), Chr8 nt 10,482,878-42,796,392 (**R**) or Chr9 nt 93,058,688-123,268,576 (**U**); Chr2 nt 885,290-43,767,987 vs Chr4 nt 9,224,896-70,058,845 (**P**), Chr8 nt 10,482,878-42,796,392 (**S**) or Chr9 nt 93,058,688-123,268,576 (**V**); Chr11 nt 3,671,083-18,452,063 vs Chr4 nt 9,224,896-70,058,845 (**Q**), Chr8 nt 10,482,878-42,796,392 (**T**) or Chr9 nt 93,058,688-123,268,576 (**W**). Panels on the right of **I**, **J**, **K**, **L**, **M**, **N**, **O**, **P**, **Q**, **R**, **S**, **T**, **U**, **V**, and **W** show the correspondent statistics of the variance by F-test with Bonferroni correction within each bin (five genes per bin). The missing point in each bin indicates the P-value close to zero. The points above the red dashed line indicate the P< 0.01.

Fig. S7

A

Number	Genome nt position	Reference	CaSki
1	131	A	G
2	350	T	G
3	1252	G	A
4	1418	C	A
5	1522	T	A
6	2457	C	T
7	2609	G	T
8	2938	A	G
9	3384	T	C
10	3410	C	T
11	3684	C	A

B

Number	Genome nt position	Reference	SiHa
1	350	T	G
2	442	A	C
3	645	A	C
4	1194	A	G
5	1842	A	G
6	3068	G	A

C

Number	Genome nt position	Reference	HeLa
1	104	T	C
2	485	T	C
3	549	C	A
4	751	C	T
5	806	G	A
6	1012	A	T
7	1194	C	A
8	1353	T	A
9	1807	T	C
10	1843	T	G
11	2269	C	T
12	5875	C	A
13	6401	A	G
14	7258	T	A
15	7486	C	T
16	7529	C	A
17	7567	A	C
18	7592	T	C
19	7670	A	T

D

HPV16 E6 protein	
HPV16 Ref.	MFQDPQERPRKLPQLCTELQTTIHDIILECVYCKQQLLRREVYDFAFRDLCIV YRDGNPYAVCDKCLKFYSKISEYRHYCYSLYGTTLEQQYNKPLCDLLIRCIN CQKPLCPEEKQRHLDKKQRFHNIRGRWTGRCMSCCRSSRTRRETQL*
CaSki	MFQDPQERPGKLPQLCTELQTTIHDIILECVYCKQQLLRREVYDFAFRDLCIV YRDGNPYAVCDKCLKFYSKISEYRHYCYSVYGTTLEQQYNKPLCDLLIRCIN CQKPLCPEEKQRHLDKKQRFHNIRGRWTGRCMSCCRSSRTRRETQL*
SiHa	MFQDPQERPRKLPQLCTELQTTIHDIILECVYCKQQLLRREVYDFAFRDLCIV YRDGNPYAVCDKCLKFYSKISEYRHYCYSVYGTTLEQQYNKPLCDLLIRCIN CQKPLCPDEKQRHLDKKQRFHNIRGRWTGRCMSCCRSSRTRRETQL*

E

HPV16 E7 protein	
HPV16 Ref.	MHGDTPTLHEYMLDLQPETTDLYCYEQLNDSSEEEDEIDGPAGQAEPDR AHYNIVTFCCKCDSTLRLCVQSTHVDIRTLEDLLMGTLGIVCPICSQKP*
SiHa	MHGDTPTLHEYMLDLQPETTDLYCYEQFNDSSEEEDEIDGPAGQAEPDR AHYNIVTFCCKCDSTLRLCVQSTHVDIRTLEDLLMGTLGIVCPICSQKP*

F

HPV18 E7 protein	
HPV18 Ref.	MHGPKATLQDIVLHLEPQNEIPVDLLCHEQLSDSEEENDEIDGVNHQHLPA RRAEPQRHTMLCMCCKCEARIELVVESSADDLRAFQQLFLNTLSFVCPWC ASQQ*
HeLa	MHGPKATLQDIVLHLEPQNEIPVDLLCHEQLSDSEEENDEIDGVNHQHLPA RRAEPQRHTMLCMCCKCEARIKLVVESSADDLRAFQQLFLNTLSFVCPWC ASQQ*

