Supplementary Information

Supplementary Methods

Patient clinical history

The patient was diagnosed with stage IB2 squamous cell carcinoma of the cervix in June 2007. She completed a course of whole-pelvic radiation for a total of 4500Gy, a pelvic side wall boost of 1080Gy, and chemotherapy with weekly cisplatin in August 2007. This was followed by a high-dose-rate brachytherapy for an additional 3000Gy. She was in her usual state of health until March 2010 when she presented to the emergency department with cough and hemoptysis. Imaging at this time revealed several lung nodules, which, upon biopsy, were significant for squamous cell carcinoma, consistent with metastasis. She was placed on carboplatin and weekly paclitaxel for a total of six cycles which she completed in June 2010. On surveillance imaging in October 2010 with PET/CT, interval enlargement of a lung nodule was noted, and was concerning for chemoresistant disease. After extensive counseling, the patient chose to continue with aggressive treatment and began receiving a combination of topotecan at 0.6mg/m2 on day 1, 2, and 3, cisplatin 50mg/m2 on day 1 and bevacizumab 15mg/kg on day 1. She went on to complete a total of 6 cycles in March 2011. Imaging at this time again showed progression. A course of albumin-bound paclitaxel at 100mg/m2 on days 1, 8, and 15 and Bevacizumab at 10mg/m2 was initiated. In April 2012, an opportunity for the patient to enroll in a study of her tumor's genome arose. Requiring a tissue sample, she underwent a left lobe VATS with lung wedge resection of the left upper lobe and chest wall. This lung biopsy was sent to CRL (Clinical Reference Laboratory) for DNA and RNA isolation and aliquots of isolates were sent to TGen for sequencing analysis. She then went on to complete a total of 17 cycles of nab-paclitaxel and bevacizumab. Unfortunately, a follow-up PET/CT in August 2012 continued to show enlarging right, middle and lower lobe mass within the lung, and newly FDG avid mediastinal lymphadenopathy. The patient was given the option of enrolling in a phase I trial versus treatment with chemoradiation. She ultimately decided to pursue the latter. To date, she has completed 7 cycles of single agent cisplatin at 40mg/m2 and received a total of 6000 cGy to her mediastinum. A follow up PET/CT in Dec 2012 demonstrated good response to treatment.

Long insert whole genome library preparation

1.1µg of each sample (tumor and germline DNA) was used to generate separate long insert whole genome libraries using Illumina's TruSeq DNA Sample Prep Kit as previously described[1]. In summary, genomic DNAs are fragmented to a target size of 900bp on the Covaries E210 (Duty cycle: 2%, Intensity: 6, Cycles/burst: 200, Time: 20 seconds, Temperature: 4°C). 100ng of the sample was run on a 1% TAE gel to verify fragmentation. Samples were end repaired and purified with Ampure XP beads using a 1:1 bead volume to sample volume ratio, and ligated with indexed adapters. Samples are gel size selected and purified using Bio-Rad Freeze 'n Squeeze columns and Ampure XP beads. Products are then amplified using PCR, cleaned using Ampure XP beads, and quantified using the Agilent Bioanalyzer and Qubit.

Whole exome library preparation

1µg of each sample (tumor and germline DNA) was used to generate separate exome libraries. Libraries were prepared using Illumina's TruSeq DNA Sample Prep Kit and Exome Enrichment Kit following the manufacturer's protocols. Final libraries were evaluated using the Agilent Bioanalyzer.

Whole transcriptome library preparation

Because normal adjacent sample was not available for RNA analyses, normal age and gender matched cervix RNA was purchased from Asterand to act as an RNA control. Stratagene MVP human adult lung total RNA (catalog#540019) was also purchased to serve as a secondary control as the biopsy is a lung metastasis. All RNA samples were analyzed on the Agilent Bioanalyzer RNA 6000 Nano Chip to validate RNA integrity (RIN ≥ 7.5). 10ng of total RNA was used to generate whole transcriptome libraries for RNA sequencing. Using the Nugen Ovation RNA-Seq System v2, total RNA was used to generate double stranded cDNA, which was amplified using Nugen's SPIA linear amplification process. Amplified cDNA was input into Illumina's TruSeq DNA Sample Preparation Kit for library preparation. In summary, 1µg of amplified cDNA was fragmented to a target insert size of 300bp and end repaired. Samples were then adenylated and indexed paired end adapters were ligated. Cleaned ligation products were input into PCR to enrich for libraries. PCR products were cleaned and quantified using the Agilent Bioanalyzer.

Paired-end next generation sequencing

Clusters were generated from libraries using Illumina's cBot and TruSeq Paired End Cluster Kit v3 and sequenced on Illumina's HiSeq 2000 using Illumina's SBS Kit v3.

Sequencing data analysis

Raw sequence data were converted to fastq files using Illumina's BCLConverter. Fastq files were validated to evaluate the distribution of quality scores and to ensure that quality scores do not drastically drop over each read. Validated fastq files for whole genome and exome data were aligned to the human reference genome (build 37) using BWA (Burrows-Wheeler Alignment)[2] and sorted with SAMtools to create binary sequence (bam) files. Lane level bam files were indel realigned and recalibrated using GATK. Lane level bam files were then merged as necessary and PCR duplicates were flagged for removal using Picard (http://picard.sourceforge.net), which was also used to evaluate GC metrics. SIFT (Sorting Tolerant From Intolerant)[3] and PolyPhen-2 (Polymorphism Phenotyping v2)[4] were used to predict potential effects of selected point mutations on protein function. The NCBI (National Center for Biotechnology Information) dbGaP accession number for sequencing data from this study is phs000628.v1.p1.

SNV calling was performed using Seurat (http://sourceforge.net/projects/seurat/) and GATK. Copy number and translocation detection from long insert whole genome data was performed as previously described [1]. In summary, copy number analysis was completed by determining the log2 difference of the normalized physical coverage (or clonal coverage) for both germline and tumor samples separately across a sliding 2kb window of the mean. To detect breakpoints (for translocation analysis) in long insert whole genome data, we first defined the range of insert sizes in the normal data, evaluated the tumor data using a window size that is 3X the insert size range of the normal data, and identified reads in each

window that map to a different location. A minimum of 8 reads mapping to a discordant location was required for a breakpoint to be called. To decrease false negatives, discordant locations to which at least 4 tumor reads map are also called. Each event was also manually inspected for confirmation.

RNAseq data was aligned against human reference genome (build 37) with Bowtie[5] using TopHat 1.2[6] and CuffDiff v1.3[6] was used to identify differentially expressed genes and isoforms. TopHat and CuffDiff were both run with an Ensembl version 63 GTF (gene transfer format) file to help guide transcript discovery. The insert size parameter for TopHat was estimated by aligning the first 2 million reads with BWA and running Picard (http://picard.sourceforge.net) CollectInsertSizeMetrics on this subset of aligned reads. For CuffDiff analysis, we additionally used a mask file containing ribosomal and mitochondrial transcripts that are masked during differential expression analysis. Differential analysis was performed on FPKM (Fragments Per Kilobase of transcript per Million fragments mapped) expression values calculated for gene and isoform. Tumor RNA reads were compared against (1) normal cervix RNA reads, and (2) merged reads from normal cervix and normal lung (treated as replicates in Cuffdiff). MetaCore from Thomson Reuters (v6.13 build 43450) was used for GeneGo pathway analysis of RNAseq data.

