Gillison et al.

SUPPLEMENTAL TABLES - TABLE OF CONTENTS

		HPV-positive HNSCC						HPV-negative HNSCC						all HNSCC	
		Ohio, n=86		TCGA, n=63		subtotal, n=149		Ohio, n=26		TCGA, n=309		subtotal, n=335		total, n=484	
category	group	n	%total	n	%total	subtotal	% subtotal	n	%total	n	%total	subtotal	% subtotal	total	% total
gender	female	8	9.3%	6	9.5%	14	9.4%	10	38.5%	105	34.0%	115	34.3%	129	26.7%
	male	78	90.7%	57	90.5%	135	90.6%	16	61.5%	204	66.0%	220	65.7%	355	73.3%
race	Caucasian	78	90.7%	61	96.8%	139	93.3%	23	88.5%	270	87.4%	293	87.5%	432	89.3%
	African American	3	3.5%	2	3.2%	5	3.4%	3	11.5%	18	5.8%	21	6.3%	26	5.4%
	Asian	0	0.0%	0	0.0%	0	0.0%	0	0.0%	9	2.9%	9	2.7%	9	1.9%
	American Indian or Alaska native	1	1.2%	0	0.0%	1	0.7%	0	0.0%	1	0.3%	1	0.3%	2	0.4%
	multiple/ other/ unknown	4	4.7%	0	0.0%	4	2.7%	0	0.0%	0	0.0%	0	0.0%	4	0.8%
	not available	0	0.0%	0	0.0%	0	0.0%	0	0.0%	11	3.6%	11	3.3%	11	2.3%
ethnicity	Hispanic or Latino	2	2.3%	2	3.2%	4	2.7%	0	0.0%	16	5.2%	16	4.8%	20	4.1%
	not Hispanic or Latino	84	97.7%	58	92.1%	142	95.3%	26	100.0%	271	87.7%	297	88.7%	439	90.7%
	not available	0	0.0%	3	4.8%	3	2.0%	0	0.0%	22	7.1%	22	6.6%	25	5.2%
age (yrs)	18-39	1	1.2%	2	3.2%	3	2.0%	2	7.7%	14	4.5%	16	4.8%	19	3.9%
	40-49	12	14.0%	11	17.5%	23	15.4%	3	11.5%	41	13.3%	44	13.1%	67	13.8%
	50-59	38	44.2%	31	49.2%	69	46.3%	10	38.5%	79	25.6%	89	26.6%	158	32.6%
	60-69	29	33.7%	12	19.0%	41	27.5%	5	19.2%	95	30.7%	100	29.9%	141	29.1%
	70-89	6	7.0%	7	11.1%	13	8.7%	6	23.1%	79	25.6%	85	25.4%	98	20.2%
primary site	alveolar ridge	0	0.0%	4	6.3%	4	2.7%	0	0.0%	14	4.5%	14	4.2%	18	3.7%
	base of tongue	27	31.4%	16	25.4%	43	28.9%	2	7.7%	11	3.6%	13	3.9%	56	11.6%
	buccal mucosa	0	0.0%	0	0.0%	0	0.0%	0	0.0%	21	6.8%	21	6.3%	21	4.3%
	floor of mouth	0	0.0%	1	1.6%	1	0.7%	0	0.0%	57	18.4%	57	17.0%	58	12.0%
	hard palate	0	0.0%	1	1.6%	1	0.7%	0	0.0%	6	1.9%	6	1.8%	7	1.4%
	oral cavity	0	0.0%	5	7.9%	5	3.4%	0	0.0%	66	21.4%	66	19.7%	71	14.7%
	oral tongue	4	4.7%	5	7.9%	9	6.0%	20	76.9%	119	38.5%	139	41.5%	148	30.6%
	oropharynx	0	0.0%	1	1.6%	1	0.7%	0	0.0%	7	2.3%	7	2.1%	8	1.7%
	tonsil	49	57.0%	30	47.6%	79	53.0%	3	11.5%	8	2.6%	11	3.3%	90	18.6%
	multiple primary site	1	1.2%	0	0.0%	1	0.7%	1	3.8%	0	0.0%	1	0.3%	2	0.4%
	unknown	5	5.8%	0	0.0%	5	3.4%	0	0.0%	0	0.0%	0	0.0%	5	1.0%
T-stage	T1	24	27.9%	8	12.7%	32	21.5%	0	0.0%	19	6.1%	19	5.7%	51	10.5%
	T2	33	38.4%	29	46.0%	62	41.6%	20	76.9%	96	31.1%	116	34.6%	178	36.8%
	T3	14	16.3%	8	12.7%	22	14.8%	5	19.2%	82	26.5%	87	26.0%	109	22.5%
	T4a/b	9	10.5%	14	22.2%	23	15.4%	1	3.8%	83	26.9%	84	25.1%	107	22.1%
	TX	0	0.0%	1	1.6%	1	0.7%	0	0.0%	7	2.3%	7	2.1%	8	1.7%
	not available	0	0.0%	0	0.0%	0	0.0%	0	0.0%	3	1.0%	3	0.9%	3	0.6%
N-stage	N0	6	7.0%	18	28.6%	24	16.1%	7	26.9%	165	53.4%	172	51.3%	196	40.5%
	N1	10	11.6%	7	11.1%	17	11.4%	7	26.9%	53	17.2%	60	17.9%	77	15.9%
	N2/a/b/c	67	77.9%	35	55.6%	102	68.5%	10	38.5%	74	23.9%	84	25.1%	186	38.4%
	N3	2	2.3%	3	4.8%	5	3.4%	2	7.7%	2	0.6%	4	1.2%	9	1.9%
	Nx	1	1.2%	0	0.0%	1	0.7%	0	0.0%	0	0.0%	0	0.0%	1	0.2%
	not available	0	0.0%	0	0.0%	0	0.0%	0	0.0%	3	1.0%	3	0.9%	3	0.6%
M-stage	M0	28	32.6%	61	96.8%	89	59.7%	1	3.8%	292	94.5%	293	87.5%	382	78.9%
	M1	1	1.2%	0	0.0%	1	0.7%	0	0.0%	1	0.3%	1	0.3%	2	0.4%
	MX	55	64.0%	2	3.2%	57	38.3%	25	96.2%	13	4.2%	38	11.3%	95	19.6%
	not available	0	0.0%	0	0.0%	0	0.0%	0	0.0%	3	1.0%	3	0.9%	3	0.6%
tumor grade	G1	0	0.0%	1	1.6%	1	0.7%	1	3.8%	53	17.2%	54	16.1%	55	11.4%
	G2	27	31.4%	23	36.5%	50	33.6%	22	84.6%	192	62.1%	214	63.9%	264	54.5%
	G3	35	40.7%	24	38.1%	59	39.6%	2	7.7%	59	19.1%	61	18.2%	120	24.8%
	GX	0	0.0%	7	11.1%	7	4.7%	0	0.0%	4	1.3%	4	1.2%	11	2.3%
	not available	24	27.9%	2	3.2%	26	17.4%	1	3.8%	1	0.3%	2	0.6%	28	5.8%
p16 assay HPV	positive	80	93.0%	27	42.9%	107	71.8%	2	7.7%	0	0.0%	2	0.6%	109	22.5%
	negative	1	1.2%	0	0.0%	1	0.7%	8	30.8%	53	17.2%	61	18.2%	62	12.8%
	not available	5	5.8%	36	57.1%	41	27.5%	16	61.5%	256	82.8%	272	81.2%	313	64.7%
ISH assay HPV	positive	77	89.5%	17	27.0%	94	63.1%	0	0.0%	0	0.0%	0	0.0%	94	19.4%
	negative	4	4.7%	0	0.0%	4	2.7%	9	34.6%	45	14.6%	54	16.1%	58	12.0%
	not available	5	5.8%	46	73.0%	51	34.2%	17	65.4%	264	85.4%	281	83.9%	332	68.6%
HPV status	positive	86	100.0%	63	100.0%	149	100.0%	0	0.0%	0	0.0%	0	0.0%	149	30.8%
	negative	0	0.0%	0	0.0%	0	0.0%	26	100.0%	309	100.0%	335	100.0%	335	69.2%

Supplemental Table S1A. Demographic and clinical characteristics of the study population.
Shown are characteristics of the entire study population (*far right*, N=484), HPV-positive (*left*, n = 149) and HPV-negative (*right*, n = 335).

dataset	CGI-WGS	Illum.-WGS	TCGA-WGS
number of cases	59	53	41
median depth of coverage (normal)	93.2	49	41.1
range of depth of coverage (normal)	80.4-105.5	42.0-64.1	30.1-52.2
median depth of coverage (tumor)	93.5	102.3	67.7
range of depth of coverage (tumor)	77.8-108.6	90.8-139.6	38.6-84.7

Supplemental Table S1B. Sequencing depth of coverage from WGS platforms.
Shown are the counts and depths of sequencing coverage for samples studied using different next-generation sequencing platforms. Ohio cohort samples were studied using CGI, Complete Genomic platform; and Illumina (Illum.) sequencing. TCGA data were downloaded from The Cancer Genome Atlas database (https://cancergenome.nih.gov/).

Sample ID	HPV	HPV type	platform	normal depth (x)	tumor depth (x)	no. SNV (WGS)	no. Indel (WGS)	no. SNV (exon)	no. Indel (exon)	SNV rate in exon (per Mbp)
GS18001	pos	HPV16	Illum.-WGS	60.4	107.8	180708	105	3959	2	78.57
GS18002	pos	HPV16	Illum.-WGS	59.9	108.3	70894	89	1301	2	25.82
GS18003	pos	HPV16	Illum.-WGS	48.8	90.8	49410	1	1190	0	23.62
GS18004	pos	HPV16	CGI-WGS	98.4	94.3	47971	214	1042	6	20.68
GS18005	pos	HPV16	Illum.-WGS	48.2	92.3	47206	80	794	1	15.76
GS18006	pos	HPV16	CGI-WGS	100.4	98.8	31331	74	755	4	14.98
GS18007	pos	HPV16	Illum.-WGS	42.8	139.6	34497	0	717	0	14.23
GS18008	pos	HPV16	Illum.-WGS	43.2	104.7	37881	79	685	2	13.59
GS18009	pos	HPV16	Illum.-WGS	46.5	107.7	26888	126	533	2	10.58
GS18010	pos	HPV16	Illum.-WGS	50.9	102.3	29857	137	491	2	9.74
GS18011	pos	HPV16	Illum.-WGS	45.2	98.3	22244	11	436	0	8.65
GS18012	pos	HPV16	Illum.-WGS	45.6	104.7	20193	271	392	5	7.78
GS18013	pos	HPV16	Illum.-WGS	51.4	103.8	18973	123	358	5	7.10
GS18014	pos	HPV16	Illum.-WGS	50.0	99.1	19546	13	355	0	7.04
GS18015	pos	HPV16	Illum.-WGS	52.7	101.1	18524	237	336	5	6.67
GS18016	pos	HPV16	CGI-WGS	105.2	108.6	13383	45	328	1	6.51
GS18017	pos	HPV16	Illum.-WGS	44.9	99.2	14722	160	308	3	6.11
GS18018	pos	HPV16	Illum.-WGS	44.4	107.5	14912	158	296	2	5.87
GS18019	pos	HPV16	CGI-WGS	93.7	93.6	13173	42	288	3	5.72
GS18020	pos	HPV35	Illum.-WGS	49.5	110.5	17152	23	286	1	5.68
GS18021	pos	HPV16	Illum.-WGS	58.2	98.9	16196	58	250	3	4.96
GS18022	pos	HPV16	CGI-WGS	104.0	102.6	12543	109	244	0	4.84
GS18023	pos	HPV16	Illum.-WGS	50.2	107.9	10833	381	203	4	4.03
GS18026	pos	HPV16	CGI-WGS	100.2	99.4	11229	148	194	1	3.85
GS18027	pos	HPV16	Illum.-WGS	50.2	94.6	12992	23	184	0	3.65
GS18028	pos	HPV16	Illum.-WGS	47.3	103.4	10269	50	165	0	3.27
GS18029	pos	HPV16	Illum.-WGS	64.1	92.2	10283	117	164	4	3.25
GS18030	pos	HPV16	Illum.-WGS	52.3	102.6	9434	103	154	2	3.06
GS18031	pos	HPV16	Illum.-WGS	48.8	95.2	9646	25	153	0	3.04
GS18033	pos	HPV16	Illum.-WGS	54.2	101.7	9961	213	152	0	3.02
GS18034	pos	HPV16	CGI-WGS	97.0	96.0	8441	115	145	3	2.88
GS18035	pos	HPV16	CGI-WGS	92.4	90.7	10454	76	143	3	2.84
GS18036	pos	HPV16	Illum.-WGS	46.3	113.2	9275	437	142	8	2.82
GS18037	pos	HPV16	CGI-WGS	92.5	94.0	6745	24	134	0	2.66
GS18038	pos	HPV16	Illum.-WGS	44.7	100.7	8851	14	134	0	2.66
GS18039	pos	HPV33	Illum.-WGS	51.6	94.9	7798	132	134	1	2.66
GS18040	pos	HPV16	CGI-WGS	89.6	81.2	6309	87	127	2	2.52
GS18041	pos	HPV16	CGI-WGS	105.5	100.5	8779	337	126	3	2.50
GS18043	pos	HPV16	Illum.-WGS	46.9	94.3	5797	100	123	3	2.44
GS18044	pos	HPV16	Illum.-WGS	48.3	99.2	6693	117	122	1	2.42
GS18046	pos	HPV16	Illum.-WGS	49.1	97.6	6596	29	107	0	2.12
GS18047	pos	HPV16	Illum.-WGS	49.0	110.1	6967	23	102	0	2.02
GS18048	pos	HPV33	Illum.-WGS	48.4	100.7	5665	1	102	0	2.02
GS18049	pos	HPV16	Illum.-WGS	54.5	111.3	6497	144	99	1	1.96
GS18051	pos	HPV16	CGI-WGS	90.3	90.7	5818	74	94	1	1.87
GS18052	pos	HPV16	Illum.-WGS	52.6	103.3	4525	27	93	1	1.85
GS18053	pos	HPV16	Illum.-WGS	47.8	94.7	6317	203	93	6	1.85
GS18055	pos	HPV16	CGI-WGS	91.8	86.5	6314	63	92	1	1.83
GS18059	pos	HPV16	CGI-WGS	88.4	93.5	4551	148	87	5	1.73
GS18061	pos	HPV16	CGI-WGS	89.2	89.5	5241	94	84	4	1.67
GS18062	pos	HPV16	Illum.-WGS	51.8	129.9	6726	58	83	2	1.65
GS18063	pos	HPV16	Illum.-WGS	45.6	101.4	6315	38	82	2	1.63
GS18065	pos	HPV16	Illum.-WGS	49.1	109.0	5672	291	81	3	1.61

GS18066	pos	HPV16	Illum.-WGS	49.7	95.9	4429	4	79	0	1.57
GS18067	pos	HPV16	Illum.-WGS	57.0	132.4	4899	115	79	5	1.57
GS18068	pos	HPV16	Illum.-WGS	47.6	117.8	6747	31	75	0	1.49
GS18069	pos	HPV16	CGI-WGS	102.1	100.4	4361	66	72	1	1.43
GS18070	pos	HPV33	CGI-WGS	102.8	101.1	3469	130	72	3	1.43
GS18071	pos	HPV16	Illum.-WGS	55.7	114.2	4737	1	72	0	1.43
GS18072	pos	HPV35	CGI-WGS	96.7	95.0	3143	26	71	0	1.41
GS18074	pos	HPV16	Illum.-WGS	42.0	99.0	4210	21	62	0	1.23
GS18076	pos	HPV16	CGI-WGS	92.0	98.5	3213	52	61	0	1.21
GS18077	pos	HPV35	Illum.-WGS	60.5	120.1	3747	9	61	0	1.21
GS18078	pos	HPV16	CGI-WGS	102.6	102.8	3824	110	60	2	1.19
GS18079	pos	HPV16	Illum.-WGS	47.3	112.4	3430	116	55	5	1.09
GS18081	pos	HPV16	Illum.-WGS	46.4	99.3	4452	123	53	1	1.05
GS18082	pos	HPV16	CGI-WGS	99.1	97.4	2666	125	52	2	1.03
GS18085	pos	HPV59	Illum.-WGS	43.5	100.0	2238	1	50	0	0.99
GS18087	pos	HPV16	CGI-WGS	80.4	90.2	1950	39	45	0	0.89
GS18088	pos	HPV16	CGI-WGS	103.8	97.5	1573	49	44	0	0.87
GS18091	pos	HPV16	CGI-WGS	91.8	89.2	2433	50	42	0	0.83
GS18092	pos	HPV16	CGI-WGS	93.2	94.7	1975	73	42	1	0.83
GS18093	pos	HPV16	Illum.-WGS	55.7	110.8	2304	163	41	5	0.81
GS18094	pos	HPV16	CGI-WGS	94.9	88.3	1522	36	39	1	0.77
GS18095	pos	HPV16	CGI-WGS	99.1	95.7	2172	76	38	5	0.75
GS18096	pos	HPV16	CGI-WGS	89.5	89.6	2036	56	37	2	0.73
GS18097	pos	HPV16	Illum.-WGS	55.0	96.6	2181	97	37	4	0.73
GS18098	pos	HPV16	Illum.-WGS	48.6	105.9	2055	0	36	0	0.71
GS18099	pos	HPV16	Illum.-WGS	52.6	108.7	2242	4	34	0	0.67
GS18101	pos	HPV16	CGI-WGS	94.2	95.8	1419	45	31	4	0.62
GS18102	pos	HPV33	CGI-WGS	101.3	97.1	1448	14	26	0	0.52
GS18103	pos	HPV18	CGI-WGS	101.9	100.9	677	5	22	0	0.44
GS18105	pos	HPV18	CGI-WGS	95.2	88.1	1086	30	16	0	0.32
GS18107	pos	HPV69	CGI-WGS	84.9	89.6	644	11	13	0	0.26
GS18108	pos	HPV16	Illum.-WGS	46.7	94.4	534	8	13	0	0.26
GS18109	pos	HPV16	CGI-WGS	82.5	80.8	232	1	10	1	0.20
TCGA-BA-4077	pos	HPV16	TCGA-WGS	41.1	44.7	26063	397	508	9	10.08
TCGA-BA-5153	pos	HPV16	TCGA-WGS	41.8	38.6	3390	28	49	0	0.97
TCGA-BB-4225	pos	HPV33	TCGA-WGS	44.3	58.2	6230	116	118	1	2.34
TCGA-CN-4741	pos	HPV16	TCGA-WGS	46.0	46.1	13249	399	178	5	3.53
TCGA-CN-5374	pos	HPV16	TCGA-WGS	36.6	75.9	21923	82	284	1	5.64
TCGA-CR-5249	pos	HPV16	TCGA-WGS	37.2	67.0	3125	33	37	1	0.73
TCGA-CR-5250	pos	HPV16	TCGA-WGS	33.8	74.2	3806	133	62	3	1.23
TCGA-CR-6470	pos	HPV16	TCGA-WGS	44.1	42.7	3168	30	62	0	1.23
TCGA-CR-6472	pos	HPV16	TCGA-WGS	43.9	45.1	44330	636	873	6	17.32
TCGA-CR-6482	pos	HPV16	TCGA-WGS	37.1	68.0	4131	17	83	1	1.65
TCGA-CR-6487	pos	HPV16	TCGA-WGS	36.2	68.0	13613	37	232	0	4.60
TCGA-CR-7385	pos	HPV16	TCGA-WGS	44.5	39.6	4312	17	66	1	1.31
TCGA-CR-7404	pos	HPV16	TCGA-WGS	43.3	45.6	12532	19	195	0	3.87
TCGA-CV-5442	pos	HPV16	TCGA-WGS	36.3	67.7	28400	451	433	6	8.59
TCGA-CV-6433	pos	HPV16	TCGA-WGS	31.0	75.8	6149	293	80	3	1.59
TCGA-CV-6961	pos	HPV16	TCGA-WGS	52.2	46.7	47608	122	994	2	19.73
TCGA-CV-7100	pos	HPV33	TCGA-WGS	39.3	43.7	3117	81	24	2	0.48

Supplemental Table S1C. Summary statistics for WGS data for individual HPV-positive OSCC. WGS data generated from tumor and matched normal control blood samples (T/N) pairs were analyzed for 103 HPV-positive OSCC. Shown for each sample from left to right are: the unique sample ID; HPV status and type as determined by WGS read alignment to a library of HPV genomes; source of genome sequencing data; depth of genome sequence coverage (normal and tumor); number of somatic variants

(SNVs and small insertion-deletion, indel polymorphisms) in the whole genome and in the exome; and rate of somatic variants per megabase pair in the exome. Somatic variants were detected in CGI data with the Complete Genomics Cancer Genome Pipeline v2. For Illumina data using 2 x 125 bp PE reads, from Ohio cohort samples or downloaded from TCGA, somatic variants were called using three analysis pipelines (Mutect, LoFreq, and Strelka), where each variant was called by at least two of the three pipelines.

Sample ID	HPV	HPV type	platform	normal coverage (x)	tumor coverage (x)	no. SNVs (WGS)	no. Indels (WGS)	no. SNVs (exon)	no. Indels (exon)	SNV rate in exon (per Mbp)
GS18024	neg	NA	Illum.-WGS	46.3	100.4	15629	46	198	2	3.93
GS18025	neg	NA	CGI-WGS	84.3	83.4	15864	476	197	13	3.91
GS18032	neg	NA	CGI-WGS	92.9	90.0	9576	150	152	1	3.02
GS18042	neg	NA	CGI-WGS	94.0	94.6	6791	157	124	4	2.46
GS18045	neg	NA	CGI-WGS	86.7	87.1	5418	89	108	4	2.14
GS18050	neg	NA	CGI-WGS	92.0	87.7	5784	126	97	0	1.92
GS18054	neg	NA	CGI-WGS	94.9	94.5	6662	146	92	2	1.83
GS18056	neg	NA	CGI-WGS	87.8	88.4	4450	71	91	3	1.81
GS18057	neg	NA	CGI-WGS	102.5	104.1	4130	102	90	9	1.79
GS18058	neg	NA	CGI-WGS	96.6	93.5	5295	111	88	7	1.75
GS18060	neg	NA	CGI-WGS	86.3	87.3	3671	45	86	1	1.71
GS18064	neg	NA	CGI-WGS	86.2	88.0	4223	63	81	2	1.61
GS18073	neg	NA	CGI-WGS	88.2	77.8	3758	137	64	3	1.27
GS18075	neg	NA	CGI-WGS	81.2	80.0	4621	74	61	1	1.21
GS18080	neg	NA	CGI-WGS	89.9	89.6	3670	58	54	0	1.07
GS18083	neg	NA	CGI-WGS	90.7	94.0	3615	87	50	1	0.99
GS18084	neg	NA	CGI-WGS	88.9	89.3	3874	49	50	1	0.99
GS18086	neg	NA	CGI-WGS	99.2	96.7	5182	76	45	1	0.89
GS18089	neg	NA	CGI-WGS	90.1	91.5	2743	98	43	3	0.85
GS18090	neg	NA	CGI-WGS	102.6	102.0	2541	115	43	4	0.85
GS18100	neg	NA	CGI-WGS	90.3	90.3	1459	29	31	1	0.62
GS18104	neg	NA	CGI-WGS	99.7	83.9	955	32	21	0	0.42
GS18106	neg	NA	CGI-WGS	88.8	88.1	748	23	13	1	0.26
GS18110	neg	NA	CGI-WGS	101.3	85.5	1606	15	8	0	0.16
GS18111	neg	NA	CGI-WGS	99.1	99.4	173	3	5	0	0.10
GS18112	neg	NA	CGI-WGS	87.8	87.4	169	9	4	0	0.08
TCGA-BA-5149	neg	NA	TCGA-WGS	40.4	67.8	17929	171	216	6	4.29
TCGA-BA-5556	neg	NA	TCGA-WGS	38.3	70.1	17895	15	219	0	4.35
TCGA-BA-6872	neg	NA	TCGA-WGS	40.0	73.9	25181	256	308	4	6.11
TCGA-BA-6873	neg	NA	TCGA-WGS	38.5	84.7	5588	99	105	2	2.08
TCGA-CN-4737	neg	NA	TCGA-WGS	35.7	82.3	5013	43	78	0	1.55
TCGA-CN-5365	neg	NA	TCGA-WGS	31.6	73.9	7674	94	94	2	1.87
TCGA-CN-6011	neg	NA	TCGA-WGS	42.9	73.2	52278	514	672	9	13.34
TCGA-CN-6994	neg	NA	TCGA-WGS	44.3	61.0	17621	267	239	3	4.74
TCGA-CQ-6225	neg	NA	TCGA-WGS	43.6	66.0	18810	1183	248	22	4.92
TCGA-CQ-6228	neg	NA	TCGA-WGS	41.6	72.4	13586	497	176	6	3.49
TCGA-CR-6467	neg	NA	TCGA-WGS	38.7	66.1	3567	132	76	3	1.51
TCGA-CR-6491	neg	NA	TCGA-WGS	42.8	68.9	17100	655	236	9	4.68
TCGA-CR-7382	neg	NA	TCGA-WGS	43.7	68.9	8579	2	147	0	2.92
TCGA-CR-7391	neg	NA	TCGA-WGS	39.2	45.4	393	0	3	0	0.06
TCGA-CV-5973	neg	NA	TCGA-WGS	40.3	70.7	7982	211	117	4	2.32
TCGA-CV-6956	neg	NA	TCGA-WGS	43.0	64.2	32697	671	388	11	7.70
TCGA-CV-7090	neg	NA	TCGA-WGS	30.3	79.1	6074	40	94	0	1.87
TCGA-CV-7180	neg	NA	TCGA-WGS	38.7	77.4	8726	25	111	1	2.20
TCGA-CV-7255	neg	NA	TCGA-WGS	30.1	77.1	10409	204	156	6	3.10
TCGA-CV-7416	neg	NA	TCGA-WGS	45.2	39.5	9187	315	121	1	2.40
TCGA-CV-7432	neg	NA	TCGA-WGS	42.4	67.0	28873	606	274	9	5.44
TCGA-CX-7086	neg	NA	TCGA-WGS	41.2	72.8	33543	528	462	10	9.17
TCGA-DQ-5625	neg	NA	TCGA-WGS	41.2	62.0	22901	165	362	4	7.18
TCGA-HD-7753	neg	NA	TCGA-WGS	44.6	61.6	14081	342	186	7	3.69

Supplemental Table S1D. Summary statistics for WGS data for individual HPV-negative OSCC.

WGS data from tumor and matched normal control blood samples (T/N) pairs were analyzed for 50 HPV-negative OSCC. Shown for each sample from *left* to *right* are: the unique sample ID, HPV status at determined from WGS data; source of genome sequencing data;, depth of genome sequence coverage (normal and tumor); number of somatic variants (SNVs and small insertion-deletion, indel polymorphisms) in the whole genome and exome; and rate of somatic variants per megabase pair in the exome. Somatic variants were detected in CGI data with the Complete Genomics Cancer Genome Pipeline v2. For Illumina data from Ohio cohort or TCGA, using 2 x 125 bp PE reads, somatic variants were called using three analysis pipelines (Mutect, LoFreq, and Strelka), where each variant was called by at least two of the three pipelines.

