SUPPLEMENTAL INFORMATION

Genome-wide analysis of HPV integration in human cancers reveals recurrent, focal genomic instability

Keiko Akagi, Jingfeng Li, Tatevik Broutian, Hesed Padilla-Nash, Weihong Xiao,
Bo Jiang, James W. Rocco, Theodoros N. Teknos, Bhavna Kumar, Danny Wangsa, Dandan He,
Thomas Ried, David E. Symer, Maura L. Gillison

SUPPLEMENTAL METHODS **PAGE NO.**

SUPPLEMENTAL FIGURES

SUPPLEMENTAL TABLES

SUPPLEMENTAL METHODS

Quantitative real-time PCR

As shown in Supplemental Fig. 8, regions of focal amplification detected by WGS were evaluated by quantitative PCR (qPCR) using the Power SYBR Green kit (ABI) on a StepOnePlus PCR machine (ABI). Genomic DNA from CAL 27 cells was used as a control in qPCR experiments, as no CNV was detected in CAL 27 using WGS data from various regions of interest identified in other samples. For each qPCR measurement, ΔCt values between amplified and flanking regions normalized to CAL 27 were determined and converted to copy number changes on a linear scale. HPV viral copy numbers were measured in samples by qPCR targeted to the HPV16 E6 or HPV18 E7 region normalized to a single copy human gene (ERV3) (Gravitt et al. 2003; D'Souza et al. 2005).

Northern blot hybridization

Templates for RNA probes were generated by PCR amplification using primers containing T7 promoter sequences and then were purified by agarose gel electrophoresis. RNA probes were generated for specific regions of *TP63*, *DIAPH2*, HPV16 E7, HPV16 E5 and a loading control (*RPLPO*) using the DIG RNA labeling kit (SP6/T7 protocol; Roche).

For Northern blots, 10-15 mcg of total RNA was separated by formaldehyde gel electrophoresis and transferred to a nitrocellulose membrane (GE Healthcare) using the Ambion NorthernMax kit (Life Technologies). RNA was cross-linked to the membrane using a UV trans-illuminator, blocked for 30 minutes at 68^{o}C with DIG Easy Hyb (Roche) and hybridized with DIG-labeled RNA probes overnight at 68^{o}C. Primary antibody was washed with the DIG Wash and Block Buffer Set (Roche). For detection, membranes were incubated at room temperature using Anti-DIG AP Conjugate (Roche) and DIG CDP-Star reagent (Roche). Membranes were

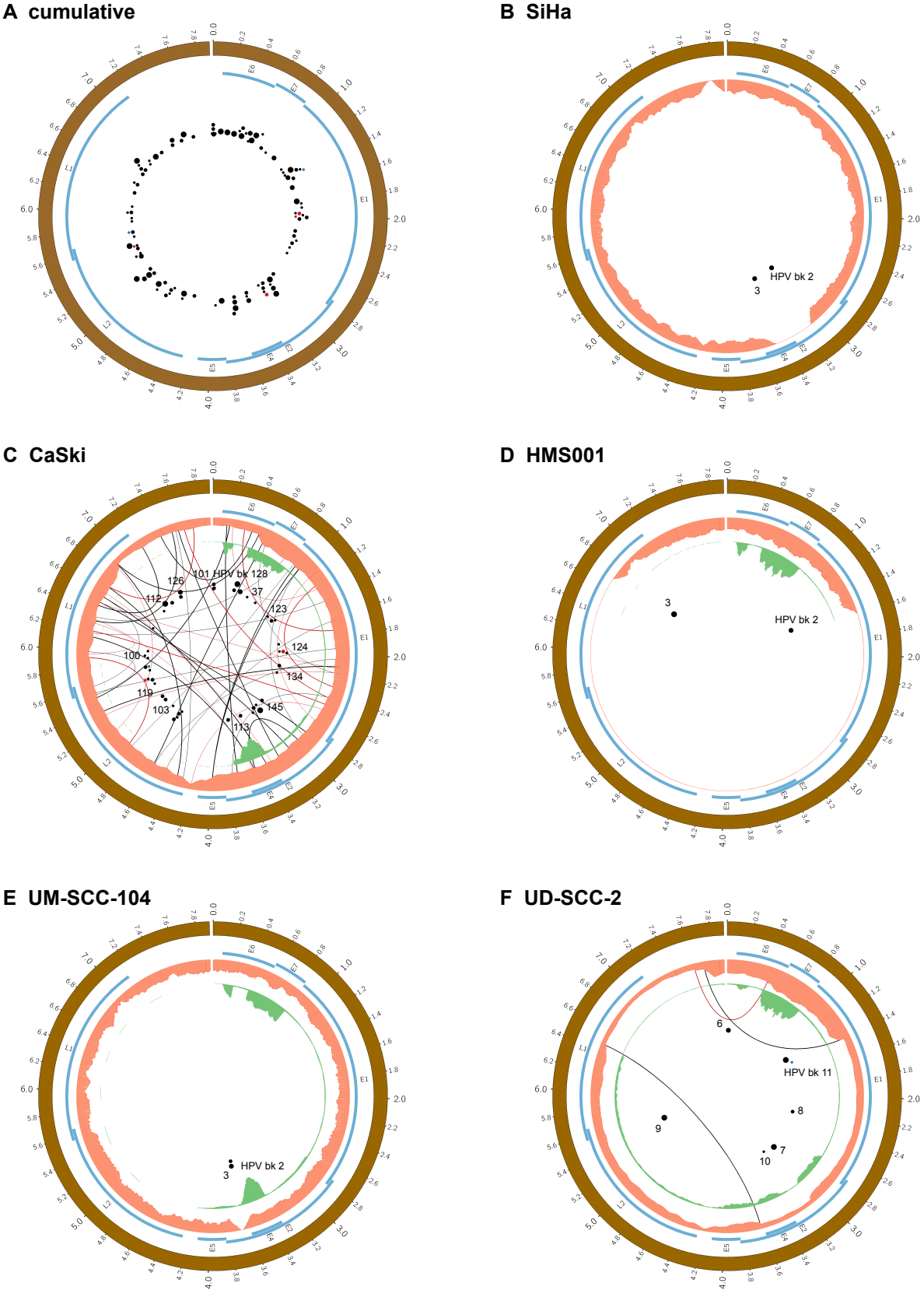
visualized with CL-XPosure film (Thermo), developed on an RP X-OMAT processor, model M6B (Kodak), and scanned using an Epson Perfection V750 pro scanner.

Western blot hybridization and peptide competition experiments

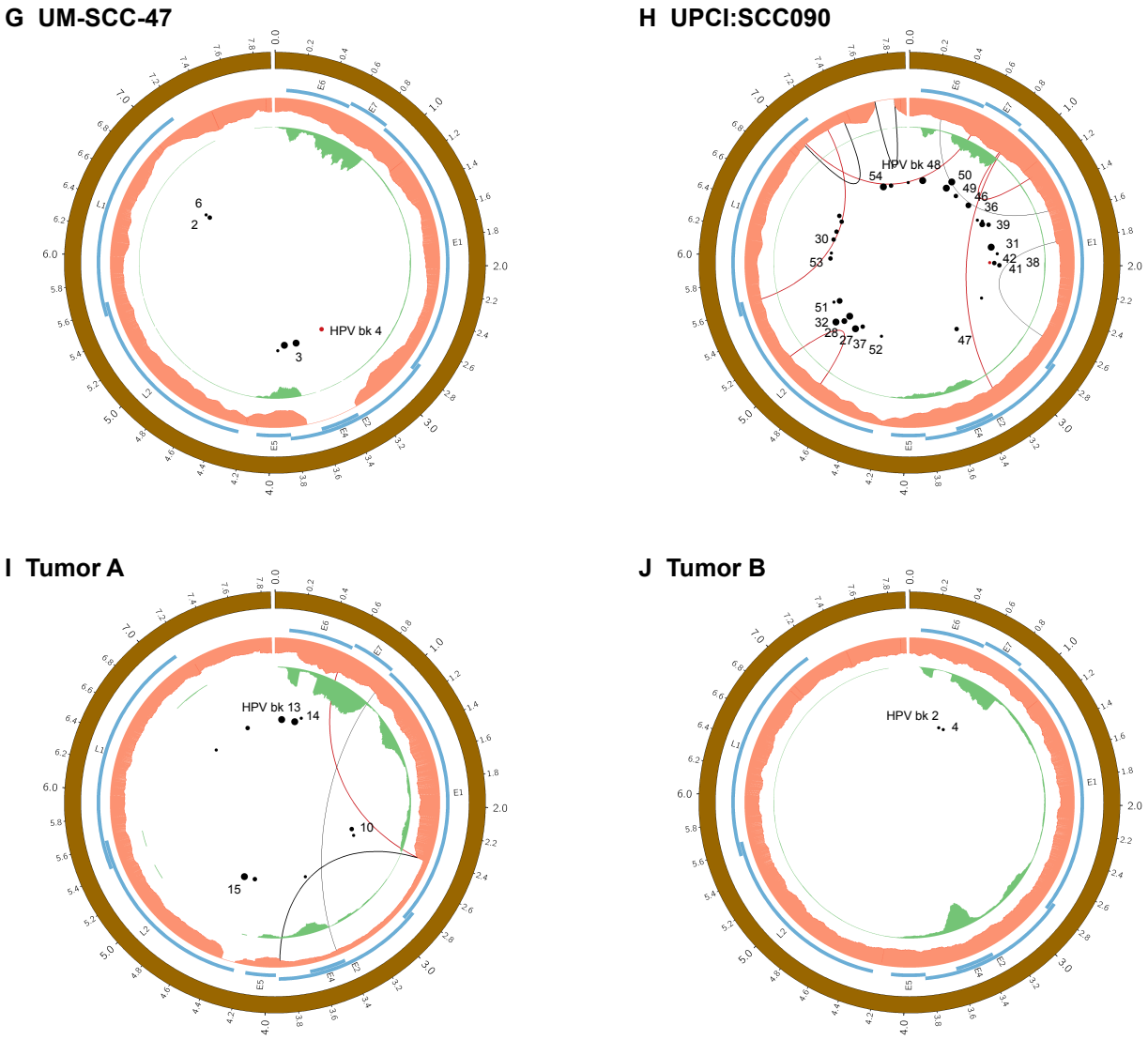
Whole cell lysates (40 mcg each) were resolved by electrophoresis in a pre-cast 10% mini-PROTEAN TGX gel (BioRad) using Tris-glycine SDS running buffer (BioRad) on the Mini-PROTEAN Tetra-cell system (BioRad). Proteins were transferred onto a PVDF membrane (BioRad) using Tris-glycine transfer buffer (BioRad) with 20% methanol and blocked with 5% milk in phosphate buffered saline with Tween buffer (PBST) for 1 hour at room temperature (RT). Membranes were incubated overnight at 4^{o}C with primary antibody or antibody mixed with peptide competitors, respectively, in 1% milk-PBST.

A mouse monoclonal antibody against amino acids 1-205 of human ΔNp63 (cat. no. SC-8431, Santa Cruz Biotechnology) or a rabbit polyclonal antibody against a synthetic peptide of human C-terminal PID domain of ΔN-P63 protein (cat. no. Ab53039, Abcam) were used (5 mcg antibody per 5mL diluent). To detect diaphanous-related formin 2 (i.e. the protein encoded by *DIAPH2*), we used a rabbit polyclonal antibody (cat. no. Ab12319, Abcam) against a synthetic peptide containing a portion of exon 3.

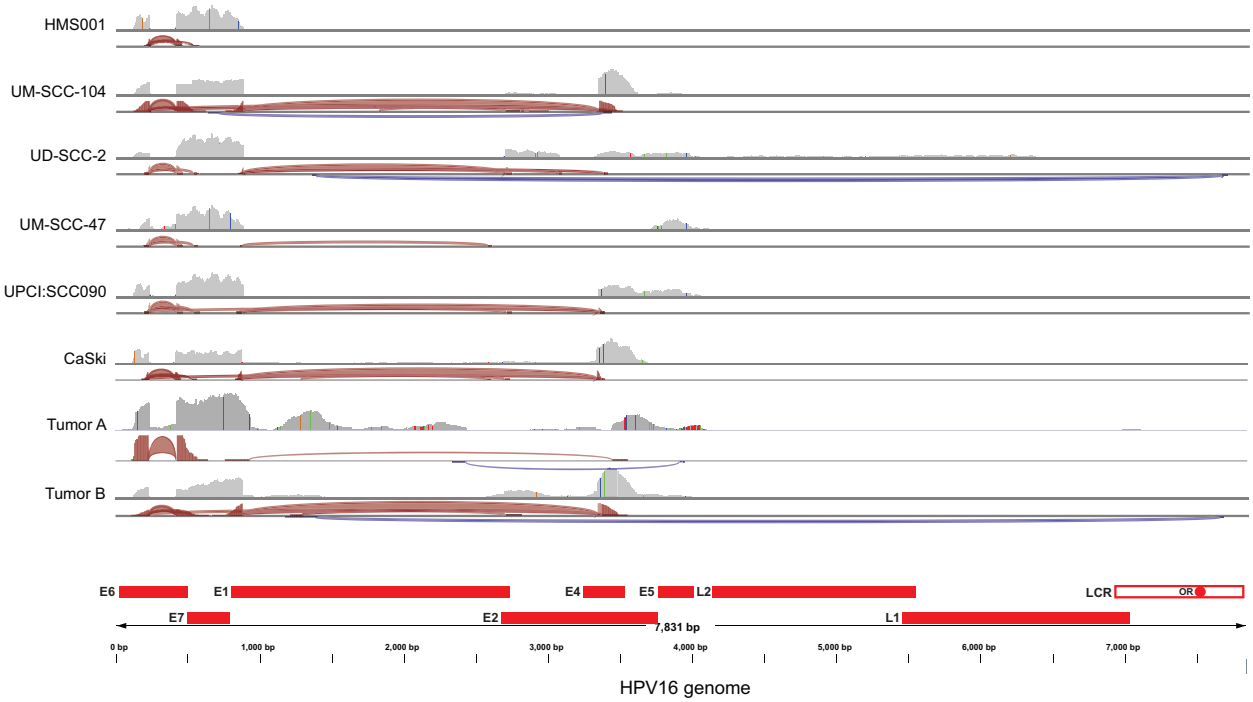
Membranes were incubated overnight at 4^{o}C with antibodies alone or in combination with serial dilutions of p63 peptide (cat. no. Ab153667, Abcam) or diaphanous-related formin 2 peptide (cat. no. Ab 156989, Abcam), respectively. Membranes were washed with PBST and incubated with respective HRP-conjugated secondary antibody. Membranes were developed with the Immun-Star HPR chemiluminescence substrate (BioRad) and visualized on CL-XPosure film (Thermo Fischer Scientific). For loading controls, membranes were probed with monoclonal anti-β-actin (cat. no. Ab82227, Abcam) or anti-GAPDH (cat. no. G8795, Sigma-Aldrich) antibodies.



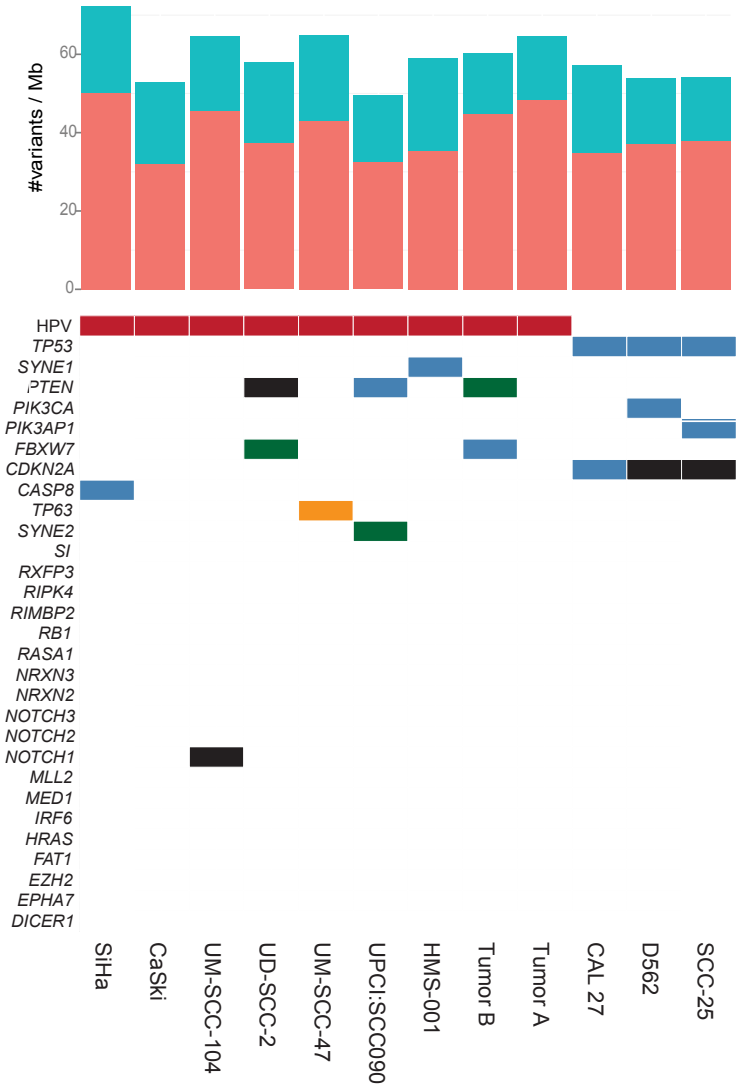
Supplemental Fig. 1. Continued on next page.