For assembly analysis of RNA reads, reads aligning to HPV18 and chromosome 6 (position 4,275,965 – 4,331,314), where HPV18 reads also mapped, were selected for assembly. The unmapped mates of these reads were also included for a paired-read assembly. We used Trinity, a de novo assembly method[7], to obtain a non-redundant set of transcripts. Trinity starts with a greedy k-mer based approach for fast and efficient transcript assembly, recovering a representative set of variants that share k-mers. In our case we used the default k-mer length (25). Next, related contigs that correspond to portions of alternatively-spliced transcripts are clustered and a de Bruijn graph for each cluster of related contigs is constructed. Lastly, paths in the de Bruijn graph are analyzed and all plausible transcript sequences are reported. To further evaluate the accuracy of the assembled transcripts, we realigned transcripts onto the human and the HPV18 genome simultaneously using BWA-SW (Smith-Waterman alignment)[2]. BWA-SW does local alignment of the input transcripts and thus can obtain alignments for the assembled contigs that span the HPV18 genome. HPV16 and HPV18 homology analysis was performed using ClustalW2 v2.1[8].

Experimental validation

To validate HPV18 integration sites and the presence of episomal HPV18, forward and reverse primers were designed upstream and downstream of four separate junctions, respectively. These junctions include: (1) E6 to LCR (episomal HPV18), (2) E2/E4 to chr6:4,328,779, (3) E2 to chr6:4,282,640, and (4) E1 to chr6:4,291,973. Primers used are as follows (numbers below correspond to previously listed numbered junctions): (1) Forward: TGTCCAGGTGCGCTACAAC, Reverse: TCCGTGCACAGATCAGGTAG; (2) Forward: GCAGGCTTCCCTGAATAAGAG, Reverse: GCACCAAGAAGTGGGTTGAC; (3) Forward: TGAACTGC-AAATGGCCCTAC, Reverse: CTATGCTCTTGGCCACCTTG; (4) Forward: TGAAATTCCGGTTGACCTTC, Reverse: CCAAGAAGTGGGTTGACAGG. PCR was performed using each primer set on cDNA that was previously generated from tumor RNA during whole transcriptome library preparation. Phusion High Fidelity PCR Master Mix with GC buffer was used to PCR 250ng of tumor cDNA per 25µL reaction. PCR conditions for

(1), (2), and (4) are: preheat (98°C), denaturation (98°C for 30 seconds), extension (98°C for 5 seconds, 62°C for 15 seconds, 72°C for 10 seconds, cycle 34X), final extension (72°C for 5 minutes), and hold at 4°C. For primer set (3), the 63°C extension step was changed to 62°C. PCR products were run on a 1.5% TAE electrophoresis gel to verify expected product sizes and Sanger sequenced at the Arizona State University DNA Sequencing lab to confirm junction sequences. Quantitative PCR (qPCR) was performed by CRL Global Services to evaluate the gene expression level of *PIK3CA*. Relative expression was measured using beta actin as a control gene.

Acknowledgments

We would like to thank the patient and her family for contributing to this study, Steve Mastrian for staff support, and Waibhav Tembe and Raghu Metpally for assistance with allele frequency counting analysis.

Supplementary Table S1. Allele frequency analysis of mutations identified through exome sequencing

Following identification of mutations through exome sequencing, read counts of supporting exome reads and supporting RNA reads were compiled to evaluate the ratio of DNA or RNA reads (out of all reads) that are found to support the event.

gene	chromosome	hg19 position	reference base	mutatnt/alternate base	mutation type	effect	codon	change	amino acid change	quality score	Total mutant allele count	Total read count	%DNA mutant allele	Total mutant allele count	Total read count	%RNA mutant allele
AIM1	chr6	106968540	G	C	SNV	NON_SYNONYMOUS_CODING	1723	MISSENSE	D745H	40.7	31	0	100.0	200	0	100.0
ANKAR	chr2	190556979	A	T	SNV	SPLICE_SITE_ACCEPTOR				38	16	69	18.8	0	0	n/a
ANKMY2	chr7	16664646	G	A	SNV	NON_SYNONYMOUS_CODING	441	MISSENSE	S111F	35.2	204	0	100.0	1	0	100.0
ANO2	chr12	5963249	C	G	SNV	NON_SYNONYMOUS_CODING	1003	MISSENSE	R198T	44.8	281	0	100.0	0	1	0
APOBEC3D	chr22	39418976	G	A	SNV	NON_SYNONYMOUS_CODING	386	MISSENSE	R56Q	40.6	190	0	100.0	0	0	n/a
ARHGAP26	chr5	142273842	G	A	SNV	NON_SYNONYMOUS_CODING	814	MISSENSE	E176K	37.6	66	0	100.0	13	0	100.0
ARID4B	chr1	235419035	C	G	SNV	NON_SYNONYMOUS_CODING	1312	MISSENSE	D72H	37	17	0	100.0	18	47	27.7
ASH1L	chr1	155449239	C	T	SNV	NON_SYNONYMOUS_CODING	2969	MISSENSE	R1141K	40.8	28	0	100.0	52	114	31.3
ASH1L	chr1	155449208	C	G	SNV	NON_SYNONYMOUS_CODING	2969	MISSENSE	M1151I	40.4	28	0	100.0	52	114	31.3
ASPSCR1	chr17	79954437	G	T	SNV	NON_SYNONYMOUS_CODING	647	MISSENSE	E216D	33.6	9	0	100.0	1	0	100.0
ATP9A	chr20	50241813	A	G	SNV	NON_SYNONYMOUS_CODING	1047	MISSENSE	L645P	37.7	13	84	13.4	2	0	100.0
BMP5	chr6	55684533	G	C	SNV	STOP_GAINED	454	NONSENSE	Y201*	39.8	21	0	100.0	0	0	n/a
C16orf62	chr16	19639965	C	G	SNV	NON_SYNONYMOUS_CODING	1052	MISSENSE	L553V	42.8	22	0	100.0	4	30	11.8
C1orf173	chr1	75038334	C	A	SNV	NON_SYNONYMOUS_CODING	1530	MISSENSE	K1020N	41.9	262	0	100.0	0	0	n/a
C20orf20	chr20	61429993	G	A	SNV	NON_SYNONYMOUS_CODING	204	MISSENSE	E109K	43.2	409	0	100.0	3	0	100.0
C3orf67	chr3	58835087	C	A	SNV	STOP_GAINED	689	NONSENSE	E423*	35.9	230	0	100.0	0	0	n/a
C5orf42	chr5	37165761	T	A	SNV	NON_SYNONYMOUS_CODING	3197	MISSENSE	R2471S	40.3	248	0	100.0	7	0	100.0
CACNA1G	chr17	48704002	G	T	SNV	NON_SYNONYMOUS_CODING	2377	MISSENSE	D2342Y	41.2	17	0	100.0	0	0	n/a
CALHM2	chr10	105206991	C	A	SNV	NON_SYNONYMOUS_CODING	323	MISSENSE	S297I	39.2	106	0	100.0	0	6	0
CD34	chr1	208062863	G	T	SNV	NON_SYNONYMOUS_CODING	385	MISSENSE	A234D	39	28	0	100.0	0	0	n/a
CDYL	chr6	4892518	G	A	SNV	NON_SYNONYMOUS_CODING	598	MISSENSE	R253K	38.2	80	0	100.0	25	0	100.0
CENPJ	chr13	25478115	G	T	SNV	NON_SYNONYMOUS_CODING	1338	MISSENSE	A925E	43.2	48	0	100.0	12	0	100.0
CENPJ	chr13	25480431	T	C	SNV	NON_SYNONYMOUS_CODING	1338	MISSENSE	E582G	39.9	48	0	100.0	12	0	100.0