HPV type	WGS	WES	Total	%Total
HPV16	90	38	128	85.90%
HPV18	2	0	2	1.30%
HPV33	6	5	11	7.40%
HPV35	3	3	6	4.00%
HPV59	1	0	1	0.70%
HPV69	1	0	1	0.70%
Total	103	46	149	100.00%

Supplemental Table S1E. HPV type distribution in HPV-positive OSCC per WGS data.
Shown is the HPV type distribution for cases identified as HPV-positive OSCC.

platform	HPV-positive	HPV-negative	total
Ohio - CGI WGS	34	25	59
Ohio - Illum. WGS	52	1	53
TCGA - WGS	17	24	41
TCGA - WES	46	285	331
total	149	335	484

Supplemental Table S1F. OSCC sample counts by HPV status and sequencing platform.
112 OSCC tumors from the Ohio cohort were evaluated by WGS, including 86 HPV-positive and 26 HPV-negative OSCC. Also included in the analysis were 372 OSCC (restricted to oropharynx and oral cavity) with WES data generated on the Illumina platform from TCGA (41 samples with WGS ~40x coverage and 331 samples sequenced by WES).

sample ID	HPV16 (ISH)	p16 (IHC)	no. HPV16 copies per cell	no. HPV18 copies per cell	no. HPV33 copies per cell	no. HPV35 copies per cell	no. HPV59 copies per cell	HPV E6/E7 mRNA detection	line Blot	clinical test (HPV ISH)	clinical test (p16)	WGS platform	HPV type by WGS	RNA-seq
GS18001	pos	pos	81.3	NA	NA	NA	NA	pos	16*	pos	pos	Illum.	16	Y
GS18002	NA	NA	NA	NA	NA	NA	NA	NA	16*	pos	pos	Illum.	16	Y
GS18003	NA	NA	NA	NA	NA	NA	NA	NA	16*	pos	pos	Illum.	16	Y
GS18004	pos	pos	1.0	NA	NA	NA	NA	pos	16	pos	pos	CGI	16	Y
GS18005	NA	NA	NA	NA	NA	NA	NA	NA	16*	pos	pos	Illum.	16	Y
GS18006	pos	pos	124.8	NA	NA	NA	NA	pos	16	pos	pos	CGI	16	Y
GS18007	NA	NA	NA	NA	NA	NA	NA	NA	16*	pos	pos	Illum.	16	Y
GS18008	NA	NA	NA	NA	NA	NA	NA	NA	16*	pos	pos	Illum.	16	Y
GS18009	NA	NA	NA	NA	NA	NA	NA	NA	16*	pos	pos	Illum.	16	Y
GS18010	pos	pos	48.6	NA	NA	NA	NA	pos	16	pos	pos	Illum.	16	Y
GS18011	NA	NA	NA	NA	NA	NA	NA	NA	16*	pos	pos	Illum.	16	Y
GS18012	pos	pos	7.5	NA	NA	NA	NA	pos	16*	pos	pos	Illum.	16	Y
GS18013	NA	NA	NA	NA	NA	NA	NA	NA	16*	pos	pos	Illum.	16	Y
GS18014	pos	pos	1.1	NA	NA	NA	NA	pos	16	pos	pos	Illum.	16	Y
GS18015	pos	pos	5.4	NA	NA	NA	NA	pos	16	pos	pos	Illum.	16	Y
GS18016	pos	pos	59.6	NA	NA	NA	NA	pos	16	neg	pos	CGI	16	Y
GS18017	pos	pos	12.0	NA	NA	NA	NA	pos	16	pos	pos	Illum.	16	Y
GS18018	pos	pos	37.3	NA	NA	NA	NA	pos	16*	pos	pos	Illum.	16	Y
GS18019	pos	pos	21.9	NA	NA	NA	NA	pos	16	pos	pos	CGI	16	Y
GS18020	neg	pos	0.0	NA	NA	115.1	NA	pos	35	pos	pos	Illum.	35	Y
GS18021	neg	pos	0.2	NA	NA	NA	NA	pos	16	unknown	unknown	Illum.	16	Y
GS18022	NA	NA	NA	NA	NA	NA	NA	NA	16*	pos	pos	CGI	16	Y
GS18023	NA	NA	NA	NA	NA	NA	NA	NA	16*	unknown	pos	Illum.	16	Y
GS18026	pos	pos	747.4	NA	NA	NA	NA	pos	16	pos	pos	CGI	16	Y
GS18027	NA	NA	NA	NA	NA	NA	NA	NA	16*	pos	pos	Illum.	16	Y
GS18028	NA	NA	NA	NA	NA	NA	NA	NA	16*	pos	pos	Illum.	16	Y
GS18029	pos	pos	49.8	NA	NA	NA	NA	pos	16	pos	pos	Illum.	16	Y
GS18030	pos	pos	20.3	NA	NA	NA	NA	pos	16*	pos	pos	Illum.	16	Y
GS18031	pos	pos	9.5	NA	NA	NA	NA	NA	16	pos	pos	Illum.	16	Y
GS18033	pos	pos	3.6	NA	NA	NA	NA	pos	16	pos	pos	Illum.	16	Y
GS18034	pos	pos	29.9	NA	NA	NA	NA	pos	16	pos	pos	CGI	16	Y

GS18035	pos	pos	2.0	NA	NA	NA	NA	pos	16	pos	pos	CGI	16	Y
GS18036	NA	NA	NA	NA	NA	NA	NA	NA	16*	unknown	unknown	Illum.	16	Y
GS18037	pos	pos	38.3	NA	NA	NA	NA	pos	16	pos	pos	CGI	16	Y
GS18038	NA	NA	NA	NA	NA	NA	NA	NA	16*	pos	pos	Illum.	16	Y
GS18039	neg	pos	0.0	NA	0.7	NA	NA	pos	33	pos	pos	Illum.	33	Y
GS18040	pos	pos	12.2	NA	NA	NA	NA	pos	16	pos	pos	CGI	16	Y
GS18041	pos	pos	25.5	NA	NA	NA	NA	pos	16	unknown	unknown	CGI	16	Y
GS18043	pos	pos	76.4	NA	NA	NA	NA	pos	16	pos	pos	Illum.	16	Y
GS18044	NA	NA	NA	NA	NA	NA	NA	NA	16*	pos	pos	Illum.	16	Y
GS18046	pos	pos	33.1	NA	NA	NA	NA	pos	16	pos	pos	Illum.	16	Y
GS18047	NA	NA	NA	NA	NA	NA	NA	NA	16*	pos	pos	Illum.	16	Y
GS18048	pos	pos	0.0	NA	NA	NA	NA	neg	33*	pos	pos	Illum.	33	Y
GS18049	pos	pos	49.3	NA	NA	NA	NA	pos	16	pos	pos	Illum.	16	Y
GS18051	pos	pos	1.7	NA	NA	NA	NA	pos	16	pos	pos	CGI	16	Y
GS18052	pos	neg	27.1	NA	NA	NA	NA	pos	16	pos	neg	Illum.	16	Y
GS18053	pos	pos	10.1	NA	NA	NA	NA	pos	16	pos	pos	Illum.	16	Y
GS18055	pos	pos	23.5	NA	NA	NA	NA	pos	16	pos	pos	CGI	16	NA
GS18059	pos	pos	20.5	NA	NA	NA	NA	pos	16	pos	pos	CGI	16	Y
GS18061	pos	pos	18.6	NA	NA	NA	NA	pos	16	pos	pos	CGI	16	Y
GS18062	NA	NA	NA	NA	NA	NA	NA	NA	16*	pos	pos	Illum.	16	Y
GS18063	NA	NA	NA	NA	NA	NA	NA	NA	16*	pos	pos	Illum.	16	Y
GS18065	NA	NA	NA	NA	NA	NA	NA	NA	16*	pos	pos	Illum.	16	Y
GS18066	pos	pos	6.8	NA	NA	NA	NA	pos	16	pos	pos	Illum.	16	Y
GS18067	NA	NA	NA	NA	NA	NA	NA	NA	16*	pos	pos	Illum.	16	Y
GS18068	NA	NA	NA	NA	NA	NA	NA	NA	16*	pos	pos	Illum.	16	Y
GS18069	pos	pos	14.8	NA	NA	NA	NA	pos	16	pos	pos	CGI	16	Y
GS18070	pos	pos	0.0	NA	15.6	NA	NA	pos	33	pos	pos	CGI	33	Y
GS18071	NA	NA	NA	NA	NA	NA	NA	NA	16*	pos	pos	Illum.	16	Y
GS18072	neg	pos	0.0	NA	NA	5.3	NA	pos	35	pos	pos	CGI	35	Y
GS18074	pos	pos	5.1	NA	NA	NA	NA	pos	16*	pos	pos	Illum.	16	Y
GS18076	pos	pos	154.1	NA	NA	NA	NA	pos	16	pos	pos	CGI	16	Y
GS18077	NA	NA	NA	NA	NA	NA	NA	NA	35*	pos	pos	Illum.	35	NA
GS18078	pos	pos	14.9	NA	NA	NA	NA	pos	16	pos	pos	CGI	16	Y
GS18079	pos	pos	58.7	NA	NA	NA	NA	pos	16	pos	pos	Illum.	16	Y
GS18081	NA	NA	NA	NA	NA	NA	NA	NA	16*	pos	pos	Illum.	16	Y
GS18082	pos	pos	53.3	NA	NA	NA	NA	pos	16	pos	pos	CGI	16	Y

GS18085	neg	neg	0.0	NA	NA	NA	363.9	pos	59	neg	pos	Illum.	59	Y
GS18087	pos	pos	60.2	NA	NA	NA	NA	pos	16	pos	pos	CGI	16	Y
GS18088	pos	pos	9.2	NA	NA	NA	NA	pos	16	pos	pos	CGI	16	Y
GS18091	pos	pos	12.8	NA	NA	NA	NA	pos	16	unknown	pos	CGI	16	Y
GS18092	pos	pos	8.4	NA	NA	NA	NA	pos	16	pos	pos	CGI	16	Y
GS18093	pos	pos	2.7	NA	NA	NA	NA	pos	16	pos	pos	Illum.	16	Y
GS18094	NA	NA	NA	NA	NA	NA	NA	NA	16*	pos	pos	CGI	16	Y
GS18095	pos	pos	1.4	NA	NA	NA	NA	pos	16	pos	pos	CGI	16	Y
GS18096	pos	pos	18.7	NA	NA	NA	NA	pos	16	pos	pos	CGI	16	Y
GS18097	pos	pos	19.8	NA	NA	NA	NA	pos	16	pos	pos	Illum.	16	Y
GS18098	NA	NA	NA	NA	NA	NA	NA	NA	16*	pos	pos	Illum.	16	Y
GS18099	NA	NA	NA	NA	NA	NA	NA	NA	16	pos	pos	Illum.	16	Y
GS18101	pos	pos	9.8	NA	NA	NA	NA	pos	16	pos	pos	CGI	16	Y
GS18102	neg	pos	0.0	NA	3.2	NA	NA	pos	33	pos	pos	CGI	33	Y
GS18103	neg	pos	0.0	3.0	NA	NA	NA	pos	18	neg	pos	CGI	18	Y
GS18105	neg	pos	0.0	2391.7	NA	NA	NA	pos	18	pos	pos	CGI	18	Y
GS18107	neg	pos	0.0	NA	NA	NA	NA	NA	69	unknown	unknown	CGI	69	Y
GS18108	pos	pos	5.5	NA	NA	NA	NA	pos	16	pos	pos	Illum.	16	Y
GS18109	pos	pos	57.7	NA	NA	NA	NA	pos	16	pos	pos	CGI	16	Y

*LB was tested with fresh tumor DNA

Supplemental Table S1G. Results of extended HPV testing of tumor specimens called HPV-positive in the Ohio cohort. Shown are extended HPV testing results for the 86 HPV-positive tumors in the Ohio cohort with WGS data. Data were generated by analysis of DNA and RNA purified from FFPE when available, with exception of last three columns on the right. From *left* to *right* these include: specimen identification; HPV16 in situ hybridization (ISH) result from research laboratory; p16 immunohistochemistry (ISH) result from research laboratory; HPV16 copies per tumor cell genome as determined by type-specific TaqMan qPCR normalized to number of diploid genomes in the PCR reaction as estimated by TaqMan qPCR to a single copy human gene, ERV3; ibid, HPV18; ibid, HPV33; ibid, HPV35; ibid HPV59; detection of HPV type-specific E6/E7 mRNA by type-specific qPCR, corresponding to DNA type in line blot; HPV type detected by Roche Linear Array; HPV ISH result performed in the clinical pathology laboratory; p16 IHC result performed in the clinical laboratory at OSU Medical Center; WGS method: HPV type detected by WGS: HPV type detected by RNA-seq. See methods for more details and references. *NA*: not available/applicable. *Pos*; positive. *Neg*; negative. *Unknown*; result not available. *Y*; yes, HPV type detected by WGS is present by RNA-seq analysis; *NA,* RNA-seq data not available. See methods for additional details.

sample ID	HPV16 (ISH)	p16 (IHC)	no. HPV16 copies / cell	HPV mRNA detection	line blot	clinical test (HPV ISH)	clinical test (p16)	WGS platform	comments
GS18024	NA	NA	NA	NA	16(2)*	unknown	neg	Illum.	low intensity band for LB
GS18025	neg	pos	0.0	neg	undetect	neg	pos	CGI	
GS18032	neg	neg	0.0	neg	16(1)	unknown	neg	CGI	low intensity band for LB
GS18042	neg	neg	0.0	neg	undetect	unknown	neg	CGI	
GS18045	neg	neg	0.0	neg	undetect	unknown	neg	CGI	
GS18050	neg	neg	0.0	neg	undetect	unknown	neg	CGI	
GS18054	neg	neg	0.0	neg	16(2)	neg	neg	CGI	low intensity band for LB
GS18056	neg	neg	0.0	neg	undetect	unknown	neg	CGI	
GS18057	neg	neg	0.0	neg	undetect	unknown	neg	CGI	
GS18058	neg	neg	0.0	neg	undetect	neg	neg	CGI	
GS18060	neg	neg	0.0	neg	16(1)	neg	neg	CGI	low intensity band for LB
GS18064	neg	neg	0.0	neg	undetect	unknown	neg	CGI	
GS18073	neg	neg	0.0	neg	undetect	neg	neg	CGI	
GS18075	neg	neg	0.0	neg	undetect	unknown	neg	CGI	
GS18080	neg	neg	0.0	neg	undetect	unknown	neg	CGI	
GS18083	neg	neg	0.0	neg	undetect	unknown	neg	CGI	
GS18084	neg	neg	0.0	neg	undetect	unknown	neg	CGI	
GS18086	neg	neg	0.0	neg	undetect	neg	neg	CGI	
GS18089	neg	neg	0.0	neg	undetect	unknown	neg	CGI	
GS18090	neg	neg	0.0	neg	undetect	unknown	neg	CGI	
GS18100	neg	neg	0.0	neg	undetect	unknown	neg	CGI	
GS18104	neg	neg	0.0	neg	undetect	unknown	neg	CGI	
GS18106	neg	neg	0.0	neg	undetect	neg	neg	CGI	
GS18110	neg	pos	0.0	neg	undetect	neg	pos	CGI	
GS18111	neg	neg	0.0	neg	undetect	neg	neg	CGI	
GS18112	neg	neg	0.0	neg	16(2)	unknown	neg	CGI	low intensity band for LB

Supplemental Table S1H. Results of extended HPV testing of tumor specimens called HPV-negative in the Ohio cohort. Data were generated by analysis of DNA and RNA purified from FFPE when available, with exception of last three columns on the right. Shown from *left* to *right* are sample identification number; HPV16 ISH from research laboratory; p16 IHC from research laboratory; HPV16 copy number by qPCR; HPV16 mRNA detection by RT-PCR; HPV type detected by Roche Linear Array; HPV ISH result from clinical pathology laboratory at OSU Medical Center; p16 IHC result from clinical laboratory at OSU medical center; WGS platform used; comments on line blot assay intensity result. See methods for additional details and references. *NA*, not applicable; *Undetect*, undetectable; *Neg*, negative; *unknown*, result not available. Low intensity bands for HPV16 on line blot could not be confirmed as present by HPV16 type-specific DNA and RNA detection.

sample ID	HPV type	HPV coverage (x)	human genome coverage (x)	HPV copy no.	platform
GS18001	HPV16	77.67	107.81	0.72	Illum.
GS18002	HPV16	6308.74	108.27	58.27	Illum.
GS18003	HPV16	793.3	90.78	8.74	Illum.
GS18004	HPV16	103.88	94.64	1.1	CGI
GS18005	HPV16	180.63	92.32	1.96	Illum.
GS18006	HPV16	10561.09	99.15	106.52	CGI
GS18007	HPV16	28.37	139.6	0.2	Illum.
GS18008	HPV16	2731.41	104.73	26.08	Illum.
GS18009	HPV16	952.59	107.66	8.85	Illum.
GS18010	HPV16	807.82	102.25	7.9	Illum.
GS18011	HPV16	3398.2	98.28	34.58	Illum.
GS18012	HPV16	2346.97	104.71	22.41	Illum.
GS18013	HPV16	52.76	103.8	0.51	Illum.
GS18014	HPV16	110.47	99.11	1.11	Illum.
GS18015	HPV16	1227.98	101.09	12.15	Illum.
GS18016	HPV16	1501.41	108.96	13.78	CGI
GS18017	HPV16	745.35	99.18	7.52	Illum.
GS18018	HPV16	3958.21	107.5	36.82	Illum.
GS18019	HPV16	744.52	93.95	7.92	CGI
GS18020	HPV35	208.82	110.46	1.89	Illum.
GS18021	HPV16	42.1	98.92	0.43	Illum.
GS18022	HPV16	52.52	102.86	0.51	CGI
GS18023	HPV16	3494.15	107.85	32.4	Illum.
GS18026	HPV16	18184.13	99.89	182.05	CGI
GS18027	HPV16	36.33	94.58	0.38	Illum.
GS18028	HPV16	2992.76	103.37	28.95	Illum.
GS18029	HPV16	279.52	92.22	3.03	Illum.
GS18030	HPV16	4115.76	102.61	40.11	Illum.
GS18031	HPV16	652.46	95.17	6.86	Illum.
GS18033	HPV16	251.51	101.65	2.47	Illum.
GS18034	HPV16	1445.21	96.29	15.01	CGI
GS18035	HPV16	26.41	91.1	0.29	CGI
GS18036	HPV16	21.41	113.18	0.19	Illum.
GS18037	HPV16	1421.43	94.3	15.07	CGI
GS18038	HPV16	1121.09	100.66	11.14	Illum.
GS18039	HPV33	179.58	94.93	1.89	Illum.
GS18040	HPV16	926.67	81.58	11.36	CGI
GS18041	HPV16	558.98	100.96	5.54	CGI
GS18043	HPV16	231.19	94.33	2.45	Illum.
GS18044	HPV16	1591.87	99.18	16.05	Illum.
GS18046	HPV16	3928.29	97.55	40.27	Illum.
GS18047	HPV16	9168.65	110.1	83.27	Illum.
GS18048	HPV33	104.12	100.65	1.03	Illum.
GS18049	HPV16	3128.92	111.31	28.11	Illum.

GS18051	HPV16	77.29	91.26	0.85	CGI
GS18052	HPV16	6.74	103.32	0.07	Illum.
GS18053	HPV16	1680.24	94.73	17.74	Illum.
GS18055	HPV16	2652.25	86.84	30.54	CGI
GS18059	HPV16	4849.33	93.91	51.64	CGI
GS18061	HPV16	36.96	89.82	0.41	CGI
GS18062	HPV16	1274.72	129.93	9.81	Illum.
GS18063	HPV16	90.58	101.45	0.89	Illum.
GS18065	HPV16	1642.39	109.05	15.06	Illum.
GS18066	HPV16	483.01	95.93	5.04	Illum.
GS18067	HPV16	2738.02	132.37	20.68	Illum.
GS18068	HPV16	97.9	117.85	0.83	Illum.
GS18069	HPV16	169.1	100.87	1.68	CGI
GS18070	HPV33	919.27	101.44	9.06	CGI
GS18071	HPV16	3251.74	114.2	28.47	Illum.
GS18072	HPV35	59.28	95.35	0.62	CGI
GS18074	HPV16	1069.76	99.03	10.8	Illum.
GS18076	HPV16	2498.92	98.87	25.27	CGI
GS18077	HPV35	3991.91	120.13	33.23	Illum.
GS18078	HPV16	153.63	103.05	1.49	CGI
GS18079	HPV16	146.24	112.37	1.3	Illum.
GS18081	HPV16	778.46	99.29	7.84	Illum.
GS18082	HPV16	3294.09	97.72	33.71	CGI
GS18085	HPV59	999.06	100.03	9.99	Illum.
GS18087	HPV16	833.84	90.5	9.21	CGI
GS18088	HPV16	439.58	98.05	4.48	CGI
GS18091	HPV16	923.98	89.63	10.31	CGI
GS18092	HPV16	554.17	94.89	5.84	CGI
GS18093	HPV16	847.8	110.83	7.65	Illum.
GS18094	HPV16	493.11	88.65	5.56	CGI
GS18095	HPV16	1361.4	96.13	14.16	CGI
GS18096	HPV16	4648.26	89.99	51.66	CGI
GS18097	HPV16	37.63	96.57	0.39	Illum.
GS18098	HPV16	1722.18	105.91	16.26	Illum.
GS18099	HPV16	585.24	108.73	5.38	Illum.
GS18101	HPV16	1185.18	96.28	12.31	CGI
GS18102	HPV33	739.38	97.61	7.58	CGI
GS18103	HPV18	59.06	101.2	0.58	CGI
GS18105	HPV18	54.11	88.51	0.61	CGI
GS18107	HPV69	11.12	89.99	0.12	CGI
GS18108	HPV16	896.56	94.39	9.5	Illum.
GS18109	HPV16	2438.13	81.11	30.06	CGI
TCGA-BA-4077	HPV16	665.15	44.23	15.04	TCGA
TCGA-BA-5153	HPV16	810.28	38.21	21.21	TCGA
TCGA-BB-4225	HPV33	1015.21	57.67	17.6	TCGA
TCGA-CN-4741	HPV16	975.29	45.66	21.36	TCGA
TCGA-CN-5374	HPV16	1052.93	75.2	14	TCGA
TCGA-CR-5249	HPV16	1151.03	66.37	17.34	TCGA
TCGA-CR-5250	HPV16	419.39	74.78	5.61	TCGA

TCGA-CR-6470	HPV16	269.64	42.32	6.37	TCGA
TCGA-CR-6472	HPV16	34.83	44.65	0.78	TCGA
TCGA-CR-6482	HPV16	16618.84	67.31	246.9	TCGA
TCGA-CR-6487	HPV16	562.68	67.31	8.36	TCGA
TCGA-CR-7385	HPV16	620.48	39.24	15.81	TCGA
TCGA-CR-7404	HPV16	3206.81	45.18	70.97	TCGA
TCGA-CV-5442	HPV16	96.1	67	1.43	TCGA
TCGA-CV-6433	HPV16	2252.19	75.09	29.99	TCGA
TCGA-CV-6961	HPV16	205.29	46.22	4.44	TCGA
TCGA-CV-7100	HPV33	1344.14	43.3	31.04	TCGA

Supplemental Table S1I. Viral genome copy number in 103 HPV-positive tumors with WGS data. The HPV copy number in each tumor was estimated using the depth of coverage of aligned reads across the viral genome using WGS data. Shown from *left* to *right* are: specimen ID; HPV type; depth of coverage for the HPV genome; the depth of coverage for the human autosome; the calculated HPV copy number (HPV coverage / autosomal human genome coverage); and data source.

HPV copy number range	Freq	Percent
0.1-0.3	3	2.91%
0.3-1	13	12.62%
1-3	9	8.74%
3-10	23	22.33%
10-30	32	31.07%
30-100	20	19.42%
>=100	3	2.91%
total	103	100.00%

Supplemental Table S1J. Ordinal distribution of estimated HPV copy number in HPV-positive OSCC.
The HPV genome copy numbers per diploid human genome ranged from 0.12 (i.e. GS18107 with HPV69 infection) to 246.9 copies (i.e.TCGA-CR-6482 with HPV16 infection). The median copy number was 11.36 copies per diploid host genome. Of the 103 samples, 16 (15.5%) had <1 copy of HPV genome per diploid host genome. However, after considering estimates for tumor content and partial deletions of the HPV genome (data not shown), most of these tumors had at least 1 copy of HPV.

integration status	CGI		Illum.		TCGA		Total
	no. samples	%total	no. samples	%total	no. samples	%total	
integrated	27	79.4%	36	69.2%	13	76.5%	73.8%
no integration sites observed (episomal)	7	20.6%	16	30.8%	4	23.5%	26.2%
total no. samples	34	100.0%	52	100.0%	17	100.0%	100.0%

Supplemental Table S1K. Detection of HPV integration in HPV-positive tumors using WGS data.
HPV breakpoints were detected using Complete Genomics protocol or Hydra protocol (cf. Akagi, et al., *Genome Research* 2014; PMID 24201445).

Transcript	median expression value (fpkm)	minimum expression value (fpkm)	maximum expression value (fpkm)	no. samples with expression (fpkm > 1)	Fraction expressed samples (%total)
HPV16_A	9.87	0.00	112.99	108	85.7%
HPV16_B	96.11	7.55	524.68	126	100.0%
HPV16_C	0.00	0.00	76.44	21	16.7%
HPV16_D	0.00	0.00	43.02	10	7.9%
HPV16_E	10.27	0.00	137.17	90	71.4%
HPV16_F	0.16	0.00	18.97	49	38.9%
HPV16_G	3.69	0.00	50.05	87	69.0%
HPV16_H	0.00	0.00	13.59	6	4.8%
HPV16_I	0.00	0.00	0.74	0	0.0%
HPV16_J	0.05	0.00	14.57	35	27.8%
HPV16_K	0.02	0.00	31.37	43	34.1%
HPV16_L	1.52	0.00	41.10	72	57.1%
HPV16_M	0.00	0.00	3.28	7	5.6%
HPV16_N	0.00	0.00	8.39	31	24.6%
HPV16_O	2.05	0.00	92.85	77	61.1%
HPV16_P	0.00	0.00	14.79	34	27.0%
HPV16_Q	0.00	0.00	118.75	32	25.4%
HPV16_R	0.00	0.00	195.32	1	0.8%
HPV16_S	0.00	0.00	2.63	1	0.8%
HPV16_T	0.00	0.00	61.68	5	4.0%

Supplemental Table S1L. Summary of expression of HPV16 transcript isoforms in 127 HPV16-positive OSCC.
Expression levels of 20 HPV16 transcripts isoforms detected in RNA-seq data based upon 20 transcript structures as reported by Zheng et al (Frontiers in Bioscience 2006; PMID 16720315) and evaluated using Cufflinks. The 20 HPV16 transcript isoforms were included as human RefSeq transcripts for quantification (cuffquant module) and normalized to expression values using the cuffnorm module. Shown are the median, minimum, and maximum expression values in FPKM (Fragments Per Kilobase of transcript per Million mapped reads) for each transcript. The number of samples with expression of >1 FPKM is also provided. All samples expressed transcript B (as per Zheng et al.), which has coding potential for E6*1, E1^E4, E5, and E7 proteins. Of 127 samples, 115 (90.5%) had detectable levels of one of three transcripts which encode the full-length E6 protein (HPV16_A > 1 FPKM, HPV16_F > 1FPKM, or HPV16_K > 1 FPKM). See also **Supplemental Fig. S1B-D**.

data set	HPV status	no. cases	median SNV rate (SNVs/Mbp)	range of SNV rate	median indel rate (indel/Mbp)	range of indel rate
Ohio CGI-WGS	pos	34	1.42	0.198-20.7	0.02	0-0.119
Ohio Illumina-WGS	pos	52	2.66	0.258-78.6	0.02	0-0.159
TCGA-WGS	pos	17	2.34	0.476-19.7	0.02	0-0.179
TCGA-WES	pos	46	1.74	0.635-19.0	0.13	0-0.516
Ohio CGI-WGS	neg	25	1.21	0.0794-3.91	0.02	0-0.258
Ohio Illumina-WGS	neg	1	3.93	3.93-3.93	0.04	0.0397-0.0397
TCGA-WGS	neg	24	3.59	0.0595-13.3	0.08	0-0.437
TCGA-WES	neg	285	2.68	0.139-79.5	0.18	0-8.037

Supplemental Table S1M. Rates of somatic SNVs and indel variants in exons, stratified by sequencing platforms.
Shown are median (and range) of somatic mutation rates in exons (per Mbp) stratified by sequencing platform for all OSCC samples included in this study. See also **Supplemental Figs. S1L-M**.