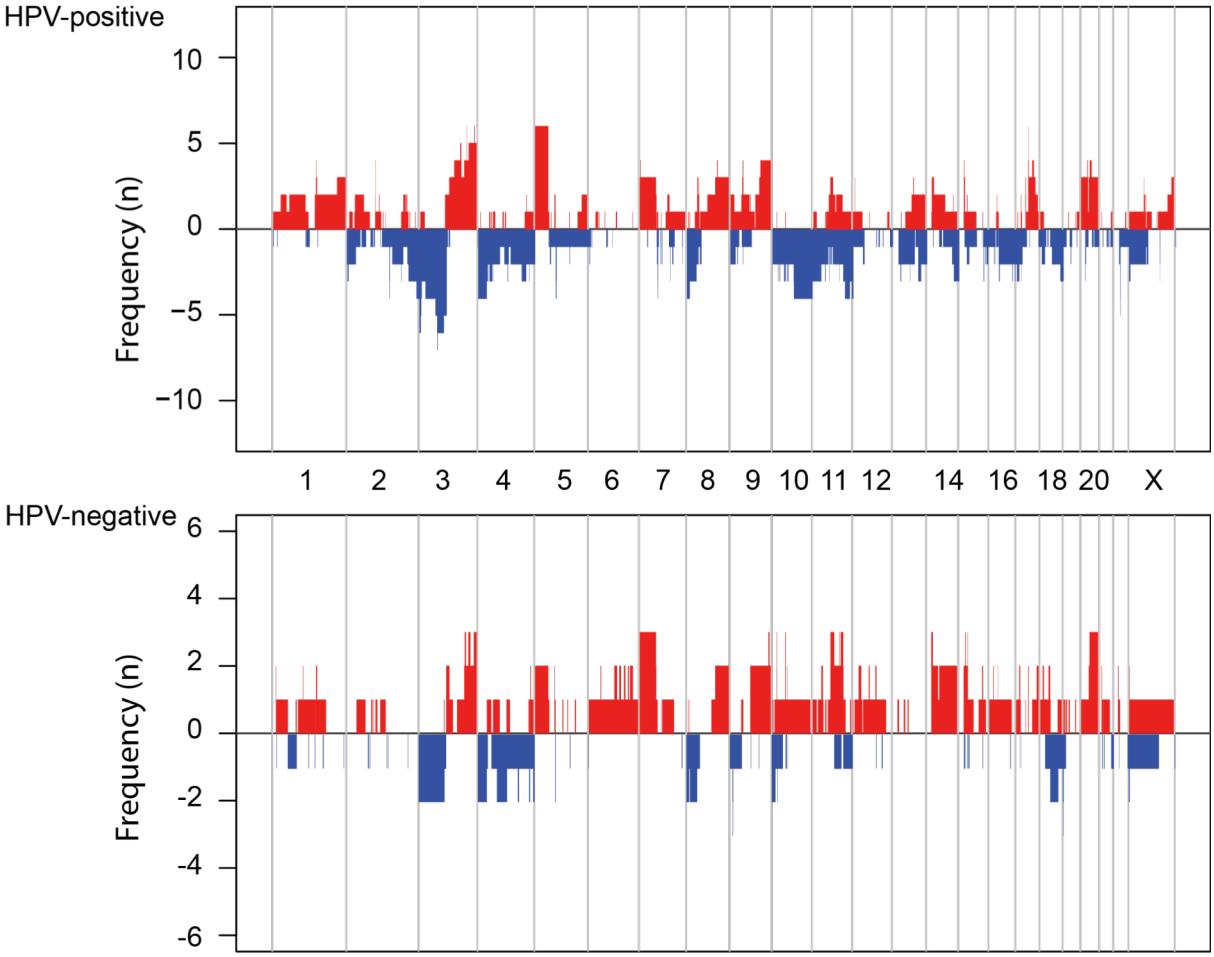
Supplemental Fig. 1. **Broad distribution of breakpoints and rearrangements mapped to the HPV genome in various cancer samples.** Circos plots depict various genomic features of HPV integrants in cell lines and primary tumor samples. **(**A) Compilation of HPV insertional breakpoints (combined from all 9 samples shown in panels B- J); (B) SiHa; (C) CaSki; (D) HMS001; (E) UM-SCC-104; (F) UD-SCC-2; (G) UM-SCC-47 (H) UPCI:SCC090 (I) Tumor A (with HPV-18); and (J) Tumor B. *Brown circles*: reference HPV genome coordinates; *blue arcs*: HPV genes; *orange histograms*: depth of coverage of HPV genome in each sample; *green histograms*: RNA-Seq viral transcripts; *black dots*: position of confirmed HPV insertional breakpoints, some of which have ID numbers shown, corresponding to those listed in Figs. 3 - 6 and Supp. Table 3; *red dots*: insertional breakpoints with failed PCR; *blue dots*: breakpoints not tested by PCR; *size of dots*: larger dots correspond to higher number of supporting discordant reads; *connecting arcs*: intraviral rearrangements; *red connecting arcs*: rearrangements in inverted orientation; *thickness of connecting arcs*: related to number of supporting reads. See Fig. 1 and Supp. Table 3.



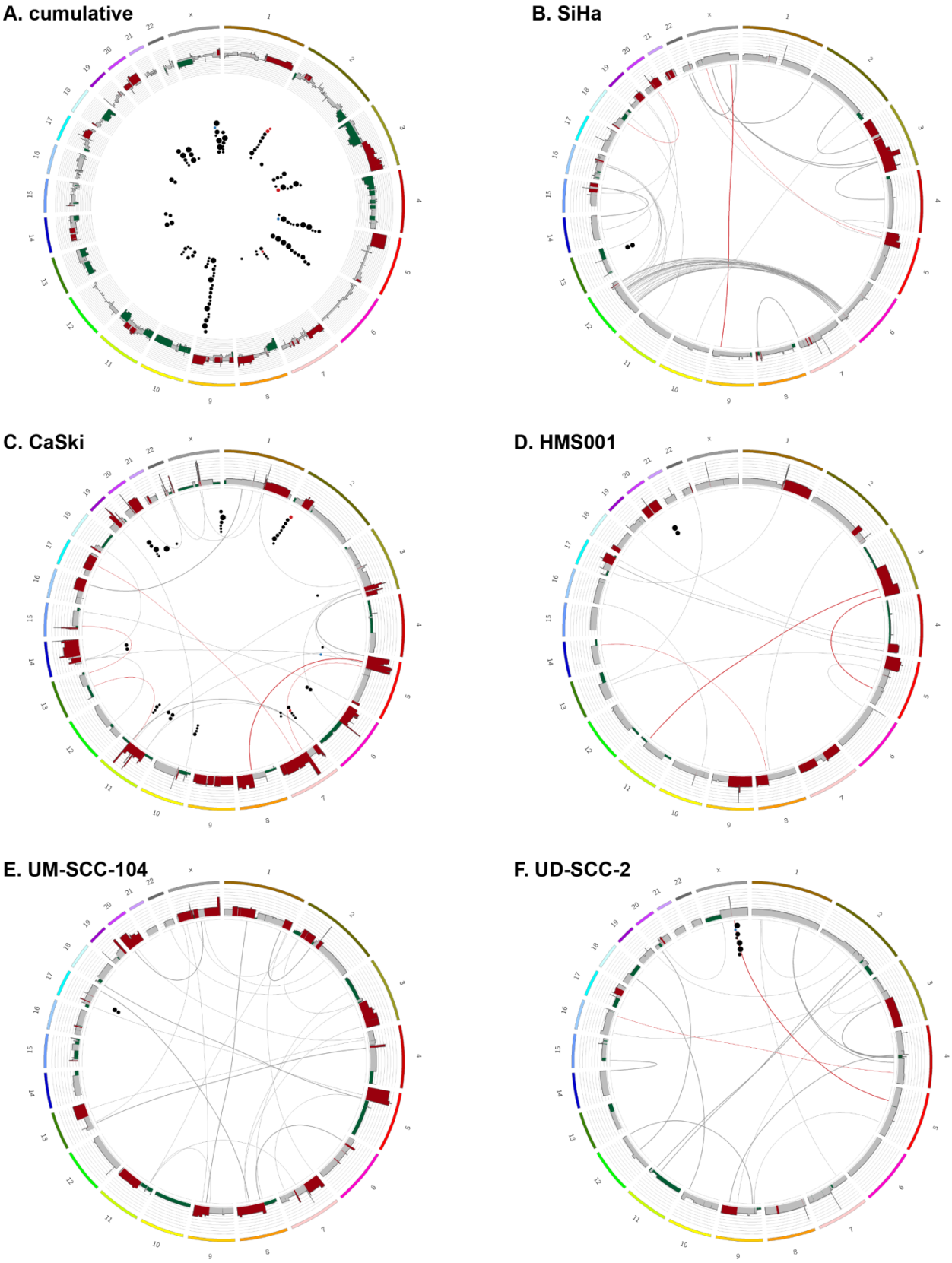
Supplemental Fig. 2. **HPV E6 and E7 oncogene transcripts are expressed in diverse cancers.** Depictions of transcript read counts (*top panel, gray histograms*) and splicing patterns (*bottom panels, salmon pink and indigo "Sashimi plot" ribbons*) from RNA-Seq data obtained from seven HPV16 positive and one HPV18-positive cancer samples, visualized by IGV (Thorvaldsdottir et al. 2012). *Vertical colored* lines: Positions of SNPs in homozygosity. *Salmon pink*, sense stranded transcripts; *indigo*, antiparallel transcripts. We did not obtain RNA-Seq data from SiHa cells. *Bottom, red rectangles*: Positions of HPV genes and long control region (LCR) including origin of replication (OR, *red* circle), shown with reference viral genome coordinates. Tumor A is HPV18-positive, where viral coordinates have been superimposed onto HPV16 genome coordinates.



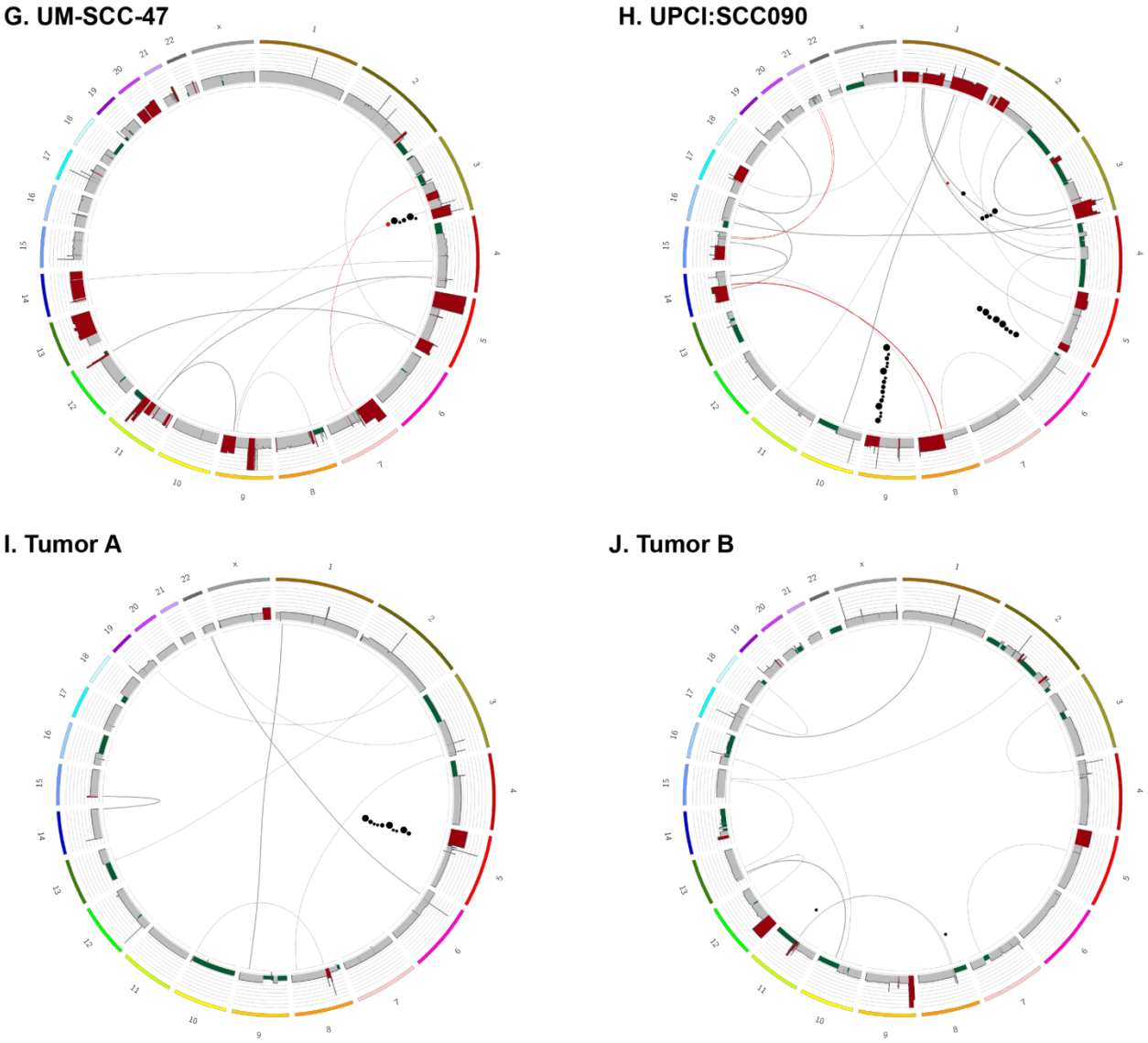
Supplemental Fig. 3. **Comparison of mutational spectra in HPV-positive vs. HPV-negative cancer samples.**
Top: Histograms of mutation rates of 12 sequenced HPV positive (*left*) and negative (*right*) cancer genomes, displayed as number of variants per megabase (Mb). *Light blue*, SNPs; *pink*: indels. Excluded from these counts are all variants identified in the dbSNP database, except those verified as somatic. *Bottom*: Mutational status of genes previously known as mutated in HNSCC cancers (Agrawal et al. 2011; Stransky et al. 2011), in each sample studied here. *Red-filled boxes*, HPV integrant present in genome*; blue*, coding SNPs and indels; *black*, large deletion including coding exon(s); *green*, chromosomal translocation in intron; *orange*, insertional mutagenesis by HPV16 viral integrant. Further details of these mutations are presented in Supplemental Table 4. Note that HPV integrant positivity and *TP53* mutations are inversely associated in these 12 samples.



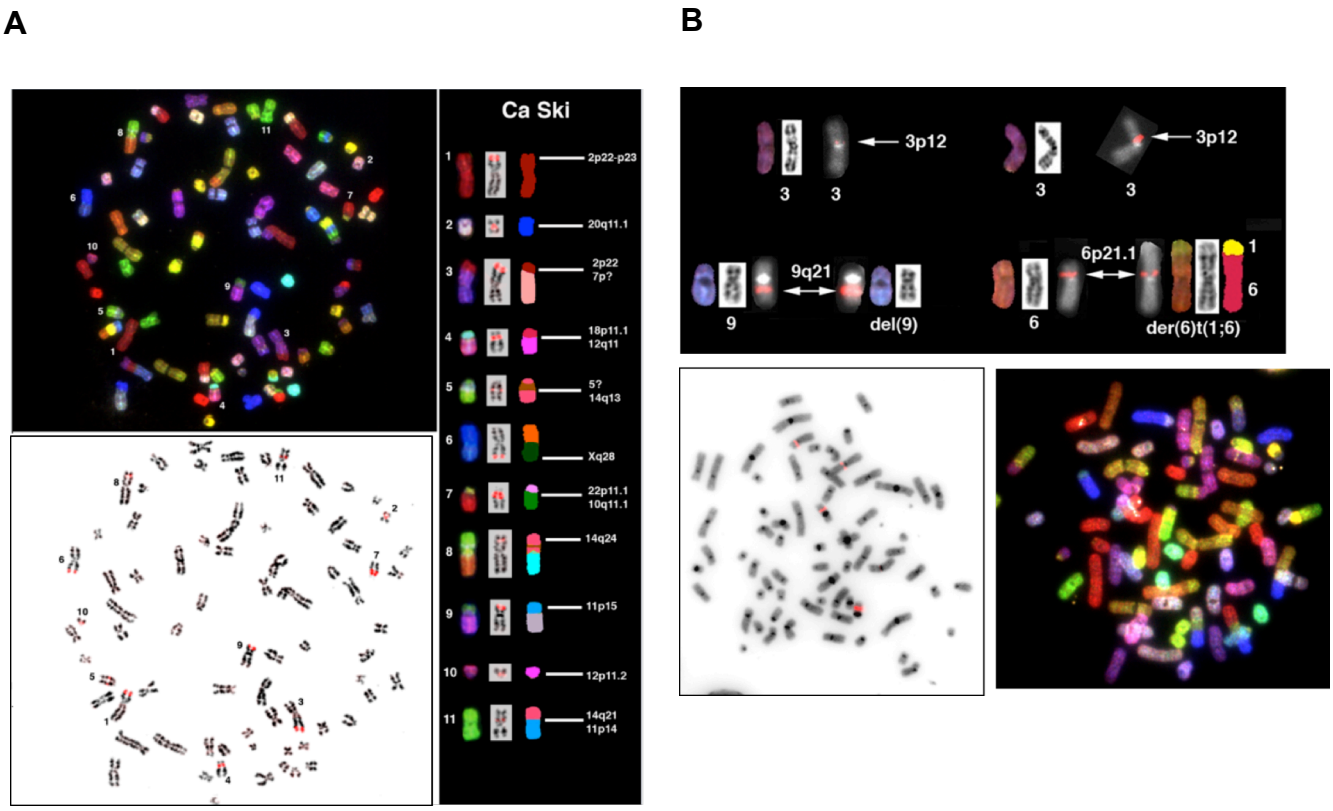
Supplemental Fig. 4. **Genomic copy number changes occur recurrently in diverse cancer samples.** Samples with DNA copy number gains or losses within 50 kb bins were counted. (*Top*) Seven HPV-positive and (*bottom*) three HPV-negative HNSCC samples were evaluated. CaSki and SiHa cells were not counted here. *Red, positive counts*: regional amplification and gain in a sample; *blue, negative counts*, loss of region in a sample; *y-axis*, cumulative number of samples showing change in copy number.



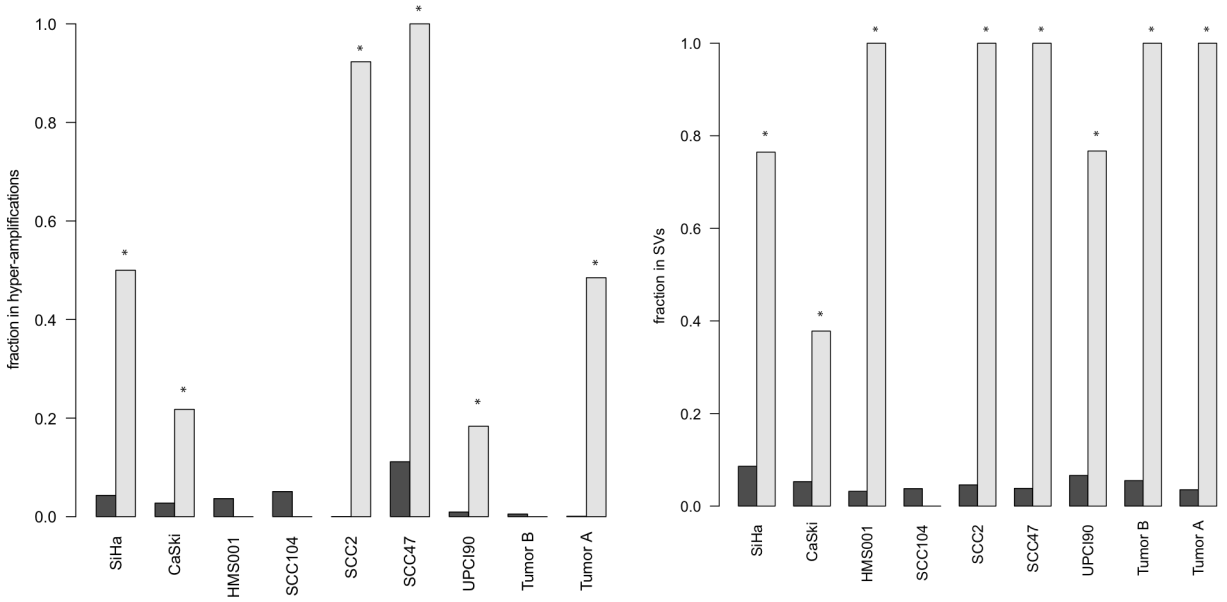
Supplemental Fig. 5. Continued on next page.