COL7A1	chr3	48611955	C	CA	insertion	FRAME_SHIFT	2944			39	15	0	100.0	0	1	0
CORIN	chr4	47655664	A	C	SNV	NON_SYNONYMOUS_CODING	1042	MISSENSE	S583R	31	10	114	8.1	0	0	n/a
CPAMD8	chr19	17015273	T	TA	insertion	SPLICE_SITE_DONOR				41.2	15	0	100.0	0	0	n/a
CPS1	chr2	211539677	G	T	SNV	NON_SYNONYMOUS_CODING	1508	MISSENSE	G1393C	41	41	0	100.0	0	0	n/a
CRAMP1L	chr16	1709887	G	T	SNV	NON_SYNONYMOUS_CODING	1269	MISSENSE	A746S	36.2	32	0	100.0	0	0	n/a
CST1	chr20	23731367	C	T	SNV	NON_SYNONYMOUS_CODING	141	MISSENSE	R46H	39.4	30	0	100.0	0	112	0
CYP2A13	chr19	41594470	A	G	SNV	NON_SYNONYMOUS_CODING	494	MISSENSE	K32E	39.1	15	160	8.6	0	0	n/a
DARC	chr1	159175468	T	C	SNV	NON_SYNONYMOUS_CODING	338	MISSENSE	L82P	42.5	29	0	100.0	2	0	100.0
DHX34	chr19	47856765	G	T	SNV	NON_SYNONYMOUS_CODING	1143	MISSENSE	A160S	39.9	17	0	100.0	1	0	100.0
DNAH14	chr1	225155250	G	C	SNV	NON_SYNONYMOUS_CODING	4515	MISSENSE	E206Q	42.3	48	0	100.0	4	0	100.0
DUSP11	chr2	74007040	G	C	SNV	NON_SYNONYMOUS_CODING	377	MISSENSE	S68C	41.5	0	0	n/a	99	0	100.0
DUSP11	chr2	74007041	A	C	SNV	NON_SYNONYMOUS_CODING	377	MISSENSE	S68A	41.1	0	0	n/a	99	0	100.0
EZH1	chr17	40864334	C	A	SNV	NON_SYNONYMOUS_CODING	750	MISSENSE	R461S	42	153	0	100.0	3	7	30.0
FAM75D1	chr9	84605947	C	T	SNV	NON_SYNONYMOUS_CODING	1576	MISSENSE	P188S	45	72	0	100.0	0	0	n/a
FMN1	chr15	33256414	T	C	SNV	NON_SYNONYMOUS_CODING	1419	MISSENSE	D1011G	40.4	89	0	100.0	2	0	100.0
FNBP4	chr11	47745674	C	G	SNV	NON_SYNONYMOUS_CODING	1017	MISSENSE	K790N	40	29	0	100.0	12	28	30.0
FNBP4	chr11	47744688	C	G	SNV	NON_SYNONYMOUS_CODING	1017	MISSENSE	G882A	39.2	29	0	100.0	12	28	30.0
FSCB	chr14	44976156	T	A	SNV	NON_SYNONYMOUS_CODING	825	MISSENSE	E12V	41.9	255	0	100.0	0	0	n/a
FUT11	chr10	75533520	C	A	SNV	STOP_GAINED	492	NONSENSE	C427*	41.5	184	0	100.0	4	3	57.1
GAB3	chrX	153940817	C	A	SNV	NON_SYNONYMOUS_CODING	587	MISSENSE	R252S	34.5	156	0	100.0	0	5	0
GIMAP6	chr7	150325106	G	A	SNV	NON_SYNONYMOUS_CODING	353	MISSENSE	R255W	44	357	0	100.0	0	0	n/a
GPR61	chr1	110085882	T	A	SNV	NON_SYNONYMOUS_CODING	579	MISSENSE	F208I	40.6	185	0	100.0	0	0	n/a
HEMK1	chr3	50617289	C	T	SNV	NON_SYNONYMOUS_CODING	338	MISSENSE	A262V	40.8	32	0	100.0	3	3	50.0
HIST1H1E	chr6	26157274	A	<DEL>	deletion	FRAME_SHIFT	219		-219	36.9	8	62	11.4	0	0	n/a
HMCN1	chr1	186106042	C	A	SNV	NON_SYNONYMOUS_CODING	5635	MISSENSE	L4519I	40.7	222	0	100.0	0	4	0
IDO1	chr8	39776416	G	C	SNV	NON_SYNONYMOUS_CODING	403	MISSENSE	C129S	41.5	122	0	100.0	14	0	100.0
IGDCC4	chr15	65676525	G	A	SNV	NON_SYNONYMOUS_CODING	1250	MISSENSE	P1192L	37.8	89	0	100.0	0	0	n/a
IGF1R	chr15	99467868	A	T	SNV	NON_SYNONYMOUS_CODING	1367	MISSENSE	N913Y	41.3	41	165	19.9	3	0	100.0
IL4I1	chr19	50399248	C	G	SNV	NON_SYNONYMOUS_CODING	589	MISSENSE	A48P	32.1	13	0	100.0	0	4	0
JUP	chr17	39914679	C	T	SNV	NON_SYNONYMOUS_CODING	745	MISSENSE	R582Q	40.3	26	0	100.0	191	176	52.0
KBTBD8	chr3	67058588	C	T	SNV	NON_SYNONYMOUS_CODING	601	MISSENSE	L529F	41.2	35	0	100.0	0	0	n/a
KIF16B	chr20	16360565	C	A	SNV	NON_SYNONYMOUS_CODING	1392	MISSENSE	K694N	41.3	198	0	100.0	7	8	46.7
KIFC1	chr6	33374620	G	A	SNV	NON_SYNONYMOUS_CODING	673	MISSENSE	E649K	39.1	70	0	100.0	1	0	100.0
KRTAP10-10	chr21	46057852	G	A	SNV	NON_SYNONYMOUS_CODING	251	MISSENSE	S173N	42.7	335	0	100.0	0	0	n/a