(i) data set	HPV status	no. cases	median SNV rate (SNV/Mbp)	range of SNV rate	median indel rate (indel/Mbp)	range of indel rate
Ohio CGI-WGS	neg	25	1.25	0.055-5.12	0.025	0.00097-0.154
Ohio Illumina-WGS	neg	1	5.05	5.05-5.05	0.015	0.015-0.015
TCGA-WGS	neg	24	4.47	0.13-16.9	0.067	0.00-0.38
Ohio CGI-WGS	pos	34	1.18	0.075-15.5	0.021	0.00032-0.109
Ohio Illumina-WGS	pos	52	2.69	0.17-58.4	0.026	0.00-0.14
TCGA-WGS	pos	17	2.01	1.01-15.4	0.026	0.0055-0.205

(ii) rate of variation (no. / Mbp)	HPV-positive	HPV-negative
median SNV rate	2.04	1.92
minimum SNV rate	0.07	0.05
maximum SNV rate	58.37	16.89
median Indel rate	0.02	0.03
minimum Indel rate	0.00	0.00
maximum Indel rate	0.21	0.38
median som var rate	2.06	1.95
min som var rate	0.07	0.05
max som var rate	58.58	17.27

(iii) type of somatic variation	HPV-positive	HPV-negative
number of samples	103	50
total number of SNVs	1332391	508294
median number of SNVs	6317	5929
minimum number of SNVs	232	169
maximum number of SNVs	180708	52278
total number of INDELs	10425	9422
median number of INDELs	74	100.5
minimum number of INDELs	0	0
maximum number of INDELs	636	1183

Supplemental Table S1N. Genome-wide somatic mutation rates and counts in samples with WGS data.
(*i*) Shown are the median and range of somatic mutation rates (per Mpb) in samples with WGS data (N=153), stratified by sequencing platform and virus status. These genome-wide mutation rates included introns and non-genic regions. (*ii*) SNV, indel and somatic mutation rates for (*left*) HPV-positive and (*right)* HPV-negative OSCC with WGS data were calculated genome-wide. (*iii*) Total, median, minimum and maximum counts of SNVs and indels were calculated genome-wide for *(left*) HPV-positive and (*right*) HPV-negative OSCC.

(i) coding consequence	**Ohio CGI-WGS (n = 34)**	**Ohio Illumina-WGS (n = 52)**	**TCGA-WGS (n = 17)**	**TCGA-WES (n = 46)**	**total (no.)**	**% total**
coding_sequence_variant	0	0	0	1	1	0.00%
frameshift_variant	25	39	22	177	263	1.13%
inframe_deletion	2	14	3	50	69	0.30%
inframe_insertion	0	1	0	5	6	0.03%
initiator_codon_variant	4	12	4	7	27	0.12%
missense_variant	1,856	6,337	1,739	4,358	14,290	61.35%
protein_altering_variant	1	0	0	0	1	0.00%
splice_acceptor_variant	16	64	14	61	155	0.67%
splice_donor_variant	14	25	7	26	72	0.31%
splice_region_variant	162	568	142	45	917	3.94%
stop_gained	215	639	146	391	1,391	5.97%
stop_lost	1	5	4	12	22	0.09%
stop_retained_variant	2	16	4	11	33	0.14%
synonymous_variant	803	2,777	674	1,790	6,044	25.95%
total (no.)	3,101	10,497	2,759	6,934	23,291	100.00%
(ii)						
coding_sequence_variant	0	0	0	1	1	0.00%
frameshift_variant	26	43	22	177	268	1.08%
inframe_deletion	3	14	4	50	71	0.29%
inframe_insertion	0	2	0	5	7	0.03%
initiator_codon_variant	4	10	5	7	26	0.10%
missense_variant	1,972	6,850	1,878	4,358	15,058	60.78%
protein_altering_variant	1	0	0	0	1	0.00%
splice_acceptor_variant	20	83	19	61	183	0.74%
splice_donor_variant	15	34	10	26	85	0.34%
splice_region_variant	198	704	192	45	1,139	4.60%
stop_gained	229	689	161	391	1,470	5.93%
stop_lost	2	9	7	12	30	0.12%
stop_retained_variant	2	18	6	11	37	0.15%
synonymous_variant	857	3,018	732	1,790	6,397	25.82%
total (no.)	3,329	11,474	3,036	6,934	24,773	100.00%

Supplemental Table S1O. Predicted functional impacts of somatic variants in exons of HPV-positive tumors.
Shown are somatic variants detected in 149 HPV-positive OSCC T/N pairs. WGS data were filtered to focus on the exome as represented in the SureSelect Human All Exon (i) v5 and (ii) v3 panels. Variant frequencies as shown in **Fig. 1A** included exonized WGS data using the v5 genomic template.

.

(i) coding consequence	Ohio CGI-WGS (n = 25)	Ohio Illumina-WGS (n = 1)	TCGA-WGS (n = 24)	TCGA-WES (n = 285)	total (no.)	% total
coding_sequence_variant	0	0	0	6	6	0.01%
frameshift_variant	27	2	55	1,885	1,969	3.56%
inframe_deletion	5	0	8	433	446	0.81%
inframe_insertion	0	0	1	58	59	0.11%
initiator_codon_variant	0	0	4	48	52	0.09%
missense_variant	718	86	2,106	31,861	34,771	62.89%
protein_altering_variant	0	0	0	0	0	0.00%
splice_acceptor_variant	8	0	30	442	480	0.87%
splice_donor_variant	7	2	20	348	377	0.68%
splice_region_variant	61	2	172	367	602	1.09%
stop_gained	64	3	147	2,455	2,669	4.83%
stop_lost	0	0	5	37	42	0.08%
stop_retained_variant	0	0	2	37	39	0.07%
synonymous_variant	256	42	779	12,702	13,779	24.92%
total (no.)	1,146	137	3,329	50,679	55,291	100.00%

(ii)

coding_sequence_variant	0	0	0	6	6	0.01%
frameshift_variant	27	2	60	1,885	1,974	3.54%
inframe_deletion	7	0	10	433	450	0.81%
inframe_insertion	0	0	2	58	60	0.11%
initiator_codon_variant	0	0	4	48	52	0.09%
missense_variant	776	89	2,281	31,861	35,007	62.78%
splice_acceptor_variant	12	0	35	442	489	0.88%
splice_donor_variant	11	2	29	348	390	0.70%
splice_region_variant	74	3	217	367	661	1.19%
stop_gained	71	3	166	2,455	2,695	4.83%
stop_lost	1	0	6	37	44	0.08%
stop_retained_variant	1	0	4	37	42	0.08%
synonymous_variant	289	42	854	12,702	13,887	24.91%
total (no.)	1,269	141	3,668	50,679	55,757	100.00%

Supplemental Table S1P. Predicted functional impacts of somatic variants in exons of HPV-negative tumors.
Shown are somatic coding variants detected in 335 HPV-negative OSCC samples including 50 samples with WGS data filtered to focus on the exome as represented in the SureSelect Human All Exon (i) v5 and (ii) v3 panels. Variant frequencies as shown in **Supplemental Fig. S1A** included exonized WGS data using the v5 genomic template.

.

data set	HPV status	no. samples	no. variants covered by RNA-seq	no. variants confirmed by RNA-seq	confirmation rate (%)
Ohio CGI-WGS	neg	25	512	439	85.7%
Ohio Illumina-WGS	neg	1	65	52	80.0%
TCGA-WGS	neg	24	1,170	1,002	85.6%
Ohio CGI-WGS	pos	33	1,562	1,304	83.5%
Ohio Illumina-WGS	pos	51	5,275	4,329	82.1%
TCGA-WGS	pos	17	1,205	1,047	86.9%
TCGA-WES	pos	46	2,765	2,006	72.5%
Total		197	12,554	10,179	81.1%

Supplemental Table S1Q. Confirmation of somatic variants by RNA-seq.
Somatic variants were identified and counted from WGS or WES data. Data were stratified by sequencing platform and HPV status. Variants called by the CGI pipeline were analyzed further when somatic quality score was ≥0. For Ohio cohort Illumina (Illum.) WGS data, somatic variants had to be detected by two or more somatic variant callers. For TGCA data, variants were called from WES data from the BAM files. SNVs were considered verified by analysis of RNA-seq data if the position of the variant had ≥ 20 aligned reads with a minimum number of 4 reads supporting the somatic allele.

HPV status	no. SNVs in target region	no. SNVs detected	no. SNVs not detected	no. SNVs not covered	%confirmed
HPV-negative	30	30	0	0	100.0%
HPV-positive	120	108	8	4	90.0%
Total	150	138	8	4	92.0%

Supplemental Table S1R. Confirmation of somatic variants by targeted re-sequencing. Recurrent SNVs detected in samples sequenced by CGI WGS helped to guide our design of Agilent SureSelect hybrid capture probes for targeted re-sequencing. A panel of 61 genes was selected for HPV-positive tumors and 58 genes for HPV-negative. Eight HPV-negative and 16 HPV-positive OSCC samples sequenced at CGI were studied in this experiment.

gene	no. mutated samples	no. non-mutated samples	mean $\log_{10}$(SNV rate) in mutated sample	mean $\log_{10}$(SNV rate) in non-mutated sample	fold increase	P-value	FDR adjusted p-value
DLG1	4	145	1.45	0.33	13.19	2.12E-03	3.70E-02
POLE	5	144	1.37	0.32	11.19	7.51E-04	1.96E-02
FREM3	3	146	1.34	0.34	10.03	4.28E-08	7.81E-06
KANSL1	4	145	1.33	0.33	9.83	7.78E-09	1.82E-06
CECR2	3	146	1.33	0.34	9.74	1.33E-06	1.63E-04
TAS1R3	6	143	1.30	0.32	9.48	3.64E-04	1.28E-02
OR8S1	3	146	1.32	0.34	9.44	2.63E-05	1.66E-03
EHMT2	5	144	1.30	0.33	9.42	2.27E-03	3.88E-02
OTOF	5	144	1.28	0.33	8.94	2.97E-03	4.60E-02
ZNF707	6	143	1.27	0.32	8.90	1.04E-03	2.50E-02
ANKRD62	3	146	1.28	0.34	8.75	3.16E-04	1.15E-02
UTRN	5	144	1.26	0.33	8.63	3.63E-03	4.95E-02
FANCA	5	144	1.26	0.33	8.59	4.27E-03	4.99E-02
HSPG2	5	144	1.26	0.33	8.57	3.90E-03	4.96E-02
VRTN	5	144	1.26	0.33	8.53	3.76E-03	4.96E-02
TOPORS	3	146	1.27	0.34	8.51	3.50E-28	5.74E-25
LTB4R2	3	146	1.27	0.34	8.46	2.63E-04	9.82E-03
MCM4	3	146	1.27	0.34	8.45	8.44E-04	2.16E-02
KLHL12	3	146	1.27	0.34	8.44	5.41E-04	1.64E-02
ABLIM1	3	146	1.25	0.34	8.15	3.09E-03	4.60E-02
GABRA3	3	146	1.25	0.34	8.15	3.09E-03	4.60E-02
IFT172	3	146	1.25	0.34	8.15	3.09E-03	4.60E-02
CREBBP	6	143	1.23	0.32	8.14	6.41E-04	1.81E-02
FKBP15	4	145	1.25	0.34	8.13	1.39E-06	1.63E-04
ANK3	6	143	1.23	0.32	8.02	1.17E-03	2.67E-02
MTAP	4	145	1.23	0.34	7.90	9.47E-05	4.86E-03
HELZ	5	144	1.23	0.33	7.88	3.95E-03	4.96E-02
PTX3	3	146	1.24	0.34	7.85	2.56E-08	5.26E-06
NUP37	3	146	1.24	0.34	7.83	3.88E-05	2.14E-03
EPCAM	4	145	1.23	0.34	7.78	4.20E-16	2.30E-13
RP1L1	5	144	1.21	0.33	7.64	4.18E-03	4.96E-02
VPS13D	7	142	1.20	0.32	7.57	4.03E-04	1.28E-02
PIGR	3	146	1.22	0.34	7.52	3.57E-03	4.95E-02
DNAH17	6	143	1.20	0.32	7.49	5.78E-06	4.12E-04
PCDHGB2	3	146	1.21	0.34	7.46	3.72E-05	2.14E-03
MAP1A	4	145	1.20	0.34	7.28	4.55E-06	3.39E-04
PUM1	4	145	1.19	0.34	7.20	1.97E-03	3.64E-02
COL24A1	3	146	1.19	0.34	7.12	3.77E-04	1.28E-02
ARFGEF1	5	144	1.18	0.33	7.12	3.91E-05	2.14E-03
PTPRQ	6	143	1.18	0.33	7.07	2.51E-03	4.15E-02
TRIP11	4	145	1.18	0.34	6.94	3.65E-03	4.95E-02
RBM25	4	145	1.18	0.34	6.94	3.86E-06	3.36E-04
DNAH8	7	142	1.16	0.32	6.93	1.39E-05	9.14E-04
HRC	6	143	1.16	0.33	6.89	5.89E-07	8.05E-05

NEK1	3	146	1.18	0.34	6.86	2.40E-03	4.02E-02
PHIP	4	145	1.17	0.34	6.84	1.14E-03	2.67E-02
ANKRD12	3	146	1.17	0.34	6.67	3.27E-03	4.76E-02
DYNC2H1	3	146	1.17	0.34	6.66	1.07E-03	2.54E-02
NBEA	7	142	1.14	0.32	6.62	1.30E-03	2.81E-02
CEP350	9	140	1.13	0.31	6.60	1.76E-04	7.62E-03
ANKRD26	5	144	1.15	0.33	6.53	2.57E-06	2.81E-04
ACIN1	3	146	1.16	0.34	6.52	1.38E-03	2.86E-02
ZFHX2	4	145	1.15	0.34	6.44	1.74E-03	3.45E-02
SAMSN1	4	145	1.14	0.34	6.41	2.17E-03	3.75E-02
ZNF547	3	146	1.15	0.34	6.35	4.02E-04	1.28E-02
GATAD2B	4	145	1.14	0.34	6.34	1.98E-04	8.13E-03
KMT2C	8	141	1.12	0.32	6.31	7.46E-04	1.96E-02
ITGB6	4	145	1.12	0.34	6.09	4.06E-04	1.28E-02
SLC8A2	4	145	1.12	0.34	6.06	1.83E-03	3.53E-02
IREB2	4	145	1.12	0.34	5.98	2.96E-03	4.60E-02
SBNO1	6	143	1.10	0.33	5.92	4.02E-03	4.96E-02
PPP1R12A	5	144	1.10	0.33	5.86	1.17E-03	2.67E-02
TIAM2	6	143	1.10	0.33	5.85	4.11E-03	4.96E-02
TLK2	3	146	1.11	0.34	5.82	2.88E-03	4.60E-02
TSR3	3	146	1.10	0.34	5.74	9.18E-04	2.32E-02
E2F7	4	145	1.10	0.34	5.73	1.22E-03	2.71E-02
MDC1	3	146	1.10	0.34	5.68	1.92E-04	8.06E-03
BIRC6	4	145	1.09	0.34	5.56	1.37E-03	2.86E-02
IFNGR1	4	145	1.08	0.34	5.53	1.51E-04	6.88E-03
FLG	13	136	1.04	0.30	5.52	1.43E-04	6.80E-03
MACF1	9	140	1.05	0.32	5.47	1.76E-04	7.62E-03
DCDC1	5	144	1.07	0.34	5.37	6.20E-04	1.78E-02
MPHOSPH8	3	146	1.07	0.35	5.34	3.86E-03	4.96E-02
RAPGEF3	7	142	1.05	0.33	5.30	7.79E-05	4.12E-03
RBL2	3	146	1.07	0.35	5.28	7.38E-04	1.96E-02
HERC1	8	141	1.04	0.32	5.26	3.83E-03	4.96E-02
ZNF33A	3	146	1.04	0.35	4.99	1.27E-03	2.77E-02
TRIP10	3	146	1.04	0.35	4.98	3.04E-03	4.60E-02
CACNA1A	9	140	1.01	0.32	4.97	7.75E-06	5.30E-04
TPR	4	145	1.04	0.34	4.94	4.09E-03	4.96E-02
MTUS1	3	146	1.04	0.35	4.91	2.12E-03	3.70E-02
SEC24C	5	144	1.02	0.34	4.78	3.90E-03	4.96E-02
MAST4	3	146	1.02	0.35	4.77	4.16E-03	4.96E-02
HMCN1	10	139	0.98	0.31	4.65	1.22E-03	2.71E-02
C6	7	142	1.00	0.33	4.65	1.98E-03	3.64E-02
DDX3X	8	141	0.99	0.32	4.64	1.45E-04	6.80E-03
VPS13C	6	143	1.00	0.33	4.61	1.22E-04	6.07E-03
PIBF1	3	146	1.00	0.35	4.52	1.49E-03	3.01E-02
RBL1	9	140	0.97	0.32	4.42	2.59E-03	4.24E-02
TTC6	8	141	0.97	0.33	4.40	3.18E-03	4.65E-02
ZNF124	3	146	0.99	0.35	4.39	3.46E-05	2.10E-03
DDX27	4	145	0.98	0.34	4.35	5.72E-04	1.71E-02
ZNF750	21	128	0.90	0.27	4.31	3.24E-06	3.13E-04

C9orf84	3	146	0.98	0.35	4.29	5.77E-11	2.37E-08
ATAD5	6	143	0.96	0.33	4.24	3.79E-03	4.96E-02
GPR98	8	141	0.95	0.33	4.22	2.21E-04	8.62E-03
MYO5A	6	143	0.95	0.33	4.17	1.89E-03	3.58E-02
DHX29	6	143	0.95	0.33	4.16	2.95E-03	4.60E-02
SCN3A	3	146	0.96	0.35	4.13	2.56E-04	9.77E-03
HERC2	11	138	0.93	0.31	4.10	1.44E-03	2.94E-02
ASTN1	4	145	0.96	0.34	4.10	3.99E-03	4.96E-02
AHNAK2	11	138	0.92	0.31	4.04	1.70E-03	3.40E-02
WIZ	4	145	0.95	0.34	4.04	1.02E-03	2.49E-02
RASGRF1	5	144	0.94	0.34	3.99	3.65E-03	4.95E-02
CTC-432M15.3	3	146	0.95	0.35	3.98	4.32E-03	4.99E-02
DYNC1H1	4	145	0.94	0.34	3.90	1.37E-03	2.86E-02
DSP	7	142	0.92	0.33	3.89	2.13E-04	8.51E-03
PKHD1L1	9	140	0.91	0.32	3.83	4.38E-04	1.36E-02
WIPF2	4	145	0.93	0.34	3.82	3.61E-03	4.95E-02
CCDC73	3	146	0.93	0.35	3.80	3.94E-06	3.36E-04
STAT2	5	144	0.92	0.34	3.76	3.43E-03	4.89E-02
MAP3K11	3	146	0.92	0.35	3.76	1.90E-03	3.58E-02
SOCS5	3	146	0.92	0.35	3.76	3.02E-03	4.60E-02
NSD1	10	139	0.89	0.32	3.73	1.00E-03	2.49E-02
LAMA3	8	141	0.90	0.33	3.72	7.12E-04	1.96E-02
COL4A3BP	4	145	0.91	0.34	3.71	5.07E-10	1.39E-07
MUC16	24	125	0.83	0.27	3.65	2.86E-06	2.93E-04
PLCB4	4	145	0.90	0.34	3.58	1.48E-10	4.87E-08
CUL1	8	141	0.88	0.33	3.54	3.69E-04	1.28E-02
ZNF354C	3	146	0.89	0.35	3.44	4.10E-06	3.36E-04
ERCC6	3	146	0.88	0.35	3.43	1.01E-17	8.27E-15
TESK1	3	146	0.88	0.35	3.38	2.11E-03	3.70E-02
COG1	4	145	0.87	0.35	3.32	2.75E-03	4.46E-02
LOXHD1	3	146	0.87	0.35	3.30	4.49E-06	3.39E-04
FSIP2	12	137	0.83	0.32	3.26	3.98E-04	1.28E-02
RYR2	13	136	0.82	0.32	3.23	3.31E-03	4.76E-02
ZNF483	3	146	0.85	0.35	3.20	3.12E-03	4.62E-02
CASZ1	14	135	0.81	0.31	3.12	7.48E-04	1.96E-02
TMCC1	5	144	0.84	0.34	3.11	4.20E-03	4.96E-02
BNC2	3	146	0.84	0.35	3.08	4.06E-03	4.96E-02
PTPN13	11	138	0.81	0.32	3.05	1.77E-03	3.46E-02
ZNF451	3	146	0.83	0.35	3.01	4.06E-03	4.96E-02
SYNE2	8	141	0.80	0.33	2.89	3.61E-03	4.95E-02
TTN	49	100	0.65	0.22	2.70	1.03E-07	1.69E-05
UBXN4	3	146	0.77	0.35	2.65	2.31E-03	3.90E-02
NAV3	3	146	0.77	0.35	2.61	3.92E-03	4.96E-02
BOD1L1	4	145	0.76	0.35	2.56	4.00E-03	4.96E-02
ASXL3	10	139	0.72	0.33	2.43	4.29E-03	4.99E-02
PIK3CA	42	107	0.63	0.25	2.38	5.60E-07	8.05E-05
CSMD3	17	132	0.67	0.32	2.24	2.11E-03	3.70E-02
FGFR3	17	132	0.67	0.32	2.23	6.13E-04	1.78E-02

KMT2D	20	129	0.65	0.31	2.16	2.04E-03	3.70E-02

Supplemental Table S1S. Genes frequently mutated in tumors with high SNV rate. Shown are 142 genes which when mutated are associated with a significantly higher overall mutation rate (variants per Mbp) than observed in tumors with wildtype alleles. Analysis was limited to genes disrupted by coding-change SNVs in three or more tumors. Listed from *left* to *right* are: gene symbols; number of mutated samples; number of non-mutated samples; mean SNV rates ($\log_{10}$ transformed) in samples with indicated gene mutated: mean SNV rates (log 10 transformed) in samples with indicated gene wildtype; fold increase of SNV rate in samples with gene mutated vs. wildtype; p-values, t-test; and FDR-adjusted p-values. Genes are sorted by fold increase in mutation rate. P-values and adjusted p-values are presented in scientific notation; E represents *10^. See also **Supplemental Fig. S1Q.**

GO ID	GO term	no. sample with mutation in high mutation group	no. sample with WT genes in high mutation group	no. high mutation group sample	Fraction in high mutation group	no. sample in low mutation group	no. sample with WT genes in low mutation group	no. low mutation group sample	Fraction in low mutation group	P-value	adjusted p-value	sim p-value
GO:0030308	negative regulation of cell growth	16	58	74	21.62%	0	75	75	0.00%	5.49E-06	8.56E-04	0.003
GO:0060333	interferon-gamma-mediated signaling pathway	22	52	74	29.73%	1	74	75	1.33%	4.73E-07	1.25E-04	0.009
GO:0030307	positive regulation of cell growth	12	62	74	16.22%	0	75	75	0.00%	1.38E-04	7.32E-03	0.011
GO:0006977	DNA damage response, signal transduction by p53 class mediator resulting in cell cycle arrest	12	62	74	16.22%	0	75	75	0.00%	1.38E-04	7.32E-03	0.011
GO:0008544	epidermis development	23	51	74	31.08%	2	73	75	2.67%	1.61E-06	3.19E-04	0.019
GO:0006351	transcription, DNA-templated	63	11	74	85.14%	26	49	75	34.67%	2.35E-10	3.73E-07	0.021
GO:0051607	defense response to virus	20	54	74	27.03%	1	74	75	1.33%	2.47E-06	4.36E-04	0.023
GO:0046777	protein autophosphorylation	21	53	74	28.38%	2	73	75	2.67%	7.86E-06	9.59E-04	0.051
GO:0009615	response to virus	17	57	74	22.97%	1	74	75	1.33%	2.69E-05	2.37E-03	0.059
GO:1901796	regulation of signal transduction by p53 class mediator	16	58	74	21.62%	1	74	75	1.33%	5.82E-05	4.01E-03	0.061
GO:0007229	integrin-mediated signaling pathway	15	59	74	20.27%	1	74	75	1.33%	1.24E-04	7.04E-03	0.069
GO:0006366	transcription from RNA polymerase II promoter	40	34	74	54.05%	9	66	75	12.00%	3.51E-08	2.25E-05	0.075
GO:0002474	antigen processing and presentation of peptide antigen via MHC class I	15	59	74	20.27%	1	74	75	1.33%	1.24E-04	7.04E-03	0.076
GO:0006260	DNA replication	15	59	74	20.27%	1	74	75	1.33%	1.24E-04	7.04E-03	0.085
GO:0007155	cell adhesion	39	35	74	52.70%	9	66	75	12.00%	8.36E-08	2.90E-05	0.094

GO:0051056	regulation of small GTPase mediated signal transduction	18	56	74	24.32%	2	73	75	2.67%	7.64E-05	4.90E-03	0.103
GO:0006281	DNA repair	22	52	74	29.73%	3	72	75	4.00%	1.87E-05	1.85E-03	0.111
GO:0042981	regulation of apoptotic process	26	48	74	35.14%	5	70	75	6.67%	1.55E-05	1.75E-03	0.111
GO:0006888	ER to Golgi vesicle-mediated transport	17	57	74	22.97%	2	73	75	2.67%	1.58E-04	7.86E-03	0.132
GO:0007605	sensory perception of sound	21	53	74	28.38%	3	72	75	4.00%	3.92E-05	2.96E-03	0.143
GO:0030198	extracellular matrix organization	21	53	74	28.38%	3	72	75	4.00%	3.92E-05	2.96E-03	0.144
GO:0008360	regulation of cell shape	17	57	74	22.97%	2	73	75	2.67%	1.58E-04	7.86E-03	0.145
GO:0006974	cellular response to DNA damage stimulus	27	47	74	36.49%	6	69	75	8.00%	2.57E-05	2.37E-03	0.15
GO:0045892	negative regulation of transcription, DNA-templated	37	37	74	50.00%	10	65	75	13.33%	1.39E-06	3.15E-04	0.158
GO:0006355	regulation of transcription, DNA-templated	53	21	74	71.62%	20	55	75	26.67%	4.26E-08	2.25E-05	0.166

Supplemental Table S1T. Gene Ontology terms enriched in highly mutated HPV-positive tumors.
Shown are gene ontology terms enriched among all genes found to be more frequently mutated in tumors with a high versus low mutation rate (≥ vs. < the median mutation rate). Analysis was limited to genes with high impact coding change mutations (i.e. stop gain, frameshift, splice donor, splice acceptor mutations) detected in three or more tumors. Fisher’s exact test for a comparison of the number of tumors in the high vs. low group with mutations in genes included in the GO term is shown. p-values were adjusted for multiple testing correction by FDR method. To account for the higher probability of mutations in tumors with a high mutation rates, a permutation analysis was performed. Shown in the far-right column is the p-value from a permutation analysis in which high impact variants were randomly distributed 1000 times across all tumors. The simulation p-value indicates the probability that the p-value for the GO term in the simulated analysis is less than the p-value of the observed data. P-values and adjusted P-values are presented in scientific notation; E represents *10^.

gene	no. mutated sample	no. non-mutated sample	mean $\log_{10}$(SNV rate) in mutated sample	mean $\log_{10}$(SNV rate) in non-mutated sample	fold change	P-value	adjusted p-value	
POLE	5	144	1.37	0.32	11.19	7.51E-04	1.43E-02	*
HUWE1	11	138	0.87	0.32	3.59	4.85E-03	4.61E-02	*
NIPBL	8	141	0.77	0.34	2.69	1.02E-02	6.48E-02	
MTOR	5	144	0.96	0.34	4.22	1.69E-02	6.54E-02	
TRRAP	10	139	0.84	0.33	3.23	1.82E-02	6.54E-02	
PRKDC	6	143	0.82	0.34	3.04	2.07E-02	6.54E-02	
ATR	4	145	1.13	0.34	6.18	3.04E-02	7.05E-02	
TIMELESS	3	146	1.40	0.34	11.46	3.31E-02	7.05E-02	
PARP1	3	146	1.39	0.34	11.16	3.34E-02	7.05E-02	
UBR5	7	142	0.65	0.35	2.00	4.27E-02	8.12E-02	
TRIP12	6	143	0.73	0.34	2.45	5.65E-02	9.02E-02	
USP47	3	146	1.27	0.34	8.41	6.02E-02	9.02E-02	
SMG1	6	143	0.92	0.34	3.81	6.17E-02	9.02E-02	
REV3L	4	145	1.04	0.34	5.00	6.90E-02	9.37E-02	
FAN1	3	146	1.14	0.34	6.24	1.25E-01	1.58E-01	
PNKP	3	146	0.81	0.35	2.88	1.36E-01	1.61E-01	
SLX4	4	145	0.61	0.35	1.83	1.64E-01	1.83E-01	
DOT1L	3	146	0.87	0.35	3.34	2.08E-01	2.20E-01	
PDS5A	3	146	0.48	0.36	1.33	2.86E-01	2.86E-01	