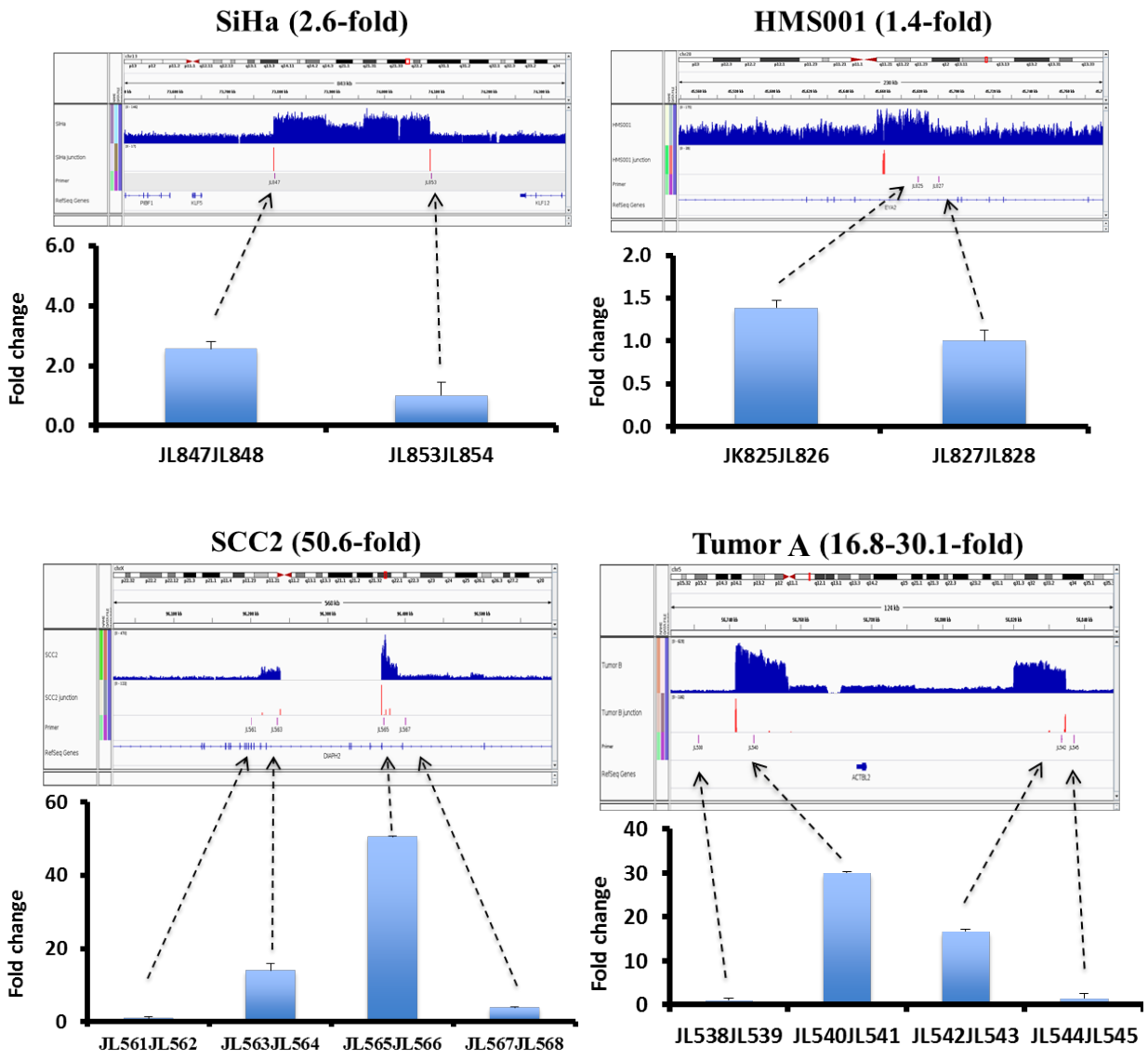
Supplemental Fig. 5. **Structural variation in various cancer samples.** Circos plots depict breakpoints, rearrangements and CNVs in diverse cancer cell lines and tumor samples. **(**A) Compilation of HPV insertional breakpoints (combined from all 9 samples shown in panels B- J), mapped to the reference human genome; (B) SiHa; (C) CaSki; (D) HMS001; (E) UM-SCC-104; (F) UD-SCC-2; (G) UM-SCC-47 (H) UPCI:SCC090 (I) Tumor A; (J) Tumor B. *Outer circles*: human chromosomes; *inner histograms*: mapped WGS depth of coverage in each sample, with *green*, loss; *red*, gain, defined in UM-SCC-47 cells as gain of ploidy N > 5, loss N < 3; defined in other samples as gain N > 4, loss N < 2; *black dots*: position of confirmed HPV insertional breakpoints; *red dots*: insertional breakpoints with failed PCR; *blue dots*: breakpoints not tested by PCR; *size of dots*: larger dots correspond to higher number of supporting discordant reads; *connecting arcs*: chromosomal translocations; *red connecting arcs*: translocation with confirming cytogenetic analysis within 5 Mbp; *thicker connecting arcs*: >10 supporting reads.



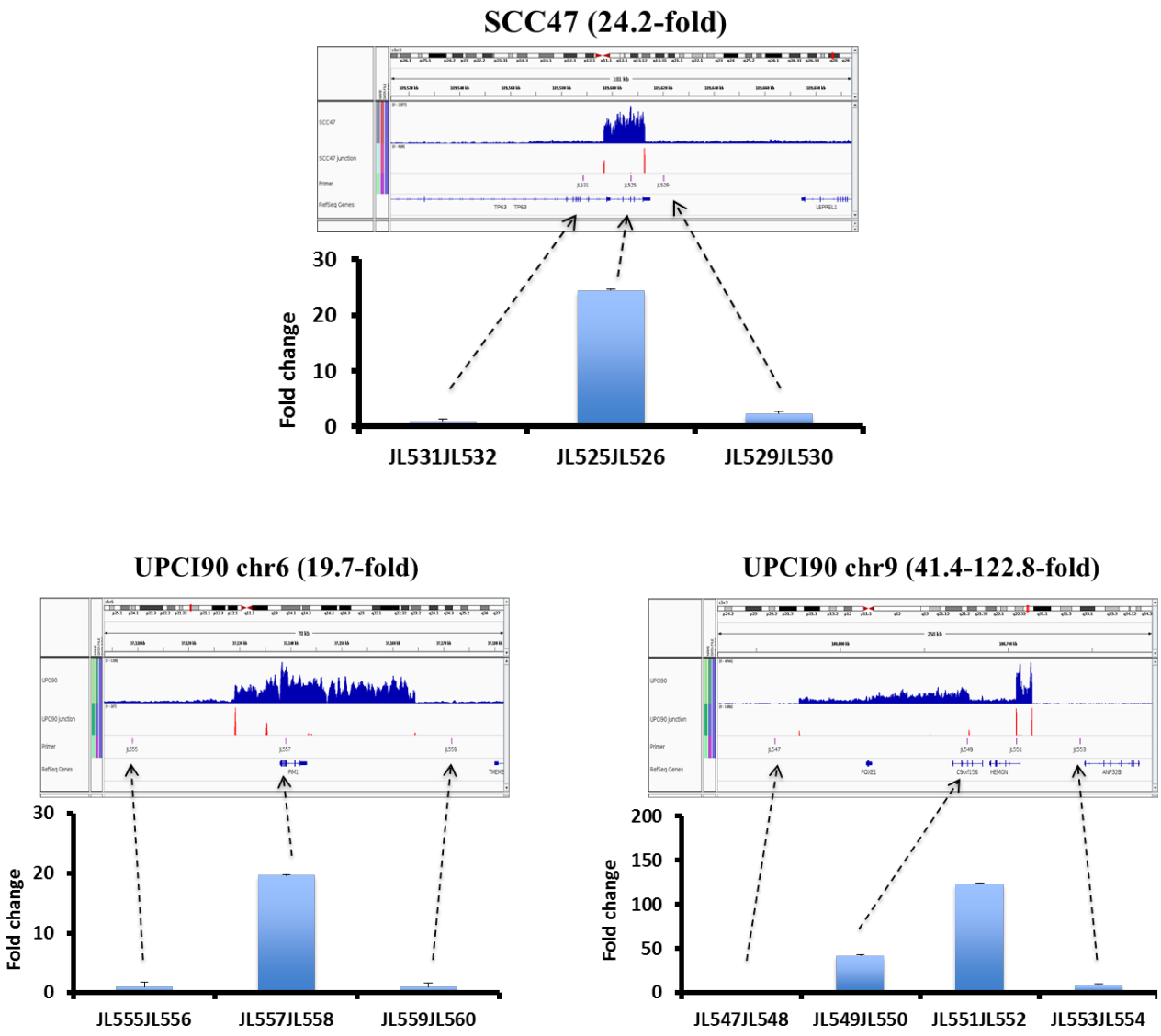
Supplemental Fig. 6. **FISH and SKY confirm HPV clusters**. FISH probe for HPV (*pink*) results are shown (*lower left*), along with SKY data showing extensive chromosomal translocations in (A) CaSki and (B) UPCI:SCC090 cells. See Fig. 2.



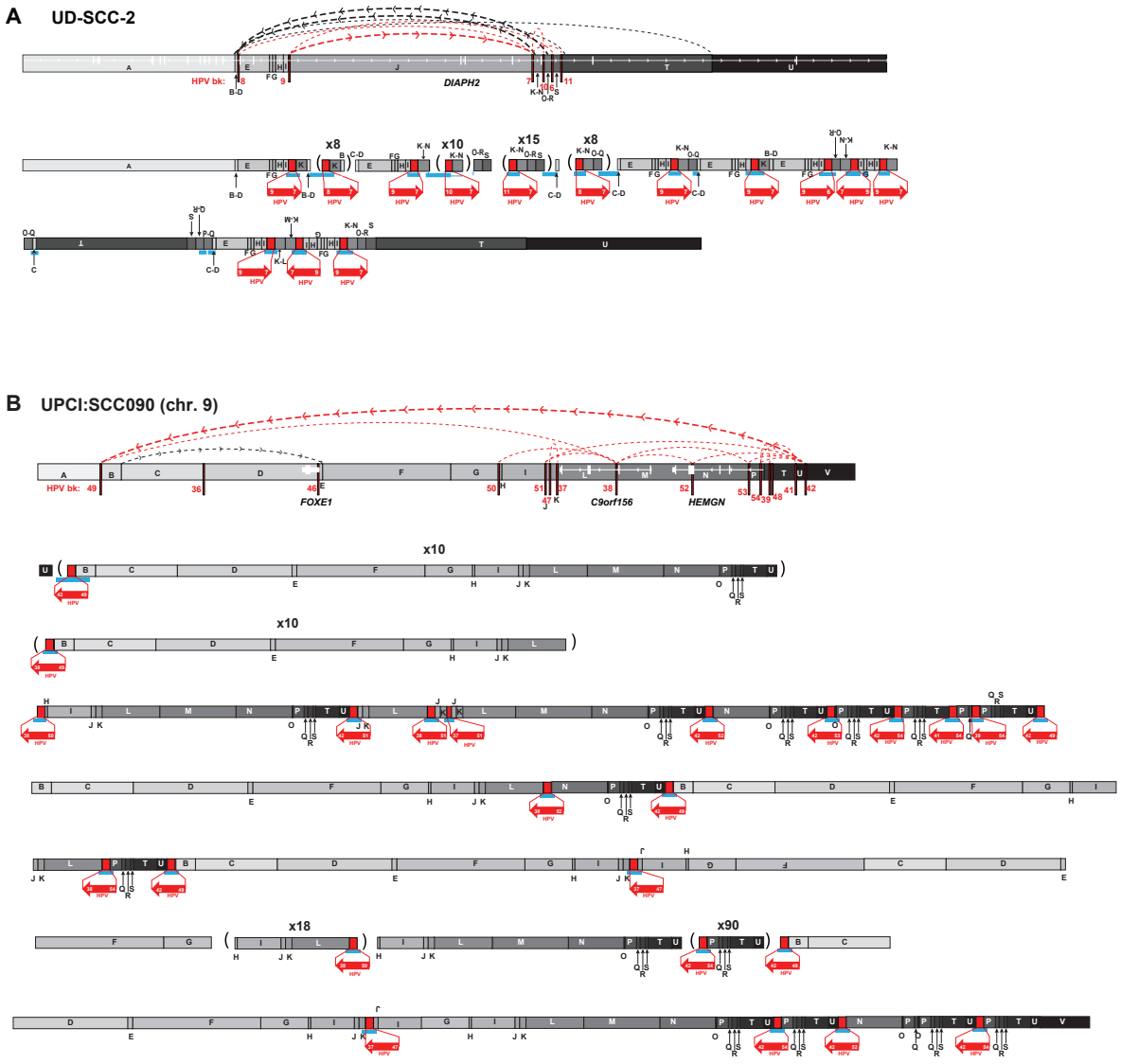
Supplemental Fig. 7. **Enrichment of hyper-amplification and structural variation near HPV integrants in diverse samples.** We assessed possible associations between (*left*) hyper-amplification or (*right*) structural variation (SV), respectively, and the proximity of HPV integrants in the 9 HPV-positive samples, i.e. SiHa, CaSki, HMS001, UM-SCC-104 (abbreviated here as SCC104), UD-SCC-2 (SCC2), UM-SCC-47 (SCC47), UPCI:SCC090 (UPCI90), Tumor B and Tumor A. Changes in DNA copy numbers were determined in 50 kb bins genome-wide, using CNAnorm (Gusnanto et al. 2012). Displayed here (*y-axis*) are the fractions of bins genome-wide (*black*) and containing HPV breakpoints (*gray*) that are associated with (*left*) hyper-amplification or (*right*) SVs in each sample (*x-axis*). *Left*: We defined hyper-amplification regions as having ploidy N > 8. We divided genomes into 50-kb bins and counted blocks containing both nearby HPV breakpoints and hyper-amplified regions (both in +/-250 kb windows). The statistical significance of hyper-amplification enrichment in bins was calculated using binomial test. *Asterisks*: samples with significant enrichment, $p < 0.05$ (i.e. SiHa, p= 3.86e-9; CaSki, p=5.81e-34; UD-UD-SCC-2, p=8.35e-83; UM-SCC-47, p= 4.08e-13; UPCI:SCC090, p=3.84e-19; Tumor B, p=7.42e-40). P-values were adjusted by Bonferroni multiple testing correction. *Right*: We detected structural variants (i.e. chromosomal translocations, deletions, inversions, intra-chromosomal translocations) using Hydra. As above, we divided genomes into 50-kb bins and counted blocks containing both nearby HPV breakpoints and SVs (both in +/-250 kb windows). *Asterisks*: samples showing significant enrichment of SVs around HPV breakpoints by binomial test, $p < 0.05$ (SiHa, p=2.15e-10; CaSki, p=6.59e-51; HMS001, p=3.08e-16; UD-SCC-2, p=1.59e-18; UM-SCC-47, p=8.79e-17; UPCI:SCC090, 4.42e-50; Tumor A, p=1.26e-34; Tumor B, p=1.74e-27). *P*-values were adjusted by Bonferroni multiple testing correction.



Supplemental Fig. 8. **Quantitative PCR confirms HPV-associated genomic amplifications**
Figure continued on next page

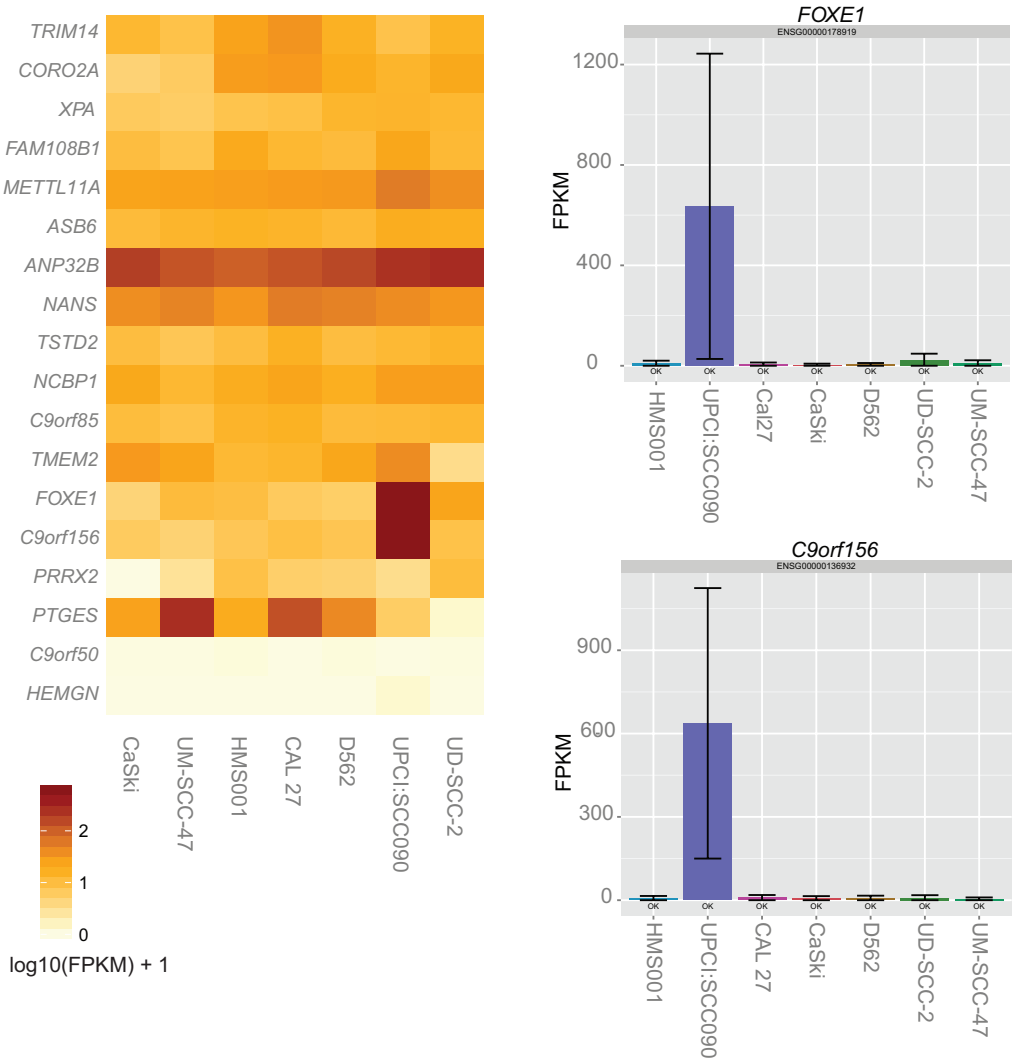


Supplemental Fig. 8. **Quantitative PCR confirms HPV-associated genomic amplifications** Copy number of amplification regions was measured using Power SYBR Green kit (ABI). Sample names were abbreviated here as follows: UD-SCC-2 (SCC2), UM-SCC-47 (SCC47), and UPCI:SCC090 (UPCI90), Locations of quantitative PCR (qPCR) amplicons to assess amplified and non-amplified flanking regions in indicated cancer samples were mapped to corresponding CNV coordinates (dashed arrows). Each measurement was first normalized to CAL 27 (–ΔCt). Fold changes (-ΔΔCt) were calculated by normalization of amplified regions to respective, flanking non-amplified regions.