KRTAP6-1	chr21	31986092	T	G	SNV	NON_SYNONYMOUS_CODING	71	MISSENSE	R44S	34.8	59	0	100.0	0	0	n/a
KRTAP9-6	chr17	39421777	C	A	SNV	NON_SYNONYMOUS_CODING	160	MISSENSE	P50T	40.1	132	0	100.0	0	0	n/a
KRTAP9-7	chr17	39432082	C	A	SNV	NON_SYNONYMOUS_CODING	169	MISSENSE	P45T	38.6	103	0	100.0	0	0	n/a
LPHN3	chr4	62936046	G	A	SNV	NON_SYNONYMOUS_CODING	1580	MISSENSE	R1388H	40.8	143	0	100.0	0	0	n/a
LRRK2	chr12	40687410	G	A	SNV	NON_SYNONYMOUS_CODING	2527	MISSENSE	R918Q	39.1	84	0	100.0	0	0	n/a
MAP3K12	chr12	53876841	C	G	SNV	NON_SYNONYMOUS_CODING	892	MISSENSE	K582N	38.6	24	0	100.0	0	3	0
MFSD11	chr17	74765828	G	T	SNV	NON_SYNONYMOUS_CODING	449	MISSENSE	G250V	42	27	0	100.0	6	0	100.0
MMP2	chr16	55539272	G	T	SNV	NON_SYNONYMOUS_CODING	660	MISSENSE	G634V	42.5	19	0	100.0	10	0	100.0
MYH2	chr17	10432195	C	T	SNV	NON_SYNONYMOUS_CODING	1941	MISSENSE	E1186K	42.6	20	0	100.0	0	0	n/a
MYLK3	chr16	46764559	T	A	SNV	NON_SYNONYMOUS_CODING	819	MISSENSE	K505M	41.1	200	0	100.0	0	0	n/a
MYO3B	chr2	171355151	G	T	SNV	STOP_GAINED	1350	NONSENSE	E1031*	42	40	0	100.0	0	0	n/a
NCKAP5	chr2	133541269	C	A	SNV	NON_SYNONYMOUS_CODING	1909	MISSENSE	V1039F	41.3	183	0	100.0	0	2	0
NGRN	chr15	90814880	G	C	SNV	NON_SYNONYMOUS_CODING	291	MISSENSE	G246R	41.1	48	0	100.0	72	0	100.0
NLRP12	chr19	54314413	G	C	SNV	STOP_GAINED	1062	NONSENSE	S167*	41.6	21	0	100.0	0	0	n/a
NRIP2	chr12	2944112	G	T	SNV	NON_SYNONYMOUS_CODING	281	MISSENSE	P13H	38.5	14	0	100.0	0	0	n/a
OR4D5	chr11	123810621	C	G	SNV	NON_SYNONYMOUS_CODING	318	MISSENSE	Q100E	41.2	25	0	100.0	0	0	n/a
OR51B6	chr11	5373650	T	TTC	insertion	FRAME_SHIFT	312			41.7	17	0	100.0	0	0	n/a
OR52E2	chr11	5080763	G	T	SNV	NON_SYNONYMOUS_CODING	325	MISSENSE	P32H	38.9	14	0	100.0	0	0	n/a
OR5A1	chr11	59210697	G	A	SNV	NON_SYNONYMOUS_CODING	315	MISSENSE	G19E	39.5	117	0	100.0	0	0	n/a
OR5K3	chr3	98109771	G	C	SNV	NON_SYNONYMOUS_CODING	321	MISSENSE	E88Q	42.8	21	0	100.0	0	0	n/a
OSBPL7	chr17	45894195	C	G	SNV	NON_SYNONYMOUS_CODING	842	MISSENSE	S249T	39.4	13	0	100.0	0	4	0
PBXIP1	chr1	154918934	C	T	SNV	NON_SYNONYMOUS_CODING	731	MISSENSE	G406R	39.4	16	0	100.0	1	13	7.1
PCDHB6	chr5	140530293	C	T	SNV	NON_SYNONYMOUS_CODING	794	MISSENSE	P152L	44.6	85	0	100.0	0	0	n/a
PCDHGA2	chr5	140720504	G	A	SNV	NON_SYNONYMOUS_CODING	932	MISSENSE	A656T	42.9	264	0	100.0	0	0	n/a
PDZK1	chr1	145748488	C	A	SNV	NON_SYNONYMOUS_CODING	519	MISSENSE	P121T	41.1	348	0	100.0	0	0	n/a
PGAM1	chr10	99190261	G	C	SNV	NON_SYNONYMOUS_CODING	254	MISSENSE	E89Q	42.6	22	0	100.0	37	0	100.0
PGR	chr11	100998759	G	T	SNV	NON_SYNONYMOUS_CODING	933	MISSENSE	A348D	29.5	6	0	100.0	0	0	n/a
PKHD1L1	chr8	110534422	G	A	SNV	NON_SYNONYMOUS_CODING	4243	MISSENSE	M4013I	42.3	257	0	100.0	0	0	n/a
PLA2G4F	chr15	42436190	C	T	SNV	NON_SYNONYMOUS_CODING	865	MISSENSE	E710K	41	36	0	100.0	0	1	0
POLA2	chr11	65035018	A	C	SNV	NON_SYNONYMOUS_CODING	598	MISSENSE	D92A	40.9	26	107	19.5	44	0	100.0
PTOV1	chr19	50360300	C	<DEL>	deletion	FRAME_SHIFT	416			43.9	78	0	100.0	0	14	0
RAD50	chr5	131925495	G	A	SNV	NON_SYNONYMOUS_CODING	1312	MISSENSE	R473K	39.4	99	0	100.0	21	0	100.0
RALYL	chr8	85441687	A	G	SNV	NON_SYNONYMOUS_CODING	304	MISSENSE	K57R	42.4	19	238	7.4	0	0	n/a
RBM27	chr5	145640327	A	T	SNV	NON_SYNONYMOUS_CODING	1060	MISSENSE	N587Y	38.6	20	51	28.2	11	0	100.0
RBM27	chr5	145640328	A	T	SNV	NON_SYNONYMOUS_CODING	1060	MISSENSE	N587I	38.6	20	51	28.2	11	0	100.0