Supplemental Table S1U. Association between mutations in DNA repair genes (GO Central) and high overall SNV rate. A list of 167 genes with annotated ontology term of DNA repair (GO:0006281) was downloaded from GO Central. Of those, 19 genes were mutated in 3 or more HPV-positive OSCC samples (*shown here*). Overall SNV rates in cancers with mutations in each of the 19 genes were compared with those in samples without mutations, using t-test (one-tailed test). P-values were adjusted for multiple testing correction using FDR method (*asterisk*, adj. $p < 0.05$). From *left* to *right* are: gene symbols; number of mutated samples; number of non-mutated samples; mean SNV rates ($\log_{10}$ transformed) in samples with mutation in indicated gene; mean SNV rates (log 10 transformed) in samples with wildtype allele in indicated gene; fold increase of SNV rate in samples with gene mutated vs. wildtype; p-values, t-test; and FDR-adjusted p-values. Listed genes are sorted by adjusted p-value. P-values and adjusted P-values are presented in scientific notation; E represents *10^.

gene	no. mutated sample	no. non-mutated sample	mean $\log_{10}$(SNV rate) in mutated sample	mean $\log_{10}$(SNV rate) in non-mutated sample	fold change	P-value	FDR adjusted p-value	
ERCC6	3	146	0.88	0.35	3.43	1.01E-17	5.84E-16	*
MDC1	3	146	1.10	0.34	5.68	1.92E-04	5.55E-03	*
POLE	5	144	1.37	0.32	11.19	7.51E-04	1.45E-02	*
HERC2	11	138	0.93	0.31	4.10	1.44E-03	2.08E-02	*
BOD1L1	4	145	0.76	0.35	2.56	4.00E-03	4.02E-02	*
FANCA	5	144	1.26	0.33	8.59	4.27E-03	4.02E-02	*
HUWE1	11	138	0.87	0.32	3.59	4.85E-03	4.02E-02	*
INO80	3	146	1.15	0.34	6.41	5.99E-03	4.34E-02	*
NIPBL	8	141	0.77	0.34	2.69	1.02E-02	6.59E-02	
SPIDR	4	145	1.04	0.34	4.95	1.58E-02	8.12E-02	
NCOA6	5	144	0.98	0.34	4.37	1.66E-02	8.12E-02	
MTOR	5	144	0.96	0.34	4.22	1.69E-02	8.12E-02	
TRRAP	10	139	0.84	0.33	3.23	1.82E-02	8.12E-02	
PRKDC	6	143	0.82	0.34	3.04	2.07E-02	8.56E-02	
ATR	4	145	1.13	0.34	6.18	3.04E-02	1.00E-01	
TOPBP1	3	146	0.94	0.35	3.88	3.07E-02	1.00E-01	
SUPT16H	4	145	1.16	0.34	6.70	3.19E-02	1.00E-01	
TIMELESS	3	146	1.40	0.34	11.46	3.31E-02	1.00E-01	
PARP1	3	146	1.39	0.34	11.16	3.34E-02	1.00E-01	
SMARCAL1	3	146	0.85	0.35	3.13	3.46E-02	1.00E-01	
RNF111	5	144	0.83	0.34	3.07	4.02E-02	1.06E-01	
UBR5	7	142	0.65	0.35	2.00	4.27E-02	1.06E-01	
INTS3	3	146	0.83	0.35	3.06	4.34E-02	1.06E-01	
TNKS1BP1	5	144	0.93	0.34	3.87	4.37E-02	1.06E-01	
RIF1	4	145	1.15	0.34	6.52	4.86E-02	1.08E-01	
TICRR	4	145	0.96	0.34	4.15	4.96E-02	1.08E-01	
DNA2	4	145	1.12	0.34	6.02	5.03E-02	1.08E-01	
TRIP12	6	143	0.73	0.34	2.45	5.65E-02	1.12E-01	
DHX9	3	146	1.28	0.34	8.72	5.92E-02	1.12E-01	
USP47	3	146	1.27	0.34	8.41	6.02E-02	1.12E-01	
TP53BP1	3	146	1.22	0.34	7.60	6.13E-02	1.12E-01	
SMG1	6	143	0.92	0.34	3.81	6.17E-02	1.12E-01	
REV3L	4	145	1.04	0.34	5.00	6.90E-02	1.20E-01	
PARP9	3	146	1.19	0.34	6.99	7.26E-02	1.20E-01	

HELQ	3	146	1.26	0.34	8.35	7.35E-02	1.20E-01
POLG	3	146	0.74	0.35	2.46	7.60E-02	1.20E-01
SHPRH	4	145	0.94	0.34	3.97	7.68E-02	1.20E-01
ESCO2	3	146	0.84	0.35	3.11	9.07E-02	1.38E-01
WDR33	3	146	0.70	0.35	2.23	9.51E-02	1.41E-01
UPF1	3	146	0.94	0.35	3.92	1.01E-01	1.46E-01
TP73	3	146	0.81	0.35	2.87	1.12E-01	1.52E-01
ERCC6L2	4	145	0.98	0.34	4.39	1.12E-01	1.52E-01
CDC5L	3	146	1.08	0.34	5.42	1.13E-01	1.52E-01
BRIP1	3	146	0.86	0.35	3.26	1.15E-01	1.52E-01
FAN1	3	146	1.14	0.34	6.24	1.25E-01	1.61E-01
PNKP	3	146	0.81	0.35	2.88	1.36E-01	1.71E-01
FANCM	4	145	0.76	0.35	2.57	1.49E-01	1.84E-01
EP300	18	131	0.48	0.34	1.38	1.58E-01	1.90E-01
SLX4	4	145	0.61	0.35	1.83	1.64E-01	1.94E-01
POLB	4	145	0.86	0.35	3.29	1.72E-01	2.00E-01
PARP4	4	145	0.59	0.35	1.71	1.81E-01	2.06E-01
DOT1L	3	146	0.87	0.35	3.34	2.08E-01	2.32E-01
USP28	3	146	0.86	0.35	3.23	2.39E-01	2.61E-01
ZFYVE26	5	144	0.52	0.35	1.45	2.65E-01	2.84E-01
POLR2B	3	146	0.74	0.35	2.47	2.83E-01	2.96E-01
PDS5A	3	146	0.48	0.36	1.33	2.86E-01	2.96E-01
TEX15	3	146	0.43	0.36	1.19	4.25E-01	4.32E-01
ASF1A	3	146	0.38	0.36	1.04	4.77E-01	4.77E-01

Supplemental Table S1V. Association between mutations in DNA repair genes (AmiGO) and high overall SNV rate. A list of 559 DNA repair genes (GO:0006281) was downloaded from AmiGO (http://amigo.geneontology.org/amigo), a database combining ontology annotations from databases including GO central, Reactome, and Ensembl. Of those, 58 genes were mutated in 3 or more samples (*shown here*). Overall SNV rates in cancers with mutations in each of the 58 genes were compared with those in samples without mutations, using t-test (one-tailed test). P-values were adjusted for multiple testing correction using FDR method (*asterisk*, adj. $p < 0.05$). From *left* to *right* are: gene symbols; number of mutated samples; number of non-mutated samples; mean SNV rates ($\log_{10}$ transformed) in samples with mutation in indicated gene; mean SNV rates ($\log_{10}$ transformed) in samples with wildtype allele in indicated gene; fold increase of SNV rate in samples with gene mutated vs. wildtype; p-values, t-test; and FDR-adjusted p-values. P-values and adjusted P-values are presented in scientific notation; E represents *10^.

pattern	mean fraction of HPV-positive samples (%total)	mean fraction of HPV-negative samples (%total)	difference	p-value	FDR adjusted p-value	
C>A	10.10	13.38	-3.28	1.53E-10	4.60E-10	*
C>G	16.81	15.25	1.55	1.24E-01	1.24E-01	
C>T	59.15	51.04	8.12	1.23E-16	7.39E-16	*
T>A	3.45	5.47	-2.02	3.69E-10	7.38E-10	*
T>C	8.18	11.47	-3.29	3.15E-07	4.73E-07	*
T>G	2.31	3.40	-1.09	4.82E-05	5.79E-05	*

Supplemental Table S2A. Fraction of six nucleotide substitution patterns genome-wide in HPV-positive and HPV-negative OSCC tumors.
Shown are mean fractions of the six possible SNVs (i.e. C>T, C>G, C>A, T>G, T>C, and T>A) identified genome-wide in 103 HPV-positive and 50 HPV-negative tumors. P-values were calculated by t-test (see also **Supplemental Fig. S2A-D**). P-values and adjusted P-values are presented in scientific notation; E represents *10^.

pattern	mean fraction of HPV-positive samples (%total)	mean fraction of HPV-negative samples (%total)	difference	p-value	FDR adjusted p-value	
C>A	12.6	18.28	-5.68	5.72E-13	1.72E-12	*
C>G	18.58	14.81	3.77	4.58E-03	4.58E-03	*
C>T	49.04	36.44	12.6	3.00E-16	1.80E-15	*
T>A	5.28	9.67	-4.39	4.13E-12	8.27E-12	*
T>C	11.15	15.25	-4.1	4.94E-05	5.92E-05	*
T>G	3.35	5.54	-2.2	8.37E-12	1.26E-11	*

Supplemental Table S2B. Fraction of six nucleotide substitution patterns in exons in HPV-positive and HPV-negative OSCC tumors.
Shown are mean fractions of the six possible SNVs (i.e. C>T, C>G, C>A, T>G, T>C, and T>A) identified in exonic regions of 149 HPV-positive and 335 HPV-negative tumors. P-values were calculated by t-test (see also **Supplemental Fig. S2A-D**). P-values and adjusted P-values are presented in scientific notation; E represents *10^.

96 patterns of tri-nucleotide substitute patterns	mean fraction of HPV-positive samples (%total)	mean fraction of HPV-negative samples (%total)	difference	P-value	FDR adjusted p-value	
C>T in T*A	13.39	5.49	7.90	2.22E-14	7.10E-13	*
C>T in T*T	7.91	4.17	3.75	5.46E-14	1.31E-12	*
C>T in T*G	3.56	2.28	1.28	1.44E-12	1.98E-11	*
T>C in A*A	1.06	2.49	-1.43	2.06E-11	1.97E-10	*
C>G in T*A	6.17	3.68	2.49	6.70E-05	1.05E-04	*
C>G in T*T	7.48	4.97	2.51	2.27E-04	3.30E-04	*
T>C in T*G	1.56	0.56	1.00	2.43E-03	3.07E-03	*

Supplemental Table S2C. Fractions of 96 tri-nucleotide substitution patterns genome-wide in HPV-positive and HPV-negative tumors.
The distribution of fractions of 96 tri-nucleotide substitution patterns genome-wide was compared in 103 HPV-positive vs. 50 HPV-negative OSCC using t-test. Eight-six of 96 patterns were significantly different in HPV-positive vs. HPV-negative tumors (adjusted p-value < 0.05). However, fractions differed in the two groups by ≤1% for the majority. Only the 7 patterns shown here demonstrated an increase or decrease in fraction of >1%. Four patterns found to be increased significantly in HPV-positive tumors were C>T or C>G mutations in the 5'-TCW tri-nucleotide context, consistent with APOBEC-associated signature mutations. On average, 34.95% of SNVs in HPV-positive tumors were APOBEC signature mutations. In contrast, 18.30% of SNVs in HPV-negative tumors were APOBEC signature mutations (see also **Supplemental Fig. S2C**). *Asterisk*, adjusted p-value by FDR method < 0.05. P-values and adjusted p-values are presented in scientific notation; E represents *10^.

96 tri-nucleotide substitution pattern	mean fraction of HPV-positive samples (%total)	mean fraction of HPV-negative samples (%total)	Difference	P-value	adjusted p-value	
C>T in T*G	8.01	5.50	2.51	7.87E-12	3.67E-10	*
C>T in T*T	6.70	3.87	2.84	1.15E-11	3.67E-10	*
C>T in T*A	10.85	5.94	4.91	1.74E-11	4.17E-10	*
C>G in T*T	5.97	4.82	1.15	2.22E-02	6.08E-02	
C>G in T*A	4.76	3.89	0.87	4.29E-02	1.03E-01	

Supplemental Table S2D. Fractions of 96 tri-nucleotide substitution patterns in exons in HPV-positive and HPV-negative tumors.
The distribution of fractions of 96 tri-nucleotide substitution patterns in exome data were compared in 149 HPV-positive vs. 335 HPV-negative OSCC using t-test. Thirty-four of 96 patterns were significantly different in HPV-positive vs. HPV-negative tumors (adjusted p-value < 0.05). However, fractions differed in the two groups by ≤1% for the majority. Only 3 patterns shown here demonstrated an increase or decrease in fraction of >1%. Two patterns found to be increased significantly in HPV-positive tumors were C>T or C>G mutations in TCW tri-nucleotide context, consistent with APOBEC signature mutations. On average, 28.29% of SNVs in HPV-positive tumors are APOBEC signature mutations. In contrast, 18.52% of SNVs in HPV-negative tumors are APOBEC signature mutations. *Asterisk*, adjusted p-value by FDR method < 0.05 (see also **Supplemental Fig. S2D**). P-values and adjusted P-values are presented in scientific notation; E represents *10^.

signature	signature fraction					number of signature SNVs				
	mean fraction of HPV-positive samples (%total)	mean fraction of HPV-negative samples (%total)	P-value	FDR-adjusted p-value		mean number of SNVs in HPV-positive samples	mean number of SNVs in HPV-negative samples	P-value	FDR-adjusted p-value	
Sig1	21.83%	16.11%	7.17E-03	1.79E-02	*	1325.21	944.78	4.64E-02	1.02E-01	
Sig2	26.06%	8.11%	2.92E-16	8.76E-15	*	5230.19	1055.00	1.71E-05	1.71E-04	*
Sig3	3.52%	7.25%	2.02E-03	6.06E-03	*	213.78	788.66	4.76E-04	2.04E-03	*
Sig4	0.32%	1.54%	4.98E-02	9.96E-02		25.44	184.14	2.12E-01	3.53E-01	
Sig5	9.26%	10.42%	5.22E-01	5.90E-01		651.19	1162.96	1.80E-01	3.37E-01	
Sig6	0.74%	0.45%	1.56E-01	2.75E-01		45.36	29.42	1.33E-02	3.37E-02	*
Sig7	1.00%	1.59%	1.98E-01	3.30E-01		98.79	115.82	5.77E-01	6.41E-01	
Sig8	4.56%	10.59%	1.99E-05	1.49E-04	*	355.75	1457.52	1.70E-04	1.28E-03	*
Sig9	1.02%	3.00%	6.10E-05	3.66E-04	*	79.28	314.98	1.13E-02	3.37E-02	*
Sig10	1.17%	0.61%	3.95E-04	1.97E-03	*	133.50	70.96	4.99E-02	1.02E-01	
Sig11	1.13%	2.76%	8.13E-07	8.13E-06	*	66.79	299.00	1.85E-07	2.77E-06	*
Sig12	2.85%	0.48%	2.81E-03	7.67E-03	*	216.06	25.08	3.33E-04	2.00E-03	*
Sig13	16.87%	10.78%	1.38E-03	4.60E-03	*	3848.83	1384.46	4.36E-01	5.23E-01	
Sig14	0.19%	0.84%	1.34E-03	4.60E-03	*	13.13	51.12	4.35E-04	2.04E-03	*
Sig15	1.15%	1.64%	2.63E-01	3.95E-01		66.99	66.24	6.89E-01	7.38E-01	
Sig16	1.83%	14.15%	3.36E-08	5.05E-07	*	106.08	1456.94	6.55E-08	1.96E-06	*
Sig17	0.20%	0.12%	6.28E-01	6.60E-01		35.19	6.32	4.24E-01	5.23E-01	
Sig18	3.18%	6.08%	1.38E-03	4.60E-03	*	191.61	456.24	5.46E-04	2.05E-03	*
Sig19	0.48%	0.60%	6.38E-01	6.60E-01		18.58	43.20	1.97E-01	3.48E-01	
Sig20	0.67%	0.01%	1.72E-02	3.97E-02	*	60.23	1.26	1.35E-02	3.37E-02	*
Sig21	0.55%	0.22%	5.61E-02	1.05E-01		42.14	14.18	3.70E-01	5.04E-01	
Sig22	0.01%	0.06%	3.24E-01	4.26E-01		0.53	1.90	3.30E-01	4.85E-01	
Sig23	0.02%	0.08%	4.45E-01	5.56E-01		1.08	8.54	5.63E-01	6.41E-01	
Sig24	0.16%	0.12%	7.14E-01	7.14E-01		17.26	1.10	2.17E-03	7.23E-03	*
Sig25	0.21%	0.72%	2.22E-02	4.76E-02	*	20.90	124.92	5.11E-02	1.02E-01	
Sig26	0.18%	0.31%	4.75E-01	5.71E-01		20.76	18.62	8.09E-01	8.37E-01	
Sig27	0.01%	0.06%	3.27E-01	4.26E-01		0.89	5.18	3.40E-01	4.85E-01	
Sig28	0.14%	0.26%	2.90E-01	4.14E-01		10.43	12.06	2.27E-01	3.58E-01	
Sig29	0.33%	0.59%	2.32E-01	3.67E-01		35.49	24.64	9.97E-01	9.97E-01	
Sig30	0.36%	0.49%	5.31E-01	5.90E-01		40.85	41.46	4.01E-01	5.23E-01	

Supplemental Table S2E. Comparison of mutational signatures in HPV-positive vs. HPV-negative tumors (WGS).
Estimated here are the fraction (*left*) and number (*right*) of mutations in the context of 30 COSMIC mutational signatures as reported by the Sanger Center (https://cancer.sanger.ac.uk/cell_lines/signatures). Shown are statistical comparisons for 103 HPV-positive vs. 50 HPV-negative tumors with WGS data by use of t-test. For the comparison of counts, data were log-transformed, $\log_{10}$(number of SNVs in signature + 1). P-values were adjusted for multiple comparisons (FDR method). *Asterisk*, FDR-adjusted p-value <0.05 (see also **Supplemental Fig. S2E - L)**. P-values and adjusted P-values are presented in scientific notation; E represents *10^.

signature	signature fraction					number of signature SNVs				
	mean fraction of HPV-positive samples (%total)	mean fraction of HPV-negative samples (%total)	P-value	FDR-adjusted p-value		mean number of SNVs in HPV-positive samples	mean number of SNVs in HPV-negative samples	P-value	FDR-adjusted p-value	
Sig1	16.03%	14.43%	2.63E-01	3.93E-01		16.07	17.32	6.00E-02	1.20E-01	
Sig2	25.61%	11.05%	4.48E-15	1.34E-13	*	87.71	22.48	2.70E-07	4.05E-06	*
Sig3	3.92%	5.04%	2.06E-01	3.54E-01		4.92	7.02	2.96E-02	9.25E-02	
Sig4	0.29%	0.92%	2.96E-02	1.78E-01		0.11	2.57	8.60E-04	5.16E-03	*
Sig5	3.78%	5.79%	4.10E-02	2.05E-01		3.94	8.10	7.94E-02	1.49E-01	
Sig6	1.61%	1.33%	5.00E-01	5.77E-01		1.53	2.50	6.62E-01	7.59E-01	
Sig7	3.13%	2.83%	5.47E-01	6.08E-01		6.21	8.97	1.26E-01	2.00E-01	
Sig8	2.25%	2.73%	4.50E-01	5.67E-01		2.47	4.21	1.09E-01	1.92E-01	
Sig9	1.01%	0.71%	4.56E-01	5.67E-01		0.67	0.80	7.33E-01	7.59E-01	
Sig10	1.90%	1.49%	1.85E-01	3.54E-01		3.44	2.03	1.55E-01	2.31E-01	
Sig11	1.36%	1.91%	1.28E-01	3.54E-01		1.52	5.04	1.57E-02	6.08E-02	
Sig12	1.84%	0.92%	9.23E-02	3.46E-01		1.93	1.34	3.41E-01	4.26E-01	
Sig13	15.72%	13.56%	1.69E-01	3.54E-01		66.53	27.68	7.32E-01	7.59E-01	
Sig14	0.24%	0.53%	4.82E-02	2.07E-01		0.27	1.07	4.53E-02	1.13E-01	
Sig15	1.01%	1.12%	6.77E-01	7.25E-01		0.81	1.77	3.90E-02	1.06E-01	
Sig16	7.85%	20.48%	5.73E-11	8.59E-10	*	7.40	30.36	1.50E-12	4.51E-11	*
Sig17	0.34%	0.56%	2.52E-01	3.93E-01		0.25	0.49	2.67E-01	3.49E-01	
Sig18	0.99%	2.24%	1.37E-03	1.03E-02	*	0.95	2.60	1.88E-04	1.41E-03	*
Sig19	0.64%	1.86%	4.35E-04	4.35E-03	*	0.54	2.57	1.58E-06	1.58E-05	*
Sig20	0.95%	0.63%	3.02E-01	4.11E-01		0.85	1.16	2.13E-01	2.91E-01	
Sig21	1.18%	1.75%	1.94E-01	3.54E-01		1.01	7.23	5.20E-02	1.20E-01	
Sig22	0.31%	0.44%	2.75E-01	3.93E-01		0.33	0.56	1.19E-01	1.99E-01	
Sig23	0.41%	0.45%	8.46E-01	8.46E-01		0.26	0.55	1.61E-01	2.31E-01	
Sig24	0.74%	0.65%	7.61E-01	7.87E-01		0.76	0.66	7.20E-01	7.59E-01	
Sig25	0.49%	0.81%	2.12E-01	3.54E-01		0.28	1.08	5.49E-03	2.75E-02	*
Sig26	1.08%	1.65%	1.09E-01	3.54E-01		0.93	5.97	3.08E-02	9.25E-02	
Sig27	0.54%	0.40%	4.72E-01	5.67E-01		0.32	0.59	5.80E-02	1.20E-01	
Sig28	0.89%	0.53%	1.66E-01	3.54E-01		0.74	0.58	5.03E-01	6.04E-01	
Sig29	0.59%	0.91%	1.79E-01	3.54E-01		0.54	1.09	1.62E-02	6.08E-02	
Sig30	3.29%	2.26%	1.62E-01	3.54E-01		2.07	2.91	8.30E-01	8.30E-01	

Supplemental Table S2F. Comparison of mutational signatures in HPV-positive vs. HPV-negative OSCC (exons).
Estimated here are the fraction and number of mutations in the context of 30 COSMIC mutational signatures as reported by the Sanger Center, limited to exons (https://cancer.sanger.ac.uk/cell_lines/signatures). Shown are statistical comparisons for 30 mutational signatures in 149 HPV-positive vs. 335 HPV-negative tumors evaluated by WES by use of t-test. For the comparison of counts, data were log-transformed as $\log_{10}$(number of SNVs in signature + 1). P-values were adjusted for multiple comparisons (FDR method). *Asterisk*, FDR-adjusted p-value <0.05 (see also **Supplemental Fig. S2E - L**). P-values and adjusted P-values are presented in scientific notation; E represents *10^.

signature	signature fraction				number of signature SNVs			
	correlation coefficient	p-value	FDR-adjusted p-value		correlation coefficient	p-value	FDR-adjusted p-value	
Sig1	-0.62	3.47E-12	5.20E-11	*	0.04	7.20E-01	8.88E-01	
Sig2	0.60	1.59E-11	1.59E-10	*	0.91	2.75E-40	8.25E-39	*
Sig3	-0.25	1.12E-02	3.04E-02	*	-0.23	2.11E-02	7.91E-02	
Sig4	-0.08	4.30E-01	6.45E-01		0.11	2.71E-01	5.42E-01	
Sig5	-0.38	8.55E-05	5.76E-04	*	-0.10	3.33E-01	5.88E-01	
Sig6	-0.19	5.86E-02	1.46E-01		0.05	5.89E-01	8.42E-01	
Sig7	-0.29	3.42E-03	1.03E-02	*	0.29	2.73E-03	2.05E-02	*
Sig8	-0.15	1.43E-01	3.29E-01		-0.08	4.19E-01	6.98E-01	
Sig9	-0.09	3.64E-01	5.84E-01		0.06	5.80E-01	8.42E-01	
Sig10	-0.09	3.70E-01	5.84E-01		0.24	1.64E-02	7.91E-02	
Sig11	-0.32	1.10E-03	4.14E-03	*	-0.12	2.10E-01	4.85E-01	
Sig12	-0.02	8.44E-01	9.09E-01		0.01	9.44E-01	9.44E-01	
Sig13	0.68	2.91E-15	8.74E-14	*	0.75	4.19E-20	6.28E-19	*
Sig14	-0.13	1.83E-01	3.66E-01		0.04	6.71E-01	8.88E-01	
Sig15	-0.36	1.91E-04	9.54E-04	*	-0.03	7.32E-01	8.88E-01	
Sig16	-0.32	8.54E-04	3.66E-03	*	-0.14	1.50E-01	4.17E-01	
Sig17	0.11	2.64E-01	4.95E-01		0.14	1.67E-01	4.17E-01	
Sig18	-0.37	9.60E-05	5.76E-04	*	-0.11	2.48E-01	5.31E-01	
Sig19	-0.31	1.28E-03	4.28E-03	*	-0.30	2.28E-03	2.05E-02	*
Sig20	0.04	6.60E-01	8.24E-01		0.03	7.40E-01	8.88E-01	
Sig21	-0.02	8.64E-01	9.09E-01		0.02	8.70E-01	9.41E-01	
Sig22	-0.01	9.09E-01	9.09E-01		-0.01	9.09E-01	9.41E-01	
Sig23	-0.01	9.09E-01	9.09E-01		-0.01	9.09E-01	9.41E-01	
Sig24	-0.09	3.64E-01	5.84E-01		0.23	1.95E-02	7.91E-02	
Sig25	0.06	5.48E-01	7.21E-01		0.15	1.28E-01	4.17E-01	
Sig26	0.07	4.56E-01	6.51E-01		0.08	4.49E-01	7.09E-01	
Sig27	-0.01	9.09E-01	9.09E-01		-0.01	9.09E-01	9.41E-01	
Sig28	-0.14	1.55E-01	3.33E-01		-0.14	1.63E-01	4.17E-01	
Sig29	0.02	8.58E-01	9.09E-01		0.25	1.18E-02	7.08E-02	
Sig30	0.06	5.53E-01	7.21E-01		0.10	3.19E-01	5.88E-01	

Supplemental Table S2G. Correlation between total number of SNVs (genome-wide) and mutational signatures in HPV-positive tumors (WGS).
Shown are Pearson's correlation coefficients for the association between the fraction (*left*) or number (*right*) of mutations in that distribute into 30 COSMIC mutational signatures (https://cancer.sanger.ac.uk/cell_lines/signatures), and the total number of SNVs identified in 103 HPV-positive tumors with WGS data. Total number of SNVs was log-transformed, $\log_{10}$(number of total SNVs), as were SNVs of a particular signature, $\log_{10}$(number of SNVs in Sig + 1). P-values were adjusted for multiple comparisons (FDR method). *Asterisk*, FDR adjusted p-value <0.05 (see also **Supplemental Fig. S2M**). P-values and adjusted P-values are presented in scientific notation; E represents *10^.