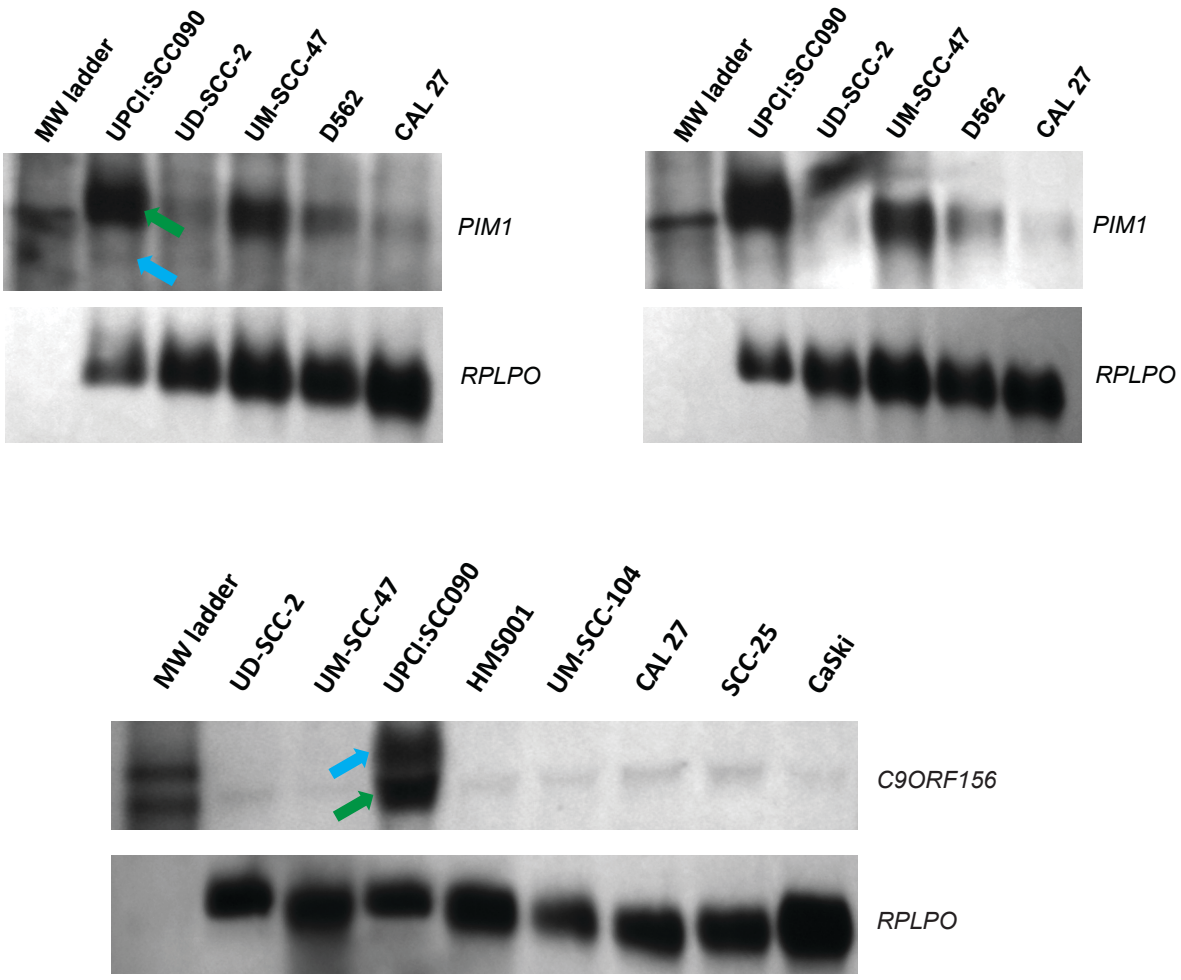
Supplemental Fig. 9. **Connectivity maps of additional HPV-associated genomic structural variants in HNSCC cell lines.** Schematics of genomic target loci before (*top*) and after (*bottom*) HPV integration in HNSCC cell lines (A) UD-SCC-2; (B) UPCI:SCC090, chr.9. Target gene schematic and connectivity map features are as described in Fig. 4 and Fig. 5 legends. HPV breakpoint numbers are listed in Supplemental Table 3. These schematics including the viral insertions are not drawn to scale. *Parentheses*, genomic segments with indicated fold amplification (*x N*) calculated from WGS data*; upside-down letters*, inverted segments.

A

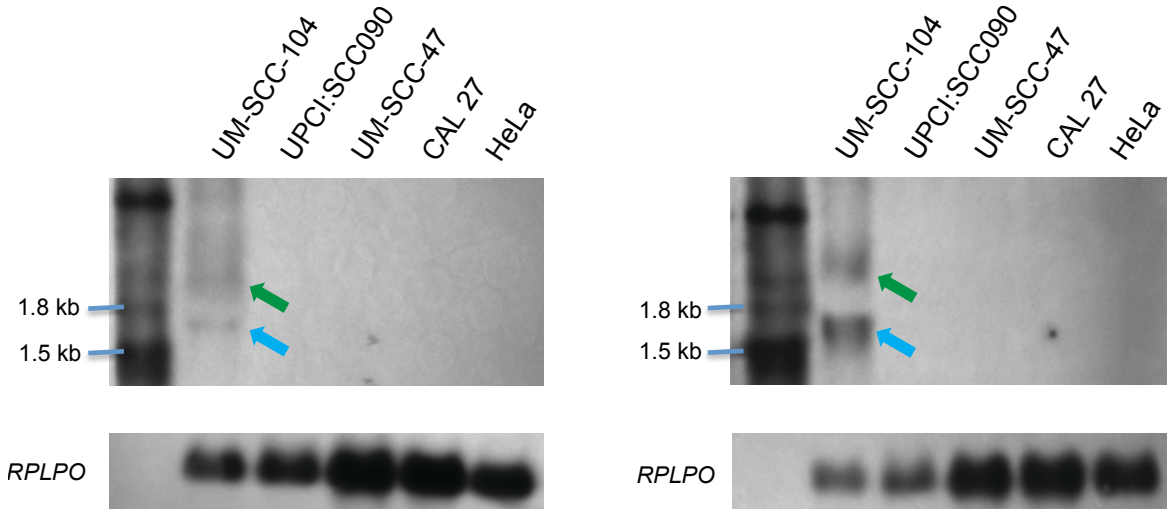


Supplemental Fig. 10. **Gene amplification from HPV-mediated structural variation results in upregulation of *FOXE1, C9ORF156* and *PIM1* transcripts in UPCI:SCC090 cells.** Figure continued on next page.

B



Supplemental Fig. 10 (continued). **Gene amplification from HPV-mediated structural variation results in upregulation of *FOXE1, C9ORF156* and *PIM1* transcripts in UPCI:SCC090 cells.**
(A) Expression levels of genes around HPV16 integrants on chr. 9 of UPCI:SCC090 cells were determined by RNA-Seq upon analysis by Cufflinks, expressed in FPKM (Fragments Per Kilobase of transcript per Million mapped reads). (*Left*) Heatmap shows the relative expression levels of genes in various samples, including UPCI:SCC090 cells bearing a nearby HPV16 integrant cluster on chr. 9. *Key, bottom left: dark brown*, strongest expression levels; *white*, weakest expression. We observed over-expression of two genes located in the center of the structural variation region. (*Right top*) RNA-Seq expression levels of *FOXE1* gene. (Right bottom) RNA-Seq expression levels of *C9ORF156* gene. **(B)** Northern blots probed for (left) *PIM1* (chr. 6) at exons 1-4, (*right*) *PIM1* exons 5-6 and (*bottom*) *C9ORF156* (chr. 9) at exon 4, each showing increased transcript levels specifically in UPCI:SCC090 cells. *Bottom of each panel*: we probed for *RPLPO* transcripts as a loading control. *Blue arrows*: viral-cellular fusion transcripts; *green arrows*: endogenous transcripts.



Supplemental Fig. 11. **Gene breakage at *SLC47A2* in UM-SCC-104 cells.** Northern blots probed with *SLC47A2* 5' (*left*) and 3' (*right*) probes demonstrate endogenous *SLC47A2* transcripts (*green arrows*) and disrupted *SLC47A2* showing evidence of HPV-mediated gene breaking (*blue arrow*). Our RACE experiments demonstrated gene breakage at *SLC47A2* in UM-SCC-104 due to HPV16 integration at intron 9. Fusion transcripts initiating from the *SLC47A2* gene terminating at the HPV16 E5 polyadenylation signal sequence site were detected. Fusion transcripts initiating from the HPV16 P97 or P670 promoter and terminating at the *SLC47A2* polyadenylation site also were detected. *Bottom*: *RPLPO* loading control.

sample name	yield (Gbp)	aligned (Gbp)	%bases >= Q30	median fragment size (nt)	mean coverage	number of SNPs	% dbSNP
SiHa	138.0	121.8	83.2%	304	40.9	3,188,808	92.8%
CaSki	139.7	116.3	86.1%	280	37.7	3,139,641	92.9%
UM-SCC-104	134.1	117.6	85.4%	296	39.0	3,264,329	93.3%
UD-SCC-2	133.4	117.0	89.0%	301	39.1	3,369,925	93.1%
UM-SCC-47	135.3	120.1	90.7%	287	40.0	3,583,093	92.8%
UPCI:SCC090	118.3	104.6	89.8%	302	35.2	3,344,202	93.4%
HMS001	125.1	109.9	89.2%	279	37.0	3,384,602	92.8%
Tumor A	120.1	105.7	86.2%	302	35.2	3,697,404	93.6%
Tumor B	117.9	102.8	87.0%	303	34.2	3,593,285	93.8%
CAL 27	121.7	107.6	89.0%	312	36.3	3,361,341	92.9%
D562	139.4	122.8	88.9%	298	41.1	3,255,895	93.4%
SCC-25	137.5	119.7	89.3%	278	40.3	3,295,979	93.5%

Supplemental Table 1. **Summary of WGS statistics for 12 cancer samples**
Some basic statistics for WGS analysis of the 12 cancer samples are provided here, including gross and aligned yields of WGS libraries, percentages of bases passing a quality filter, median library fragment size, mean fold coverage, numbers of SNPs detected, and percentage of SNPs found in the dbSNP database. The yield in Gbases was determined after quality filtering. The number of bases after alignment to the human genome (hg19) by CASAVA pipeline is shown in the fourth column. The size of fragments for paired-end sequencing is ~300 bp. Greater than 90% of detected SNPs were reported SNPs in dbSNP Build 131. Please see also Table 1 in main paper.

Supplemental Table 2. **Clonal single nucleotide polymorphisms found in HPV16 viral genomes in cancer samples**

CHROM	POS	REF	ALT	SiHa	UM-SCC-104	Tumor B	CaSki	UD-SCC-2	SCC47	UPCI: SCC090	HMS001
HPV16	12	T	C	.	.	.	C	.	.	C	.
HPV16	31	C	T	.	.	.	.	.	T	.	.
HPV16	109	T	C	.	.	.	.	.	C	.	.
HPV16	131	A	G	.	.	.	G	.	.	G	.
HPV16	132	G	T	.	.	.	.	.	T	.	.
HPV16	178	T	G	.	.	.	.	.	.	.	G
HPV16	188	G	C	.	.	.	.	.	C	.	.
HPV16	285	C	G	.	.	.	.	.	G	.	.
HPV16	286	T	A	.	.	.	.	.	A	.	.
HPV16	289	A	G	.	.	.	.	.	G	.	.
HPV16	335	C	T	.	.	.	.	.	T	.	.
HPV16	350	T	G	G	G	.	G	.	.	G	.
HPV16	403	A	G	.	.	.	.	.	G	.	.
HPV16	442	A	C	C	.	.	.	.	.	.	.
HPV16	645	A	C	C	.	.	.	.	.	.	.
HPV16	647	A	G	.	.	.	.	.	G	.	G
HPV16	712	C	A	.	.	.	.	A	.	.	.
HPV16	789	T	C	.	.	.	.	.	C	.	.
HPV16	795	T	G	.	.	.	.	.	G	.	.
HPV16	846	T	C	.	.	.	.	.	.	.	C
HPV16	921	T	C	.	.	.	.	.	C	.	.
HPV16	945	A	G	.	.	.	.	.	.	G	.
HPV16	1035	T	G	.	.	.	.	.	.	.	G
HPV16	1096	C	G	.	.	.	.	.	G	.	.
HPV16	1136	TGG	TGGG	TGGG	TGGG	.	.	.	.	.	TGGG
HPV16	1162	G	A	.	.	.	.	.	A	.	.
HPV16	1193	A	G	G	.	.	.	.	.	.	.
HPV16	1199	T	C	.	.	.	.	.	C	.	.
HPV16	1251	G	A	.	.	.	A	.	.	.	.
HPV16	1365	T	A	.	.	.	.	.	A	.	.
HPV16	1376	C	T	.	.	.	.	.	T	.	.
HPV16	1400	T	C	.	C	.	.	.	.	.	.
HPV16	1415	C	T	.	.	.	.	.	T	.	.
HPV16	1417	C	A	.	.	.	A	.	.	.	.
HPV16	1425	C	G	.	.	.	.	.	G	.	.
HPV16	1485	T	C	.	.	.	.	.	C	.	.
HPV16	1514	G	A	.	A	.	.	.	.	.	A
HPV16	1521	T	A	.	.	A	A	A	.	A	.
HPV16	1623	C	T	.	.	.	.	.	T	.	.
HPV16	1743	C	A	.	.	.	.	.	A	.	.
HPV16	1836	G	A	.	.	.	.	A	.	.	.
HPV16	1841	A	G	G	.	.	.	.	G	.	.
HPV16	1991	A	T	.	.	.	.	.	T	.	.
HPV16	2040	C	T	.	.	.	.	.	T	.	.
HPV16	2219	G	C	.	.	.	.	.	C	.	.
HPV16	2236	C	G	.	.	.	.	.	G	.	.
HPV16	2248	G	A	.	.	.	.	.	A	.	.
HPV16	2261	C	T	.	.	.	.	.	T	.	.
HPV16	2286	C	T	.	.	.	.	.	T	.	.

HPV16	2343	C	T	.	.	.	.	.	T	.	.
HPV16	2354	T	C	.	.	.	.	.	C	.	.
HPV16	2375	T	G	.	.	G	.	.	.	.	.
HPV16	2456	C	T	.	.	T	T	T	.	T	.
HPV16	2585	T	C	.	.	.	.	.	C	.	.
HPV16	2607	A	C	.	.	.	.	.	C	.	.
HPV16	2630	T	A	.	.	.	.	.	A	.	.
HPV16	2764	C	G	.	.	.	.	.	G	.	.
HPV16	2859	C	A	.	.	.	.	.	A	.	.
HPV16	2925	A	G	G	G	G	G	G	G	G	.
HPV16	2937	A	G	.	.	G	G	G	.	G	.
HPV16	3042	C	T	.	.	.	.	.	T	.	.
HPV16	3067	G	A	A	.	.	.	.	.	.	.
HPV16	3158	C	A	.	.	.	.	.	A	.	.
HPV16	3160	C	T	.	.	.	.	.	T	.	.
HPV16	3181	G	A	.	.	.	.	.	A	.	.
HPV16	3185	C	A	.	.	.	.	.	A	.	.
HPV16	3248	G	A	.	.	.	.	.	A	.	.
HPV16	3364	G	A	.	.	A	.	.	.	.	.
HPV16	3383	T	C	.	.	C	C	C	.	C	.
HPV16	3409	C	T,A	.	T	A	T	T	.	T	
HPV16	3588	G	T	.	.	.	.	T	.	.	.
HPV16	3634	G	T	.	.	.	.	T	.	.	.
HPV16	3683	C	A	A	.	A	A	A	.	A	.
HPV16	3777	G	T	.	.	.	.	.	T	.	.
HPV16	3786	C	A	.	.	.	.	.	A	.	.
HPV16	3804	T	G	.	.	.	.	.	G	.	.
HPV16	3839	T	A	.	.	.	.	A	.	.	.
HPV16	3857	T	C	.	.	.	.	.	C	.	.
HPV16	3871	C	A	.	.	.	A	.	.	.	.
HPV16	3978	A	C	C	C	C	C	C	C	C	.
HPV16	3990	C	T,G	G	.	.	.	.	T	.	
HPV16	4041	A	G,T	G	G	G	G	G	T	G	
HPV16	4088	T	C	.	.	.	.	.	C	.	.
HPV16	4113	T	A	.	A	.	.	.	.	.	.
HPV16	4148	A	C	.	.	.	.	.	C	.	.
HPV16	4175	C	T	T	.	T	T	T	.	T	.
HPV16	4194	TG	GG,T	T	.	.	GG	.	.	GG	
HPV16	4199	G	A	.	.	.	.	.	A	.	.
HPV16	4225	A	C	.	.	.	C	.	.	.	.
HPV16	4227	T	C	C	C	C	C	C	.	C	.
HPV16	4280	T	C	.	.	.	.	.	C	.	.
HPV16	4364	A	T	T	T	T	T	T	T	T	.
HPV16	4427	G	T	.	.	.	.	.	T	.	.
HPV16	4457	A	C	C	.	.	.	.	.	.	.
HPV16	4460	G	A	.	.	.	.	.	A	.	.
HPV16	4517	A	G	.	.	.	.	.	G	.	.
HPV16	4544	T	G	.	.	.	.	.	G	.	.
HPV16	4599	T	C	.	.	.	.	.	C	.	.
HPV16	4613	T	C	.	.	.	C	.	.	C	C
HPV16	4643	T	A	.	.	.	.	.	A	.	.
HPV16	4724	C	T	.	.	.	.	.	T	.	.
HPV16	4853	C	T	.	.	.	.	.	T	.	.
HPV16	4886	A	G	.	.	.	.	.	G	.	.