G

Number	HPV16 Genome nt position	Reference	A1BJ	A1M9	A1M8
1	109	T	C	C	T
2	178	T	T	T	G
3	350	T	G	T	T
4	647	A	A	A	G
5	840	C	C	C	T
6	846	T	T	T	C

H

Number	HPV18 Genome nt position	Reference	A3HE	A1MI
1	104	T	G	G
2	485	T	C	C
3	549	C	A	A
4	751	C	T	T

I

Number	HPV16 Genome nt position	Reference	T1074	T6050	T6709	T5350	T3023
1	178	T	G	G	G	G	T
2	256	C	C	C	C	C	T
3	350	T	T	T	T	T	G
4	647	A	G	G	G	G	A
5	843	T	C	C	T	T	T
4	846	T	C	C	C	C	T

J

HPV16 E6 protein	
HPV16 Ref.	MFQDPQERPRKLPQLCTELQTTIHDIILECVYCKQQLLRREVYDFAFRDLCIVY RDGNPYAVCDKCLKFYSKISEYRHYCYSLYGTTLEQQYNKPLCDLLIRCINCQ KPLCPEEKQRHLDKKQRFHNIRGRWTGRCMSCCRSSRTRRETQL*
A1M8	MFQDPQERPRKLPQLCTELQTTIHEIILECVYCKQQLLRREVYDFAFRDLCIVY RDGNPYAVCDKCLKFYSKISEYRHYCYSLYGTTLEQQYNKPLCDLLIRCINCQ KPLCPEEKQRHLDKKQRFHNIRGRWTGRCMSCCRSSRTRRETQL*
A1BJ	MFQDPQERPRKLPQLCTELQTTIHDIILECVYCKQQLLRREVYDFAFRDLCIVY RDGNPYAVCDKCLKFYSKISEYRHYCYSVYGTTLEQQYNKPLCDLLIRCINCQ KPLCPEEKQRHLDKKQRFHNIRGRWTGRCMSCCRSSRTRRETQL*

K

HPV16 E7 protein	
HPV16 Ref.	MHGDTPTLHEYMLDLQPETTDLYCYEQLNDSSEEEDEIDGPAGQAEPDR AHYNIVTFCCKCDSTLRLCVQSTHVDIRTLEDLLMGTLGIVCPICSQKP*
A1M8	MHGDTPTLHEYMLDLQPETTDLYCYEQLSDSSEEEDEIDGPAGQAEPDR AHYNIVTFCCKCDSTLRLCVQSTHVDIRTLEDLLMGTLGIVCPICSQKP*

L

HPV16 E6 protein	
HPV16 Ref.	MFQDPQERPRKLPQLCTELQTTIHDIILECVYCKQQLLRREVYDFAFRDLCIV YRDGNPYAVCDKCLKFYSKISEYRHYCYSLYGTTLEQQYNKPLCDLLIRCIN CQKPLCPEEKQRHLDKKQRFHNIRGRWTGRCMSCCRSSRTRRETQL*
T1074 T6050 T6709 T5350	MFQDPQERPRKLPQLCTELQTTIHEIILECVYCKQQLLRREVYDFAFRDLCIVY RDGNPYAVCDKCLKFYSKISEYRHYCYSLYGTTLEQQYNKPLCDLLIRCINCQ KPLCPEEKQRHLDKKQRFHNIRGRWTGRCMSCCRSSRTRRETQL*
T3023	MFQDPQERPRKLPQLCTELQTTIHDIILECVYCKQQLLRREVYDFAFRDLCIVY RDGNPYAVCDKCLKFYSKISEYRHYCYSVYGTTLEQQYNKPLCDLLIRCINCQ KPLCPEEKQRHLDKKQRFHNIRGRWTGRCMSCCRSSRTRRETQL*

M

HPV16 E7 protein	
HPV16 Ref.	MHGDTPTLHEYMLDLQPETTDLYCYEQLNDSSEEEDEIDGPAGQAEPDR AHYNIVTFCCKCDSTLRLCVQSTHVDIRTLEDLLMGTLGIVCPICSQKP*
T1074 T6050 T6709 T5350	MHGDTPTLHEYMLDLQPETTDLYCYEQLSDSSEEEDEIDGPAGQAEPDR AHYNIVTFCCKCDSTLRLCVQSTHVDIRTLEDLLMGTLGIVCPICSQKP*