signature	signature fraction				number of signature SNVs			
	correlation coefficient	p-value	FDR-adjusted p-value		correlation coefficient	p-value	FDR-adjusted p-value	
Sig1	-0.38	1.95E-06	1.95E-05	*	0.00	9.77E-01	9.99E-01	
Sig2	0.63	4.92E-18	7.38E-17	*	0.89	1.48E-51	4.43E-50	*
Sig3	-0.09	2.80E-01	3.24E-01		0.09	3.01E-01	6.08E-01	
Sig4	-0.17	4.09E-02	8.18E-02		-0.14	9.01E-02	3.86E-01	
Sig5	-0.22	6.55E-03	3.27E-02	*	0.00	9.94E-01	9.99E-01	
Sig6	-0.18	3.08E-02	8.02E-02		0.04	6.55E-01	7.86E-01	
Sig7	-0.10	2.47E-01	2.96E-01		0.43	3.19E-08	2.39E-07	*
Sig8	-0.13	1.01E-01	1.44E-01		0.00	9.99E-01	9.99E-01	
Sig9	-0.23	5.04E-03	3.02E-02	*	-0.08	3.04E-01	6.08E-01	
Sig10	-0.02	8.27E-01	8.27E-01		0.44	2.28E-08	2.28E-07	*
Sig11	-0.18	3.11E-02	8.02E-02		0.10	2.06E-01	5.48E-01	
Sig12	-0.07	3.79E-01	4.06E-01		-0.06	4.61E-01	7.02E-01	
Sig13	0.70	1.48E-23	4.45E-22	*	0.83	6.64E-40	9.95E-39	*
Sig14	-0.11	1.65E-01	2.15E-01		0.06	4.34E-01	7.02E-01	
Sig15	-0.18	3.21E-02	8.02E-02		-0.10	2.37E-01	5.48E-01	
Sig16	-0.20	1.55E-02	6.64E-02		-0.08	3.36E-01	6.29E-01	
Sig17	-0.13	1.10E-01	1.50E-01		-0.10	2.24E-01	5.48E-01	
Sig18	-0.14	9.51E-02	1.44E-01		0.00	9.85E-01	9.99E-01	
Sig19	-0.10	2.38E-01	2.96E-01		-0.05	5.15E-01	7.02E-01	
Sig20	-0.14	9.95E-02	1.44E-01		-0.04	6.10E-01	7.62E-01	
Sig21	-0.18	3.10E-02	8.02E-02		-0.02	8.52E-01	9.83E-01	
Sig22	-0.05	5.52E-01	5.71E-01		-0.04	5.98E-01	7.62E-01	
Sig23	-0.19	2.04E-02	7.66E-02		-0.10	2.21E-01	5.48E-01	
Sig24	-0.16	4.64E-02	8.70E-02		0.12	1.33E-01	4.42E-01	
Sig25	-0.17	3.72E-02	8.18E-02		-0.13	1.12E-01	4.20E-01	
Sig26	-0.16	4.95E-02	8.74E-02		-0.05	5.11E-01	7.02E-01	
Sig27	-0.17	3.84E-02	8.18E-02		-0.15	6.97E-02	3.48E-01	
Sig28	-0.14	8.45E-02	1.41E-01		-0.07	3.81E-01	6.72E-01	
Sig29	-0.08	3.31E-01	3.68E-01		-0.06	4.68E-01	7.02E-01	
Sig30	-0.33	4.73E-05	3.54E-04	*	-0.20	1.35E-02	8.12E-02	

Supplemental Table S2H. Correlation between total number of SNVs and mutational signatures in HPV-positive tumors (exons).
Shown are Pearson's correlation coefficients for the association between the fraction (*left*) or number (*right*) of mutations that distribute into 30 COSMIC mutational signatures (https://cancer.sanger.ac.uk/cell_lines/signatures) and the total number of SNVs in 149 HPV-positive tumors with exome data (WES and exonized WGS data). Total number of SNVs was log-transformed, $\log_{10}$(number of total SNVs), as were SNVs of a particular signature, $\log_{10}$(number of SNVs in Sig + 1). P-values were adjusted for multiple comparisons (FDR method). *Asterisk*, FDR adjusted p-value <0.05. (see also **Supplemental Fig. S2M**). P-values and adjusted P-values are presented in scientific notation; E represents *10^.

	signature fraction				number of signature SNVs			
signature	correlation coefficient	p-value	FDR-adjusted p-value		correlation coefficient	p-value	FDR-adjusted p-value	
Sig1	-0.62	1.79E-06	2.69E-05	*	0.23	1.04E-01	2.61E-01	
Sig2	0.22	1.20E-01	3.32E-01		0.89	1.27E-17	3.82E-16	*
Sig3	0.14	3.23E-01	5.70E-01		0.22	1.19E-01	2.76E-01	
Sig4	0.15	2.85E-01	5.65E-01		0.15	2.90E-01	5.38E-01	
Sig5	0.12	4.23E-01	5.78E-01		0.28	5.13E-02	1.40E-01	
Sig6	-0.12	3.97E-01	5.78E-01		0.00	9.93E-01	9.93E-01	
Sig7	-0.16	2.58E-01	5.53E-01		0.21	1.43E-01	3.06E-01	
Sig8	0.50	2.02E-04	2.02E-03	*	0.75	5.15E-10	5.15E-09	*
Sig9	0.22	1.22E-01	3.32E-01		0.44	1.58E-03	7.88E-03	*
Sig10	0.23	1.03E-01	3.32E-01		0.45	1.18E-03	7.09E-03	*
Sig11	0.23	1.08E-01	3.32E-01		0.72	3.68E-09	2.76E-08	*
Sig12	-0.21	1.47E-01	3.69E-01		-0.08	5.65E-01	7.95E-01	
Sig13	0.10	5.00E-01	6.25E-01		0.77	4.97E-11	7.46E-10	*
Sig14	-0.12	4.24E-01	5.78E-01		0.07	6.10E-01	7.95E-01	
Sig15	-0.67	8.66E-08	2.60E-06	*	-0.08	5.79E-01	7.95E-01	
Sig16	0.06	6.85E-01	7.61E-01		0.40	4.52E-03	1.94E-02	*
Sig17	-0.01	9.20E-01	9.36E-01		-0.01	9.20E-01	9.51E-01	
Sig18	-0.15	3.01E-01	5.65E-01		0.30	3.69E-02	1.11E-01	
Sig19	-0.14	3.42E-01	5.70E-01		0.08	5.88E-01	7.95E-01	
Sig20	0.12	4.22E-01	5.78E-01		0.12	4.22E-01	7.03E-01	
Sig21	-0.10	4.81E-01	6.25E-01		0.09	5.43E-01	7.95E-01	
Sig22	-0.19	1.86E-01	4.30E-01		-0.05	7.41E-01	8.56E-01	
Sig23	0.07	6.44E-01	7.43E-01		0.07	6.44E-01	8.05E-01	
Sig24	-0.38	6.21E-03	3.73E-02	*	-0.31	2.60E-02	8.65E-02	
Sig25	0.34	1.72E-02	8.59E-02		0.35	1.39E-02	5.22E-02	
Sig26	-0.09	5.26E-01	6.32E-01		-0.03	8.52E-01	9.13E-01	
Sig27	-0.01	9.36E-01	9.36E-01		0.03	8.52E-01	9.13E-01	
Sig28	-0.30	3.42E-02	1.47E-01		-0.05	7.31E-01	8.56E-01	
Sig29	-0.38	5.98E-03	3.73E-02	*	-0.15	2.89E-01	5.38E-01	
Sig30	0.03	8.15E-01	8.73E-01		0.15	3.05E-01	5.38E-01	

Supplemental Table S2I. Correlation between total number of SNVs (genome-wide) and mutational signatures in HPV-negative tumors (WGS).
Shown are Pearson's correlation coefficients for the association between the fraction (*left*) or number (*right*) of mutations that distribute into 30 COSMIC mutational signatures (https://cancer.sanger.ac.uk/cell_lines/signatures) and the total number of SNVs in 50 HPV-negative tumors with WGS data. Total number of SNVs was log-transformed, $\log_{10}$(number of total SNVs), as were SNVs of a particular signature, $\log_{10}$(number of SNVs in Sig + 1). P-values were adjusted for multiple comparisons (FDR method). Asterisk, FDR adjusted p-value <0.05 (see also **Supplemental Fig. S2M**). P-values and adjusted P-values are presented in scientific notation; E represents *10^.

signature	signature fraction				number of signature SNVs			
	correlation coef	p-value	FDR-adjusted p-value		correlation coef	p-value	FDR-adjusted p-value	
Sig1	-0.31	1.18E-08	1.18E-07	*	0.19	5.39E-04	2.69E-03	*
Sig2	0.41	5.20E-15	1.56E-13	*	0.58	8.90E-32	2.67E-30	*
Sig3	-0.07	2.25E-01	3.78E-01		0.12	2.45E-02	4.90E-02	*
Sig4	0.12	2.95E-02	9.19E-02		0.20	2.51E-04	1.51E-03	*
Sig5	0.01	8.56E-01	8.86E-01		0.06	3.13E-01	4.09E-01	
Sig6	-0.12	3.06E-02	9.19E-02		0.12	2.24E-02	4.81E-02	*
Sig7	-0.05	3.34E-01	4.78E-01		0.23	2.09E-05	2.09E-04	*
Sig8	0.04	4.26E-01	5.35E-01		0.13	1.58E-02	3.65E-02	*
Sig9	-0.07	1.82E-01	3.41E-01		0.01	8.05E-01	8.05E-01	
Sig10	-0.04	5.14E-01	6.17E-01		0.18	8.30E-04	3.56E-03	*
Sig11	-0.01	9.06E-01	9.06E-01		0.17	1.30E-03	4.88E-03	*
Sig12	-0.02	7.39E-01	8.21E-01		0.10	6.04E-02	1.13E-01	
Sig13	0.39	1.30E-13	1.95E-12	*	0.52	1.97E-24	2.95E-23	*
Sig14	0.05	3.91E-01	5.34E-01		0.15	6.91E-03	2.07E-02	*
Sig15	-0.10	7.46E-02	1.72E-01		0.14	9.94E-03	2.71E-02	*
Sig16	0.01	8.34E-01	8.86E-01		0.22	4.69E-05	3.52E-04	*
Sig17	-0.22	3.96E-05	2.38E-04	*	-0.07	1.76E-01	2.40E-01	
Sig18	-0.21	1.25E-04	6.25E-04	*	0.05	3.99E-01	4.98E-01	
Sig19	-0.06	2.51E-01	3.78E-01		0.07	1.76E-01	2.40E-01	
Sig20	-0.06	2.52E-01	3.78E-01		0.08	1.26E-01	1.99E-01	
Sig21	-0.19	3.98E-04	1.70E-03	*	0.14	1.18E-02	2.94E-02	*
Sig22	-0.07	2.32E-01	3.78E-01		0.03	6.31E-01	6.53E-01	
Sig23	-0.14	1.25E-02	4.69E-02	*	0.03	5.78E-01	6.20E-01	
Sig24	-0.26	1.55E-06	1.16E-05	*	0.03	5.69E-01	6.20E-01	
Sig25	-0.09	1.15E-01	2.46E-01		0.09	1.05E-01	1.75E-01	
Sig26	0.04	4.28E-01	5.35E-01		0.15	4.96E-03	1.65E-02	*
Sig27	0.02	7.18E-01	8.21E-01		0.10	7.33E-02	1.29E-01	
Sig28	-0.11	4.88E-02	1.33E-01		-0.04	4.55E-01	5.25E-01	
Sig29	-0.11	5.38E-02	1.35E-01		0.04	4.22E-01	5.06E-01	
Sig30	-0.08	1.24E-01	2.49E-01		0.08	1.44E-01	2.16E-01	

Supplemental Table S2J. Correlation between total number of SNVs and mutational signatures in HPV-negative tumors (exons).
Shown are Pearson's correlation coefficients for the association between the fraction (*left*) or number (*right*) of mutations that distribute into 30 COSMIC mutational signatures (https://cancer.sanger.ac.uk/cell_lines/signatures) and the total number of SNVs in 335 HPV-negative tumors with WES data (WES and exonized WGS data). Total number of SNVs was log-transformed, $\log_{10}$(number of total SNVs), as were SNVs of a particular signature, $\log_{10}$(number of SNVs in Sig + 1). P-values were adjusted for multiple comparisons (FDR method). *Asterisk*, FDR adjusted p-value <0.05 (see also **Supplemental Fig. S2M**). P-values and adjusted P-values are presented in scientific notation; E represents *10^.

	signature fraction					number of signature SNVs				
signature	mean fraction <10 pack-years (%total)	mean fraction ≥ 10 pack-years (%total)	p-value	FDR-adjusted p-value		mean number of SNVs <10 pack-years	mean number of SNVs ≥10 pack-years	p-value	FDR-adjusted p-value	
Sig1	23.60%	12.35%	9.15E-04	1.37E-02	*	1061.47	901.57	4.42E-01	6.51E-01	
Sig2	8.97%	7.69%	4.32E-01	5.63E-01		451.27	1361.10	3.66E-01	6.47E-01	
Sig3	5.05%	8.08%	1.01E-01	3.92E-01		277.53	898.23	6.31E-01	7.89E-01	
Sig4	0.97%	2.08%	3.25E-01	5.28E-01		74.40	269.70	6.67E-01	8.01E-01	
Sig5	9.84%	9.83%	9.98E-01	9.98E-01		558.40	1338.73	8.40E-01	8.68E-01	
Sig6	0.45%	0.52%	8.22E-01	8.51E-01		24.93	36.57	5.65E-01	7.70E-01	
Sig7	2.11%	1.39%	4.14E-01	5.63E-01		91.47	112.90	2.31E-01	6.47E-01	
Sig8	6.95%	11.53%	9.12E-02	3.92E-01		493.80	1719.27	2.38E-02	2.01E-01	
Sig9	3.92%	2.15%	1.04E-01	3.92E-01		264.20	251.00	5.90E-01	7.70E-01	
Sig10	0.49%	0.72%	4.03E-01	5.63E-01		30.13	96.57	4.88E-02	2.44E-01	
Sig11	2.37%	3.06%	2.87E-01	5.28E-01		119.80	365.80	3.35E-02	2.01E-01	
Sig12	0.67%	0.40%	5.10E-01	6.12E-01		30.27	24.50	3.66E-01	6.47E-01	
Sig13	12.34%	10.48%	5.45E-01	6.26E-01		499.27	1842.97	2.42E-01	6.47E-01	
Sig14	1.21%	0.79%	3.31E-01	5.28E-01		64.67	52.87	4.33E-01	6.51E-01	
Sig15	2.48%	1.29%	2.70E-01	5.28E-01		45.33	66.50	7.46E-01	8.57E-01	
Sig16	6.23%	19.82%	4.36E-04	1.31E-02	*	240.87	2178.90	1.77E-04	5.32E-03	*
Sig17	0.00%	0.20%	3.26E-01	5.28E-01		0.00	10.53	3.26E-01	6.47E-01	
Sig18	8.69%	4.41%	4.67E-02	2.80E-01		416.07	426.87	7.71E-01	8.57E-01	
Sig19	0.39%	0.70%	4.18E-01	5.63E-01		10.93	60.13	9.30E-01	9.30E-01	
Sig20	0.00%	0.01%	3.26E-01	5.28E-01		0.00	2.10	3.26E-01	6.47E-01	
Sig21	0.15%	0.26%	5.95E-01	6.37E-01		10.20	14.17	3.10E-01	6.47E-01	
Sig22	0.00%	0.10%	2.48E-01	5.28E-01		0.00	3.17	1.65E-01	6.47E-01	
Sig23	0.27%	0.00%	3.34E-01	5.28E-01		28.47	0.00	3.34E-01	6.47E-01	
Sig24	0.23%	0.09%	5.63E-01	6.26E-01		3.40	0.13	4.53E-01	6.51E-01	
Sig25	0.33%	0.80%	1.70E-01	5.11E-01		18.00	136.27	4.56E-01	6.51E-01	
Sig26	0.09%	0.47%	1.88E-01	5.14E-01		3.20	29.43	2.39E-01	6.47E-01	
Sig27	0.00%	0.09%	1.62E-01	5.11E-01		0.00	8.63	1.77E-01	6.47E-01	
Sig28	0.51%	0.05%	3.50E-02	2.62E-01		32.07	3.20	3.12E-02	2.01E-01	
Sig29	1.41%	0.06%	1.82E-02	1.82E-01		38.47	4.20	1.05E-02	1.57E-01	
Sig30	0.34%	0.58%	4.94E-01	6.12E-01		21.00	52.40	8.36E-01	8.68E-01	

Supplemental Table S2K. Comparison of mutational signatures in non-/low-smokers vs. heavy-smokers in HPV-negative tumors (WGS).
Shown are statistical comparisons of the fraction (*left*) and number (*right*) of SNVs that distribute into 30 COSMIC mutational signatures (https://cancer.sanger.ac.uk/cell_lines/signatures) between non-/low-smokers (<10 pack-years, n=15) vs. heavy smokers (≥10 pack-years, n=30) among 45 HPV-negative tumors with WGS and available smoking data (t-test, p-values adjusted for multiple comparisons, FDR method). The most significant difference was observed in Sig16. Among HPV-positive tumors, none of the 30 mutational signatures were significantly different in non/light vs. heavy smokers (data not shown). *Asterisk*, FDR adjusted p-value <0.05 (see also **Fig. 2, Supplemental Fig. S2N**). P-values and adjusted P-values are presented in scientific notation; E represents *10^.

signature	signature fraction					number of signature SNVs				
	mean fraction, <10 pack years (%total)	mean fraction, >=10 pack years (%total)	p-value	FDR-adjusted p-value		mean number of SNVs, <10 pack years	mean number of SNVs, >=10 pack years	p-value	FDR-adjusted p-value	
Sig1	20.28%	12.00%	3.61E-07	5.41E-06	*	21.80	14.93	2.17E-05	3.25E-04	*
Sig2	13.61%	9.10%	1.68E-03	1.68E-02	*	25.96	19.94	1.12E-01	2.80E-01	
Sig3	4.72%	4.94%	8.37E-01	8.66E-01		5.01	7.25	8.03E-01	8.10E-01	
Sig4	0.69%	0.89%	6.76E-01	8.43E-01		2.30	2.25	7.29E-01	8.10E-01	
Sig5	4.81%	6.07%	3.64E-01	5.74E-01		5.97	8.97	5.99E-01	7.82E-01	
Sig6	1.74%	0.95%	1.03E-01	2.74E-01		4.33	1.15	9.31E-02	2.54E-01	
Sig7	3.03%	2.83%	8.04E-01	8.66E-01		3.71	14.29	8.10E-01	8.10E-01	
Sig8	2.37%	2.93%	5.21E-01	6.80E-01		2.80	4.49	3.30E-01	4.97E-01	
Sig9	0.65%	0.89%	5.06E-01	6.80E-01		0.51	1.13	4.05E-01	5.53E-01	
Sig10	1.49%	1.61%	7.17E-01	8.43E-01		1.76	2.19	3.09E-01	4.97E-01	
Sig11	1.39%	2.22%	8.65E-02	2.74E-01		1.36	8.28	4.94E-03	4.94E-02	*
Sig12	1.42%	0.75%	1.53E-01	3.53E-01		1.93	1.22	5.18E-02	2.48E-01	
Sig13	15.31%	12.34%	1.03E-01	2.74E-01		29.51	26.33	6.47E-01	8.08E-01	
Sig14	0.41%	0.43%	8.90E-01	8.90E-01		0.77	0.53	3.86E-01	5.51E-01	
Sig15	1.42%	1.08%	3.08E-01	5.19E-01		2.72	1.31	2.46E-01	4.89E-01	
Sig16	7.55%	28.27%	2.69E-14	8.07E-13	*	10.48	40.52	1.29E-13	3.88E-12	*
Sig17	1.09%	0.34%	1.08E-01	2.74E-01		0.60	0.40	7.35E-02	2.48E-01	
Sig18	2.76%	2.14%	4.02E-01	6.04E-01		2.74	2.55	7.26E-01	8.10E-01	
Sig19	1.91%	2.11%	7.30E-01	8.43E-01		1.78	3.32	3.04E-01	4.97E-01	
Sig20	0.89%	0.29%	1.10E-01	2.74E-01		0.66	0.44	2.61E-01	4.89E-01	
Sig21	2.44%	0.97%	2.13E-02	1.49E-01		19.50	1.10	3.18E-02	1.91E-01	
Sig22	0.36%	0.57%	1.96E-01	4.20E-01		0.45	0.71	7.45E-02	2.48E-01	
Sig23	0.29%	0.51%	2.32E-01	4.35E-01		0.37	0.46	7.71E-01	8.10E-01	
Sig24	0.83%	0.47%	3.11E-01	5.19E-01		0.60	0.46	2.20E-01	4.71E-01	
Sig25	1.02%	0.68%	4.38E-01	6.25E-01		0.94	1.03	3.32E-01	4.97E-01	
Sig26	1.65%	1.53%	8.30E-01	8.66E-01		14.09	1.97	7.78E-01	8.10E-01	
Sig27	0.24%	0.48%	9.75E-02	2.74E-01		0.32	0.64	9.20E-02	2.54E-01	
Sig28	0.73%	0.44%	2.29E-01	4.35E-01		0.71	0.50	1.47E-01	3.40E-01	
Sig29	1.54%	0.64%	2.98E-02	1.49E-01		1.65	0.84	7.19E-02	2.48E-01	
Sig30	3.33%	1.53%	2.58E-02	1.49E-01		4.40	2.10	2.68E-02	1.91E-01	

Supplemental Table S2L. Comparison of mutational signatures between heavy-smokers and non-/low-smokers in HPV-negative tumors (exon).
Shown are statistical comparisons of the fraction (*left*) and number (*right*) of SNVs that distribute into 30 COSMIC mutational signatures (https://cancer.sanger.ac.uk/cell_lines/signatures) between non-/low-smokers (<10 pack-years, n=105) vs. heavy smokers (≥10 pack-years, n=167) among 272 HPV-negative tumors with WES and available smoking data (t-test, p-values adjusted for multiple comparisons, FDR method). Number of SNVs were log-transformed, $\log_{10}$(number of signature SNVs + 1). The most significant difference was observed in Sig16. None of 30 mutational signatures show significant difference by smoking status in HPV-positive tumors (data not shown). *Asterisk*, FDR adjusted p-value <0.05 (see also **Fig.2, Supplemental Fig. S2N**). P-values and adjusted P-values are presented in scientific notation; E represents *10^.

signature	signature fraction				number of signature SNVs			
	correlation coefficient	p-value	FDR-adjusted p-value		correlation coefficient	p-value	FDR-adjusted p-value	
Sig1	-0.45	1.98E-03	2.98E-02	*	0.01	9.54E-01	9.84E-01	
Sig2	-0.13	3.95E-01	7.50E-01		0.20	1.92E-01	5.64E-01	
Sig3	0.02	8.90E-01	9.53E-01		-0.09	5.58E-01	8.33E-01	
Sig4	0.10	5.31E-01	7.50E-01		0.01	9.28E-01	9.84E-01	
Sig5	0.07	6.67E-01	8.00E-01		0.08	6.09E-01	8.33E-01	
Sig6	0.05	7.61E-01	8.46E-01		-0.03	8.41E-01	9.48E-01	
Sig7	0.01	9.26E-01	9.53E-01		-0.18	2.26E-01	5.64E-01	
Sig8	0.35	1.84E-02	1.10E-01		0.44	2.67E-03	1.60E-02	*
Sig9	-0.32	3.09E-02	1.18E-01		-0.15	3.29E-01	6.17E-01	
Sig10	0.12	4.50E-01	7.50E-01		0.30	4.44E-02	1.90E-01	
Sig11	0.21	1.64E-01	4.62E-01		0.44	2.66E-03	1.60E-02	*
Sig12	-0.24	1.11E-01	3.69E-01		-0.31	4.01E-02	1.90E-01	
Sig13	-0.14	3.46E-01	7.41E-01		0.14	3.68E-01	6.50E-01	
Sig14	-0.07	6.31E-01	7.89E-01		-0.03	8.22E-01	9.48E-01	
Sig15	-0.21	1.69E-01	4.62E-01		0.08	6.11E-01	8.33E-01	
Sig16	0.48	9.60E-04	2.88E-02	*	0.62	6.36E-06	1.91E-04	*
Sig17	0.10	5.25E-01	7.50E-01		0.10	5.25E-01	8.29E-01	
Sig18	-0.32	3.14E-02	1.18E-01		0.05	7.50E-01	9.37E-01	
Sig19	-0.11	4.61E-01	7.50E-01		-0.15	3.14E-01	6.17E-01	
Sig20	0.10	5.25E-01	7.50E-01		0.10	5.25E-01	8.29E-01	
Sig21	0.11	4.57E-01	7.50E-01		0.28	6.59E-02	2.20E-01	
Sig22	-0.08	5.79E-01	7.55E-01		0.00	9.84E-01	9.84E-01	
Sig23	-0.18	2.47E-01	5.70E-01		-0.18	2.47E-01	5.70E-01	
Sig24	-0.18	2.37E-01	5.70E-01		-0.19	2.17E-01	5.64E-01	
Sig25	0.32	3.08E-02	1.18E-01		0.28	5.94E-02	2.20E-01	
Sig26	0.01	9.53E-01	9.53E-01		0.03	8.53E-01	9.48E-01	
Sig27	0.05	7.33E-01	8.45E-01		0.07	6.52E-01	8.50E-01	
Sig28	-0.43	3.28E-03	3.28E-02	*	-0.47	1.19E-03	1.60E-02	*
Sig29	-0.41	4.87E-03	3.65E-02	*	-0.44	2.52E-03	1.60E-02	*
Sig30	-0.09	5.50E-01	7.50E-01		-0.15	3.18E-01	6.17E-01	

Supplemental Table S2M. Correlation between mutational signatures and smoking (cigarette pack-years) in HPV-negative tumors (WGS).
Shown are Pearson's correlation coefficients for associations between cigarette smoking (cigarette pack-years) as a continuous variable and the fraction (*left*) or number (*right*) of SNVs that distribute into 30 COSMIC mutational signatures (https://cancer.sanger.ac.uk/cell_lines/signatures) among 45 HPV-negative tumors with WGS data and available smoking data. Number of SNVs of a particular signature were log-transformed, $\log_{10}$(no. Signature SNVs + 1). P-values were adjusted for multiple comparisons (FDR method). None of the 30 mutational signatures showed a significant correlation with pack-years of smoking in HPV-positive tumors (data not shown). *Asterisk*, FDR adjusted p-value <0.05 (see also **Fig.2, Supplemental Fig. S2N**). P-values and adjusted P-values are presented in scientific notation; E represents *10^.

signature	fraction				number of SNVs			
	correlation coefficient	p-value	FDR-adjusted p-value		correlation coefficient	p-value	FDR-adjusted p-value	
Sig1	-0.20	1.05E-03	1.05E-02	*	-0.08	2.02E-01	5.50E-01	
Sig2	-0.16	7.78E-03	5.83E-02		-0.05	4.19E-01	6.62E-01	
Sig3	-0.03	6.36E-01	9.08E-01		-0.02	6.93E-01	8.89E-01	
Sig4	0.02	7.80E-01	9.08E-01		0.02	7.12E-01	8.89E-01	
Sig5	0.02	7.56E-01	9.08E-01		0.01	8.41E-01	9.18E-01	
Sig6	-0.12	5.76E-02	2.19E-01		-0.12	5.70E-02	3.42E-01	
Sig7	-0.05	3.90E-01	6.88E-01		0.02	8.05E-01	9.18E-01	
Sig8	0.00	9.62E-01	9.84E-01		0.06	3.53E-01	6.62E-01	
Sig9	0.00	9.36E-01	9.84E-01		0.00	9.56E-01	9.56E-01	
Sig10	0.01	9.11E-01	9.84E-01		0.05	3.70E-01	6.62E-01	
Sig11	0.03	6.43E-01	9.08E-01		0.10	1.04E-01	4.21E-01	
Sig12	-0.07	2.40E-01	5.14E-01		-0.09	1.35E-01	4.21E-01	
Sig13	-0.10	1.17E-01	3.39E-01		0.00	9.48E-01	9.56E-01	
Sig14	-0.02	7.87E-01	9.08E-01		-0.04	5.13E-01	7.69E-01	
Sig15	-0.11	5.85E-02	2.19E-01		-0.09	1.40E-01	4.21E-01	
Sig16	0.37	5.38E-10	1.61E-08	*	0.39	2.27E-11	6.80E-10	*
Sig17	-0.06	2.88E-01	5.42E-01		-0.05	4.04E-01	6.62E-01	
Sig18	-0.10	1.13E-01	3.39E-01		-0.06	2.99E-01	6.62E-01	
Sig19	0.00	9.84E-01	9.84E-01		0.03	5.89E-01	8.41E-01	
Sig20	-0.09	1.24E-01	3.39E-01		-0.09	1.19E-01	4.21E-01	
Sig21	-0.12	4.14E-02	2.07E-01		-0.10	8.44E-02	4.21E-01	
Sig22	0.08	1.97E-01	4.92E-01		0.14	2.41E-02	2.41E-01	
Sig23	-0.04	5.10E-01	8.05E-01		-0.03	6.63E-01	8.89E-01	
Sig24	-0.06	2.89E-01	5.42E-01		-0.06	3.23E-01	6.62E-01	
Sig25	-0.04	4.91E-01	8.05E-01		-0.05	4.17E-01	6.62E-01	
Sig26	-0.02	7.75E-01	9.08E-01		-0.01	8.57E-01	9.18E-01	
Sig27	0.20	1.04E-03	1.05E-02	*	0.22	2.59E-04	3.89E-03	*
Sig28	-0.07	2.34E-01	5.14E-01		-0.05	3.72E-01	6.62E-01	
Sig29	-0.02	7.47E-01	9.08E-01		-0.02	7.93E-01	9.18E-01	
Sig30	-0.13	3.52E-02	2.07E-01		-0.13	3.41E-02	2.56E-01	