HPV16	4937	G	A	A	A	A	A	A	A	A	.
HPV16	4962	G	A	.	A	.	.	.	.	.	.
HPV16	5040	T	C,G	G	.	C	C	C	C	.	
HPV16	5041	C	A	A	.	.	.	.	.	.	.
HPV16	5141	G	A	.	.	.	.	.	A	.	.
HPV16	5225	A	G,T,C	C	T	G	G	G	.	G	
HPV16	5235	G	A	A	.	.	.	.	A	.	.
HPV16	5258	A	G	.	.	.	.	.	G	.	.
HPV16	5289	A	C	.	.	.	.	.	C	.	.
HPV16	5309	T	C	.	.	.	.	.	C	.	.
HPV16	5327	A	C	.	.	.	C	.	.	.	.
HPV16	5368	C	T	.	.	.	.	.	T	.	.
HPV16	5378	G	A	.	.	.	.	.	A	.	.
HPV16	5388	G	A	.	.	.	.	.	A	.	.
HPV16	5402	T	C	.	.	.	.	.	C	.	.
HPV16	5486	C	T	.	.	.	.	.	T	.	.
HPV16	5494	T	C	.	.	.	.	.	C	.	.
HPV16	5505	G	A	.	.	.	.	.	A	.	.
HPV16	5506	C	G	.	.	G	.	.	.	.	.
HPV16	5563	C	G	.	.	.	.	.	G	.	.
HPV16	5697	G	A	.	.	.	.	.	A	.	.
HPV16	5834	C	G	.	.	.	.	.	.	C/G	.
HPV16	5835	A	G	G	.	.	.	.	.	.	.
HPV16	5863	C	T	.	.	.	.	.	T	.	.
HPV16	5910	T	C	.	.	.	.	.	C	.	.
HPV16	6164	C	A	.	.	.	.	.	A	.	.
HPV16	6241	C	G	G	G	G	G	G	G	G	.
HPV16	6246	T	C	.	.	.	.	.	C	.	.
HPV16	6303	G	A	.	.	A	.	.	.	.	.
HPV16	6315	A	G	.	.	.	.	.	G	.	.
HPV16	6433	A	G	G	G	G	G	G	G	G	.
HPV16	6481	T	C	.	.	.	.	.	C	.	.
HPV16	6558	C	T	.	.	.	.	.	T	.	.
HPV16	6694	A	C	.	.	.	.	.	C	.	.
HPV16	6720	G	A	.	.	.	.	.	A	.	.
HPV16	6853	C	T	.	.	.	.	.	T	.	.
HPV16	6861	T	C	.	.	.	C	.	.	C	.
HPV16	6864	C	T	.	.	.	.	.	T	.	.
HPV16	6901	AC	ACATC	ACATC	ACATC	.	.	.	ACATC	.	ACATC
HPV16	6927	T	C	.	.	.	.	.	C	.	.
HPV16	6948	AGATGAT	AGAT	AGAT	AGAT	.	.	.	AGAT	.	AGAT
HPV16	6969	C	T	.	.	.	.	.	T	.	.
HPV16	6993	G	A	.	.	.	.	.	A	.	.
HPV16	7059	G	A,T	.	.	.	.	.	T	.	A
HPV16	7083	A	G	.	G	.	.	.	.	.	.
HPV16	7174	A	C	.	.	.	.	.	.	.	C
HPV16	7176	T	C	.	.	.	.	.	.	.	C
HPV16	7185	T	A	.	.	.	.	A	.	.	.
HPV16	7192	G	T	T	T	T	T	T	T	T	T
HPV16	7200	T	C	.	.	.	.	.	.	.	C
HPV16	7232	A	C	.	.	.	.	.	C	.	.
HPV16	7262	A	C	.	.	.	.	.	C	.	.
HPV16	7269	C	T	.	.	.	.	.	.	.	T
HPV16	7286	A	C	.	.	.	.	.	.	.	C

HPV16	7288	A	C	.	.	.	.	.	.	.	C
HPV16	7309	C	G	.	.	G	.	G	.	G	.
HPV16	7328	T	G	.	.	.	.	.	G	.	.
HPV16	7393	C	T	.	.	.	T	.	.	.	.
HPV16	7432	G	GC	GC	GC	.	.	.	.	.	GC
HPV16	7483	A	C	.	.	.	.	.	C	.	.
HPV16	7487	G	A	.	.	.	.	.	A	.	.
HPV16	7519	G	A	A	A	A	A	A	A	A	A
HPV16	7667	C	T	.	.	.	.	.	T	.	.
HPV16	7687	C	A	.	.	.	.	.	A	.	.
HPV16	7728	A	C	.	.	.	.	.	.	.	C
HPV16	7762	C	T	.	.	.	.	.	T	.	.
HPV16	7779	T	C	.	.	.	.	.	.	.	C
HPV16	7784	C	T	.	.	.	.	.	T	.	.
HPV16	7824	G	A	.	.	.	.	.	A	.	.
HPV16	7832	G	T	.	.	.	.	.	T	.	.
HPV16	7840	G	A	.	.	.	.	.	.	.	A
HPV16	7860	CG	C	.	.	.	.	C	.	.	.
HPV16	7861	GA	G	G	G	G	A	.	A	G	G
HPV16	7867	G	A	.	.	.	.	.	.	A	.

Supplemental Table 3. **HPV breakpoints confirmed by PCR amplification and Sanger sequencing.**
Please note, counts of breakpoints presented in Table 1 are slightly higher in some cases; those listed here have been experimentally validated.

Sample Name	BreakId	# reads	junction sequence	description	HPV start	HPV stop	HPV gene	chrom #	chrom start	chrom stop
SiHa	3	42	TGCCACTACACCCAGCTAATTTTTGTATTAGGCAGCACTTGGCCAACCACTCCGCCGCG	Chr. 13 (q22.1) at position 73,789,077-73,788,866 to HPV16 nt 3384-3417 (share TTGT)	3384	3417	E2/E4	13 (q22.1)	73,789,077	73,788,866
SiHa	2	44	GGAAGTGCAGTTTGATGGAGACATATGCTGCTCAGCTTAGCTGACTAACC	HPV16 nt 3105-3132 to chr.13 (q22.1) at position 74,087,562-74,087,537 (share ATGC)	3105	3132	E2	13 (q22.1)	74,087,562	74,087,537
CaSki	97	3	GAATACATTTACCTGACCCCAATAAGTTTGATATGTTGAAAATACAATGCATAAA	HPV16 nt 5858-5887 to chr. 10 (p14) at position 11,689,965-11,689,939	5858	5887	L1	10 (p14)	11,689,965	11,689,939
CaSki	98	3	ATTGCAGCCTCCAACATGCAGAAACTGGATGTGTGCGTGGTGGTGTG	HPV16 nt 4739-4769 to chr. 10 at position 11,725,462-11,725,486 (share ACTGGA)	4739	4769	L2	10 (p14)	11,725,462	11,725,486
CaSki	99	8	TACAATGGGCCTACGATAATGACATAGTTTTCTTCATAACGTGTTATTTCATC	HPV16 nt 1945-1972 to chr. 10 (p14) 11,742,450-11,742,424 (share GT)	1945	1972	E1	10(p14)	11,742,450	11,742,424
CaSki	100	22	TGACCCAGTCAATGTTATCCCAGAAAATACTGTCTACTTGCCTCCTGTCCCAGTATCTAAGGTTGTAAGCACGGATGAAT	Chr. 10 (p14) at position 11,742,513 – 11,742,484 to HPV16 nt 5665-5716 (share AC)	5665	5716	L1	10 (p14)	11,742,513	11,742,484
CaSki	101	11	CAAGCAACTTATATAATAATACTAAACTACAATAGCATAGCAATAGAGTCTTCAGGT	HPV16 nt 7881-7905/1-9 to Chr. 11 6,762,248 – 67,622,24 (share TA)	7881	7905	LCR	11(p15.4)	6,762,248	6,762,224
CaSki	105	13	CCTTTGAGGTTAAATTCTGAAACAACCCTCCATCCCCCCAGATGTATCAGGATTTAG	Chr. 11 (p15.4) 6,762,290-6,762,261 to HPV16 nt 4608-4636 (share TC)	4608	4636	L2	11 (p15.4)	6,762,290	6,762,261
CaSki	106	12	ACTGTTATCTGCTCTTTTCTGGTATGTCAGAAATTctggtatgtaaatttaatcAGAACGCAAATTTTTACTACAAGCAGGATTGAAGGCC	Chr. 11 (p13) at position 32,277,860 – 32,277,839 to Δ13 bp (tatgtcagaaatt) insertion to Chr 11 (p13) at position 32,389,325 – 32,389,303 to HPV16 nt 7029-7065 (share AGAA)	7029	7065	L1	11 (p13) 11 (p13)	32,277,860 32,389,325	32,277,839 32,389,303
CaSki	104	9	TTTTGACCTCATCATTACTTCGGCATTAGGAAgtaagctccTAATGGCATATATGATATTTATGCAGATG	Chr. 11 (q22.1) at position 100,637,470 – 100,637,501 and 100,637,834-100,637,842 to HPV 16 nt 5321-5349	5321	5349	L2	11 (q22.1) 11 (q22.1)	100,637,470 100,637,834	100,637,501 100,637,842

CaSki	103	24	GATAATCCTGCATATGAAGGTATAGATG TAAAATGAGAATAAATAGGAAGAAGATC A	HPV16 nt 4992-5020 to chr. 11 (q22.1) at position 100,637,500 -100,637,527	4992	5020	L2	11 (q22.1)	100,637,500	100,637,527
CaSki	102	3	ATTTTATAATCCAGATACACAGCGGCTT TGGGAGGCCAAGGCAGGTGGATCA	HPV16 nt 5904-5930 to chr. 11 (q22.1) at position 100,642,594 – 100,642,568 (share CT)	5904	5930	L1	11 (q22.1)	100,642,594	100,642,568
CaSki	76y	2	TTTGCTCTTGTTGCCCAGGCTGGAGTG CAATGGCACTATATGGTGCAGCTAACAC AGGTAAATCATTATT	Chr. 11 (q22.1) at position 100,677,954-100,677,989 to HPV16 nt 2285-2320 (share AC)	2285	2320	E1	11 (q22.1)	100,677,954	100,677,989
CaSki	109	23	GGGTGTTGGGGAGCTACAGTGCCAAGA AAATACTTGTAAATATTAAGTTGTATGTG TGTTTGTATGTATG	Chr. 14 (q12) at position 25,288,208 – 25,288,180 to Δ3 bp (ATA) insertion to HPV16 nt 7219-7256	7219	7256	LCR	14 (q12)	25,288,208	25,288,180
CaSki	107	13	AAAGACATCTGGACAAAAAGCAAAGTGC TCCTAGACTTGCACAGCAGACTAG	HPV16 nt 450-474 to chr. 14 (q21.3) at position 47,984,368 – 47,984,397 (share AG)	450	474	E6	14 (q21.3)	47,984,368	47,984,397
CaSki	116	6	TGTTTTTTATTTTTAGAGATGGGCATTGA TTATATTGGTATAAAACAGGT	Chr. 19 (q13.42) at position 55,804,440-55,804,413 to HPV16 nt 1817-1838	1817	1838	E1	19 (q13.42)	55,804,440	55,804,413
CaSki	117	6	GAGGTAATCCAGAGCCCCAAAGGCGCG TGCCCTTCAATGTTCATGAAGGAATACG AACATATTTTGTGCA	Chr. 19 (q13.42) at position 55,816,527-55,816,490 to HPV16 nt 3230-3263 (share AT)	3230	3263	E2	19 (q13.42)	55,816,527	55,816,490
CaSki	112	82	ACTAGAAGATACTTATAGGTTTGTAACA TCAAAGCCTCGCGTGGTGGTGGGCTCC TGTAATCCCAGCTACTCAGGAGGCTGA GGCAG	HPV16 nt 6876-6902 to Δ4bp (ATCA) insertion to chr. 19 at position 55,818,769-55,818,824	6876	6902	L1	19 (q13.42)	55,818,769	55,818,824
CaSki	114	39	TTTTCGTACAGTCCTGGAGGTCAGAAGT CCCTGTATTGCATGAATATATGTTAGATT TGC	Chr. 19 (q13.42) at position 55,821,546 – 55,821,576 to Δ3 bp (TGT) insertion to HPV16 nt 582-607	582	607	E7	19 (q13.42)	55,821,546	55,821,576
CaSki	113	28	GAGGTCAGAAGTCCCACATGGTCTCTT GCCAGACACCGGAAACCCCTGCCACAC C	Chr. 19 (q13.42) at position 55,821,562 -55,821,587 to Δ1 bp (T) insertion to HPV16 nt 3486-3513	3486	3513	E4	19 (q13.42)	55,821,562	55,821,587
CaSki	118	6	GACATACATACATTCTATGAATTCCACC TGAGACCGAAACTGGGAAAAGACAC	HPV 16 nt 6801-6827 to chr. 2 (p24.1) at position 22,639,092 - 22,639,065 (share AC)	6801	6827	L1	2 (p24.1)	22,639,092	22,639,065

CaSki	126	46	ATTTCTATCACTGATGGGCATTTGGGTTGATAATAAACACGTGTGTATGTGTTTTTAAATGCTTGTGTAACTATTGTGTGATGCAACATAAATAAACTTA	Chr. 2 (p24.1) at position 22,639,132 – 22,639,102 to HPV16 nt 7259-7329 (share AT)	7259	7359	LCR	2 (p24.1)	22,639,132	22,639,102
CaSki	119	26	TACCATCTGTACCCTCTACATCTTTGCTAATATTGGCTCTAGAATTACAAA	HPV16 nt 5380-5404 to chr. 2 (p24.1) at position 22,645,646 – 22,645,618 (share TTT)	5380	5404	L2	2 (p24.1)	22,645,646	22,645,618
CaSki	122	9	GTCTTTGTCTTATTTCATTTAATATATAATAGGTCGTGGTCAGCCATTAGGTGTGGGCAT	Chr. 2 (p24.1) at position 22,662,013 – 22,662,035 to Δ6bp (ATATAA) insertion to HPV16 nt 5957-5987	5957	5987	L1	2 (p24.1)	22,662,013	22,662,035
CaSki	120	8	AAGTGCCTTTCTAAAAAGTAATTCACAGAGGAGGACATTAAACAATGAATGGAGAAAA	HPV16 nt 2030-2057 to chr. 2 (p24.1) at position 22,702,082 – 22,702,053	2030	2057	E1	2 (p24.1)	22,702,082	22,702,053
CaSki	127	11	AGAACAGTTTCCCCGCACCCACAGCCAAATAAAGTTGTAGACCCTGCTTTTG	Chr. 2 (p23.3) at position 27,318,140 – 27,318,113 to Δ1bp (A) insertion to HPV16 nt 4940-4962	4940	4962	L2	2 (p23.3)	27,318,140	27,318,113
CaSki	123	16	CCTCTCCGTTTCCTTTGACACCACCCTCCAGGGTCTTTTTTATTTTTTGCAGTCAGTACAGTAGTGGAAGTGGGG	Chr. 2 (p23.3) at position 27,358,998 – 27,359,046 to HPV16 nt 1364-1392 (share TTG)	1364	1392	E1	2 (p23.3)	27,358,998	27,359,046
CaSki	121	12	ATTTACAAGCAACTTATATAATAATAAGCCAAGGCCTGGGCTTCAGGGAGGTAGTG	HPV16 nt 7876-7901 to Δ3bp (AGC) insertion to chr. 2 (p23.3) at position 27,360,017-27,360,043	7876	7901	LCR	2(p23.3)	27,360,017	27,360,043
CaSki	128	121	GCAAAGATTCCATAATATAAGGGGTCGAGGCCCATGGTATCGATCTGTGCTGCA	HPV16 nt 469-495 to Δ3bp (AGG) insertion to chr. 20 (p11.1) at position 26,257,343 – 26,257,366	469	495	E6	20 (p11.1)	26,257,343	26,257,366
CaSki	129	8	TACTCCAAAGATGATATACGTTTTCAAGGCAATGATGGTTTGAAAGCAAAGACAGCGGGTATGGCAATA	Chr. 21 (q21.1) at position 21,954,751 – 21954788 to HPV16 nt 1242-1272	1242	1272	E1	21 (q21.1)	21,954,751	21,954,788
CaSki	92X	4	TGCACCAAAAGAGAACTGCAATGTGTTGCATGGGCTCCCAGTGCTGGACAGGTTGATGCCAGGGACATAGGCTACATAACCCTTTCCTATGAAGAGATTA	HPV16 nt 87-107 to chr. 3 (q23) at position 140,480,013 - 140,479,936 (share GT)	84	107	E6	3 (q23)	140,480,013	140,479,936
CaSki	133	18	CTACATGGCATTGGACAGGACATAAAATCCGGGAGTGCTGTGTTTCATGGT	HPV16 nt 3704-3728 to chr. 6 (p21.1) at position 45,659,121 - 45,659,097	3704	3728	E2	6 (p21.1)	45,659,121	45,659,097