N

SiHa virus-host fusion protein 1 Transcript-3: HPV16:97-880/Chr13:73456962-2979

Coding region: HPV16 E1:865-880/Chr13:73456962-73456814 **(165 nts / 54 aa)**

Protein: MADPADRKLDVCLANWIKGQDPSVCCIQETHLTCKDTHRIKIKGWRKIYQANGK*

O

SiHa virus-host fusion protein 2 Transcript-4: HPV16:97-226/Chr13:73456962-2979

Coding region: HPV16 E6:104-226/Chr13:73456962-73456814 **(159 nts / 52 aa)**

Protein: MFQDPQERPRKLPQLCTELQTTIHDIILECVYCKQQLLRREIENWMSAWQIG*

P

HeLa virus-host fusion protein 1 Transcript-2: Chr8:127219268-18810/HPV18:24-233/416-929/Chr8:127229130-28559

Coding region: Chr8:127218909-127218810/HPV18:24-118 **(195 nts / 64 aa)**

Protein: MNFGLLCNPSGAFTDNIDFEVALTWPALPLTRLVYKKGSNRKRSGPKTVYIKDVRNTPQYHGAL*

Q

HeLa truncated HPV18 E1 protein Transcript-3: HPV18:102-233/416-2497/Chr8:127229301-8559

Coding region: HPV18:914-2497/Chr8:127229301-127229299 **(1587 nts / 528 aa)**

Protein:MADPEGTDGEGTGCNGWFYVQAIVDKKTGDVISDDEDENATDTGSDMVDFIDTQGTFCEQAELETAQALFHAQEVH
NDAQVLHVLKRKFAGGSKENSPLGERLEVDTELSPRLQEISLNSGQKKAKRRLFTISDSGYGCSEVEATQNQVTTNGEHGG
NVCSGGSTEAIDNGGTEGNNSSVDGTSDNSNIENVNPQCTIAQLKDLLKVNNKQGAMLAVFKDTYGLSFTDLVRNFKSDKTT
CTDWVTAIFGVNPTIAEGFKTLIQPFILYAHIQCLDCKWGVLILALLRYKCGKSRLTVAKGLSTLLHVPETCMLIQPPKLRSSVA
ALYWYRTGISNISEVMGDTPEWIQRLTIIQHGIDDSNFDLSEMVQWAFDNELTDESDMAFEYALLADSNSNAAAFLKSNCQAK
YLKDCATMCKHYRRAQKRQMNMSQWIRFRCSKIDEGGDWRPIVQFLRYQQIEFITFLGALKSFLKGTPKKNCLVFCGPANTG
KSYFGMSFIHFIQGAVISFVNSTSHFWLEPLTDTKVAML*

R

A1BJ virus-host fusion protein 1 Transcript-4: HPV16:97-226/Chr4:73777731-73774860

Coding region: HPV16 E6:104-226/Chr4:73777731-73777699 **(150 nts / 49 aa)**

Protein: MFQDPQERPRKLPQLCTELQTTIHDIILECVYCKQQLLRREGQDDNEPS*

S

A2RE virus-host fusion protein 1 Transcript-4: HPV16:97-226/Chr1:38022699-38023808

Coding region: HPV16 E6:104-226/Chr1:38022699-38022809/38023545-38023625 **(318 nts / 105 aa)**

Protein:
MFQDPQERPRKLPQLCTELQTTIHDIILECVYCKQQLLRRERIAKERQKQYNCLTQRIEREKKLFVIAQKIQTRKDLMDKTQKV
KVKKETVNSPAIYKFQSRRKR*

T

T1074 host-virus fusion protein 1 **Transcript-1: Chr17:73644445-73348599/HPV16: 5639-7321**

Coding region: Chr17:73644088-73348599/HPV16 L1: 5639-7156 **(7683 nts / 2560 aa)**

Protein:

MWGLLIWTLLALHQIRAARAQDDVSPYFKTEPVRTQVHLEGNRLVLTCMAEGSWPLEFKWLHNNRELTKFSLEYRYMITSLD
RTHAGFYRCIVRNRMGALLQRQTEVQVAYMGSFEEGEKHQSVSHGEAAVIRAPRIASFPQPQVTWFRDGRKIPPSSRIAITLE
NTLVILSTVAPDAGRYYVQAVNDKNGDNKTSQPITLTVENVGGPADPIAPTIIIPPKNTSVVAGTSEVTLECVANARPLIKLHIIWK
KDGVLLSGGISDHNRRLTIPNPTGSDAGYYECEAVLRSSSVPSVVRGAYLSVLEPPQFVKEPERHITAEMEKVVDIPCQAKGVP
PPSITWYKDAAVVEVEKLTRFRQRNDGGLQISGLVPDDTGMFQCFARNAAGEVQTSTYLAVTSIAPNITRGPLDSTVIDGMSVV
LACETSGAPRPAITWQKGERILASGSVQLPRFTPLESGSLLISPTHISDAGTYTCLATNSRGVDEASADLVVWARTRITKPPQD
QSVIKGTQASMVCGVTHDPRVTIRYIWEKDGATLGTESHPRIRLDRNGSLHISQTWSGDIGTYTCRVISAGGNDSRSAHLRVRQ
LPHAPEHPVATLSTVERRAINLTWTKPFDGNSPLIRYILEMSENNAPWTVLLASVDPKATSVTVKGLVPARSYQFRLCAVNDVG
KGQFSKDTERVSLPEEPPTAPPQNVIASGRTNQSIMIQWQPPPESHQNGILKGYIIRYCLAGLPVGYQFKNITDADVNNLLLEDL
IIWTNYEIEVAAYNSAGLGVYSSKVTEWTLQGVPTVPPGNVHAEATNSTTIRFTWNAPSPQFINGINQGYKLIAWEPEQEEEVT
MVTARPNFQDSIHVGFVSGLKKFTEYFTSVLCFTTPGDGPRSTPQLVRTHEDVPGPVGHLSFSEILDTSLKVSWQEPGEKNGIL
TGYRISWEEYNRTNTRVTHYLPNVTLEYRVTGLTALTTYTIEVAAMTSKGQGQVSASTISSGVPPELPGPPTNLGISNIGPRSVT
LQFRPGYDGKTSISRWLVEAQVGVVGEGEEWLLIHQLSNEPDARSMEVPDLNPFTCYSFRMRQVNIVGTSPPSQPSRKIQTLQ
APPDMAPANVSLRTASETSLWLRWMPLPEMEYNGNPESVGYKIKYSRSDGHGKTLSHVVQDRVERDYTIEDLEEWTEYRVQ
VQAFNAIGSGPWSQTVVGRTRESVPSSGPTNVSALATTSSSMLVRWSEVPEADRNGLVLGYKVMYKEKDSDTQPRFWLVEG
NSSRSAQLTGLGKYVLYEVQVLAFTRIGDGSPSHPPILERTLDDVPGPPMGILFPEVRTTSVRLIWQPPAAPNGIILAYQITHRLN
TTTANTATVEVLAPSARQYTATGLKPESVYLFRITAQTRKGWGEAAEALVVTTEKRDRPQPPSRPMVQQEDVRARSVLLSWE
PGSDGLSPVRYYTIQTRELPSGRWALHSASVSHNASSFIVDRLKPFTSYKFRVKATNDIGDSEFSEESESLTTLQAAPDEAPTIL
SVTPHTTTSVLIRWQPPAEDKINGILLGFRIRYRELLYEGLRGFTLRGINNPGATWAELTSMYSMRNLSRPSLTQYELDNLNKH
RRYEIRMSVYNAVGEGPSSPPQEVFVGEAVPTAAPRNVVVHGATATQLDVTWEPPPLDSQNGDIQGYKIYFWEAQRGNLTER
VKTLFLAENSVKLKNLTGYTAYMVSVAAFNAAGDGPRSTPTQGQTQQAAPSAPSSVKFSELTTTSVNVSWEAPQFPNGILEG
YRLVYEPCSPVDGVSKIVTVDVKGNSPLWLKVKDLAEGVTYRFRIRAKTFTYGPEIEANVTTGPGEGAPGPPGVPIIVRYSSAIA
IHWSSGDPGKGPITRYVIEARPSDEGLWDILIKDIPKEVSSYTFSMDILKPGVSYDFRVIAVNDYGFGTPSSPSQSVPAQKANPF
YEEWWFLVVIALVGLIFILLLVFVLIIRGQSKKYAKKTDSGNSAKSGALGHSEMMSLDESSFPALELNNRRLSVKNSFCRKNGL
YTRSPPRPSPGSLHYSDEDVTKYNDLIPAESSSLTEKPSEISDSQMSLWLPSEATVYLPPVPVSKVVSTDEYVARTNIYYHAGT
SRLLAVGHPYFPIKKPNNNKILVPKVSGLQYRVFRIHLPDPNKFGFPDTSFYNPDTQRLVWACVGVEVGRGQPLGVGISGHPL
LNKLDDTENASAYAANAGVDNRECISMDYKQTQLCLIGCKPPIGEHWGKGSPCTNVAVNPGDCPPLELINTVIQDGDMVDTGF
GAMDFTTLQANKSEVPLDICTSICKYPDYIKMVSEPYGDSLFFYLRREQMFVRHLFNRAGTVGENVPDDLYIKGSGSTANLASS
NYFPTPSGSMVTSDAQIFNKPYWLQRAQGHNNGICWGNQLFVTVVDTTRSTNMSLCAAISTSETTYKNTNFKEYLRHGEEYDL
QFIFQLCKITLTADVMTYIHSMNSTILEDWNFGLQPPPGGTLEDTYRFVTSQAIACQKHTPPAPKEDPLKKYTFWEVNLKEKFS
ADLDQFPLGRKFLLQAGLKAKPKFTLGKRKATPTTSSTSTTAKRKKRKL*