Supplemental Table S2N. Correlation between mutational signatures and smoking (cigarette pack-years) in HPV-negative tumors (exon).
Shown are Pearson's correlation coefficients for associations between cigarette smoking (pack-years) as a continuous variable and the fraction (*left*) or number (*right*) of SNVs that distribute into 30 COSMIC mutational signatures (https://cancer.sanger.ac.uk/cell_lines/signatures) among 272 HPV-negative tumors with WES data and available smoking data. Number of SNVs of a particular signature were log-transformed, $\log_{10}$(number of signature SNVs + 1). P-values were adjusted for multiple comparisons (FDR method). None of the 30 mutational signatures showed a significant correlation with pack-years of smoking in HPV-positive tumors (data not shown). *Asterisk*, FDR adjusted p-value <0.05 (see also **Fig.2, Supplemental Fig. S2N**). P-values and adjusted P-values are presented in scientific notation; E represents *10^.

signature	fraction				number of SNVs			
	correlation coefficient	p-value	adjusted p-value		correlation coefficient	p-value	adjusted p-value	
Sig1	-0.22	5.23E-03	5.23E-02		-0.13	9.89E-02	7.42E-01	
Sig2	-0.11	1.63E-01	8.13E-01		-0.03	6.66E-01	9.21E-01	
Sig3	-0.05	5.04E-01	9.49E-01		-0.06	4.90E-01	9.21E-01	
Sig4	-0.04	6.10E-01	9.49E-01		-0.03	7.07E-01	9.21E-01	
Sig5	0.04	6.56E-01	9.49E-01		0.01	9.21E-01	9.21E-01	
Sig6	-0.07	3.79E-01	9.49E-01		-0.09	2.58E-01	9.21E-01	
Sig7	-0.07	3.53E-01	9.49E-01		-0.02	8.24E-01	9.21E-01	
Sig8	-0.04	5.79E-01	9.49E-01		-0.04	5.97E-01	9.21E-01	
Sig9	0.01	8.97E-01	9.61E-01		0.02	7.90E-01	9.21E-01	
Sig10	-0.02	7.91E-01	9.49E-01		-0.02	8.38E-01	9.21E-01	
Sig11	0.02	7.78E-01	9.49E-01		0.02	7.60E-01	9.21E-01	
Sig12	-0.06	4.47E-01	9.49E-01		-0.07	3.72E-01	9.21E-01	
Sig13	-0.08	3.34E-01	9.49E-01		0.01	8.65E-01	9.21E-01	
Sig14	0.12	1.28E-01	7.66E-01		0.11	1.85E-01	9.21E-01	
Sig15	-0.03	7.17E-01	9.49E-01		-0.10	2.24E-01	9.21E-01	
Sig16	0.27	5.95E-04	1.79E-02	*	0.29	2.23E-04	6.69E-03	*
Sig17	0.00	9.64E-01	9.77E-01		0.02	8.18E-01	9.21E-01	
Sig18	0.05	5.09E-01	9.49E-01		-0.04	6.60E-01	9.21E-01	
Sig19	0.00	9.77E-01	9.77E-01		0.01	9.13E-01	9.21E-01	
Sig20	-0.01	8.72E-01	9.61E-01		-0.01	8.99E-01	9.21E-01	
Sig21	-0.03	6.78E-01	9.49E-01		-0.04	6.51E-01	9.21E-01	
Sig22	-0.02	7.66E-01	9.49E-01		0.02	8.47E-01	9.21E-01	
Sig23	0.25	1.80E-03	2.70E-02	*	0.19	1.90E-02	2.85E-01	
Sig24	-0.04	6.03E-01	9.49E-01		-0.08	2.92E-01	9.21E-01	
Sig25	-0.03	7.10E-01	9.49E-01		0.05	4.99E-01	9.21E-01	
Sig26	-0.04	5.88E-01	9.49E-01		-0.03	6.69E-01	9.21E-01	
Sig27	-0.08	3.28E-01	9.49E-01		-0.05	5.22E-01	9.21E-01	
Sig28	-0.07	4.04E-01	9.49E-01		-0.08	3.40E-01	9.21E-01	
Sig29	0.01	8.62E-01	9.61E-01		0.05	5.60E-01	9.21E-01	
Sig30	-0.14	7.14E-02	5.36E-01		-0.16	4.41E-02	4.41E-01	

Supplemental Table S2O. Correlation between mutational signatures and amount of alcohol consumption per day in HPV-negative tumors (exon).
Shown are Pearson's correlation coefficients for the associations between alcohol consumption (current drinks per day) and the fraction (*left*) and number (*right*) of SNVs that distribute into 30 COSMIC mutational signatures (https://cancer.sanger.ac.uk/cell_lines/signatures) among 156 HPV-negative tumors with WES or exonized WGS data and available alcohol use history. Number of SNVs was log-transformed, $\log_{10}$(number of signature SNVs + 1). P-values were adjusted for multiple comparisons (FDR method). The most significant correlation was observed in Sig16. None of 30 mutational signatures were correlated with alcohol in HPV-positive tumors (data not shown). *Asterisk*, FDR adjusted p-value <0.05 (see also **Supplemental Fig. S2N**). P-values and adjusted P-values are presented in scientific notation; E represents *10^.

gene	description	no. variants with coding change	no. unique variants with coding change	no. synonymous variants	no. patients with coding change variant	p	q
FGFR3	fibroblast growth factor receptor 3	19	7	2	17	0.00E+00	0.00E+00
NFKBIA	Nuclear Factor Of Kappa Light Polypeptide Gene Enhancer In B-Cells Inhibitor, Alpha	4	4	0	4	0.00E+00	0.00E+00
PIK3CA	phosphatidylinositol-4,5-bisphosphate 3-kinase, catalytic subunit alpha	45	9	0	42	1.46E-13	9.17E-10
PTEN	phosphatase and tensin homolog	13	12	0	13	4.55E-11	2.14E-07
KMT2D	lysine (K)-specific methyltransferase 2D	22	22	4	20	2.52E-09	9.51E-06
CYLD	cylindromatosis (turban tumor syndrome)	11	11	1	11	4.45E-09	1.40E-05
RB1	retinoblastoma 1	10	10	0	9	8.22E-09	2.21E-05
DDX3X	DEAD (Asp-Glu-Ala-Asp) box helicase 3, X-linked	9	8	0	8	2.74E-08	6.45E-05
ZNF750	zinc finger protein 750	23	22	1	21	7.08E-07	1.48E-03
IFNGR1	Interferon Gamma Receptor 1	4	4	0	4	7.82E-07	1.48E-03
EP300	E1A Binding Protein P300	18	12	0	18	1.63E-06	2.79E-03
FBXW7	F-box and WD repeat domain containing 7, E3 ubiquitin protein ligase	11	7	0	11	2.00E-06	3.14E-03
CASZ1	castor zinc finger 1	15	14	2	14	3.06E-06	4.44E-03
TRAF3	TNF receptor-associated factor 3	5	5	0	5	1.05E-05	1.33E-02
RBL1	Retinoblastoma-Like 1	9	9	0	9	3.01E-05	3.55E-02
TAF5	TAF5 RNA polymerase II, TATA box binding protein (TBP)-associated factor, 100kDa	5	5	0	4	7.08E-05	7.03E-02
RIPK4	receptor interacting serine/threonine kinase 4	6	6	0	6	8.17E-05	7.70E-02
HIST1H2AE	histone cluster 1 H2A family member e	4	4	0	4	1.77E-04	1.43E-01
STAT1	signal transducer and activator of transcription 1	4	4	1	4	1.82E-04	1.43E-01
NSD1	Nuclear Receptor Binding SET Domain Protein 1	10	10	1	10	2.41E-04	1.63E-01
ASAP1	ArfGAP with SH3 domain, ankyrin repeat and PH domain 1	10	10	0	9	2.21E-04	1.63E-01
BBX	BBX, HMG-box containing	12	12	2	5	2.31E-04	1.63E-01
SLTM	SAFB like transcription modulator	6	6	0	5	2.33E-04	1.63E-01
TGFBR2	transforming growth factor beta receptor 2	5	5	0	5	2.83E-04	1.84E-01

Supplemental Table S3A. Most significantly mutated genes in 149 HPV-positive OSCC samples. We identified significant somatic mutations in 149 HPV-positive tumors using MutSig. Genes with recurrent mutations (q-value <0.2) and with expression in OSCC tumors (median RNA-seq expression > 1 fpkm) were tabulated. Shown here are the number of sample alleles with coding change variants, the number of unique variants among these coding change mutations, the number of patients with coding change mutations, and the number of synonymous variants, with (*right*) p-values and adjusted p-values. *Bold text*, genes with mutations at q-value < 0.1. See also **Fig. 1.** P-values and adjusted P-values are presented in scientific notation; E represents *10^.

gene1	gene2	no. sample mutated in gene1	no. sample mutated in gene2	no. sample mutated in both gene1 and gene2	no. total sample	no. expected	fold increase	p-value	FDR adjusted p-value	
ZNF750	*DDX3X*	21	8	6	149	1.13	5.32	9.17E-05	2.53E-02	*
ZNF750	*NSD1*	21	10	6	149	1.41	4.26	5.68E-04	7.84E-02	
ZNF750	*ASAP1*	21	9	5	149	1.27	3.94	3.03E-03	1.78E-01	
ZNF750	*FGFR3*	21	17	7	149	2.40	2.92	3.21E-03	1.78E-01	
FGFR3	*FBXW7*	17	11	5	149	1.26	3.98	3.23E-03	1.78E-01	
FGFR3	*STAT1*	17	4	3	149	0.46	6.57	4.67E-03	2.15E-01	
CASZ1	*BBX*	14	5	3	149	0.47	6.39	6.00E-03	2.36E-01	
ASAP1	*DDX3X*	9	8	3	149	0.48	6.21	7.44E-03	2.57E-01	
EP300	*PTEN*	18	13	5	149	1.57	3.18	1.02E-02	2.99E-01	
FGFR3	*BBX*	17	5	3	149	0.57	5.26	1.08E-02	2.99E-01	
EP300	*BBX*	18	5	3	149	0.60	4.97	1.29E-02	3.23E-01	
DDX3X	*IFNGR1*	8	4	2	149	0.21	9.31	1.44E-02	3.32E-01	
RBL1	*IFNGR1*	9	4	2	149	0.24	8.28	1.84E-02	3.50E-01	
RBL1	*STAT1*	9	4	2	149	0.24	8.28	1.84E-02	3.50E-01	
ZNF750	*BBX*	21	5	3	149	0.70	4.26	2.03E-02	3.50E-01	
ZNF750	*TGFBR2*	21	5	3	149	0.70	4.26	2.03E-02	3.50E-01	
ZNF750	*RBL1*	21	9	4	149	1.27	3.15	2.33E-02	3.79E-01	
CASZ1	*DDX3X*	14	8	3	149	0.75	3.99	2.82E-02	4.04E-01	
ZNF750	*CASZ1*	21	14	5	149	1.97	2.53	2.94E-02	4.04E-01	
RBL1	*BBX*	9	5	2	149	0.30	6.62	2.96E-02	4.04E-01	
PIK3CA	*KMT2D*	42	20	10	149	5.64	1.77	3.07E-02	4.04E-01	
NSD1	*TGFBR2*	10	5	2	149	0.34	5.96	3.65E-02	4.58E-01	
CASZ1	*RBL1*	14	9	3	149	0.85	3.55	3.99E-02	4.67E-01	
RBL1	*RIPK4*	9	6	2	149	0.36	5.52	4.30E-02	4.67E-01	
FBXW7	*TRAF3*	11	5	2	149	0.37	5.42	4.40E-02	4.67E-01	
FBXW7	*BBX*	11	5	2	149	0.37	5.42	4.40E-02	4.67E-01	

Supplemental Table S3B. Co-occurrence of somatic mutations in gene pairs in HPV-positive OSCC.
Shown are number of tumors with co-occurring mutations (SNVs and small Indels with coding-change consequences) in gene pairs among 24 recurrently mutated genes as detected by MutSig in 149 HPV-positive OSCC. All possible two-gene combinations were evaluated, of which 26 pairs with co-occurring mutations were observed in the same tumor. After applying multiple testing correction, one pair, i.e. *ZNF750* and *DDX3X*, showed significant increase in co-occurring mutations by Fisher's exact test with unadjusted p-value < 0.05. We did not observe a combination of gene pairs with exclusive mutation patterns at statistically significant level (data not shown). P-values and adjusted P-values are presented in scientific notation; E represents *10^.

gene	description	no. variants with coding change	no. unique variants with coding change	no. synony-mous variants	no. patients with coding change variant	p	q
TP53	tumor protein p53	324	208	2	267	0.00E+00	0.00E+00
FAT1	FAT atypical cadherin 1	119	111	8	91	0.00E+00	0.00E+00
CASP8	caspase 8	60	51	0	52	0.00E+00	0.00E+00
HRAS	Harvey rat sarcoma viral oncogene homolog	30	12	0	28	0.00E+00	0.00E+00
TGFBR2	transforming growth factor beta receptor II	19	15	2	18	0.00E+00	0.00E+00
HLA-B	major histocompatibility complex, class I, B	19	16	1	17	0.00E+00	0.00E+00
CDKN2A	cyclin-dependent kinase inhibitor 2A	86	36	1	83	2.89E-15	7.78E-12
AJUBA	ajuba LIM protein	20	18	2	17	8.72E-13	2.06E-09
PIK3CA	phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit alpha	52	25	1	51	1.14E-12	2.38E-09
NOTCH1	notch 1	72	71	7	64	2.80E-12	5.28E-09
EPHA2	EPH receptor A2	21	18	2	17	1.01E-09	1.58E-06
FBXW7	F-box and WD repeat domain containing 7	24	21	0	22	8.58E-09	1.24E-05
NSD1	nuclear receptor binding SET domain protein 1	31	31	0	27	2.81E-08	3.78E-05
RASA1	RAS p21 protein activator 1	15	15	0	14	6.35E-07	7.49E-04
KMT2D	lysine methyltransferase 2D	51	51	4	41	1.32E-06	1.47E-03
FOSL2	FOS like antigen 2	8	8	1	8	3.30E-06	3.46E-03
NF2	neurofibromin 2	8	8	0	8	5.13E-06	5.09E-03
CTCF	CCCTC-binding factor	16	15	1	14	1.08E-05	1.02E-02
NOTCH2	notch 2	19	19	1	18	5.28E-05	4.33E-02
SMAD4	SMAD family member 4	10	10	1	9	8.25E-05	6.48E-02
PSIP1	PC4 and SFRS1 interacting protein 1	8	8	0	8	8.69E-05	6.56E-02
NFE2L2	nuclear factor, erythroid 2 like 2	19	16	2	18	9.23E-05	6.69E-02
RAC1	ras-related C3 botulinum toxin substrate 1	1]1	7	0	10	1.12E-04	7.83E-02
EIF2S2	eukaryotic translation initiation factor 2 subunit beta	7	6	1	7	1.81E-04	1.18E-01
PODXL	podocalyxin like	8	7	0	7	2.73E-04	1.66E-01

Supplemental Table S3C. Most significantly mutated genes in 335 HPV-negative OSCC samples. Somatic mutations were identified in 335 HPV-negative tumors using MutSig, including 50 samples with WGS data and 285 samples from TCGA with WES data. Genes with recurrent mutations (q-value <0.2) and with expression in OSCC tumors (median RNA-seq expression >1 fpkm) were tabulated. Shown from *left* to *right* are: gene symbol; gene name; number of sample alleles with coding change variants; number of unique variants among these coding change mutations; number of synonymous variants; and number of patients with coding change mutations. *Bold text*, genes with mutations at q-value < 0.1. See **Supplemental Fig. S1A**. P-values and adjusted P-values (q-values) are presented in scientific notation; E represents *10^.

gene	HPV-positive no. patients	HPV-positive % patients	HPV-negative no. patients	HPV-negative % patients	fold change in HPV-positive patients	p-value	FDR adjusted p-value	
ZNF750	21	14.1%	4	1.2%	11.8	2.46E-08	5.90E-07	*
FGFR3	17	11.4%	4	1.2%	9.5	1.84E-06	4.23E-05	*
CYLD	11	7.4%	2	0.6%	12.3	7.73E-05	1.70E-03	*
CASZ1	14	9.4%	5	1.5%	6.3	1.11E-04	2.32E-03	*
PTEN	13	8.7%	5	1.5%	5.8	2.82E-04	5.64E-03	*
DDX3X	8	5.4%	1	0.3%	17.9	4.76E-04	9.04E-03	*
PIK3CA	42	28.2%	51	15.3%	1.8	1.16E-03	2.09E-02	*
ASAP1	9	6.0%	4	1.2%	5.0	4.45E-03	7.56E-02	
NFKBIA	4	2.7%	0	0.0%	NaN	8.81E-03	1.41E-01	
TRAF3	5	3.4%	1	0.3%	11.2	1.20E-02	1.80E-01	
RBL1	9	6.0%	6	1.8%	3.4	2.05E-02	2.87E-01	
SLTM	5	3.4%	2	0.6%	5.6	3.15E-02	4.09E-01	
RB1	9	6.0%	7	2.1%	2.9	4.92E-02	5.90E-01	
EP300	18	12.1%	22	6.6%	1.8	4.98E-02	5.90E-01	
RIPK4	6	4.0%	4	1.2%	3.4	7.60E-02	7.60E-01	
IFNGR1	4	2.7%	2	0.6%	4.5	7.61E-02	7.60E-01	
TAF5	4	2.7%	2	0.6%	4.5	7.61E-02	7.60E-01	
HIST1H2AE	4	2.7%	2	0.6%	4.5	7.61E-02	7.60E-01	
STAT1	4	2.7%	4	1.2%	2.2	2.59E-01	1.00E+00	
BBX	5	3.4%	6	1.8%	1.9	3.27E-01	1.00E+00	
TGFBR2	5	3.4%	18	5.4%	0.6	4.88E-01	1.00E+00	
NSD1	10	6.7%	27	8.1%	0.8	7.12E-01	1.00E+00	
KMT2D	20	13.4%	41	12.3%	1.1	7.67E-01	1.00E+00	
FBXW7	11	7.4%	22	6.6%	1.1	8.45E-01	1.00E+00	

Supplemental Table S3D. Comparison of mutation frequencies for 24 HPV-positive MutSig genes in HPV-positive vs. HPV-negative OSCC. The frequency of mutations in the 24 most significantly mutated genes, originally identified using MutSig in 149 HPV-positive OSCC tumors, was compared in 149 HPV-positive vs. 335 HPV-negative OSCC. The significance of the comparison was assessed using Fisher's Exact test. We applied the multiple testing correction on p-values using FDR method.* indicates FDR adjusted p-value <0.05. P-values and adjusted P-values are presented in scientific notation; E represents *10^.

gene	HPV-positive no. patients	HPV-positive % patients	HPV-negative no. patients	HPV-negative % patients	fold change in HPV-negative patients	p-value	FDR adjusted p-value	
TP53	2	1.3%	267	79.9%	59.6	7.74E-68	1.93E-66	*
CDKN2A	0	0.0%	83	24.9%	NaN	1.82E-15	4.36E-14	*
FAT1	4	2.7%	91	27.2%	10.1	3.43E-12	7.90E-11	*
CASP8	2	1.3%	52	15.6%	11.6	4.46E-07	9.81E-06	*
NOTCH1	9	6.0%	64	19.2%	3.2	1.02E-04	2.15E-03	*
PIK3CA	42	28.2%	51	15.3%	0.5	1.16E-03	2.32E-02	*
HRAS	2	1.3%	28	8.4%	6.2	1.83E-03	3.48E-02	*
RASA1	0	0.0%	14	4.2%	NaN	7.07E-03	1.27E-01	
AJUBA	2	1.3%	17	5.1%	3.8	7.28E-02	1.00E+00	
PODXL	0	0.0%	7	2.1%	NaN	1.06E-01	1.00E+00	
NF2	0	0.0%	8	2.4%	NaN	1.14E-01	1.00E+00	
PSIP1	0	0.0%	8	2.4%	NaN	1.14E-01	1.00E+00	
NOTCH2	3	2.0%	18	5.4%	2.7	1.45E-01	1.00E+00	
NFE2L2	4	2.7%	18	5.4%	2.0	2.41E-01	1.00E+00	
FOSL2	1	0.7%	8	2.4%	3.6	2.86E-01	1.00E+00	
CTCF	3	2.0%	14	4.2%	2.1	2.93E-01	1.00E+00	
HLA-B	4	2.7%	17	5.1%	1.9	3.34E-01	1.00E+00	
EPHA2	4	2.7%	17	5.1%	1.9	3.34E-01	1.00E+00	
RAC1	2	1.3%	10	3.0%	2.2	3.59E-01	1.00E+00	
EIF2S2	1	0.7%	7	2.1%	3.1	4.45E-01	1.00E+00	
TGFBR2	5	3.4%	18	5.4%	1.6	4.88E-01	1.00E+00	
SMAD4	2	1.3%	9	2.7%	2.0	5.16E-01	1.00E+00	
NSD1	10	6.7%	27	8.1%	1.2	7.12E-01	1.00E+00	
KMT2D	20	13.4%	41	12.3%	0.9	7.67E-01	1.00E+00	
FBXW7	11	7.4%	22	6.6%	0.9	8.45E-01	1.00E+00	

Supplemental Table S3E. Comparison of mutation frequencies for 25 HPV-negative OSCC MutSig genes in HPV-positive vs. HPV-negative OSCC. The frequency of mutations in the 25 most significantly mutated genes, originally identified using MutSig in 335 HPV-negative OSCC, were compared in 335 HPV-negative vs. 149 HPV-positive OSCC. The significance of the comparison was assessed using Fisher's Exact test. We applied multiple testing correction on p-values using the FDR method. NaN, undefined. *Asterisk*, FDR adjusted p-value <0.05. P-values and adjusted P-values are presented in scientific notation; E represents *10^.

	hg19		hg38	
gene	no. somatic variants	no. non-synonymous variants	no. somatic variants	no. non-synonymous variants
TP53	216	203	216	205
FAT1	98	86	96	84
CDKN2A	41	40	42	42
NOTCH1	54	47	53	48
CASP8	48	45	48	45
PIK3CA	37	35	37	36
KMT2D	42	38	48	43
HRAS	22	22	12	12
NSD1	25	24	24	23
FBXW7	19	19	20	20
TGFBR2	13	13	14	13
NOTCH2	26	17	21	19
NFE2L2	17	14	17	14
HLA-B	17	16	9	4
AJUBA	14	11	15	12
EPHA2	18	15	18	15
RASA1	11	7	13	8
CTCF	10	8	10	8
RAC1	10	9	10	9
SMAD4	8	6	9	6
FOSL2	8	7	8	7
NF2	6	6	6	6
PSIP1	5	2	5	2
EIF2S2	5	4	5	4
PODXL	4	3	8	4

Supplemental Table S3F. Comparison of mutation counts for 25 HPV-negative OSCC MutSig genes using hg19 vs. hg38 reference genome assemblies. Mutation counts for the 25 most significantly mutated genes, originally identified using MutSig in HPV-negative cancers, were determined by analysis of 217 HPV-negative WES data downloaded from TCGA. As described in Supplemental Materials, WES sequencing reads were aligned to both the hg19 and hg38 human reference genome assemblies using BWA-0.7.15, using the alternate locus-aware alignment protocol for hg38. Somatic variants were called using Mutect2 and were filtered at a population allele frequency of <0.01 and for non-synonymous variants using VEP v. 92.

gene	**less than 10 packs**			**greater than 10 packs**			**P-value**	**adjusted p-value**
	no. mutated	no. total	%total	no. mutated	no. total	%total		
AJUBA	2	66	3.03%	0	69	0.00%	0.24	1
CASP8	1	66	1.52%	1	69	1.45%	1	1
CTCF	1	66	1.52%	2	69	2.90%	1	1
EIF2S2	1	66	1.52%	0	69	0.00%	0.49	1
EPHA2	2	66	3.03%	2	69	2.90%	1	1
FAT1	3	66	4.55%	1	69	1.45%	0.36	1
FBXW7	4	66	6.06%	5	69	7.25%	1	1
FOSL2	0	66	0.00%	1	69	1.45%	1	1
HLA-B	3	66	4.55%	1	69	1.45%	0.36	1
HRAS	0	66	0.00%	2	69	2.90%	0.5	1
KMT2D	8	66	12.12%	11	69	15.94%	0.62	1
NFE2L2	3	66	4.55%	1	69	1.45%	0.36	1
NOTCH1	6	66	9.09%	2	69	2.90%	0.16	1
NOTCH2	2	66	3.03%	1	69	1.45%	0.61	1
NSD1	5	66	7.58%	4	69	5.80%	0.74	1
PIK3CA	20	66	30.30%	19	69	27.54%	0.85	1
RAC1	0	66	0.00%	2	69	2.90%	0.5	1
SMAD4	2	66	3.03%	0	69	0.00%	0.24	1
TGFBR2	2	66	3.03%	2	69	2.90%	1	1
TP53	0	66	0.00%	2	69	2.90%	0.5	1
AJUBA, CASP8, CTCF, EIF2S2, EPHA2, FAT1, FBXW7, FOSL2, HLA-B, HRAS, KMT2D, NFE2L2, NOTCH1, NOTCH2, NSD1, PIK3CA, RAC1, SMAD4, TGFBR2, TP53	39	66	59.09%	38	69	55.07%	0.73	1

Supplemental Table S3G. Frequency of mutations in HPV-negative OSCC MutSig genes among HPV-positive heavy smokers.
The frequency of mutations in genes, originally identified as significantly mutated among HPV-negative OSCC patients MutSig HPV-negative, were compared among non/light vs. heavy smokers in HPV-positive OSCC. P-values from Fisher's Exact test, and FDR-adjusted p-value. P-values and adjusted P-values are presented in scientific notation; E represents *10^.