RPS6KA5	chr14	91360762	G	A	SNV	NON_SYNONYMOUS_CODING	802	MISSENSE	P547S	40.9	166	0	100.0	6	0	100.0
RYR3	chr15	33954961	C	T	SNV	NON_SYNONYMOUS_CODING	4870	MISSENSE	R1744W	34.9	44	0	100.0	0	0	n/a
SEPSECS	chr4	25156660	G	C	SNV	NON_SYNONYMOUS_CODING	501	MISSENSE	H221D	38.3	16	0	100.0	6	0	100.0
SERPINA5	chr14	95057110	A	T	SNV	NON_SYNONYMOUS_CODING	406	MISSENSE	K305N	38.4	20	64	23.8	0	0	n/a
SNX9	chr6	158330728	C	T	SNV	NON_SYNONYMOUS_CODING	595	MISSENSE	P246L	41.4	45	0	100.0	2	13	13.3
SPEN	chr1	16261251	C	T	SNV	NON_SYNONYMOUS_CODING	3664	MISSENSE	P2839L	39.1	22	0	100.0	57	138	29.2
SULT1C3	chr2	108863787	C	G	SNV	NON_SYNONYMOUS_CODING	304	MISSENSE	P46R	40.7	20	0	100.0	0	0	n/a
TCEB3	chr1	24078344	C	T	SNV	NON_SYNONYMOUS_CODING	798	MISSENSE	R443W	44.6	101	0	100.0	11	6	64.7
TEP1	chr14	20850166	C	T	SNV	NON_SYNONYMOUS_CODING	2627	MISSENSE	G1444R	39.8	13	0	100.0	4	11	26.7
TEX2	chr17	62232317	G	A	SNV	NON_SYNONYMOUS_CODING	1134	MISSENSE	R946W	39	89	0	100.0	15	0	100.0
TRPV2	chr17	16321012	C	G	SNV	NON_SYNONYMOUS_CODING	764	MISSENSE	F10L	40.9	20	0	100.0	0	3	0
UNC5C	chr4	96104080	G	A	SNV	STOP_GAINED	931	NONSENSE	Q807*	41.1	200	0	100.0	7	0	100.0
UPF2	chr10	11973732	C	T	SNV	STOP_GAINED	1272	NONSENSE	W1199*	40.4	21	0	100.0	3	13	18.8
USP33	chr1	78177530	C	T	SNV	NON_SYNONYMOUS_CODING	942	MISSENSE	G801R	40.1	23	0	100.0	2	4	33.3
WRN	chr8	30942680	A	T	SNV	SPLICE_SITE_ACCEPTOR				40.6	15	188	7.4	0	0	n/a
WRN	chr8	30999279	T	C	SNV	NON_SYNONYMOUS_CODING	1432	MISSENSE	L1074S	39.7	15	188	7.4	0	0	n/a
XIRP2	chr2	168103907	G	A	SNV	NON_SYNONYMOUS_CODING	3549	MISSENSE	G2002E	39.9	107	0	100.0	0	0	n/a
XRCC1	chr19	44050099	G	T	SNV	NON_SYNONYMOUS_CODING	647	MISSENSE	Q512K	38.4	11	0	100.0	16	0	100.0
ZDHHC9	chrX	128957684	G	A	SNV	NON_SYNONYMOUS_CODING	364	MISSENSE	S153F	42.3	186	0	100.0	78	0	100.0
ZFYVE1	chr14	73490915	A	T	SNV	NON_SYNONYMOUS_CODING	789	MISSENSE	L101Q	40.2	91	408	18.2	1	0	100.0
ZNF556	chr19	2877531	C	T	SNV	NON_SYNONYMOUS_CODING	456	MISSENSE	T192M	41	28	0	100.0	0	0	n/a
ZNF778	chr16	89293283	A	T	SNV	NON_SYNONYMOUS_CODING	729	MISSENSE	Q168L	43.6	26	424	5.8	1	0	100.0
AGAP1	chr2	236877171	G	C	SNV	NON_SYNONYMOUS_CODING	1069	MISSENSE	D729H	29.3	20	87	23.0	0	1	0
C10orf53	chr10	50916481	T	C	SNV	NON_SYNONYMOUS_CODING	157	MISSENSE	C98R	42.2	54	228	24.0	0	0	0
GAGE2D	chrX	49208335	A	G	SNV	NON_SYNONYMOUS_CODING	116	MISSENSE	M22V	31.4	47	1142	4.0	0	0	0
HLA-A	chr6	29912386	G	C	SNV	NON_SYNONYMOUS_CODING	371	MISSENSE	K335N	40	22	99	22.0	112	892	13.0
ITPRIP	chr10	106075338	C	A	SNV	NON_SYNONYMOUS_CODING	547	MISSENSE	A158S	37.3	14	145	10.0	1	12	8.0
MAP1A	chr15	43821113	G	A	SNV	NON_SYNONYMOUS_CODING	3041	MISSENSE	R2719Q	40.9	35	244	14.0	0	5	0.0
OPN1MW2	chrX	153490458	T	C	SNV	NON_SYNONYMOUS_CODING	364	MISSENSE	I65T	28.8	32	237	14.0	0	0	0.0
PCDHB11	chr5	140580033	G	T	SNV	NON_SYNONYMOUS_CODING	797	MISSENSE	R229M	36.4	11	145	8.0	0	1	0.0
PLXNA2	chr1	208272313	A	C	SNV	NON_SYNONYMOUS_CODING	1894	MISSENSE	C537G	35.7	43	89	48.0	0	24	0.0
PREX2	chr8	69104723	G	C	SNV	NON_SYNONYMOUS_CODING	1606	MISSENSE	A1523P	30.9	32	86	37.0	0	2	0.0
SCN4A	chr17	62029022	T	C	SNV	NON_SYNONYMOUS_CODING	1836	MISSENSE	E872G	35.6	47	113	42.0	0	0	0.0
SKI	chr1	2161016	A	C	SNV	NON_SYNONYMOUS_CODING	728	MISSENSE	T271P	30.7	24	58	41.0	5	41	12.0
TMEM11	chr17	21101941	C	G	SNV	NON_SYNONYMOUS_CODING	192	MISSENSE	C92S	27	4	99	4.0	0	26	0.0

UBC	chr12	125396952	C	A	SNV	NON_SYNONYMOUS_CODING	685	MISSENSE	G456W	38.9	49	439	11.0	161	2068	8.0
ZNF30	chr19	35434684	T	A	SNV	NON_SYNONYMOUS_CODING	624	MISSENSE	F273I	40.3	33	203	16.0	0	6	0.0
ZNF552	chr19	58319497	G	A	SNV	NON_SYNONYMOUS_CODING	407	MISSENSE	P379S	40.9	41	255	16.0	4	34	12.0

Supplementary Table S2. Differentially expressed COSMIC genes identified through RNAseq

All differentially expressed COSMIC genes identified through at least 1 RNAseq analysis are listed. Two types of analyses were performed: (1) tumor versus normal cervix, and (2) tumor versus normal cervix and normal lung.

	Normal cervix control		Normal cervix and normal lung control	
Gene	**log2 fold change**	**Corrected P-value**	**log2 fold change**	**Corrected P-value**
ACSL3	1.28557	2.29E-02	-	-
ARNT	-1.37262	9.01E-03	-1.33323	1.98E-02
ATM	-	-	-2.37699	3.36E-03
BCL3	3.14247	1.70E-06	-	-
BLM	4.59462	5.25E-06	3.78613	1.20E-03
BRCA1	1.92904	7.33E-03	-	-
BRIP1	4.04964	1.71E-03	-	-
BUB1B	5.41674	3.94E-06	4.55721	3.65E-04
C16orf75	3.38104	5.01E-03	-	-
CARD11	-2.78111	2.98E-02	-	-
CBLC	3.41076	6.12E-03	-	-
CCND2	-2.86513	5.40E-03	-3.45665	3.45E-03
CD274	-	-	-3.33033	2.25E-04
CDH1	2.10353	4.60E-06	-	-
CHEK2	1.56057	2.82E-02	-	-
COL1A1	1.89305	1.33E-02	2.19945	2.69E-02
COX6C	-	-	-2.15385	4.97E-02
CREB3L1	-	-	-2.60398	2.14E-02
CREB3L2	-1.76013	9.43E-03	-2.44424	2.24E-03
DDIT3	-1.87368	3.04E-02	-	-
DDX6	-	-	-1.32943	3.09E-02
DEK	1.6274	1.39E-02	1.9604	2.30E-02
EBF1	-2.44396	4.94E-04	-	-
ECT2L	-5.0741	3.86E-06	-4.9536	1.54E-03
EGFR	1.17464	3.06E-02	-	-
ELN	-2.69268	4.13E-03	-3.23491	4.30E-03
EPS15	-	-	-1.57701	2.26E-02
ERG	-2.1528	2.73E-03	-	-
EZH2	2.69625	6.01E-08	-	-
FAM46C	-4.13499	7.07E-04	-4.60063	4.00E-03
FANCA	3.29746	1.44E-04	-	-
FANCD2	3.27265	1.82E-08	2.87824	1.26E-02
FBXO11	-0.958008	3.90E-02	-2.3396	2.74E-03
FGFR1	-1.5683	1.21E-04	-	-
FGFR3	2.92142	9.53E-04	-	-
FHIT	-4.6421	4.28E-06	-4.41071	3.03E-03
GATA2	-3.70734	3.67E-06	-	-
GATA3	2.33746	2.05E-02	2.7358	4.22E-02
HERPUD1	-2.47673	1.39E-02	-3.57125	5.69E-04
HEY1	-	-	-2.84168	1.38E-02
HIST1H4I	3.26703	4.95E-02	-	-
HLF	-6.54802	1.25E-12	-6.38574	1.09E-08