CaSki	134	27	AAGGCCACCACTTTGCCTTGTTTATAAG ATTTTTGCAAGGCATACCTAAAAAAAATT GCA	Chr. 6 (p21.1) at position 45,659,179 – 45,659,154 to HPV16 nt 2246-2280 (share A)	2246	2280	E1	6 (p21.1)	45,659,179	45,659,154
CaSki	140	19	CCTGGAGGAGGCATTTCAGCCTGAGGG TCAAGTTGACTATTATGGTTTAT	Chr. 7 (p22.1) at position 6,964,845 - 6,964,819 to HPV16 nt 3199-3226 (share GAGGG)	3199	3226	E2	7 (p22.1)	6,964,845	6,964,819
CaSki	135	4	CTTTTGGTGGTGCATACAATATTCCTTTT AATATTCCTTTTAATAAATGCCTTTTTAA AATATTTTAAATGATTATGGATATTGCTT TGAAATTTGTAATAGATGCAGACTTCTTA GTTAAATAAAtcggtgtaaaacccgattgtatattcc	HPV16 nt 5437-5464 to Δ1bp (T) insertion to HPV16 nt 5454-5464 to chr. 7 (p21.3) at position 7,628,821 – 7,628,904 to Δ2bp (AA) insertion to chr.7 (p11.2) at position 7,041,846-7,041,820	5437	5464	L2	7 (p21.3)	7,628,821	7,628,904
					5454	5464	L2	7 (p11.2)	7,041,846	7,041,820
CaSki	136	3	TGGTGGTGCATACAATATTCCTTTTAATA TTCCTTTTAATAAATGCCTTTTTAAAATA TTTT	HPV16 nt 5441-5464 to chr. 7 (p21.3) at position 7,628,821-7,628,846	5441	5464	L2	7 (p21.3)	7,628,821	7,628,846
CaSki	139	9	AAGGAGAACTGGTTGCTAGAATATCTGG GTCCTCCTTCTATAGTTTCTTTAGTGGA AGAAACTAGTTTTA	Chr. 7 (p21.3) at position 7,679,158-7,679,190 to HPV16 nt 4541-4578 (share T)	4541	4578	L2	7(p21.3)	7,679,158	7,679,190
CaSki	137	3	AAGAAACTAGTTTTATTGATGCTGGTGC ACTTAAGGATGGGTAACGGGAAGAACC AC	HPV16 nt 4564-4593 to chr. 7 (p21.3) at position 7,680,926-7,680,957 (share GCAC)	4564	4593	L2	7(p21.3)	7,680,926	7,680,957
CaSki	138	2	GCTGGACAAGCAGAACCGGACAGAGCC CATTACAAGCTACTGCTTGTGCTATAAT TGCTAGAGT	HPV16 nt 685-719 to chr. 7 (p11.2) at position 54,144,011 – 54,144,039	685	719	E7	7 (p11.2)	54,144,011	54,144,039
CaSki	103x	2	GTTACGAAAACGACGTAAACGTTTAGAG TCGGAGAGTGGCTGGGCACAGTG	HPV16 nt 5597-5621 to chr. 7 (q11.21) at position 62,984,329 -62,984,355 (share A)	5597	5621	L2	7 (q11.21)	62,984,329	62,984,355
CaSki	107x	2	CATGGATATACAGTGGAAGTGTATATGT CCCCAATAGTTATTTTTT	HPV16 nt 3091-3111 to Δ5bp insertion to chr. X (q27.3) at position 144,770,260-144,770,277	3091	3111	E2	X (q27.3)	144,770,261	144,770,277
CaSki	108x	4	ACCACTTACAAATATTTTAAATGTACAGG GAAATCAGATTGACCATAGAAATAT	HPV16 nt 1430-1455 to chr. X (q27.3) 144,775,156 - 144,775,186 (share TAC)	1430	1455	E1	X (q27.3)	144,775,156	144,775,186
CaSki	142	4	CAGGGGATGCTATATCAGATGACGAGC TCAATTTGGAAGAAAAGAACCAGTG	HPV16 nt 950-975 to chr. X (q27.3) 144,778,296-144,778,324 (share GAG)	950	975	E1	X (q27.3)	144,778,296	144,778,324

CaSki	143	23	CAAGCAATTGAACTGCAACTAATGGCTA TATATATATTTGGCTATA	HPV16 nt 2965-2986 to chr. X (q27.3) 144,789,749 – 144,789,773 (share A)	2965	2986	E2	X (q27.3)	144,789,749	144,789,773
CaSki	145	73	CTCTAGCAGGGACTCTGGCCTTGCTAG ACACATATATATATTTGTGAAGAAGC	Chr. X (q27.3) at position 144,798,010-144,798,032 to Δ3bp (CTA) insertion to HPV16 nt 3156-3182	3156	3182	E2	X (q27.3)	144,798,010	144,798,032
CaSki	144	11	ATCTGTACCTTCCATTGTACACAGTGCT CAAATGTACATATGTGTCCCCTTTTCTTC CTTTTCACCCAATGAAACCCTGTCTTAC TATTC	HPV16 nt 4598-4613 to chr. X to Δ11bp (GTACACAGTGC) insertion to HPV16 nt 1930-1920 to Δ7bp (TATGTGT) insertion to chr. X (q27.3) at position 144,798,049-144,798,093	4598 1930	4613 1920	L2 E1	X (q27.3)	144,798,049	144,798,093
UM-SCC-104	2	21	ATGTTTAAGATATAGATTTAAAAAGCATT GTAAGAGTGAGAGATTCAGTTGGTTTGC GGG	HPV16 nt 3651-3682 to Δ 1bp (A) insertion to chr. 17 (p11.2) at position 19,609,637-19,609,611	3651	3682	E2	17 (p11.2)	19,609,637	19,609,611
UM-SCC-104	3	30	GCAGGTGCCCGTGGGAGGTGGCTGAG GCAGTGTCGTCTACATGGCATTGGACA GG	Chr. 17 (p11.2) at position 19,609,671-19,609,643 to HPV nt 3694-3722 (share GCA)	3694	3722	E2	17 (p11.2)	19,609,671	19,609,643
UD-SCC-2	8	14	TCAGGCTCTTGTGTGGTTCCATATGAAT TTTAGGATTACATATGATAATCCTGCATA TGAAGGTATAGATGTGGATAtgagtttaatga aatttctgcaagggtctgtaatatgttt	Chr. X (q21.33) at position 96,215,084-96,215,121 to HPV 16 nt 4983-5025 (share ATTA) to HPV 2324-2365 (share TA)	4983 2324	5025 2365	L2 E1	X (q21.33)	96,215,084	96,215,121
UD-SCC-2	9	59	GATTTGCATTTTGTTCCAGTGATAGTTA GTTATCCTTTAGTATCAGGTCCTGATAT ACCC	Chr. X (q21.33) at position 96,238,745-96,238,769 to Δ 6bp (TAGTTA) insertion to HPV 16 nt 5459-5486	5459	5486	L2	X (q21.33)	96,238,745	96,238,769
UD-SCC-2	7	288	ACAGGATGTATAAAAAAACATGGATATA CAGTTAGTTCTTCTTGTTGGAAACTA	HPV16 nt 3073-3104 to chr. X (q21.2) at position 96,369,879 to 96,369,903 (share AGT)	3073	3104	E2	X (q21.33)	96,369,879	96,369,903
UD-SCC-2	10	7	ATGCCCTAAAGGGTGCTGTTACCAAAG GTTACGAACATATTTTGTGCAGTTTAAA GATG	Chr. X (q21.33) at position 96,373,444-96,373,470 to Δ 2bp (GT) insertion to HPV16 nt 3245-3274	3245	3274	E2	X (q21.33)	96,373,444	96,373,470
UD-SCC-2	6	48	TTTTGGGTTACACATTTACAAGCAACTT GAGCTAAAAAAAAAAAAAAAAAAAAAC CCACAAAAAACTC	HPV16 nt 7863-7890 to chr. X (q21.33) at position 96,375,642-96,375,600 to (share T)	7863	7890	LCR	X (q21.33)	96,375,642	96,375,600
UD-SCC-2	11	64	TTAACAAAAATACATTTTTTTTGTTCTTG ATGGTTGTAGTCAGTACAGTAGTGGAAG TGGGGGAGAGGGT	Chr. X(q21.33) at position 96,380,863-96,380,894 to HPV16 nt 1361-1400 (share TG)	1361	1400	E1	X (q21.33)	96,380,863	96,380,894

UM-SCC-47	3	309	AGTAATTTAATCTATCTGAGTAATACAAA CATATTTTGTGCAGT	Chr. 3 (q28) at position 189,596,834-189,596,814 to Δ 6bp (AATACA) insertion to HPV16 nt 3249-3265	3249	3265	E2	3 (q28)	189,596,834	189,596,814
UM-SCC-47	3x	N/A	CTTGATACTGCATCCACAACATTACTGG CGTGCTTTTTGCTTTGCTTGATCTACCTT ATCTCAGTGGTCCTCGACTGAA	HPV16 nt 3858-3901 to chr. 3 (q28) at position 189,597,479 - 189,597,513	3858	3901	E5	3 (q28)	189,597,479	189,597,513
UM-SCC-47	6	6	AGGGTAGGTCAGAAACCACAGATGAAG AAAGAAGTGCTGCCATATCTACTCAGAA CCTAC	Chr. 3 (q28) at position 189,601,595-189,601,562 to HPV nt 6671-6698 (share G)	6671	6698	L1	3 (q28)	189,601,595	189,601,562
UM-SCC-47	2	29	TGTTGATACTACACGCAGTACAAATATG TCAacagacactttcccttgtttccatgc	HPV16 nt 6636-6666 to chr. 3 (q28) at position 189,607,490-189,607,462 (share TCA)	6636	6666	L1	3 (q28)	189,607,490	189,607,462
UM-SCC-47	4	705	GAGTGACGATGATGTACAGATTCTTCAG GACATAATGTAAAACATAAAAGTG	Chr. 3 (q28) at position 189,612,825-189,612,849 to HPV16 nt 3719-3745	3719	3745	E2	3 (q28)	189,612,825	189,612,849
UM-SCC-47	5	9	aaagaagaaccaggtgatccagacttcccTGTCTA CATACACATCATTAATA	Chr. 3(q28) at position 189,620,958-189,620,988 to HPV16 nt 3955-3977 (share T)	3955	3977	E5	3 (q28)	189,620,958	189,620,988
UPCI:SCC090	21	17	TTTCCACCATGAAAATCCTTCTTTCAGC ACTGGCCACAATAATGGCATTTGTTGGG GTAACCAACTATTT	Chr. 2 (q23.3) at position 151,992,545-151,992,576 to HPV16 nt 6589-6627	6589	6627	L1	2 (q23.3)	151,992,545	151,992,576
UPCI:SCC090	22	14	GTTGATACTACACGCAGTACAATACTA TGTCGTTAGGCAGAAGAAATTCA	HPV16 nt 6636-6660 to chr. 3 (p12.2) at position 81,834,929-81,834,959 (share AATA)	6636	6660	L1	3 (p12.2)	81,834,929	81,834,959
UPCI:SCC090	22y	18	TCTCTTCAATAAGTCATGTTGGGAAAAA GGGCATCCACAGGCAGTGGATGTTACC ATTCCATTGTTTTTTACACTGCACTATGT GCAACTACTGAATCACTATGTACATTGT GTCAtacattgttcatttgtaaaactgcacatgggtgtgtg caaaccgttttaggttacacatttacaagcaacttatataat aatacta	Chr. 3 (p12.2) at position 82,041,238-82,041,264 to Δ 4bp (AGGG) insertion to chr. 3 (p12.2) at position 82,041,270-82,041,282 to Δ 8bp (TGGATGTT) insrtion to HPV16 nt 7591-7656/7814-7904 (share GTCAT)	7591	7656	LCR	3 (p12.2)	82,041,238	82,041,264
					7814	7904	LCR		82,041,270	82,041,282
UPCI:SCC090	24	7	GAAAACTGGACATCCACAGGCAGTGGA TGTTACCATTCCATTGTTTTTTACACTGC ACTATGTGCAACTACTGAATCACTATGT ACATTGTGTCATAcattgttcatttgtaaaa	Chr. 3 (p12.2) at position 82,041,260-82,041,282 to Δ8bp (TGGATGTT) insertion to HPV16 nt 7591-7656/7814-7837 (share GTCAT)	7591	7656	LCR	3 (p12.2)	82,041,260	82,041,282
					7814	7837	LCR			

UPCI:SCC090	23	31	GCAGAAACAGAGACAGCACATGCGTCT CTGTTGCTTCCAAGTTTTGGCAAT	HPV16 nt 1048-1072 to 3bp insertion (CTC) to chr. 3 (p12.2) at position 82,275,717-82,275,739	1048	1072	E1	3 (p12.2)	82,275,717	82,275,739
UPCI:SCC090	26	307	GGGTGGTGGCTGATGCTTGTAATCCCA GCTAAACTTATTACATATGATAATCCTG	Chr. 6 (p21.2) at position 36,878,615-36,878,572 to HPV16 nt 4970-5004 (share TCCCA, insert Δ1bp G)	4970	5004	L2	6 (p21.2)	36,878,615	36,878,572
UPCI:SCC090	34	40	GTGGCTGATGCTTGTAATCCCAGCTAAA CTTATTACATATGATAATCCTGCATATGA AGGTATAGATGTGGATAATACATTATATT TTCCTAGTAATGATAATAGTATTAATATA GCTCCAGATCCTGACTTTTTGGATATAG TTGCTTTACATAGGCAGGTAGCCtctgcag gtgttggaggctgcaatacagatgggtcagtgaaagtgg gattattatgtgtaa	Chr. 6 (p21.2) at position 36,878,593-36,878,572 to HPV16 nt HPV16 nt 4970-5109 (share TCCCA, insert Δ1bp G) to Δ7bp (AGGTAGCC) insertion to HPV16 nt 4762-4703	4970 4762	5109 4703	L2 L2	6(p21.2)	36,878,593	36,878,572
UPCI:SCC090	29	20	AAACACAGCACTAATCTAAAAACCAACT TGACATTTATTTAATAGGGCTGGTGCTG TTGG	Chr. 6 (p21.2) at position 36,904,431-36,904,461 to HPV16 nt 6410-6440 (share GA)	6410	6440	L1	6 (p21.2)	36,904,431	36,904,461
UPCI:SCC090	27	163	CGCACAACACAACAAGTTAAAGTTGTAG ACAAAGTAATACATTGAAGGCC	HPV16 nt 4923-4952 to chr. 6 (p21.2) at position 37,129,026 – 37,129,048 (share the GAC)	4923	4952	L2	6 (p21.2)	37,129,026	37,129,048
UPCI:SCC090	28	154	GCATATGAAGGTATAGATGTGGATAATA GATGCTGTTATTATCTCCATTCTGT	HPV16 nt 5001-5031 to chr. 6 (p21.2) position 37,135,218-37,135,244 (share TA)	5001	5028	L2	6 (p21.2)	37,135,218	37,135,244
UPCI:SCC090	31	23	TGGCCACCTCATATAAAACATCATTGCA GAAAGTAATGCAAAGGCAGCAA	Chr. 6 (p21.2) at position 37,143,517 – 37,143,550 to HPV16 nt 1464-1482 (share AGT)	1464	1482	E1	6 (p21.2)	37,143,517	37,143,550
UPCI:SCC090	30	19	GGATTACAGGCATGTGCCACCAGGCCT GTAACAGGCTAACAAAAGTGAAG	Chr. 6 (p22.3) at position 37,144,190-37,144,215 to 4bp insertion (TGTA) to HPV16 nt 6276-6295	6276	6295	L1	6 p(22.3)	37,144,190	37,144,215
UPCI:SCC090	32	30	CTCTGCCTCCCAGGTTCGAACAATTCTC CTATTACAAACTATAACACCTTCTAC	Chr. 6 (p22.1) at position 37,164,521 – 37,164,552 to HPV16 nt 5255-5278 (share AT)	5255	5278	L2	6 (p22.1)	37,164,521	37,164,552
UPCI:SCC090	49	58	TAAAAACCACAATGAGATATCATTTCAC ATTTTTAATTTTACCTTTTCCAGCTGGAC AAG	Chr. 9 (q22.33) at position 100,575,645-100,575,617 to Δ16bp (ttttaattttaccttt) insertion to HPV16 nt 681-684	681	694	E7	9 (q22.33)	100,575,645	100,575,617
UPCI:SCC090	36	2	ACTCAGCAGATGTTACAGGTAGAAGCCT CCAAATCAGGCTTTGGTTGAATG	HPV16 nt 1284-1308 to chr. 9 (q22.33) at position 100,596,047-100,596,022	1284	1308	E1	9 (q22.33)	100,596,047	100,596,022

UPCI:SCC090	46	16	GGCCTCCCACCGGCCGACGACACCAG GAATTGTGTGCCCCATCTGTTCTC	Chr. 9 (q22.33) at position 100,620,041-100,620,065 (share A)	822	847	E7	9 (q22.33)	100,620,041	100,620,065
UPCI:SCC090	43	3	AAGACAAACAGTATTACAACATAGTTGT AATTCTGCCGAATTACAACTTTACACTG CAATTGCACTGCAATTGTTGATGAGCTT TG	HPV16 nt 1886-1916 to Δ16bp (TCTGCCGAATTACAAC) to HPV16 nt 7606-7616 to Δ 2bp (AT) to chr. 9 (q22.33) at position 100,638,320 – 100,638,345	1886 7606	1916 7616	E1 LCR	9 (q22.33)	100,638,320	100,638,345
UPCI:SCC090	44	2	TTTCCTAATGAGTTTCCATTTGACGAAAA CCCAGCCCATAGTCATTTATTATCAAGT ATT	HPV16 2625-2654 to chr. 9 (q22.33) at position 100,647,590 – 100,647,621 (share AC)	2625	2654	E1	9 (q22.33)	100,647,590	100,647,621
UPCI:SCC090	50	55	CAGGCAGAGGAGCTGTGGAAGTTAAAG GAGGAGGATGAAATAGATGGTCCAGCT GGACA	Chr. 9 (q22.33) at position 100,653,571-100,653,544 to HPV16 nt 659-692 (share AGG)	659	692	E7	9 (q22.33)	100,653,571	100,653,544
UPCI:SCC090	51	8	AACGCCCATTCCCACAGGATCTGCATG GAACTACATATACTACCACTTCACATGC AGCCTCACCTACTTC	Chr. 9 (q22.33) at position 100,662,726 – 100,662,696 to HPV 16 nt 5275-5314 (share C)	5275	5314	L2	9 (q22.33)	100,662,726	100,662,696
UPCI:SCC090	47	21	CTTGTTGACACAATCCCTTTTCCAAAAC TCTGGCCATGTTCATGAAGGAATACGAA CATATTTTGTGCAG	Chr. 9 (q22.33) at position 100,663,585-100,663,620 to HPV16 nt 3230-3264	3230	3264	E2	9 (q22.33)	100,663,585	100,663,620
UPCI:SCC090	37	25	ACTACTGTTACTACACATAATAATCAGA GGTTCACAGGAAATT	HPV16 nt 4692-4716 to chr. 9 (q22.33) at position 100,665,235-100,665,211 (share AATAATC)	4692	4716	L2	9 (q22.33)	100,665,235	100,665,211
UPCI:SCC090	38	15	ATAGTAGACGATAGTGAAATTGCATATA AATATGCACAATTGGCAGACACTAATAG TAATGCAAGTGCCTTTCTAAAAAGTAATT CACAGGCAAAGAGTCCAGGGCTGCGTG ATTATAATAACTGTGACAAACCCCCT	HPV16 nt 1968-2062 to Δ 18bp (GAGTCCAGGGCTGCGTGA) to chr. 9 (q22.33) at position 100,676,884-100,676,860	1968	2062	E1	9 (q22.33)	100,676,884	100,676,860
UPCI:SCC090	52	3	ATAAAAGCTGTAAAATATAATTTTTACTA TTGCTGATCAAATATTACAATATG	Chr. 9 (q22.33) at position 100,691,903-100,691,782 to Δ 4bp (TTTT) to HPV16 nt 4353-4380	4353	4380	L2	9 (q22.33)	100,691,903	100,691,782
UPCI:SCC090	53	22	TACAGGGCACAGGGCTCAGGAACTGCA CCATGCCATTAGGTGTGGGCATTAGTG GCCATC	Chr. 9 (q22.33) at position 100,702,536-100,702,506 to HPV16 nt 5970-5998	5970	5998	L1	9 (q22.33)	100,702,536	100,702,506
UPCI:SCC090	54	1089	GCAACTCTCTCCCTCAATACCTAGTTTT GTAGCTTCAACCGAATTCGGTTGCATGC	Chr. 9 (q22.33) at position 100,704,916-100,704,890 to HPV16 nt 7438-7469 (share TTT)	7438	7469	LCR	9 (q22.33)	100,704,916	100,704,890