	HPV-positive oral cancer			Cervical cancer					
gene	no. samples with mutation	no. total samples	fraction of mutated samples (%total)	no. samples with mutation	no. total samples	fraction of mutated samples (%total)	p-value	FDR adjusted p-value	
PIK3CA	42	149	28.2%	98	305	32.1%	4.49E-01	6.73E-01	
ZNF750	21	149	14.1%	8	305	2.6%	1.37E-05	1.64E-04	*
KMT2D	20	149	13.4%	43	305	14.1%	8.86E-01	9.25E-01	
EP300	18	149	12.1%	37	305	12.1%	1.00E+00	1.00E+00	
FGFR3	17	149	11.4%	0	305	0.0%	3.12E-09	7.49E-08	*
CASZ1	14	149	9.4%	5	305	1.6%	2.36E-04	1.89E-03	*
PTEN	13	149	8.7%	20	305	6.6%	4.43E-01	6.73E-01	
CYLD	11	149	7.4%	3	305	1.0%	4.89E-04	2.94E-03	*
FBXW7	11	149	7.4%	35	305	11.5%	1.89E-01	4.14E-01	
NSD1	10	149	6.7%	19	305	6.2%	8.40E-01	9.16E-01	
RB1	9	149	6.0%	16	305	5.2%	8.27E-01	9.16E-01	
RBL1	9	149	6.0%	7	305	2.3%	5.64E-02	1.69E-01	
ASAP1	9	149	6.0%	6	305	2.0%	4.56E-02	1.56E-01	
DDX3X	8	149	5.4%	10	305	3.3%	3.10E-01	5.32E-01	
RIPK4	6	149	4.0%	1	305	0.3%	5.93E-03	2.85E-02	*
TRAF3	5	149	3.4%	7	305	2.3%	5.40E-01	7.63E-01	
BBX	5	149	3.4%	4	305	1.3%	1.62E-01	4.14E-01	
SLTM	5	149	3.4%	2	305	0.7%	4.11E-02	1.56E-01	
TGFBR2	5	149	3.4%	8	305	2.6%	7.66E-01	9.16E-01	
NFKBIA	4	149	2.7%	3	305	1.0%	2.24E-01	4.14E-01	
IFNGR1	4	149	2.7%	7	305	2.3%	7.56E-01	9.16E-01	
TAF5	4	149	2.7%	3	305	1.0%	2.24E-01	4.14E-01	
HIST1H2AE	4	149	2.7%	3	305	1.0%	2.24E-01	4.14E-01	
STAT1	4	149	2.7%	6	305	2.0%	7.35E-01	9.16E-01	

Supplemental Table S3H. Comparison of mutation frequencies in 24 HPV-positive OSCC MutSig genes (HPV-positive OSCC vs. cervical cancer).
Shown are the number of HPV-positive OSCC (*left*) and cervical cancers (*right*) with mutations in 24 genes, originally identified as significantly mutated in HPV-positive OSCC by MutSig. The list of somatic variants in cervical cancers was downloaded from TCGA. Comparisons were evaluated using Fisher's Exact test; p-values were corrected for multiple testing using FDR method. *Asterisk*, FDR adjusted p-value <0.05 (see also **Supplemental Fig S3**). P-values and adjusted P-values are presented in scientific notation; E represents *10^.

	HPV-positive oral cancer			Cervical cancer				
gene	no. samples with mutation	no. total samples	fraction of mutated samples (%total)	no. samples with mutation	no. total samples	fraction of mutated samples (%total)	p-value	FDR adjusted p-value
PIK3CA	42	149	28.2%	98	305	32.1%	4.49E-01	7.86E-01
EP300	18	149	12.1%	37	305	12.1%	1.00E+00	1.00E+00
FBXW7	11	149	7.4%	35	305	11.5%	1.89E-01	6.63E-01
PTEN	13	149	8.7%	20	305	6.6%	4.43E-01	7.86E-01
HLA-A	3	149	2.0%	18	305	5.9%	9.32E-02	4.35E-01
ARID1A	5	149	3.4%	19	305	6.2%	2.65E-01	7.42E-01
NFE2L2	4	149	2.7%	16	305	5.2%	3.29E-01	7.68E-01
HLA-B	4	149	2.7%	11	305	3.6%	7.82E-01	9.96E-01
KRAS	1	149	0.7%	17	305	5.6%	9.59E-03	1.34E-01
ERBB3	5	149	3.4%	14	305	4.6%	6.25E-01	9.73E-01
MAPK1	6	149	4.0%	12	305	3.9%	1.00E+00	1.00E+00
CASP8	2	149	1.3%	15	305	4.9%	6.76E-02	4.35E-01
TGFBR2	5	149	3.4%	8	305	2.6%	7.66E-01	9.96E-01
SHKBP1	1	149	0.7%	3	305	1.0%	1.00E+00	1.00E+00

Supplemental Table S3I. Comparison of mutation frequencies in 14 genes recurrently mutated in cervical cancer as determined by MutSig in HPV-positive OSCC vs. cervical cancer.
Shown are the number of HPV-positive OSCC (left) and cervical cancers (right) with mutations in 14 genes, originally identified as significantly mutated by MutSig in cervical cancers available in TCGA. The list of somatic variants in cervical cancers was downloaded from TCGA. Comparisons were evaluated using Fisher's Exact test; p-values were corrected for multiple testing using FDR method. * indicates FDR adjusted p-value <0.05 (see also **Supplemental Fig S3**).

sample ID	**chr.**	**position**	**reference allele**	**alternative allele**	**gene**	**transcript ID**	**amino acid position**	**amino acid change**	**consequence**	**annotated variant**	novel In OSCC?
GS18026	1	10,700,041	C	T	*CASZ1*	ENST00000377022.3	1413	G/D	missense	NA	Y
TCGA-CR-6481	1	10,702,951	G	A	*CASZ1*	ENST00000377022.3	1376	S/F	missense	NA	Y
TCGA-CR-6472	1	10,707,968	C	T	*CASZ1*	ENST00000377022.3	1129	M/I	missense	NA	Y
GS18015	1	10,710,804	G	C	*CASZ1*	ENST00000377022.3	942	S/*	stop gained	NA	Y
GS18021	1	10,711,096	C	CG	*CASZ1*	ENST00000377022.3	906	P/PX	frameshift	NA	Y
GS18001	1	10,714,023	G	C	*CASZ1*	ENST00000377022.3	697	I/M	missense	rs763476319	Y
TCGA-CR-7369	1	10,715,709	C	CA	*CASZ1*	ENST00000377022.3	554	M/IX	frameshift	NA	Y
GS18034	1	10,715,713	C	A	*CASZ1*	ENST00000377022.3	553	C/F	missense	COSM4976501	Y
TCGA-CV-7406	1	10,715,844	GATCACGTCC TGCT	G	*CASZ1*	ENST00000377022.3	505-509	KQDVI/X	frameshift	NA	Y
GS18012	1	10,715,871	C	A	*CASZ1*	ENST00000377022.3	NA	NA	splice acceptor	NA	Y *
TCGA-BB-4223	1	10,715,871	C	T	*CASZ1*	ENST00000377022.3	NA	NA	splice acceptor	NA	Y
TCGA-HD-7754	1	10,715,871	C	A	*CASZ1*	ENST00000377022.3	NA	NA	splice acceptor	NA	Y *
TCGA-CR-7369	1	10,720,212	G	A	*CASZ1*	ENST00000377022.3	296	S/L	missense	NA	Y
TCGA-P3-A6SW	1	10,720,504	C	T	*CASZ1*	ENST00000377022.3	199	E/K	missense	COSM6620722	Y
TCGA-HL-7533	1	10,725,203	C	G	*CASZ1*	ENST00000377022.3	148	E/Q	missense	COSM5599247	Y
GS18008	2	191,859,791	G	A	*STAT1*	ENST00000361099.3	314	Q/*	stop gained	NA	Y
GS18007	2	191,859,923	G	C	*STAT1*	ENST00000361099.3	270	L/V	missense	NA	Y
TCGA-CV-5442	2	191,873,691	G	A	*STAT1*	ENST00000361099.3	91	Q/*	stop gained	NA	Y
TCGA-CV-7406	2	191,874,723	G	A	*STAT1*	ENST00000361099.3	3	Q/*	stop gained	NA	Y
TCGA-CV-6961	3	30,691,947	C	G	*TGFBR2*	ENST00000359013.4	175	S/*	stop gained	NA	Y
GS18028	3	30,713,559	C	G	*TGFBR2*	ENST00000359013.4	320	S/*	stop gained	COSM730357	Y
GS18021	3	30,715,678	G	A	*TGFBR2*	ENST00000359013.4	471	D/N	missense	COSM1593641	Y
TCGA-DQ-7596	3	30,732,969	C	T	*TGFBR2*	ENST00000359013.4	553	R/C	missense	COSM1650139	Y
TCGA-P3-A5QF	3	30,733,072	C	T	*TGFBR2*	ENST00000359013.4	587	S/F	missense	COSM2983548	Y
GS18012	3	107,491,505	G	T	*BBX*	ENST00000325805.8	313	E/*	stop gained	NA	Y
GS18015	3	107,491,607	G	C	*BBX*	ENST00000325805.8	347	E/Q	missense	NA	Y
GS18001	3	107,491,637	G	C	*BBX*	ENST00000325805.8	357	E/Q	missense	NA	Y
TCGA-BA-4077	3	107,491,933	G	C	*BBX*	ENST00000325805.8	455	K/N	missense	NA	Y
TCGA-BA-4077	3	107,492,069	G	C	*BBX*	ENST00000325805.8	501	E/Q	missense	COSM159345	Y
TCGA-BA-4077	3	107,492,081	G	C	*BBX*	ENST00000325805.8	505	D/H	missense	NA	Y
TCGA-BA-4077	3	107,492,134	G	C	*BBX*	ENST00000325805.8	522	L/F	missense	NA	Y
TCGA-BA-4077	3	107,492,151	AG	A	*BBX*	ENST00000325805.8	528	K/X	frameshift	NA	Y

TCGA-BA-4077	3	107,492,169	G	C	*BBX*	ENST00000325805.8	534	R/T	missense	NA	Y
TCGA-BA-4077	3	107,492,201	G	A	*BBX*	ENST00000325805.8	545	D/N	missense	NA	Y
GS18007	3	107,492,217	C	G	*BBX*	ENST00000325805.8	550	S/*	stop gained	NA	Y
GS18001	3	107,519,974	C	T	*BBX*	ENST00000325805.8	862	L/F	missense	COSM1417581	Y
TCGA-CR-6471	3	178,916,876	G	A	*PIK3CA*	ENST00000263967.3	88	R/Q	missense	COSM271684	N
GS18070	3	178,936,082	G	A	*PIK3CA*	ENST00000263967.3	542	E/K	missense	COSM125369	N
GS18033	3	178,936,082	G	A	*PIK3CA*	ENST00000263967.3	542	E/K	missense	COSM125369	N
GS18018	3	178,936,082	G	A	*PIK3CA*	ENST00000263967.3	542	E/K	missense	COSM125369	N
GS18009	3	178,936,082	G	A	*PIK3CA*	ENST00000263967.3	542	E/K	missense	COSM125369	N
GS18013	3	178,936,082	G	A	*PIK3CA*	ENST00000263967.3	542	E/K	missense	COSM125369	N
TCGA-CR-5249	3	178,936,082	G	A	*PIK3CA*	ENST00000263967.3	542	E/K	missense	COSM125369	N
TCGA-BB-4223	3	178,936,082	G	A	*PIK3CA*	ENST00000263967.3	542	E/K	missense	COSM125369	N
TCGA-CR-5243	3	178,936,082	G	A	*PIK3CA*	ENST00000263967.3	542	E/K	missense	COSM125369	N
TCGA-HD-7754	3	178,936,082	G	A	*PIK3CA*	ENST00000263967.3	542	E/K	missense	COSM125369	N
TCGA-HD-7832	3	178,936,082	G	A	*PIK3CA*	ENST00000263967.3	542	E/K	missense	COSM125369	N
TCGA-TN-A7HI	3	178,936,082	G	A	*PIK3CA*	ENST00000263967.3	542	E/K	missense	COSM125369	N
GS18059	3	178,936,091	G	A	*PIK3CA*	ENST00000263967.3	545	E/K	missense	COSM125370	N
GS18051	3	178,936,091	G	A	*PIK3CA*	ENST00000263967.3	545	E/K	missense	COSM125370	N
GS18016	3	178,936,091	G	A	*PIK3CA*	ENST00000263967.3	545	E/K	missense	COSM125370	N
GS18010	3	178,936,091	G	A	*PIK3CA*	ENST00000263967.3	545	E/K	missense	COSM125370	N
GS18038	3	178,936,091	G	A	*PIK3CA*	ENST00000263967.3	545	E/K	missense	COSM125370	N
GS18023	3	178,936,091	G	A	*PIK3CA*	ENST00000263967.3	545	E/K	missense	COSM125370	N
GS18028	3	178,936,091	G	A	*PIK3CA*	ENST00000263967.3	545	E/K	missense	COSM125370	N
GS18003	3	178,936,091	G	A	*PIK3CA*	ENST00000263967.3	545	E/K	missense	COSM125370	N
GS18009	3	178,936,091	G	A	*PIK3CA*	ENST00000263967.3	545	E/K	missense	COSM125370	N
GS18014	3	178,936,091	G	A	*PIK3CA*	ENST00000263967.3	545	E/K	missense	COSM125370	N
GS18029	3	178,936,091	G	A	*PIK3CA*	ENST00000263967.3	545	E/K	missense	COSM125370	N
GS18031	3	178,936,091	G	A	*PIK3CA*	ENST00000263967.3	545	E/K	missense	COSM125370	N
GS18020	3	178,936,091	G	A	*PIK3CA*	ENST00000263967.3	545	E/K	missense	COSM125370	N
GS18048	3	178,936,091	G	A	*PIK3CA*	ENST00000263967.3	545	E/K	missense	COSM125370	N
GS18012	3	178,936,091	G	A	*PIK3CA*	ENST00000263967.3	545	E/K	missense	COSM125370	N
GS18030	3	178,936,091	G	C	*PIK3CA*	ENST00000263967.3	545	E/Q	missense	COSM27133	N
GS18002	3	178,936,091	G	A	*PIK3CA*	ENST00000263967.3	545	E/K	missense	COSM125370	N
TCGA-BA-4077	3	178,936,091	G	A	*PIK3CA*	ENST00000263967.3	545	E/K	missense	COSM125370	N
TCGA-CN-4741	3	178,936,091	G	A	*PIK3CA*	ENST00000263967.3	545	E/K	missense	COSM125370	N
TCGA-CR-6472	3	178,936,091	G	A	*PIK3CA*	ENST00000263967.3	545	E/K	missense	COSM125370	N

TCGA-CR-6487	3	178,936,091	G	A	*PIK3CA*	ENST00000263967.3	545	E/K	missense	COSM125370	N
TCGA-CR-7385	3	178,936,091	G	A	*PIK3CA*	ENST00000263967.3	545	E/K	missense	COSM125370	N
TCGA-CR-7404	3	178,936,091	G	A	*PIK3CA*	ENST00000263967.3	545	E/K	missense	COSM125370	N
TCGA-CV-5442	3	178,936,091	G	A	*PIK3CA*	ENST00000263967.3	545	E/K	missense	COSM125370	N
TCGA-CV-7406	3	178,936,091	G	A	*PIK3CA*	ENST00000263967.3	545	E/K	missense	COSM125370	N
TCGA-HL-7533	3	178,936,091	G	A	*PIK3CA*	ENST00000263967.3	545	E/K	missense	COSM125370	N
TCGA-MZ-A6I9	3	178,936,091	G	A	*PIK3CA*	ENST00000263967.3	545	E/K	missense	COSM125370	N
TCGA-P3-A5QF	3	178,936,091	G	A	*PIK3CA*	ENST00000263967.3	545	E/K	missense	COSM125370	N
GS18010	3	178,938,922	G	A	*PIK3CA*	ENST00000263967.3	722	E/K	missense	COSM1192505	Y
GS18021	3	178,938,934	G	A	*PIK3CA*	ENST00000263967.3	726	E/K	missense	COSM446019	N
GS18008	3	178,943,756	G	A	*PIK3CA*	ENST00000263967.3	808	R/Q	missense	rs780837825	Y
TCGA-CR-6471	3	178,952,072	A	G	*PIK3CA*	ENST00000263967.3	1043	M/V	missense	COSM94983	N
GS18078	3	178,952,085	A	G	*PIK3CA*	ENST00000263967.3	1047	H/R	missense	COSM94986	N
GS18007	4	1,803,117	G	A	*FGFR3*	ENST00000340107.4	157	E/K	missense	NA	Y
GS18038	4	1,803,564	C	T	*FGFR3*	ENST00000340107.4	248	R/C	missense	COSM1133721	N
TCGA-QK-A6V9	4	1,803,564	C	T	*FGFR3*	ENST00000340107.4	248	R/C	missense	COSM1133721	N
GS18016	4	1,803,568	C	G	*FGFR3*	ENST00000340107.4	249	S/C	missense	COSM1149892	N
GS18034	4	1,803,568	C	G	*FGFR3*	ENST00000340107.4	249	S/C	missense	COSM1149892	N
GS18015	4	1,803,568	C	G	*FGFR3*	ENST00000340107.4	249	S/C	missense	COSM1149892	N
GS18044	4	1,803,568	C	G	*FGFR3*	ENST00000340107.4	249	S/C	missense	COSM1149892	N
GS18008	4	1,803,568	C	G	*FGFR3*	ENST00000340107.4	249	S/C	missense	COSM1149892	N
GS18029	4	1,803,568	C	G	*FGFR3*	ENST00000340107.4	249	S/C	missense	COSM1149892	N
GS18046	4	1,803,568	C	G	*FGFR3*	ENST00000340107.4	249	S/C	missense	COSM1149892	N
GS18012	4	1,803,568	C	G	*FGFR3*	ENST00000340107.4	249	S/C	missense	COSM1149892	N
TCGA-BA-A4IH	4	1,803,568	C	G	*FGFR3*	ENST00000340107.4	249	S/C	missense	COSM1149892	N
TCGA-CN-A6V7	4	1,803,568	C	G	*FGFR3*	ENST00000340107.4	249	S/C	missense	COSM1149892	N
TCGA-CR-6481	4	1,803,568	C	G	*FGFR3*	ENST00000340107.4	249	S/C	missense	COSM1149892	N
TCGA-P3-A5QF	4	1,803,568	C	G	*FGFR3*	ENST00000340107.4	249	S/C	missense	COSM1149892	N
TCGA-DQ-7594	4	1,806,153	C	A	*FGFR3*	ENST00000340107.4	393	A/E	missense	COSM5945996	Y
TCGA-QK-A6V9	4	1,808,029	C	G	*FGFR3*	ENST00000340107.4	671	R/G	missense	NA	Y
TCGA-CV-5442	4	1,808,852	G	C	*FGFR3*	ENST00000340107.4	764	D/H	missense	NA	Y
TCGA-CV-5442	4	1,808,924	G	A	*FGFR3*	ENST00000340107.4	788	D/N	missense	rs548817695	Y
GS18034	4	153,247,288	C	T	*FBXW7*	ENST00000281708.4	505	R/H	missense	COSM287373	N
GS18096	4	153,247,289	G	T	*FBXW7*	ENST00000281708.4	505	R/S	missense	COSM732382	Y
GS18010	4	153,247,289	G	A	*FBXW7*	ENST00000281708.4	505	R/C	missense	COSM108571	N
GS18018	4	153,247,289	G	C	*FBXW7*	ENST00000281708.4	505	R/G	missense	COSM99605	N

TCGA-BB-4228	4	153,247,289	G	C	*FBXW7*	ENST00000281708.4	505	R/G	missense	COSM99605	N
TCGA-CN-A6V7	4	153,247,289	G	C	*FBXW7*	ENST00000281708.4	505	R/G	missense	COSM99605	N
GS18076	4	153,247,366	C	T	*FBXW7*	ENST00000281708.4	479	R/Q	missense	COSM22974	N
GS18067	4	153,247,366	C	T	*FBXW7*	ENST00000281708.4	479	R/Q	missense	COSM22974	N
TCGA-P3-A5QF	4	153,247,366	C	T	*FBXW7*	ENST00000281708.4	479	R/Q	missense	COSM22974	N
GS18015	4	153,247,367	G	C	*FBXW7*	ENST00000281708.4	479	R/G	missense	COSM22966	N
GS18007	4	153,247,367	G	A	*FBXW7*	ENST00000281708.4	479	R/*	stop gained	COSM206697	Y
TCGA-CV-6961	5	176,562,391	C	G	*NSD1*	ENST00000439151.2	96	S/C	missense	NA	Y
GS18002	5	176,637,154	C	T	*NSD1*	ENST00000439151.2	585	S/L	missense	CM056665	Y
GS18009	5	176,637,651	C	G	*NSD1*	ENST00000439151.2	751	L/V	missense	rs775181583	Y
GS18018	5	176,637,673	C	G	*NSD1*	ENST00000439151.2	758	S/*	stop gained	NA	Y
TCGA-CR-6481	5	176,673,765	G	A	*NSD1*	ENST00000439151.2	1489	D/N	missense	NA	Y
TCGA-P3-A6T6	5	176,675,230	G	T	*NSD1*	ENST00000439151.2	1516	E/*	stop gained	NA	Y
TCGA-CR-7385	5	176,687,121	C	T	*NSD1*	ENST00000439151.2	1700	R/*	stop gained	rs587784137	Y
GS18004	5	176,696,674	G	C	*NSD1*	ENST00000439151.2	1792	G/A	missense	CM030077	Y
GS18036	5	176,700,729	C	T	*NSD1*	ENST00000439151.2	1856	Q/*	stop gained	rs587784151	Y
TCGA-P3-A5QF	5	176,710,792	G	A	*NSD1*	ENST00000439151.2	2005	R/Q	missense	rs587784174	Y
TCGA-RS-A6TP	6	26,217,210	G	A	*HIST1H2AE*	ENST00000303910.2	3	G/E	missense	NA	Y
GS18017	6	26,217,216	G	A	*HIST1H2AE*	ENST00000303910.2	5	G/E	missense	COSM1292266	Y
GS18014	6	26,217,272	C	G	*HIST1H2AE*	ENST00000303910.2	24	L/V	missense	NA	Y
GS18062	6	26,217,427	G	C	*HIST1H2AE*	ENST00000303910.2	75	K/N	missense	COSM740883	Y
GS18005	6	137,519,355	G	C	*IFNGR1*	ENST00000367739.4	428	S/*	stop gained	NA	Y
TCGA-CR-5248	6	137,519,504	ACT	A	*IFNGR1*	ENST00000367739.4	378	S/X	frameshift	COSM1073785	Y
GS18006	6	137,524,735	G	A	*IFNGR1*	ENST00000367739.4	212	Q/*	stop gained	NA	Y
TCGA-CR-6481	6	137,527,279	G	A	*IFNGR1*	ENST00000367739.4	123	R/*	stop gained	NA	Y
GS18004	8	131,073,109	C	T	*ASAP1*	ENST00000357668.1	970	D/N	missense	NA	Y
TCGA-DQ-7594	8	131,073,207	C	G	*ASAP1*	ENST00000357668.1	937	G/A	missense	rs750552166	Y
GS18004	8	131,073,232	C	T	*ASAP1*	ENST00000357668.1	929	E/K	missense	NA	Y
TCGA-BA-5559	8	131,127,917	C	T	*ASAP1*	ENST00000357668.1	710	R/Q	missense	NA	Y
TCGA-CR-6471	8	131,130,412	T	G	*ASAP1*	ENST00000357668.1	625	Q/H	missense	NA	Y
TCGA-BA-5153	8	131,130,883	T	G	*ASAP1*	ENST00000357668.1	549	D/A	missense	NA	Y
GS18005	8	131,146,584	C	G	*ASAP1*	ENST00000357668.1	392	R/T	missense	COSM749390	Y
TCGA-HD-7832	8	131,172,152	T	C	*ASAP1*	ENST00000357668.1	323	Y/C	missense	COSM3896762	Y
TCGA-HL-7533	8	131,179,784	C	G	*ASAP1*	ENST00000357668.1	303	E/Q	missense	NA	Y
GS18012	8	131,414,177	C	G	*ASAP1*	ENST00000357668.1	5	A/P	missense	NA	Y
GS18072	10	89,624,275	C	T	*PTEN*	ENST00000371953.3	17	Q/*	stop gained	COSM5153	N

GS18044	10	89,624,275	C	T	*PTEN*	ENST00000371953.3	17	Q/*	stop gained	COSM5153	N
GS18001	10	89,653,790	C	T	*PTEN*	ENST00000371953.3	30	P/S	missense	COSM6503486	Y
TCGA-CN-A499	10	89,653,856	G	GA	*PTEN*	ENST00000371953.3	52	D/EX	frameshift	NA	Y
TCGA-CR-7369	10	89,692,794	A	G	*PTEN*	ENST00000371953.3	93	H/R	missense	COSM5060	Y
GS18027	10	89,692,896	G	T	*PTEN*	ENST00000371953.3	127	G/V	missense	COSM5285	Y
TCGA-T2-A6X0	10	89,692,904	C	T	*PTEN*	ENST00000371953.3	130	R/*	stop gained	COSM5152	N
TCGA-BB-A6UM	10	89,692,905	G	A	*PTEN*	ENST00000371953.3	130	R/Q	missense	COSM5033	N
GS18079	10	89,692,911	G	T	*PTEN*	ENST00000371953.3	132	G/V	missense	COSM5229	Y
TCGA-BA-5559	10	89,692,944	G	A	*PTEN*	ENST00000371953.3	143	G/D	missense	rs786202047	Y
GS18033	10	89,717,777	G	T	*PTEN*	ENST00000371953.3	NA	NA	splice donor	COSM5970	Y
GS18101	10	89,720,757	TTTGCAGTATAGAGCGTG	T	*PTEN*	ENST00000371953.3	303-309	ICSIERA/IX	frameshift	NA	Y
GS18004	10	89,725,179	G	A	*PTEN*	ENST00000371953.3	388	E/K	missense	NA	Y
TCGA-P3-A6T6	10	105,127,981	GC	G	*TAF5*	ENST00000369839.3	79	A/X	frameshift	NA	Y
GS18006	10	105,138,176	C	T	*TAF5*	ENST00000369839.3	328	Q/*	stop gained	NA	Y
GS18006	10	105,138,303	C	G	*TAF5*	ENST00000369839.3	370	S/*	stop gained	NA	Y
TCGA-CR-6472	10	105,139,422	G	T	*TAF5*	ENST00000369839.3	391	E/*	stop gained	NA	Y
TCGA-CQ-5323	10	105,146,990	G	T	*TAF5*	ENST00000369839.3	630	V/L	missense	NA	Y
TCGA-CR-5243	12	49,418,485	C	T	*KMT2D*	ENST00000301067.7	5310	G/R	missense	NA	Y
GS18023	12	49,420,322	C	T	*KMT2D*	ENST00000301067.7	5143	E/K	missense	NA	Y
GS18027	12	49,420,670	G	A	*KMT2D*	ENST00000301067.7	5027	R/*	stop gained	COSM6917461	Y
GS18006	12	49,423,015	C	A	*KMT2D*	ENST00000301067.7	4694	E/*	stop gained	NA	Y
TCGA-RS-A6TP	12	49,423,200	G	A	*KMT2D*	ENST00000301067.7	4687	Q/*	stop gained	NA	Y
GS18027	12	49,425,638	G	A	*KMT2D*	ENST00000301067.7	4284	Q/*	stop gained	COSM459483	Y
GS18046	12	49,425,644	G	A	*KMT2D*	ENST00000301067.7	4282	R/*	stop gained	COSM3461536	Y
TCGA-CN-A6V6	12	49,428,219	G	A	*KMT2D*	ENST00000301067.7	3494	P/L	missense	rs746854405	Y
GS18035	12	49,432,429	GTC	G	*KMT2D*	ENST00000301067.7	2903	R/X	frameshift	NA	Y
TCGA-DQ-7593	12	49,433,051	G	A	*KMT2D*	ENST00000301067.7	2774	R/W	missense	rs779166504	Y
GS18063	12	49,433,388	G	A	*KMT2D*	ENST00000301067.7	2687	R/*	stop gained	COSM220686	Y
TCGA-P3-A5QF	12	49,433,935	G	A	*KMT2D*	ENST00000301067.7	2540	Q/*	stop gained	NA	Y
TCGA-CV-5442	12	49,436,060	G	A	*KMT2D*	ENST00000301067.7	1974	T/M	missense	rs777415982	Y
GS18009	12	49,438,578	C	T	*KMT2D*	ENST00000301067.7	1638	D/N	missense	COSM4990265	Y
TCGA-CR-6487	12	49,438,595	G	C	*KMT2D*	ENST00000301067.7	1632	S/*	stop gained	CD114126	Y
TCGA-CN-4741	12	49,443,899	C	A	*KMT2D*	ENST00000301067.7	1158	E/*	stop gained	COSM693654	Y
TCGA-BA-4077	12	49,443,980	G	C	*KMT2D*	ENST00000301067.7	1131	P/A	missense	NA	Y
GS18014	12	49,445,251	C	T	*KMT2D*	ENST00000301067.7	739	E/K	missense	rs587783706	Y