HMGA1	1.68257	1.54E-02	-	-
HMGA2	-1.97415	1.65E-02	-	-
HOXA13	-	-	-2.82653	2.51E-02
HOXD11	3.16112	2.95E-02	-	-
HRAS	1.73148	2.06E-02	-	-
IDH1	-	-	-1.84508	4.87E-02
IDH2	2.37824	1.66E-03	2.45094	4.00E-03
IL6ST	-	-	-2.2078	7.80E-05
IRF4	-2.50508	3.02E-02	-	-
JAZF1	-2.62242	4.21E-02	-3.5536	1.96E-02
KDR	-2.79933	2.49E-03	-3.59782	1.56E-03
KIT	-3.76571	1.32E-05	-3.70096	8.84E-04
KTN1	1.30076	2.75E-03	-	-
LIFR	-2.52004	1.44E-06	-	-
LMO2	-3.84801	2.18E-06	-4.26898	4.92E-04
MDM4	-1.82362	7.62E-06	-1.88634	5.00E-04
MET	1.62344	4.32E-04	-	-
MITF	-5.63339	0.00E+00	-5.40875	3.34E-12
MLLT10	-	-	-1.59951	4.43E-02
MYB	1.92429	3.22E-05	-	-
MYC	-	-	1.42483	3.56E-02
MYCN	-2.71175	4.25E-02	-	-
MYH11	-4.06939	1.60E-05	-3.96242	3.91E-04
MYH9	1.30656	4.95E-04	-	-
NACA	-0.995925	2.23E-02	-	-
NCOA1	-1.53963	9.15E-03	-1.5856	3.21E-02
NDRG1	1.92879	4.56E-07	2.04017	9.98E-08
NFKB2	1.46101	4.04E-02	-	-
NIN	0.961703	3.10E-02	-	-
NR4A3	-4.0802	3.48E-05	-3.9297	1.70E-02
NSD1	-1.38569	1.40E-02	-1.74835	3.22E-02
PAX8	-3.29217	8.70E-09	-3.01918	1.12E-03
PDE4DIP	-0.746779	3.71E-02	-	-
PDGFRA	-1.96051	1.79E-03	-	-
PDGFRB	-1.98957	1.73E-03	-1.98147	1.48E-02
PER1	-3.00964	7.79E-03	-	-
PIK3R1	-1.61945	1.89E-02	-	-
PLAG1	-1.84164	1.09E-02	-	-
POU2AF1	-2.22812	1.62E-02	-	-
POU5F1	-3.32112	1.16E-05	-3.04518	1.00E-02
PPARG	-2.15037	6.02E-04	-2.56852	1.00E-03
PRKAR1A	-	-	-2.1682	1.37E-03
RARA	-1.55825	4.69E-02	-	-
RECQL4	2.68667	3.86E-03	3.13109	1.38E-02
ROS1	-	-	-6.91617	1.88E-13
RUNX1	0.957039	3.68E-02	-	-
SLC34A2	-11.0341	0.00E+00	-12.1246	0.00E+00
SMO	-3.09773	3.44E-02	-	-
SOCS1	-3.86699	1.62E-02	-	-
SOX2	4.7022	1.00E-04	4.39721	1.22E-02
TAL1	-6.33292	2.24E-07	-6.64583	2.69E-05

TMPRSS2	-	-	-3.87307	3.02E-03
TP53	1.44729	6.70E-03	-	-
TPM3	1.26275	1.25E-03	-	-
WT1	-3.18541	3.70E-03	-	-
ZNF331	-2.42427	2.24E-10	-	-

Supplementary Table S3. Expression changes identified in COSMIC genes overlapping with regions of copy number change

Gene	**Tumor vs normal cervix log2 fold** log2 fold	**Tumor vs normal cervix** corrected P-value	**Tumor vs normal cervix & normal lung log2 fold** log2 fold	**Tumor vs normal cervix & normal lung** corrected P-value
FANCD2	3.27	1.82E-08	2.88	1.26E-02
PPARG	-2.15	6.02E-04	-2.57	1.00E-03
SOX2	4.70	1.00E-04	4.40	1.22E-02
LIFR	-2.52	1.44E-06	n/a	n/a

Supplementary Table S4. QPCR validation of *PIK3CA* overexpression

Quantitative PCR was performed to validate overexpression of *PIK3CA* using beta-actin (*ACTB*) as a control gene. Results from this analysis are shown below.

Gene	Ct	Average Ct	Delta Ct
PIK3CA	30.8 30.7	30.7	8.7
ACTB	22.3 21.8	22	

Supplementary Table S5. Pathway analysis of RNAseq data

Differentially expressed genes (corrected P<0.05) identified from each of two RNAseq analyses were analyzed to identify the molecular pathways that are most likely to be affected. The top 10 pathways for each of the 2 analyses are listed.

Comparison	Rank	Pathway	P-value
tumor vs cervix	1	Cell cycle_Chromosome condensation in prometaphase	1.427E-11
tumor vs cervix	2	Cytoskeleton remodeling_Keratin filaments	2.542E-11
tumor vs cervix	3	Hedgehog signaling in gastric cancer	1.626E-10
tumor vs cervix	4	Cell cycle_Role of APC in cell cycle regulation	1.564E-09
tumor vs cervix	5	Cell cycle_Spindle assembly and chromosome separation	3.064E-09
tumor vs cervix	6	Development_Regulation of epithelial-to-mesenchymal transition (EMT)	3.773E-09
tumor vs cervix	7	Role of DNA methylation in progression of multiple myeloma	6.024E-09
tumor vs cervix	8	Stromal-epithelial interaction in Prostate Cancer	8.977E-09
tumor vs cervix	9	Cell cycle_The metaphase checkpoint	1.893E-08
tumor vs cervix	10	DNA damage_ATM / ATR regulation of G2 / M checkpoint	2.336E-08
tumor vs cervix & lung	1	Cytoskeleton remodeling_Keratin filaments	5.915E-08
tumor vs cervix & lung	2	Hedgehog signaling in gastric cancer	1.614E-07
tumor vs cervix & lung	3	Role of DNA methylation in progression of multiple myeloma	2.957E-07
tumor vs cervix & lung	4	Stromal-epithelial interaction in Prostate Cancer	6.516E-07
tumor vs cervix & lung	5	Cell cycle_Role of APC in cell cycle regulation	6.526E-07
tumor vs cervix & lung	6	Cell cycle_Spindle assembly and chromosome separation	9.954E-07
tumor vs cervix & lung	7	Cell adhesion_ECM remodeling	2.719E-06
tumor vs cervix & lung	8	Development_Regulation of epithelial-to-mesenchymal transition (EMT)	2.792E-06
tumor vs cervix & lung	9	Main genetic and epigenetic alterations in lung cancer	4.491E-06
tumor vs cervix & lung	10	Immune response_Alternative complement pathway	8.809E-06

Supplementary Table S6. Alignment of WGS reads against HPV viral genomes

All tumor WGS reads were aligned against all 42 HPV viral genomes listed in NCBI's Entrez Genome Viral Genome database. Reads aligning to HPV genomes only aligned to HPV16 and HPV18 (bold).