UPCI:SCC090	40	2	TGAAAGACACACTATATGCCAAACACCACTGATTTATGAGAGAATGAGTGAAAGAAAAAG	HPV16 nt 1406-1435 to chr. 9 (q22.33) at position 100,706,106 – 100,706,076 (share T)	1406	1435	E1	9 (q22.33)	100,706,106	100,706,076
UPCI:SCC090	45	2	GCAACTCTCTCCCTCAATACCTAGTTTTGTAGCTTCAACCGAATTCGGTTGCATGCTTTT	Chr. 9 (q22.33) at position 100,704,916 – 100,704,890 to HPV16 nt 7438-7473 (share TTT)	7438	7473	LCR	9 (q22.33)	100,704,916	100704890
UPCI:SCC090	39	47	AAGACACACTATATGCCAAACACCACTTGATTTATGAGAGAATGAGTGAA	HPV16 nt 1409-1436 to chr. 9 (q22.33) at position 100,706,106 – 100,706,084	1409	1436	E1	9 (q22.33)	100,706,106	100,706,084
UPCI:SCC090	48	52	TGATGGTTTCCAGCTTCATCCACTTCCCTGCAAAGGACGTGAACTCATCCTTTTTTATGGCTGCATAGTATTCCATGGCATATGCTGTATGTGATAAATGTTTAAAGTTTTATTCTA	Chr. 9 (q22.33) at position 100,707,913-100,707,995 to HPV16 nt 279-317 (share CATAT)	279	317	E6	9 (q22.33)	100,707,913	100,707,995
UPCI:SCC090	41	15	ATTGGCAGACACTAATAGTAATGCAAGTGACCTGATGGAACTGAAAACCATGGCATGAGA	HPV16 nt 2006-2034 to chr. 9 (q22.33) at position 100,712,492 – 100,712,460	2006	2034	E1	9 (q22.33)	100,712,492	100,712,460
UPCI:SCC090	42	1194	TTATGTGTGTCTCCAATGTGTATGATGATAGAGCCTCCAAAATTGCGTAGTACAGCAGCAGCATTATATTCACATCTACTAGAAAAGGTAAAATGTGAAAAATaacatttagaaatattgatattaaa	HPV16 nt 1755-1825 to chr. 9 (q22.33) at position 100,575,623 – 100,575,598 to Δ 9bp (GTGAAAAAT) to Chr 9 (q22.3) at position 100,714,372-100,714,348	1755	1824	E1	9 (q22.33) 9 (q22.33)	100,575,623 100,714,372	100,575,598 100,714,348
HMS001	2	42	GTATTGCTGCATTTGGACTTACACCCATGACTTTTCAACAAATTGTGCAGGTG	HPV16 nt 1573-1599 to chr. 20 (q13.12) at position 45,660,419 to 45,660,393 (share A)	1573	1599	E1	20 (q13.12)	45,660,419	45,660,393
HMS001	3	68	AAAAATCAGGAATTCCCTGCATGTTCCTACATGTGTTACTATTAGTAAGAaactacataataaaatacta	Chr. 20 (q13.12) at position 45,660,938-45,660,912 to Δ 4bp (TACA) insertion to HPV nt 1685-1703/6690-6712 (share AGA)	1685 6690	1703 6712	E1 L1	20 (q13.12)	45,660,938	45,660,912
Tumor A	15	296	ATCAAGTAACATTGGCAGGAATAAGGGAAGGCAGATAAAATATTGCAATGGTCAAGCCTT	Chr. 5 (q11.2) at position 56,741,700-56,741,670 to HPV18 nt 4364-4393 (share G)	4364	4393	L2	5 (q11.2)	56,741,700	56,741,670
Tumor A	7	15	TTTCAAGTTATAAAACTGCACACCTTACAGCAGTAACTTAGAAAGTTACTTTACCTTTCT	HPV18 nt 7390-7421 to chr. 5 (q11.2) at position 56,751,290 -56,751,262	7390	7421	LCR	5 (q11.2)	56,751,290	56,751,263
Tumor A	11	7	CAGATACTAAGGTGGCCATGTTAGATGTTAGATGATGCAACGACCACGTGTTGGACCTTAGATTTTGAGAAGTCCAAA	HPV18 nt 2475-2501 to Δ 7bp (TTAGATG) to HPV18 nt 2502-2523 to chr. 5 (q11.2) at position 56,757,184-56,757,210 (share TTGGAC)	2475 2502	2501 2523	E1 E1	5 (q11.2)	56,757,184	56,757,210

Tumor A	8	3	TTATGTCCTATATTCATAGTATGAATAGC TACTCGACATTGTTTATTTTTT	HPV18 nt 6775-6803 to chr. 5 (q11.2) at position 56,771,953-56,771,928 to (share TAGC)	6775	6803	L1	5 (q11.2)	56,771,953	56,771,928
Tumor A	9	22	GTATTTTTGTAATAAAATTATGGTATCCC ACAAAATACATTACTAGTCATCAGGGCA ATG	HPV18 nt 4225-4255 to Δ 5bp (AAAAT) insertion to chr. 5 (q11.2) at position 56,830,246 to 56,830,223	4225	4255	L2	5 (q11.2)	56,830,246	56,830,223
Tumor A	13	94	ACCCACAGCTCTCTCTAGACATGCAAAC TGAAATAACCTGTGTATAT	Chr. 5 (q11.2) at position 56,834,681-56,834,712 to HPV18 nt 189-206 (share GAA)	189	206	E6	5 (q11.2)	56,834,681	56,834,712
Tumor A	12	8	GCTTGTTAAACAGCTACAGCACACCCCC TCACCATGTCCTCTGCAGCAACATGGAT GCAA	HPV18 nt 3464-3496 to chr. 5 (q11.2) at position 58,773,704-58,773,731 (share C)	3464	3496	E2/E4	5 (q11.2)	58,773,704	58,773,731
Tumor A	13x	5	GACTCTGTGTATGGAGACACATTGGAAA AACTCCACACTTCTGTGGTCTCCTCCAA CATG	HPV18 nt 351-382 to chr. 5 (q11.2) at position 58,777,461-58,777,490 (share CT)	351	382	E6	(q11.2)	58,777,461	58,777,490
Tumor A	14	72	AGAGCTAGCAGAAAAAAAGAAATAACT AACTAACACTGGGTTATACAAT	HPV18 nt 379-404 to chr. 5 (q11.2) at position 58,795,217 -58,795,246 (share AACTAA)	379	404	E6	(q11.2)	58,795,217	58,795,246
Tumor A	10	27	ACCTAGAGCTAGCAGAAAAAAAGAAAT AACTAACTAACACTGGGTTATACAATTTA TTAATAAGGTGCC	Chr. 5 (q11.2) at position 58,795,213-58,795, 246 to HPV18 nt 379-420 // 549 /2437+ (share AACTAA)	379	420	E6	(q11.2)	58,795,213	58,795,246
Tumor B	2y	2	GGGTCATATATTACATGCTTTGCTCGAT GTATGTCTTGTTGCAGATCATC	Chr. 11 (q13.4) at position 71,162,944-71,162,968 to HPV16 nt 505-531 (share TC)	505	531	E6	11 (q13.4)	71,162,944	71,162,968
Tumor B	4y	9	AGCTGTAATCATGCATGGAGATACACCT AAACTCGCCACCATTATTTTCCTT	HPV16 nt 552-580 to chr. 8 (p11.23) at position 37,346,181-37,346,204	552	580	E7	8 (p11.23)	37,346,181	37,346,204

Supplemental Table 3. **HPV breakpoints confirmed by PCR amplification and Sanger sequencing**

Red letters, HPV sequence (Alignment to NC_001526.2 for HPV16 and NC_001357.1 for HPV18); *blue /"Δ"*, Nucleotides that did not align to HPV or cellular sequence (untemplated sequence); *black,* Cellular sequence (Alignment to Hg19 human reference genome); *green,* Nucleotides shared between virus and host genomes.

NOTE: In some cases, HPV^HPV or cellular^cellular sequence rearrangements were detected close to the viral-cellular DNA junction. These rearranged sequences are represented as uppercase and lowercase letters. Viral and cellular sequence coordinates are based on alignment analysis to respective reference genomes. Some integration events contain nucleotides shared between both viral and cellular sequence. Coordinates provided reflect alignment to their respective reference genomes since we cannot assign these shared nucleotides to a particular genomic sequence.

N/A, not reliably detected in WGS data but found by PCR amplification.

A

gene	sample	chrom	position	ref	alt	function	reported ID
TP53	CAL 27	chr17	7,578,271	T	A	NON_SYNONYMOUS_CODING	COSM11066
TP53	D562	chr17	7,578,406	C	T	NON_SYNONYMOUS_CODING	COSM10648
TP53	SCC-25	chr17	7,578,221	TTC	T	FRAMESHIFT_CODING	COSM13120
SYNE1	HMS001	chr6	152,861,106	G	A	NON_SYNONYMOUS_CODING	rs141719907
PTEN	UPCI: SCC090	chr10	89,692,895	G	A	NON_SYNONYMOUS_CODING	COSM249825
PIK3CA	D562	chr3	178,952,085	A	G	NON_SYNONYMOUS_CODING	COSM94986
PIK3AP1	SCC-25	chr10	98,469,642	G	A	NON_SYNONYMOUS_CODING	--
FBXW7	Tumor B	chr4	153,303,408	T	C	NON_SYNONYMOUS_CODING	rs201852130
CDKN2A	CAL 27	chr9	21,971,153	C	A	STOP_GAINED	COSM13281
CASP8	SiHa	chr2	202,131,411	C	T	STOP_GAINED	COSM252414

B

gene	sample	chrom	start	stop	type	size
PTEN	UD-SCC-2	chr10	89,537,435	89,727,525	DEL	190 kb
CDKN2A	D562	chr9	21,970,825	21,985,209	DEL	14kb
CDKN2A	SCC-25	chr9	21,911,000	21,983,000	DEL	~72kb
NOTCH1	UM-SCC-104	chr9	139,407,761	139,408,464	DEL	778 bp

C

gene	sample	chrom 1	start1	stop1	strand 1	chrom 2	start2	stop2	strand 2
PTEN	Tumor B	chr10	89,630,271	89,630,505	-	chr13	31,760,064	31,760,376	+
FBXW7	UD-SCC-2	chr4	153,292,076	153,292,301	+	chr16	76,779,678	76,779,877	-
SYNE2	UPCI: SCC090	chr8	71,976,377	71,976,608	-	chr14	64,609,009	64,609,655	+

Supplemental Table 4. **Mutations in genes previously identified as mutated in HNSCC**. (A) List of SNPs and small indels with deleterious coding mutations in genes previously identified as mutated in HNSCC. We used 29 genes reported by two groups (Agrawal et al. 2011; Stransky et al. 2011). These genes also are listed in Supplemental Fig. 1. (B) List of large deletions affecting exons of HNSCC cancer genes. (C) List of chromosome translocations observed in introns of HNSCC cancer genes.

sample	chrom	control unamplified neighborhood		amplified region		fold change
		sampled coordinates	mean coverage	sampled coordinates	mean coverage	
SiHa	chr13	chr13:73700000-73750000	30.74	chr13:73800000-73850000	100.09	3.26
CaSki	chrX	chrX:144800000-144810000	21.86	chrX:144789952-144798190	139.22	6.37
CaSki	chr19	chr19:55830000-55840000	27.87	chr19:55820989-55821587	230.13	8.26
CaSki	chr20	chr20:26230000-26240000	71.99	chr20:26257342-26260000	277.92	3.86
UD-SCC-2	chrX	chrX:9615000000-96200000	25.11	chrX:96369878-96375636	328.79	13.10
SCC47	chr3	chr3:189540000-189560000	58.44	chr3:189597649-189620988	1327.88	22.72
UPCI: SCC090	chr3	chr3:82450000-82500000	17.56	chr3:82000000-82041000	54.60	3.11
UPCI: SCC090	chr6	chr6:37170000-37180000	40.34	chr6:37129025-37164294	504.30	12.50
UPCI: SCC090	chr9	chr9:100715000-100725000	49.44	chr9:100712492-100714372	2908.51	58.83
HMS001	chr20	chr20:45620000-45630000	77.84	chr20:45660000-45670000	120.11	1.54
Tumor A	chr5	chr5:56720000-56730000	47.94	chr5:56741669-56751290	469.65	9.80
Tumor B	chr8	chr8:37000000-37100000	25.71	chr8:37500000-37600000	57.43	2.23
Tumor B	chr11	chr11:71500000-71600000	20.19	chr11:71000000-71100000	57.38	2.84

Supplemental Table 5. **Summary of coverage depths surrounding HPV-mediated amplification regions** The table shows the mean coverage of alignment from WGS for HPV-mediated amplification region. We calculated the mean depth of coverage for the sampled coordinates in amplification regions and their neighboring unamplified control regions as displayed by IGV genome browser.

location	all breaks		non-clustered breaks		in-silico
	count	percent	count	percent	percent
5 prime	14	12.6%	5	12.5%	12.0%
3 prime	17	15.3%	3	7.5%	10.8%
exon	3	2.7%	1	2.5%	2.1%
intron	44	39.6%	12	30.0%	36.6%
intergenic	33	29.7%	19	47.5%	38.5%
total	111	100.0%	40	100.0%	100.0%

category	#overlap to fragile sites	total N	%total	p-value
all HPV breaks	17	111	15.3%	0.0205
non-clustered HPV breaks	7	40	17.5%	0.36
in-silico	248,354	1,000,000	24.8%	NA

category	#overlap to DNaseI sensitivity sites	total N	%total	p-value
all HPV breaks	30	111	27.0%	0.003
non-clustered HPV breaks	8	40	20.0%	0.515
in-silico	159761	1000000	16.0%	NA

Supplemental Table 6. **Genomic characteristics of HPV insertional breakpoints**
(*Top*) Location of HPV breakpoints relative to neighboring genes: We determined the relative location of HPV integration relative to NCBI refSeq genes. All breaks column shows numbers for all WGS-detected HPV breakpoints. Non-clustered breaks column shows numbers after reducing the number of breakpoints located in HPV breakpoint clusters. In-silico column shows numbers obtained from randomly chosen 1 million genomic locations in human genome. Using all HPV breaks, we observed modest enrichment of HPV breakpoints within 50 kb from RefSeq genes (78 out of 111, p-value=0.063 by binomial test). However, the enrichment is not significant after adjusting the over-representation of HPV breaks in clusters (21 out of 40, p-value=0.25).
(Middle) HPV breakpoints and genomic fragile sites: We determined the overlap between HPV breakpoints and reported genomic fragile sites. HPV breakpoints were more frequently observed in genomic fragile sites than expected by chance (p=0.020). However, the bias of HPV breakpoints in the genomic fragile sites was not significant after adjusting the over-representation of HPV clusters.
(*Bottom*) HPV breakpoints and DNaseI hypersensitivity sites: We compared the overlap between the HPV breakpoints and the DNaseI hypersensitivity sites reported by ENCODE project. When we included all HPV breakpoints, we observed an enrichment of HPV breakpoints in DNaseI hypersensitivity sites (p=0.003). However, the enrichment was not significant after adjusting the over-representation of breakpoints in breakpoint clusters.