U

T1074 virus-host fusion protein 2 **Transcript-5: HPV16:97-226/Chr17:73338940-73334403**

Coding region: HPV16 E6:104-226/Chr13:733338940-73338587 **(477 nts / 158 aa)**

Protein:

MFQDPQERPRKLPQLCTELQTTIHEIILECVYCKQQLLRREGSDSEYEVDSNHQKAHSFVNHYISDPTYYNSWRRQQKGISRA
QAYSYTESDSGEPDHTTVTNSTSTQQGSLFRPKASRTPTPQNPPNPPSQQSTLYRPPSSLAPGSRAPIAGFSSFV*

V

T5350 virus-host fusion protein 1 **Transcript-4: HPV16:97-226/Chr3:1899355002-189937565**

Coding region: HPV16 E6:104-226/Chr3:1899355002-1899355193 **(318 nts / 105 aa)**

Protein:

MFQDPQERPRKLPQLCTELQTTIHEIILECVYCKQQLLRREIQKKCNPRRKGQMSQSPVSSWVCMCPGSLGTSLNLYALDM
WLGLLRHKVGELDERLQTLSRMAL*

Figure S7. Point mutations that cause missense mutations in E6 and E7 proteins from an integrated, expressible viral DNA and predicted virus-host fusion proteins from the transcribed virus-host chimeric transcripts in cervical cancer cell lines and tissues. (**A**-**C**) Point mutations in an integrated, expressible viral DNA identified in CaSki (**A**), SiHa (**B**), and HeLa (**C)** cells and ten cervical cancer tissues (**G-I**) identified by RNA-seq. Most mutations are silent. The missense mutations that cause amino acid changes in E6 and/or E7 proteins are indicated in red for CaSki (**D**), SiHa (**E**) and HeLa (**F**) cells, and eight cervical cancer tissues (**J**-**M**). The normal amino acid residues in the reference proteins are underlined. (**N**-**V**) Amino acid sequences of the predicted viral-host fusion proteins expressed from the chimeric virus-host fusion RNA transcripts detected in SiHa (**N**, **O**) and HeLa (**P**, **Q**) cells and in the cervical cancer tissues A1BJ (**R**), A2RE (**S**), T1074 (**T**, **U**) and T5350 (**V**). In HeLa cells, transcript-3 (**Q**) retains a truncated E1 ORF (528 aa) using a host stop codon at the integration junction site. A normal full size HPV18 E1 has 658 aa residues.