Viral reference	Sequence length (bp)	Number of mapped reads
gi_9626041_ref_NC_001354_1_Human_papillomavirus_type_41_complete_genome	7614	0
gi_9626053_ref_NC_001355_1_Human_papillomavirus_type_6b_complete_genome	7902	0
gi_9626597_ref_NC_001457_1_Human_papillomavirus_type_4_complete_genome	7353	0
gi_9626605_ref_NC_001458_1_Human_papillomavirus_type_63_complete_genome	7348	0
gi_9627145_ref_NC_001531_1_Human_papillomavirus_5_complete_genome	7746	0
gi_9627257_ref_NC_001576_1_Human_papillomavirus_type_10_complete_genome	7919	0
gi_9627305_ref_NC_001583_1_Human_papillomavirus_type_26_complete_genome	7855	0
gi_9627327_ref_NC_001586_1_Human_papillomavirus_type_32_complete_genome	7961	0
gi_9627334_ref_NC_001587_1_Human_papillomavirus_type_34_complete_genome	7723	0
gi_9627363_ref_NC_001591_1_Human_papillomavirus_type_49_complete_genome	7560	0
gi_9627377_ref_NC_001593_1_Human_papillomavirus_type_53_complete_genome	7856	0
gi_9627389_ref_NC_001595_1_Human_papillomavirus_type_7_complete_genome	8027	0
gi_9627396_ref_NC_001596_1_Human_papillomavirus_type_9_complete_genome	7434	0
gi_9628542_ref_NC_001690_1_Human_papillomavirus_type_48_complete_genome	7100	0
gi_9628550_ref_NC_001691_1_Human_papillomavirus_type_50_complete_genome	7184	0
gi_9628566_ref_NC_001693_1_Human_papillomavirus_type_60_complete_genome	7313	0
gi_22138122_ref_NC_004104_1_Human_papillomavirus_type_90_complete_genome	8033	0
gi_27531786_ref_NC_004500_1_Human_papillomavirus_type_92_complete_genome	7461	0
gi_50253426_ref_NC_005134_2_Human_papillomavirus_type_96_complete_genome	7438	0
gi_109390382_ref_NC_008188_1_Human_papillomavirus_type_103_complete_genome	7263	0
gi_109390389_ref_NC_008189_1_Human_papillomavirus_type_101_complete_genome	7259	0
gi_167600365_ref_NC_010329_1_Human_papillomavirus_type_88_complete_genome	7326	0
gi_224983322_ref_NC_012213_1_Human_papillomavirus_type_108_complete_genome	7149	0
gi_225927560_ref_NC_012485_1_Human_papillomavirus_type_109_complete_genome	7346	0
gi_225927568_ref_NC_012486_1_Human_papillomavirus_type_112_complete_genome	7227	0
gi_254810663_ref_NC_013035_1_Human_papillomavirus_116_complete_genome	7184	0
gi_297342356_ref_NC_014185_1_Human_papillomavirus_121_complete_genome	7342	0
gi_310698439_ref_NC_001526_2_Human_papillomavirus_type_16_complete_genome	7905	0
gi_319962660_ref_NC_014956_1_Human_papillomavirus_type_134_complete_genome	7309	0
gi_319962668_ref_NC_014955_1_Human_papillomavirus_type_132_complete_genome	7125	0
gi_319976668_ref_NC_014954_1_Human_papillomavirus_type_131_complete_genome	7182	0
gi_319976676_ref_NC_014953_1_Human_papillomavirus_type_129_complete_genome	7219	0
gi_319976684_ref_NC_014952_1_Human_papillomavirus_type_128_complete_genome	7259	0
gi_358356460_ref_NC_016157_1_Human_papillomavirus_type_126_complete_genome	7326	0
gi_389656400_ref_NC_017993_1_Human_papillomavirus_type_135_complete_genome	7293	0
gi_389656408_ref_NC_017994_1_Human_papillomavirus_type_136_complete_genome	7319	0
gi_389656416_ref_NC_017995_1_Human_papillomavirus_type_137_complete_genome	7236	0
gi_389656424_ref_NC_017996_1_Human_papillomavirus_type_140_complete_genome	7341	0
gi_389656432_ref_NC_017997_1_Human_papillomavirus_type_144_complete_genome	7271	0
gi_335334258_emb_FR872717_1_Human_papillomavirus_type_11_complete_genome	7933	0
gi_6002612_gb_AF092932_1_Human_papillomavirus_type_6_complete_genome	**8012**	**20**
gi_9626069_gb_NC_001357_1_Human_papillomavirus_18_complete_genome	**7857**	**19013**

Supplementary Table S7. RNA reads aligning to HPV18

The number of RNA reads mapping to HPV18 genes are listed.

HPV18 gene	Start position	End position	Number of mapped reads
E6	105	581	3774
E7	590	907	2922
E1	914	2887	6167
E2	2817	3914	4103
E4	3418	3684	924
E5	3936	4157	923
L2	4244	5632	294
L1	5430	7136	147

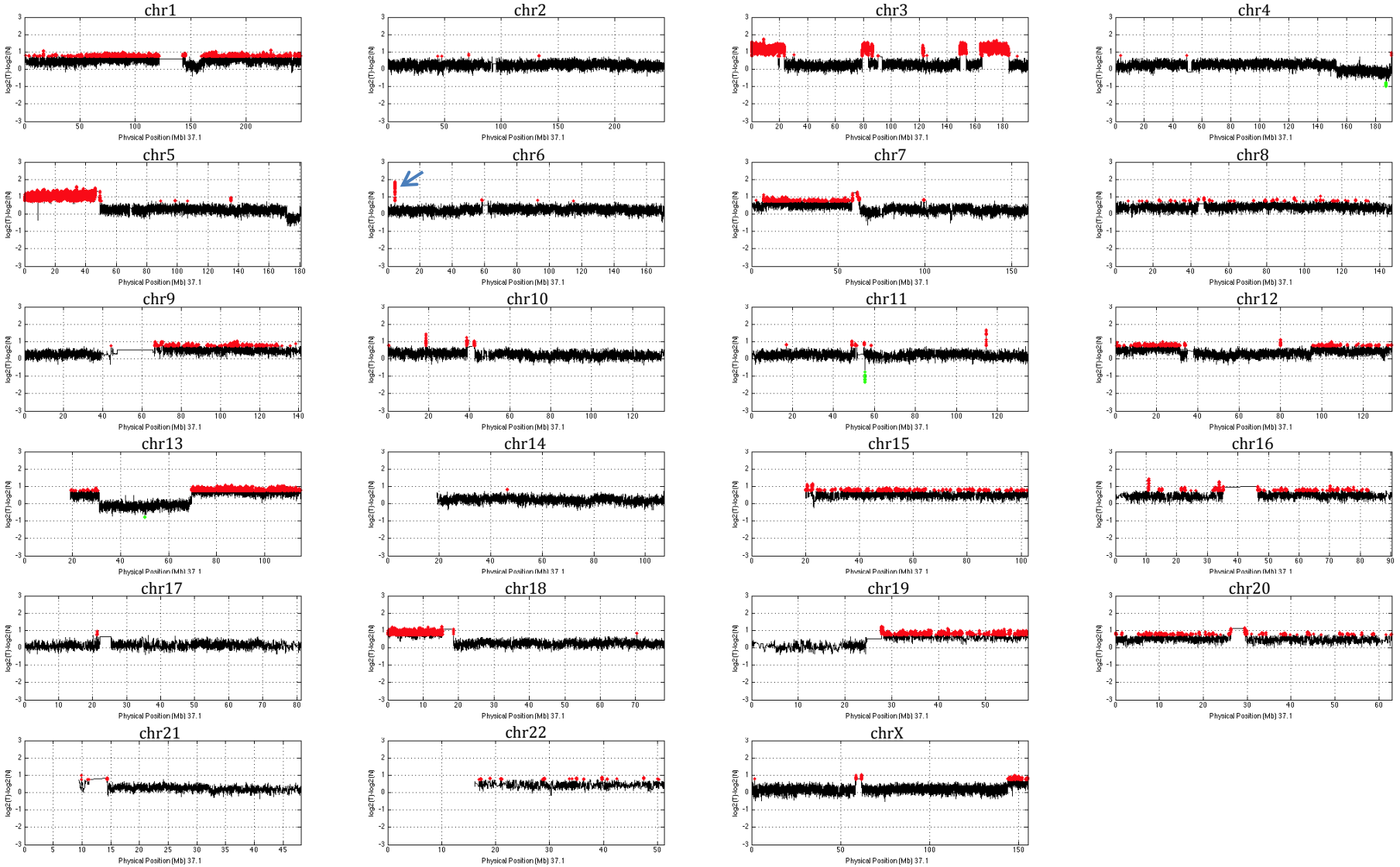
Supplementary Table S8. Expression of genes whose products are affected by HPV18 proteins