Supplemental Table 7. **HPV breakpoints and their neighboring genes**

sample	breakId	#reads	chrom	start	stop	PCR	neighboring genes
SiHa	3	42	chr13	73788865	73789077	Yes	No_gene_within_50kb
SiHa	2	44	chr13	74087301	74087562	Yes	No_gene_within_50kb
CaSki	97	3	chr10	11689807	11689956	Yes	USP6NL(USP6 N-terminal like),5'=36.3kb(SE);
CaSki	98	3	chr10	11725462	11725630	Yes	No_gene_within_50kb
CaSki	99	8	chr10	11742222	11742450	Yes	ECHDC3(enoyl CoA hydratase domain containing 3),5'=41.9kb(AS);
CaSki	100	22	chr10	11742483	11742710	Yes	ECHDC3(enoyl CoA hydratase domain containing 3),5'=41.9kb(AS);
CaSki	101	11	chr11	6762046	6762248	Yes	GVINP1(GTPase, very large interferon inducible pseudogene 1),5'=19.1kb(SE); OR2AG2(olfactory receptor, family 2, subfamily AG, member 2),3'=27.0kb(SE); OR2AG1(olfactory receptor, family 2, subfamily AG, member 1),5'=44.0kb(AS);
CaSki	105	13	chr11	6762260	6762462	Yes	GVINP1(GTPase, very large interferon inducible pseudogene 1),5'=19.1kb(SE); OR2AG2(olfactory receptor, family 2, subfamily AG, member 2),3'=27.0kb(SE); OR2AG1(olfactory receptor, family 2, subfamily AG, member 1),5'=44.0kb(AS);
CaSki	106	12	chr11	32277838	32277999	Yes	No_gene_within_50kb
CaSki	104	9	chr11	100637347	100637456	Yes	ARHGAP42(Rho GTPase activating protein 42),inside=intron1(SE);
CaSki	103	24	chr11	100637502	100637735	Yes	ARHGAP42(Rho GTPase activating protein 42),inside=intron1(SE);
CaSki	102	3	chr11	100642470	100642580	Yes	ARHGAP42(Rho GTPase activating protein 42),inside=intron2(AS);
CaSki	76y	2	chr11	100677878	100677946	Yes	ARHGAP42(Rho GTPase activating protein 42),inside=intron3(SE);
CaSki	109	23	chr14	25288179	25288383	Yes	STXBP6(syntaxin binding protein 6 (amisyn)),inside=intron5(SE);
CaSki	107	13	chr14	47984372	47984564	Yes	MDGA2(MAM domain containing glycosylphosphatidylinositol anchor 2),inside=intron1(AS);
CaSki	116	6	chr19	55804412	55804586	Yes	HSPBP1(HSPA (heat shock 70kDa) binding protein, cytoplasmic cochaperone 1),5'=12.7kb(SE); PPP6R1(protein phosphatase 6, regulatory subunit 1),5'=34.4kb(SE);
CaSki	117	6	chr19	55816498	55816670	Yes	HSPBP1(HSPA (heat shock 70kDa) binding protein, cytoplasmic cochaperone 1),5'=24.7kb(SE); PPP6R1(protein phosphatase 6, regulatory subunit 1),5'=46.5kb(SE);
CaSki	112	82	chr19	55818768	55818989	Yes	HSPBP1(HSPA (heat shock 70kDa) binding protein, cytoplasmic cochaperone 1),5'=27.0kb(AS); PPP6R1(protein phosphatase 6, regulatory subunit 1),5'=48.7kb(AS);
CaSki	113	28	chr19	55821374	55821587	Yes	BRSK1(BR serine/threonine kinase 1),inside=intron18(SE); HSPBP1(HSPA (heat shock 70kDa) binding protein, cytoplasmic cochaperone 1),5'=29.8kb(AS);
CaSki	114	39	chr19	55821409	55821576	Yes	BRSK1(BR serine/threonine kinase 1),inside=intron18(SE); HSPBP1(HSPA (heat shock 70kDa) binding protein, cytoplasmic cochaperone 1),5'=29.8kb(AS);
CaSki	118	6	chr2	22638845	22639092	Yes	No_gene_within_50kb
CaSki	126	46	chr2	22639101	22639296	Yes	No_gene_within_50kb
CaSki	119	26	chr2	22645475	22645646	Yes	No_gene_within_50kb
CaSki	122	9	chr2	22661861	22661981	Yes	No_gene_within_50kb
CaSki	120	8	chr2	22701935	22702082	Yes	No_gene_within_50kb
CaSki	127	11	chr2	27318115	27318343	Yes	OST4(oligosaccharyltransferase 4 homolog (S. cerevisiae)),5'=23.5kb(SE); AGBL5(ATP/GTP binding protein-like 5),3'=24.6kb(AS);
CaSki	123	16	chr2	27358821	27359046	Yes	KHK(ketohexokinase (fructokinase)),3'=35.4kb(SE); EMILIN1(elastin microfibril interfacer 1),3'=49.8kb(SE);
CaSki	121	12	chr2	27360016	27360200	Yes	CGREF1(cell growth regulator with EF-hand domain 1),5'=26.0kb(AS); KHK(ketohexokinase

							(fructokinase)),3'=36.4kb(SE);
CaSki	124	20	chr2	33141303	33141688	No	LINC00486(long intergenic non-protein coding RNA 486),inside=intron3(SE);
CaSki	128	121	chr20	26257342	26257562	Yes	No_gene_within_50kb
CaSki	129	8	chr21	21954641	21954788	Yes	No_gene_within_50kb
CaSki	92x	4	chr3	140479835	140480005	Yes	No_gene_within_50kb
CaSki	132	5	chr5	4906889	4906967	No	No_gene_within_50kb
CaSki	96x	6	chr5	46292375	46292548	ND	No_gene_within_50kb
CaSki	133	18	chr6	45658917	45659125	Yes	No_gene_within_50kb
CaSki	134	27	chr6	45659153	45659377	Yes	No_gene_within_50kb
CaSki	140	19	chr7	6964824	6965045	Yes	No_gene_within_50kb
CaSki	135	4	chr7	7041728	7041846	No	No_gene_within_50kb
CaSki	136	3	chr7	7628820	7629001	Yes	MIOS(missing oocyte, meiosis regulator, homolog (Drosophila)),inside=intron8(SE); RPA3(replication protein A3, 14kDa),3'=47.8kb(AS);
CaSki	139	9	chr7	7679054	7679190	Yes	RPA3(replication protein A3, 14kDa),inside=intron5(AS); LOC729852(hypothetical LOC729852),5'=1.2kb(SE); MIOS(missing oocyte, meiosis regulator, homolog (Drosophila)),3'=32.1kb(SE);
CaSki	137	3	chr7	7680996	7681104	Yes	RPA3(replication protein A3, 14kDa),inside=intron4(AS); LOC729852(hypothetical LOC729852),inside=intron1(SE); MIOS(missing oocyte, meiosis regulator, homolog (Drosophila)),3'=33.9kb(SE);
CaSki	138	2	chr7	54144048	54144194	Yes	No_gene_within_50kb
CaSki	103x	2	chr7	62984333	62984468	Yes	No_gene_within_50kb
CaSki	107x	2	chrX	144770251	144770444	Yes	No_gene_within_50kb
CaSki	108x	4	chrX	144775155	144775268	Yes	No_gene_within_50kb
CaSki	142	4	chrX	144778295	144778420	Yes	No_gene_within_50kb
CaSki	143	23	chrX	144789748	144789952	Yes	No_gene_within_50kb
CaSki	145	73	chrX	144797797	144798032	Yes	No_gene_within_50kb
CaSki	144	11	chrX	144798048	144798190	Yes	No_gene_within_50kb
UM-SCC-104	2	21	chr17	19609450	19609637	Yes	SLC47A2(solute carrier family 47, member 2),inside=intron9(SE); ALDH3A2(aldehyde dehydrogenase 3 family, member A2),3'=28.7kb(AS);
UM-SCC-104	3	30	chr17	19609642	19609917	Yes	SLC47A2(solute carrier family 47, member 2),inside=intron9(SE); ALDH3A2(aldehyde dehydrogenase 3 family, member A2),3'=28.7kb(AS);
UD-SCC-2	8	14	chrX	96214932	96215121	Yes	DIAPH2(diaphanous homolog 2 (Drosophila)),inside=intron16(SE);
UD-SCC-2	9	59	chrX	96238540	96238769	Yes	DIAPH2(diaphanous homolog 2 (Drosophila)),inside=intron17(SE);
UD-SCC-2	7	288	chrX	96369878	96370121	Yes	DIAPH2(diaphanous homolog 2 (Drosophila)),inside=exon21(SE);
UD-SCC-2	10	7	chrX	96373247	96373464	Yes	DIAPH2(diaphanous homolog 2 (Drosophila)),inside=intron21(SE);
UD-SCC-2	6	48	chrX	96375411	96375636	Yes	DIAPH2(diaphanous homolog 2 (Drosophila)),inside=intron21(AS);
UD-SCC-2	7y	2	chrX	96375559	96375636	ND	DIAPH2(diaphanous homolog 2 (Drosophila)),inside=intron21(AS);
UD-SCC-2	11	64	chrX	96380685	96380894	Yes	DIAPH2(diaphanous homolog 2 (Drosophila)),inside=intron21(SE);
UM-SCC-47	3	309	chr3	189596808	189597649	Yes	TP63(tumor protein p63),inside=intron10(SE);
UM-SCC-47	6y	10	chr3	189596840	189597063	No	TP63(tumor protein p63),inside=intron10(AS);
UM-SCC-47	6	6	chr3	189601574	189601729	Yes	TP63(tumor protein p63),inside=intron10(AS);
UM-SCC-47	2	29	chr3	189607269	189607490	Yes	TP63(tumor protein p63),inside=intron12(AS);
UM-SCC-47	4	705	chr3	189612619	189612899	Yes	TP63(tumor protein p63),inside=exon13(SE);
UM-SCC-47	5	9	chr3	189620828	189620988	Yes	TP63(tumor protein p63),3'=5.9kb(SE);
UPCI:SCC090	20	9	chr2	33141303	33141692	No	LINC00486(long intergenic non-protein coding RNA 486),inside=intron3(SE);
UPCI:SCC090	21	17	chr2	151992389	151992569	Yes	No_gene_within_50kb
UPCI:SCC090	22	14	chr3	81834933	81835154	Yes	GBE1(glucan (1,4-alpha-), branching enzyme 1),5'=24.0kb(AS);

UPCI:SCC090	22y	18	chr3	82041035	82041270	Yes	No_gene_within_50kb
UPCI:SCC090	24	7	chr3	82041216	82041282	Yes	No_gene_within_50kb
UPCI:SCC090	23	31	chr3	82275716	82275924	Yes	No_gene_within_50kb
UPCI:SCC090	26	307	chr6	36878571	36878664	Yes	C6orf89(chromosome 6 open reading frame 89),inside=intron3(SE); PPIL1(peptidylprolyl isomerase (cyclophilin)-like 1),5'=35.8kb(AS); PI16(peptidase inhibitor 16),5'=43.6kb(SE);
UPCI:SCC090	34	40	chr6	36878620	36878832	Yes	C6orf89(chromosome 6 open reading frame 89),inside=intron3(AS); PPIL1(peptidylprolyl isomerase (cyclophilin)-like 1),5'=35.8kb(SE); PI16(peptidase inhibitor 16),5'=43.6kb(AS);
UPCI:SCC090	29	20	chr6	36904218	36904436	Yes	C6orf89(chromosome 6 open reading frame 89),3'=7.7kb(SE); PI16(peptidase inhibitor 16),5'=11.6kb(SE);
UPCI:SCC090	27	163	chr6	37129025	37129269	Yes	PIM1(pim-1 oncogene),5'=8.9kb(SE);
UPCI:SCC090	28	154	chr6	37135217	37135469	Yes	PIM1(pim-1 oncogene),5'=2.7kb(SE); TMEM217(transmembrane protein 217),3'=44.7kb(AS);
UPCI:SCC090	31	23	chr6	37143344	37143550	Yes	PIM1(pim-1 oncogene),3'=0.3kb(SE); TMEM217(transmembrane protein 217),3'=36.4kb(AS);
UPCI:SCC090	30	19	chr6	37144001	37144215	Yes	PIM1(pim-1 oncogene),3'=1.0kb(SE); TMEM217(transmembrane protein 217),3'=35.7kb(AS);
UPCI:SCC090	32	30	chr6	37164294	37164468	Yes	TMEM217(transmembrane protein 217),3'=15.5kb(AS); PIM1(pim-1 oncogene),3'=21.3kb(SE);
UPCI:SCC090	49	58	chr9	100575616	100575840	Yes	FOXE1(forkhead box E1 (thyroid transcription factor 2)),5'=39.9kb(AS);
UPCI:SCC090	36	2	chr9	100595823	100596047	Yes	FOXE1(forkhead box E1 (thyroid transcription factor 2)),5'=19.5kb(AS);
UPCI:SCC090	46	16	chr9	100619827	100620065	Yes	FOXE1(forkhead box E1 (thyroid transcription factor 2)),3'=1.1kb(SE); C9orf156(chromosome 9 open reading frame 156),3'=46.7kb(AS);
UPCI:SCC090	43	3	chr9	100638333	100638472	Yes	FOXE1(forkhead box E1 (thyroid transcription factor 2)),3'=19.3kb(SE); C9orf156(chromosome 9 open reading frame 156),3'=28.4kb(AS);
UPCI:SCC090	44	2	chr9	100647628	100647748	Yes	C9orf156(chromosome 9 open reading frame 156),3'=19.1kb(AS); FOXE1(forkhead box E1 (thyroid transcription factor 2)),3'=28.6kb(SE); HEMGN(hemogen),3'=41.4kb(AS);
UPCI:SCC090	50	55	chr9	100653543	100653795	Yes	C9orf156(chromosome 9 open reading frame 156),3'=13.2kb(SE); FOXE1(forkhead box E1 (thyroid transcription factor 2)),3'=34.5kb(AS); HEMGN(hemogen),3'=35.5kb(SE);
UPCI:SCC090	51	8	chr9	100662706	100662889	Yes	C9orf156(chromosome 9 open reading frame 156),3'=4.1kb(SE); HEMGN(hemogen),3'=26.4kb(SE); FOXE1(forkhead box E1 (thyroid transcription factor 2)),3'=43.7kb(AS);
UPCI:SCC090	47	21	chr9	100663410	100663620	Yes	C9orf156(chromosome 9 open reading frame 156),3'=3.2kb(AS); HEMGN(hemogen),3'=25.5kb(AS); FOXE1(forkhead box E1 (thyroid transcription factor 2)),3'=44.6kb(SE);
UPCI:SCC090	37	25	chr9	100664986	100665235	Yes	C9orf156(chromosome 9 open reading frame 156),3'=1.5kb(SE); HEMGN(hemogen),3'=23.8kb(SE); FOXE1(forkhead box E1 (thyroid transcription factor 2)),3'=46.2kb(AS);
UPCI:SCC090	38	15	chr9	100676665	100676888	Yes	C9orf156(chromosome 9 open reading frame 156),inside=intron2(SE); HEMGN(hemogen),3'=12.2kb(SE); HEMGN(hemogen),inside=intron3(SE);
UPCI:SCC090	52	3	chr9	100691921	100692037	Yes	C9orf156(chromosome 9 open reading frame 156),5'=7.1kb(SE); HEMGN(hemogen),inside=intron1(SE);
UPCI:SCC090	53	22	chr9	100702505	100702757	Yes	C9orf156(chromosome 9 open reading frame 156),5'=17.7kb(SE); HEMGN(hemogen),inside=intron1(SE);
UPCI:SCC090	54	1089	chr9	100704889	100705146	Yes	C9orf156(chromosome 9 open reading frame 156),5'=20.0kb(SE); HEMGN(hemogen),inside=intron1(SE);
UPCI:SCC090	40	2	chr9	100705415	100705523	Yes	C9orf156(chromosome 9 open reading frame 156),5'=20.7kb(SE); HEMGN(hemogen),inside=intron1(AS);
UPCI:SCC090	45	2	chr9	100705480	100705636	Yes	C9orf156(chromosome 9 open reading frame 156),5'=20.6kb(AS);

UPCI:SCC090	39	47	chr9	100705860	100706106	Yes	HEMGN(hemogen),inside=intron1(SE); C9orf156(chromosome 9 open reading frame 156),5'=21.3kb(SE);
UPCI:SCC090	48	52	chr9	100707746	100707995	Yes	HEMGN(hemogen),5'=0.9kb(AS); C9orf156(chromosome 9 open reading frame 156),5'=23.1kb(AS);
UPCI:SCC090	41	15	chr9	100712273	100712492	Yes	HEMGN(hemogen),5'=5.4kb(SE); C9orf156(chromosome 9 open reading frame 156),5'=27.6kb(SE);
UPCI:SCC090	42	1194	chr9	100714137	100714372	Yes	HEMGN(hemogen),5'=7.2kb(SE); C9orf156(chromosome 9 open reading frame 156),5'=29.5kb(SE);
HMS001	2	42	chr20	45660215	45660419	Yes	EYA2(eyes absent homolog 2 (Drosophila)),inside=intron5(AS);
HMS001	3	68	chr20	45660911	45661128	Yes	EYA2(eyes absent homolog 2 (Drosophila)),inside=intron5(AS);
Tumor A	15	296	chr5	56741669	56742016	Yes	ACTBL2(actin, beta-like 2),3'=34.2kb(SE);
Tumor A	7	15	chr5	56751010	56751290	Yes	ACTBL2(actin, beta-like 2),3'=24.6kb(SE);
Tumor A	11	7	chr5	56757183	56757426	Yes	ACTBL2(actin, beta-like 2),3'=18.7kb(AS);
Tumor A	8	3	chr5	56771733	56771953	Yes	ACTBL2(actin, beta-like 2),3'=3.9kb(SE);
Tumor A	9	22	chr5	56829992	56830246	Yes	No_gene_within_50kb
Tumor A	13	94	chr5	56834325	56834712	Yes	No_gene_within_50kb
Tumor A	12	8	chr5	58773703	58774026	Yes	PDE4D(phosphodiesterase 4D, cAMP-specific),inside=intron1(AS);
Tumor A	13x	5	chr5	58777475	58777715	Yes	PDE4D(phosphodiesterase 4D, cAMP-specific),inside=intron1(AS);
Tumor A	14	72	chr5	58794914	58795246	Yes	PDE4D(phosphodiesterase 4D, cAMP-specific),inside=intron1(AS);
Tumor A	10	27	chr5	58795095	58795246	Yes	PDE4D(phosphodiesterase 4D, cAMP-specific),inside=intron1(SE);
Tumor B	2y	2	chr11	71162651	71162875	Yes	DHCR7(7-dehydrocholesterol reductase),5'=3.4kb(AS); LOC339902(hCG1813818),5'=28.5kb(AS);
Tumor B	4y	9	chr8	37346184	37346551	Yes	No_gene_within_50kb

SUPPLEMENTAL REFERENCES

Agrawal N, Frederick MJ, Pickering CR, Bettegowda C, Chang K, Li RJ, Fakhry C, Xie TX, Zhang J, Wang J et al. 2011. Exome sequencing of head and neck squamous cell carcinoma reveals inactivating mutations in NOTCH1. *Science* 333(6046): 1154-1157.

D'Souza G, Sugar E, Ruby W, Gravitt P, Gillison M. 2005. Analysis of the effect of DNA purification method on the detection of human papillomavirus in oral rinses by PCR. *J Clin Microbiol* 43(11).

Gravitt PE, Peyton C, Wheeler C, Apple R, Higuchi R, Shah KV. 2003. Reproducibility of HPV 16 and HPV 18 viral load quantitation using TaqMan real-time PCR assays. *J Virol Methods* 112(1-2): 23-33.

Gusnanto A, Wood HM, Pawitan Y, Rabbitts P, Berri S. 2012. Correcting for cancer genome size and tumour cell content enables better estimation of copy number alterations from next-generation sequence data. *Bioinformatics* 28(1): 40-47.

Stransky N, Egloff AM, Tward AD, Kostic AD, Cibulskis K, Sivachenko A, Kryukov GV, Lawrence MS, Sougnez C, McKenna A et al. 2011. The mutational landscape of head and neck squamous cell carcinoma. *Science* 333(6046): 1157-1160.

Thorvaldsdottir H, Robinson JT, Mesirov JP. 2012. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform*. 14(2): 178-192.