		Tumor vs normal cervix		Tumor vs normal cervix & lung	
Gene Symbol	Gene Product	Log2 fold change	Corrected P-value	Log 2 fold change	Corrected P-value
TP53	p53	1.45	6.70E-03	0.64	4.76E-01
WAF1/CDKN1A	p21	0.40	7.22E-01	0.51	7.27E-01
CDKN1B	p27	-0.85	5.06E-01	-0.83	6.24E-01
CCNE1	cyclin E1	2.75	7.99E-02	2.85	3.29E-01
CCNE2	cyclin E2	3.04	4.88E-05	3.53	3.27E-04
CCNA1	cyclin A1	-1.36	3.59E-01	-2.11	3.21E-01
CDK2	cdk2	1.77	9.02E-02	1.48	3.63E-01
RB1	Rb	-0.41	6.55E-01	-0.37	7.76E-01
RBL1	p107	1.23	2.30E-01	1.04	5.27E-01
RBL2	p130	-0.34	7.67E-01	-0.47	7.04E-01
E2F1	E2F	6.13	3.42E-05	6.00	1.70E-03
CDKN2A	p14	5.89	0[a]	6.41	1.07E-10
CDC25A	cdc25a	5.01	6.40E-03	4.09	2.70E-02
BRD4	brd4	0.46	5.79E-01	0.27	8.26E-01
E6AP/UBE3A	E6-AP	0.24	8.47E-01	0[a]	1.00E+00

[a]A corrected p-value of zero is indicative of a p-value of less than 1.0E-09.

Supplementary Figure S1. Somatic copy number analysis

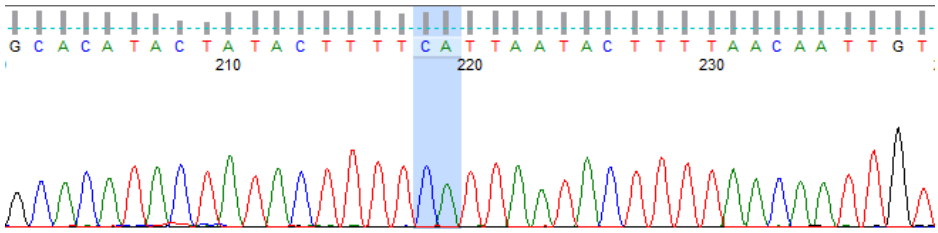
Copy number analysis was performed using WGS data and results are shown below. Each plot represents an individual chromosome and chromosome numbers are shown above each plot. The y-axis of each plot is the copy number change (log scale) between tumor and normal, while the x-axis is the physical position. Copy number gains are marked in red and losses in green. The red peak located on chromosome 6p (marked by the blue arrow) demarcates the location where HPV18 integrated into the genome and where copy gains were identified.



Supplementary Figure S2. Sanger chromatograms for validation of HPV18 insertion

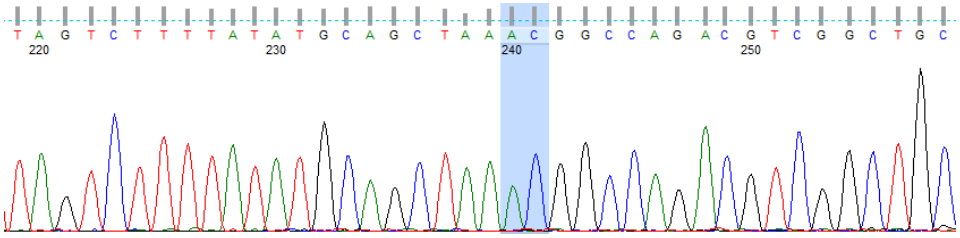
A. Validation of episomal HPV18 DNA

The junction between E6 and LCR occurs between the highlighted C and A bases.



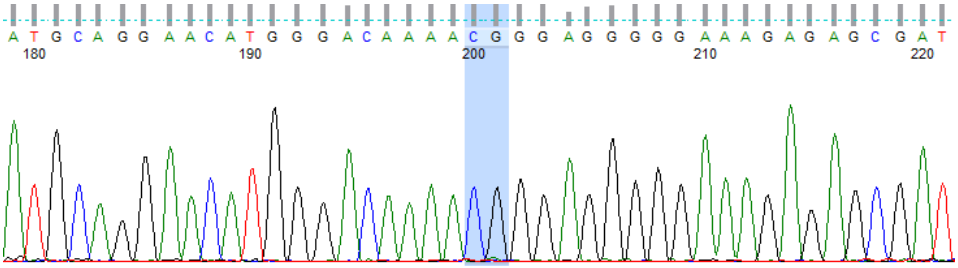
B. Validation of the junction between HPV18 E2/E4 and chromosome 6

The junction between HPV18 E2/E4 and chromosome 6 (HPV18:3,535-chr6:4,328,779) occurs between the highlighted A and C bases



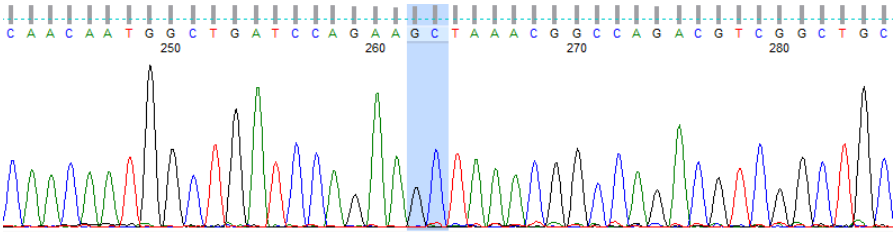
C. Validation of the junction between HPV18 E2 and chromosome 6

The junction between HPV18 E2 and chromosome 6 (HPV18:3,274-chr6:4,282,640) occurs between the highlighted C and G bases.



D. Validation of the junction between HPV18 E1 and chromosome 6

The junction between HPV18 E1 and chromosome 6 (HPV18:929-chr6:4,291,973) occurs between the highlighted G and C bases.



Supplementary references

1. Liang, W.S. et al. Long insert whole genome sequencing for copy number variant and translocation detection. *Nucleic Acids Res* **25**, 25 (2013).
2. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics.* **25**, 1754-60. doi: 10.1093/bioinformatics/btp324. Epub 2009 May 18. (2009).
3. Kumar, P., Henikoff, S. & Ng, P.C. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc.* **4**, 1073-81. doi: 10.1038/nprot.2009.86. Epub 2009 Jun 25. (2009).
4. Adzhubei, I.A. et al. A method and server for predicting damaging missense mutations. *Nat Methods.* **7**, 248-9. doi: 10.1038/nmeth0410-248. (2010).
5. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S.L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25. doi: 10.1186/gb-2009-10-3-r25. Epub 2009 Mar 4. (2009).
6. Trapnell, C. et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc.* **7**, 562-78. doi: 10.1038/nprot.2012.016. (2012).
7. Haas, B.J. et al. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc.* **8**, 1494-512. doi: 10.1038/nprot.2013.084. Epub 2013 Jul 11. (2013).
8. Larkin, M.A. et al. Clustal W and Clustal X version 2.0. *Bioinformatics.* **23**, 2947-8. Epub 2007 Sep 10. (2007).