GS18028	12	49,445,257	G	A	*KMT2D*	ENST00000301067.7	737	R/W	missense	COSM5675122	Y
GS18009	12	49,445,734	C	T	*KMT2D*	ENST00000301067.7	578	E/K	missense	NA	Y
TCGA-CQ-5323	12	49,447,022	C	A	*KMT2D*	ENST00000301067.7	308	E/*	stop gained	NA	Y
GS18001	12	49,447,041	G	C	*KMT2D*	ENST00000301067.7	301	F/L	missense	COSM6921946	Y
GS18036	13	48,916,765	TG	T	*RB1*	ENST00000267163.4	99	W/X	frameshift	CM951102	Y
TCGA-KU-A6H7	13	48,936,983	C	T	*RB1*	ENST00000267163.4	251	R/*	stop gained	COSM1367181	N
GS18036	13	48,939,078	G	T	*RB1*	ENST00000267163.4	304	G/*	stop gained	COSM6965278	Y
TCGA-CN-A499	13	48,941,648	C	T	*RB1*	ENST00000267163.4	320	R/*	stop gained	COSM1152653	Y
GS18077	13	48,941,654	G	A	*RB1*	ENST00000267163.4	322	E/K	missense	COSM1367200	Y
TCGA-DQ-7594	13	48,947,540	G	A	*RB1*	ENST00000267163.4	NA	NA	splice acceptor	COSM5752556	Y
TCGA-CR-6487	13	48,954,379	T	G	*RB1*	ENST00000267163.4	NA	NA	splice donor	CS973319	Y
TCGA-QK-A6V9	13	49,027,168	C	T	*RB1*	ENST00000267163.4	579	R/*	stop gained	COSM1756816	N
TCGA-CV-6939	13	49,039,219	CA	C	*RB1*	ENST00000267163.4	766	T/X	frameshift	NA	Y
GS18007	13	49,047,507	C	G	*RB1*	ENST00000267163.4	834	S/*	stop gained	COSM1942437	Y
GS18039	14	35,871,987	ACATCAG	A	*NFKBIA*	ENST00000216797.5	207-209	ADV/V	inframe deletion	NA	Y
GS18098	14	35,872,355	C	A	*NFKBIA*	ENST00000216797.5	NA	NA	splice donor	NA	Y
GS18079	14	35,872,460	A	AG	*NFKBIA*	ENST00000216797.5	148	L/PX	frameshift	NA	Y
TCGA-MZ-A5BI	14	35,873,703	G	A	*NFKBIA*	ENST00000216797.5	50	Q/*	stop gained	NA	Y
GS18003	14	103,352,541	G	A	*TRAF3*	ENST00000560371.1	196	D/N	missense	rs142439625	Y
TCGA-CR-5250	14	103,352,584	TC	T	*TRAF3*	ENST00000560371.1	210	V/X	frameshift	NA	Y
TCGA-BB-4228	14	103,363,706	C	T	*TRAF3*	ENST00000560371.1	310	R/*	stop gained	COSM144682	Y
GS18096	14	103,363,739	G	A	*TRAF3*	ENST00000560371.1	NA	NA	splice donor	NA	Y
GS18019	14	103,369,592	C	T	*TRAF3*	ENST00000560371.1	321	R/*	stop gained	NA	Y
GS18103	15	59,179,314	C	T	*SLTM*	ENST00000380516.2	899	E/K	missense	NA	Y
TCGA-CV-6961	15	59,181,735	C	T	*SLTM*	ENST00000380516.2	700	E/K	missense	NA	Y
GS18010	15	59,182,521	C	G	*SLTM*	ENST00000380516.2	680	E/Q	missense	NA	Y
GS18010	15	59,185,210	C	G	*SLTM*	ENST00000380516.2	595	R/T	missense	COSM6538321	Y
TCGA-CR-5248	15	59,186,128	CT	C	*SLTM*	ENST00000380516.2	513	K/X	frameshift	NA	Y
TCGA-KU-A6H7	15	59,192,011	C	T	*SLTM*	ENST00000380516.2	239	E/K	missense	NA	Y
TCGA-CN-5374	16	50,788,321	A	G	*CYLD*	ENST00000427738.3	300	N/S	missense	rs772008589	Y
GS18078	16	50,810,125	G	T	*CYLD*	ENST00000427738.3	320	A/S	missense	NA	Y
TCGA-P3-A5QE	16	50,811,808	C	G	*CYLD*	ENST00000427738.3	365	S/*	stop gained	COSM1737934	Y
TCGA-BA-5559	16	50,811,826	C	A	*CYLD*	ENST00000427738.3	371	S/*	stop gained	COSM6454818	N
TCGA-CR-6470	16	50,818,266	A	C	*CYLD*	ENST00000427738.3	618	D/A	missense	NA	Y
TCGA-CN-A6UY	16	50,818,322	C	CA	*CYLD*	ENST00000427738.3	637	Q/QX	frameshift	NA	Y

GS18093	16	50,820,802	GAGGAAA	G	*CYLD*	ENST00000427738.3	663-664	RK/-	inframe deletion	NA	Y	
TCGA-BA-5153	16	50,820,854	A	T	*CYLD*	ENST00000427738.3	680	K/*	stop gained	NA	Y	
GS18053	16	50,825,494	T	C	*CYLD*	ENST00000427738.3	712	C/R	missense	NA	Y	
GS18094	16	50,828,333	C	T	*CYLD*	ENST00000427738.3	894	R/W	missense	COSM4604929	N	
TCGA-DQ-7593	16	50,830,344	A	AGG	*CYLD*	ENST00000427738.3	932-933	-/X	frameshift	NA	Y	
TCGA-CR-6481	17	80,789,280	C	T	*ZNF750*	ENST00000269394.3	351	E/K	missense	COSM7170686	Y	
GS18109	17	80,789,483	TGTCTCGGGTCTTGGGTTCCGTAGACTGACAGCAGGGGTGC	T	*ZNF750*	ENST00000269394.3	270-283	APLLSVYGTQDPRH/X	frameshift	NA	Y	
GS18008	17	80,789,502	C	CGTAGACTG	*ZNF750*	ENST00000269394.3	276-277	-/QSX	frameshift	NA	Y	
GS18020	17	80,789,502	C	A	*ZNF750*	ENST00000269394.3	277	G/*	stop gained	NA	Y	
TCGA-P3-A5QF	17	80,789,699	G	C	*ZNF750*	ENST00000269394.3	211	S/*	stop gained	NA	Y	
GS18004	17	80,789,889	CGAGG	C	*ZNF750*	ENST00000269394.3	146-147	AL/X	frameshift	NA	Y	
GS18021	17	80,789,924	G	T	*ZNF750*	ENST00000269394.3	136	S/*	stop gained	COSM4833863	Y	*
GS18007	17	80,789,924	G	T	*ZNF750*	ENST00000269394.3	136	S/*	stop gained	COSM4833863	Y	*
GS18012	17	80,789,963	T	TGG	*ZNF750*	ENST00000269394.3	123	H/PX	frameshift	NA	Y	
GS18001	17	80,790,024	G	A	*ZNF750*	ENST00000269394.3	103	Q/*	stop gained	COSM708549	Y	
GS18005	17	80,790,032	G	T	*ZNF750*	ENST00000269394.3	100	S/*	stop gained	COSM4426914	Y	
GS18018	17	80,790,042	CAG	C	*ZNF750*	ENST00000269394.3	96	S/X	frameshift	COSM6841808	N	
TCGA-HL-7533	17	80,790,115	G	GTC	*ZNF750*	ENST00000269394.3	72	D/EX	frameshift	NA	Y	
TCGA-HL-7533	17	80,790,149	C	T	*ZNF750*	ENST00000269394.3	61	R/Q	missense	rs375467494	Y	
TCGA-CR-6481	17	80,790,150	G	A	*ZNF750*	ENST00000269394.3	61	R/*	stop gained	COSM5285736	Y	
GS18029	17	80,790,178	GTTTT	G	*ZNF750*	ENST00000269394.3	50-51	KN/X	frameshift	NA	Y	
TCGA-CV-6961	17	80,790,178	GT	G	*ZNF750*	ENST00000269394.3	51	N/X	frameshift	NA	Y	
GS18019	17	80,790,189	GACCATACTTC	G	*ZNF750*	ENST00000269394.3	44-47	MKYG/X	frameshift	NA	Y	
GS18017	17	80,790,190	ACC	A	*ZNF750*	ENST00000269394.3	47	G/X	frameshift	NA	Y	
GS18009	17	80,790,216	GTGACT	G	*ZNF750*	ENST00000269394.3	37-38	KS/X	frameshift	NA	Y	
TCGA-CR-7368	17	80,790,225	C	CA	*ZNF750*	ENST00000269394.3	35-36	-/X	frameshift	NA	Y	
TCGA-KU-A6H7	17	80,790,229	G	T	*ZNF750*	ENST00000269394.3	34	C/*	stop gained	NA	Y	
TCGA-DQ-7594	17	80,790,328	C	T	*ZNF750*	ENST00000269394.3	1	M/I	initiator codon	COSM5440618	Y	
GS18005	20	35,627,333	C	G	*RBL1*	ENST00000373664.3	NA	NA	splice acceptor	NA	Y	
TCGA-TN-A7HI	20	35,635,822	C	T	*RBL1*	ENST00000373664.3	955	D/N	missense	NA	Y	
GS18001	20	35,661,233	C	T	*RBL1*	ENST00000373664.3	739	M/I	missense	NA	Y	

GS18008	20	35,661,259	C	T	*RBL1*	ENST00000373664.3	731	E/K	missense	NA	Y	
TCGA-CV-7406	20	35,668,562	G	A	*RBL1*	ENST00000373664.3	633	R/*	stop gained	rs201261568	Y	
GS18029	20	35,672,548	C	G	*RBL1*	ENST00000373664.3	571	D/H	missense	NA	Y	
GS18002	20	35,690,618	C	G	*RBL1*	ENST00000373664.3	318	E/Q	missense	NA	Y	
TCGA-CR-5248	20	35,696,531	C	G	*RBL1*	ENST00000373664.3	117	E/Q	missense	NA	Y	
GS18015	20	35,717,416	CTGG	C	*RBL1*	ENST00000373664.3	88-89	TR/R	inframe deletion	NA	Y	
TCGA-BB-4228	21	43,161,351	A	G	*RIPK4*	ENST00000332512.3	668	S/P	missense	NA	Y	
GS18015	21	43,161,417	C	G	*RIPK4*	ENST00000332512.3	646	E/Q	missense	NA	Y	
GS18044	21	43,162,082	TTCA	T	*RIPK4*	ENST00000332512.3	423-424	MK/K	inframe deletion	NA	Y	
TCGA-CR-5248	21	43,165,984	C	T	*RIPK4*	ENST00000332512.3	291	D/N	missense	COSM1751660	Y	
GS18017	21	43,171,401	G	C	*RIPK4*	ENST00000332512.3	160	S/C	missense	NA	Y	
TCGA-CR-5249	21	43,187,153	AGGTGCGCAG CAGCGCCAGG GCCCAT	A	*RIPK4*	ENST00000332512.3	8-16	PWALAL LRT/X	frameshift	NA	Y	
TCGA-CR-6487	22	41,531,855	C	T	*EP300*	ENST00000263253.7	523	Q/*	stop gained	NA	Y	
GS18007	22	41,531,880	C	G	*EP300*	ENST00000263253.7	531	S/*	stop gained	NA	Y	
GS18001	22	41,536,229	G	T	*EP300*	ENST00000263253.7	616	E/*	stop gained	NA	Y	
GS18003	22	41,543,849	C	T	*EP300*	ENST00000263253.7	714	Q/*	stop gained	NA	Y	
TCGA-BA-5559	22	41,553,402	G	A	*EP300*	ENST00000263253.7	1164	C/Y	missense	COSM5487677	Y	
GS18051	22	41,564,814	G	A	*EP300*	ENST00000263253.7	1372	C/Y	missense	COSM4387468	Y	
GS18033	22	41,565,529	G	A	*EP300*	ENST00000263253.7	1399	D/N	missense	COSM122851	N	
GS18023	22	41,565,529	G	A	*EP300*	ENST00000263253.7	1399	D/N	missense	COSM122851	N	
GS18079	22	41,565,529	G	A	*EP300*	ENST00000263253.7	1399	D/N	missense	COSM122851	N	
TCGA-CR-6470	22	41,565,529	G	A	*EP300*	ENST00000263253.7	1399	D/N	missense	COSM122851	N	
TCGA-CV-7100	22	41,565,529	G	A	*EP300*	ENST00000263253.7	1399	D/N	missense	COSM122851	N	
TCGA-BB-A6UM	22	41,565,529	G	A	*EP300*	ENST00000263253.7	1399	D/N	missense	COSM122851	N	
TCGA-DQ-7593	22	41,565,529	G	A	*EP300*	ENST00000263253.7	1399	D/N	missense	COSM122851	N	
GS18085	22	41,566,424	A	G	*EP300*	ENST00000263253.7	1434	H/R	missense	COSM3740445	Y	
TCGA-CR-6471	22	41,566,460	A	G	*EP300*	ENST00000263253.7	1446	Y/C	missense	NA	Y	
GS18055	22	41,566,511	T	C	*EP300*	ENST00000263253.7	1463	L/P	missense	COSM4104436	Y	
TCGA-CN-A6V6	22	41,568,590	G	A	*EP300*	ENST00000263253.7	1514	E/K	missense	COSM1716512	N	
GS18012	22	41,572,904	C	T	*EP300*	ENST00000263253.7	1730	S/F	missense	COSM320129	Y	
GS18004	X	41,193,508	G	C	*DDX3X*	ENST00000399959.2	1	M/I	initiator codon	NA	Y	*
TCGA-RS-A6TP	X	41,193,508	G	C	*DDX3X*	ENST00000399959.2	1	M/I	initiator codon	NA	Y	*
TCGA-HL-7533	X	41,200,777	AG	A	*DDX3X*	ENST00000399959.2	65	D/X	frameshift	NA	Y	

GS18005	X	41,202,539	CAGTGCAAAA GCATGCTATT CCTATTATCAA AGAGAAAAG	C	*DDX3X*	ENST00000399959.2	205-218	PVQKHA IPIIKEK R/P	inframe deletion	NA	Y
GS18021	X	41,205,622	CAGTTCCGCT CAGGAAAAAG CCCA	C	*DDX3X*	ENST00000399959.2	486-493	QFRSG KSP/X	frameshift	NA	Y
GS18007	X	41,205,783	C	G	*DDX3X*	ENST00000399959.2	508	S/*	stop gained	COSM6847254	Y
GS18022	X	41,205,855	C	T	*DDX3X*	ENST00000399959.2	532	T/M	missense	COSM4855955	Y
TCGA-CR-6481	X	41,206,168	G	C	*DDX3X*	ENST00000399959.2	558	D/H	missense	NA	Y
TCGA-CR-6481	X	41,206,210	G	C	*DDX3X*	ENST00000399959.2	572	E/Q	missense	NA	Y

Supplemental Table S4A. Somatic variants observed in the most significantly mutated genes in HPV-positive tumors. Presented here is a list of somatic variants in the 24 most significantly mutated genes identified in 149 HPV-positive OSCC using MutSig (q<0.2). Shown from left to right are: sample ID; chromosome; chromosomal position;, reference allele; alternative allele; gene symbol; transcript ID; amino acid position; amino acid change; consequence of mutation predicted by VEP v74; annotated variant ID; and previously unreported in OSCC (Y, yes; N, no; see also **Fig. 4, Supplemental Fig. S4**). *Asterisk* (in previously unreported column), previously unreported, recurrent variant.

GO biological process complete	REFLIST (20972)	observed (24)	expected	fold enrichment	P-value	genes
apoptotic process (GO:0006915)	908	10	1.08	9.24	3.49E-04	*DDX3X, EP300, FGFR3, NFKBIA, PTEN, RB1, SLTM, STAT1, TGFBR2, TRAF3*
regulation of sequence-specific DNA binding transcription factor activity (GO:0051090)	378	7	0.45	15.53	1.88E-03	*CYLD, EP300, NFKBIA, PTEN, RB1, RIPK4, TRAF3*
viral process (GO:0016032)	632	8	0.75	10.62	3.90E-03	*DDX3X, EP300, FBXW7, NFKBIA, RB1, RBL1, STAT1, TAF5*
regulation of signal transduction (GO:0009966)	2793	14	3.33	4.2	4.83E-03	*CYLD, DDX3X, EP300, FBXW7, FGFR3, IFNGR1, KMT2D, NFKBIA, PIK3CA, PTEN, RB1, STAT1, TAF5, TGFBR2*
regulation of cell cycle G1/S phase transition (GO:1902806)	154	5	0.18	27.24	8.43E-03	*DDX3X, EP300, FBXW7, PTEN, RB1*
cellular macromolecule metabolic process (GO:0044260)	6817	20	8.13	2.46	1.22E-02	*BBX, CASZ1, CYLD, DDX3X, EP300, FBXW7, FGFR3, KMT2D, NSD1, PIK3CA, PTEN, RB1, RBL1, RIPK4, SLTM, STAT1, TAF5, TGFBR2, TRAF3, ZNF750*
positive regulation of transcription, DNA-templated (GO:0045893)	1356	10	1.62	6.19	1.42E-02	*CASZ1, DDX3X, EP300, KMT2D, NFKBIA, NSD1, RB1, RBL1, STAT1, ZNF750*
macromolecule modification (GO:0043412)	3180	14	3.79	3.69	2.39E-02	*CYLD, EP300, FBXW7, FGFR3, KMT2D, NSD1, PIK3CA, PTEN, RB1, RBL1, RIPK4, TAF5, TGFBR2, TRAF3*
egulation of multicellular organismal process (GO:0051239)	2720	13	3.24	4.01	2.77E-02	*ASAP1, CASZ1, CYLD, DDX3X, EP300, FBXW7, NFKBIA, PIK3CA, PTEN, RB1, STAT1, TGFBR2, TRAF3*
regulation of innate immune response (GO:0045088)	361	6	0.43	13.94	2.92E-02	*CYLD, EP300, IFNGR1, NFKBIA, STAT1, TRAF3*
regulation of transcription from RNA polymerase II promoter (GO:0006357)	1859	11	2.22	4.96	3.05E-02	*DDX3X, EP300, FBXW7, KMT2D, NFKBIA, NSD1, RB1, RBL1, SLTM, STAT1, ZNF750*
positive regulation of molecular function (GO:0044093)	1874	11	2.23	4.92	3.30E-02	*ASAP1, DDX3X, EP300, FBXW7, FGFR3, NFKBIA, PIK3CA, PTEN, RB1, RIPK4, TGFBR2*

negative regulation of developmental process (GO:0051093)	852	8	1.02	7.88	3.62E-02	*ASAP1, CYLD, FBXW7, FGFR3, NFKBIA, PTEN, STAT1, TGFBR2*

Supplemental Table S4B. PANTHER biological processes over-represented in HPV-positive OSCC.
Shown are significantly increased biological processes represented in the 24 most significantly mutated genes in 149 HPV-positive tumors identified using PANTHER Overrepresentation Test (release 20160715). Shown here from left to right are: the GO term; the number of genes associated with the GO term in the reference human genome; the number of observed genes in our list; the number of expected genes based on the number of genes in reference genome; fold enrichment; Bonferroni adjusted p-values; and genes associated with the GO term. The most descendent node based on the tree structure of PANTHER database was chosen in cases where multiple, similar GO terms were found at a significance level of $p<0.05$.

GO molecular function complete	REFLIST (20972)	observed (24)	expected	Fold Enrichment	P-value	genes
core promoter sequence-specific DNA binding (GO:0001046)	107	4	0.13	31.36	2.19E-02	*EP300, NSD1, STAT1, ZNF750*
androgen receptor binding (GO:0050681)	41	3	0.05	61.38	4.64E-02	*EP300, NSD1, RB1*

Supplemental Table S4C. PANTHER molecular function ontology terms over-represented in HPV-positive OSCC.
Significantly enriched molecular functions of 24 significantly mutated genes were identified in 149 HPV-positive tumors using PANTHER Overrepresentation Test (release 20160715). We used complete GO terms for molecular function, and the p-values for molecular function terms were adjusted by Bonferroni multiple testing correction. Shown here are: the GO term; the number of genes associated with the GO term in reference human genome; the number of observed genes in our list; the number of expected genes based on the number of genes in reference genome; fold enrichment; Bonferroni adjusted p-values; and genes associated with the GO term. We have chosen the most descendent node based on the tree structure of PANTHER database in cases where multiple similar GO terms were found at a significance level of $p<0.05$.

PANTHER Pathways	REFLIST (20972)	observed (24)	expected	Fold Enrichment	P-value	genes
p53 pathway feedback loops 2 (P04398)	53	5	0.06	79.14	8.30E-07	*PIK3CA, PTEN, RB1, RBL1, STAT1*
Inflammation mediated by chemokine and cytokine signaling pathway (P00031)	261	5	0.31	16.07	2.04E-03	*IFNGR1, NFKBIA, PIK3CA, PTEN, STAT1*
p53 pathway (P00059)	88	3	0.1	28.6	2.51E-02	*EP300, PIK3CA, PTEN*

Supplemental Table S4D. **PANTHER functional pathways over-represented in HPV-positive OSCC.** We identified significantly enriched biological processes by analysis of the 24 most significantly mutated genes in 149 HPV-positive tumors using PANTHER Overrepresentation Test (release 20160715). We used the PANTHER pathway database, and the p-values for pathways are adjusted by Bonferroni multiple testing correction. Shown here are: the pathway term; the number of genes associated with Pathway term in reference human genome; the number of observed genes in our list; the number of expected genes based on the number of genes in reference genome; fold enrichment (observed/expected); Bonferroni adjusted p-values, and genes associated with the pathway term.

gene	mutation (SNVs & Indels)		CNV (gain)		CNV (loss)		CNV (gain or loss)		any alteration (mutation or CNV)	
	no. mutated	%total	no. gain	%total	no. loss	%total	no. altered	%total	no. altered	%total
PIK3CA	42	28.2%	100	67.1%	0	0.0%	100	67.1%	106	71.1%
ZNF750	21	14.1%	7	4.7%	17	11.4%	24	16.1%	42	28.2%
KMT2D	20	13.4%	5	3.4%	0	0.0%	5	3.4%	25	16.8%
EP300	18	12.1%	3	2.0%	5	3.4%	8	5.4%	25	16.8%
FGFR3	17	11.4%	7	4.7%	10	6.7%	17	11.4%	31	20.8%
CASZ1	14	9.4%	6	4.0%	12	8.1%	18	12.1%	28	18.8%
PTEN	13	8.7%	1	0.7%	28	18.8%	29	19.5%	37	24.8%
CYLD	11	7.4%	1	0.7%	32	21.5%	33	22.1%	37	24.8%
FBXW7	11	7.4%	5	3.4%	10	6.7%	15	10.1%	24	16.1%
NSD1	10	6.7%	6	4.0%	7	4.7%	13	8.7%	20	13.4%
RB1	9	6.0%	2	1.3%	51	34.2%	53	35.6%	60	40.3%
RBL1	9	6.0%	23	15.4%	7	4.7%	30	20.1%	37	24.8%
ASAP1	9	6.0%	41	27.5%	1	0.7%	42	28.2%	47	31.5%
DDX3X	8	5.4%	13	8.7%	3	2.0%	16	10.7%	23	15.4%
RIPK4	6	4.0%	8	5.4%	7	4.7%	15	10.1%	19	12.8%
TRAF3	5	3.4%	2	1.3%	34	22.8%	36	24.2%	38	25.5%
BBX	5	3.4%	67	45.0%	1	0.7%	68	45.6%	70	47.0%
SLTM	5	3.4%	1	0.7%	5	3.4%	6	4.0%	11	7.4%
TGFBR2	5	3.4%	3	2.0%	24	16.1%	27	18.1%	32	21.5%
NFKBIA	4	2.7%	1	0.7%	16	10.7%	17	11.4%	18	12.1%
IFNGR1	4	2.7%	4	2.7%	13	8.7%	17	11.4%	21	14.1%
TAF5	4	2.7%	2	1.3%	18	12.1%	20	13.4%	24	16.1%
HIST1H2AE	4	2.7%	8	5.4%	11	7.4%	19	12.8%	23	15.4%
STAT1	4	2.7%	3	2.0%	5	3.4%	8	5.4%	9	6.0%

Supplemental Table S5A. Gene disruptions by SNVs and/or CNVs in HPV-positive OSCC. Counts and frequencies of gene disruptions by somatic SNVs/Indels and/or CNVs in the 24 most frequently mutated genes identified by MutSig in 149 HPV-positive OSCC tumors are shown here **(see also Fig. 1, Supplemental Fig. S5B).**

gene	mutation (SNVs & Indels)		CNV (gain)		CNV (loss)		CNV (gain/loss)		any alteration (mutation /CNV)	
	no. mutated	%total	no. gain	%total	no. loss	%total	no. altered	%total	no. altered	%total
TP53	267	81.2%	7	2.1%	15	4.6%	22	6.7%	270	82.1%
FAT1	91	27.7%	4	1.2%	40	12.2%	44	13.4%	122	37.1%
CDKN2A	83	25.2%	7	2.1%	133	40.4%	140	42.6%	213	64.7%
NOTCH1	64	19.5%	27	8.2%	17	5.2%	44	13.4%	99	30.1%
CASP8	52	15.8%	14	4.3%	10	3.0%	24	7.3%	75	22.8%
PIK3CA	51	15.5%	117	35.6%	0	0.0%	117	35.6%	153	46.5%
KMT2D	41	12.5%	4	1.2%	6	1.8%	10	3.0%	49	14.9%
HRAS	28	8.5%	12	3.6%	13	4.0%	25	7.6%	47	14.3%
NSD1	27	8.2%	6	1.8%	29	8.8%	35	10.6%	57	17.3%
FBXW7	22	6.7%	2	0.6%	15	4.6%	17	5.2%	38	11.6%
TGFBR2	18	5.5%	1	0.3%	60	18.2%	61	18.5%	77	23.4%
NOTCH2	18	5.5%	7	2.1%	27	8.2%	34	10.3%	50	15.2%
NFE2L2	18	5.5%	33	10.0%	3	0.9%	36	10.9%	51	15.5%
HLA-B	17	5.2%	7	2.1%	6	1.8%	13	4.0%	29	8.8%
AJUBA	17	5.2%	28	8.5%	11	3.3%	39	11.9%	50	15.2%
EPHA2	17	5.2%	7	2.1%	8	2.4%	15	4.6%	32	9.7%
RASA1	14	4.3%	2	0.6%	29	8.8%	31	9.4%	42	12.8%
CTCF	14	4.3%	6	1.8%	5	1.5%	11	3.3%	25	7.6%
RAC1	10	3.0%	47	14.3%	0	0.0%	47	14.3%	57	17.3%
SMAD4	9	2.7%	2	0.6%	50	15.2%	52	15.8%	57	17.3%
FOSL2	8	2.4%	8	2.4%	2	0.6%	10	3.0%	18	5.5%
NF2	8	2.4%	21	6.4%	9	2.7%	30	9.1%	36	10.9%
PSIP1	8	2.4%	29	8.8%	28	8.5%	57	17.3%	65	19.8%
EIF2S2	7	2.1%	48	14.6%	1	0.3%	49	14.9%	55	16.7%
PODXL	7	2.1%	9	2.7%	11	3.3%	20	6.1%	26	7.9%

Supplemental Table S5B. Gene disruptions by SNVs and/or CNVs in HPV-negative OSCC. Frequencies of gene disruptions by somatic SNVs/Indels and/or CNVs in 329 HPV-negative OSCC tumors are shown here. We used 50 WGS samples (26 cases in Ohio cohort + 24 in TCGA) and 279 TCGA samples having both copy number alteration (microarray) and mutation data (WES) available. Counts of SNVs and/or CNVs in the 25 most frequently mutated genes identified by MutSig analysis are shown (see also **Supplemental Fig. S5B**).

gene	HPV-positive no. patients	HPV-positive % patients	HPV-negative no. patients	HPV-negative % patients	fold change in HPV-positive patients	p-value	FDR adjusted p-value	
RB1	54	36.24%	35	10.45%	3.47	1.04E-10	2.49E-09	**
ZNF750	38	25.50%	16	4.78%	5.34	2.09E-10	4.80E-09	**
BBX	61	40.94%	54	16.12%	2.54	9.08E-09	2.00E-07	**
CYLD	35	23.49%	17	5.07%	4.63	1.05E-08	2.21E-07	**
PTEN	35	23.49%	18	5.37%	4.37	2.11E-08	4.21E-07	**
PIK3CA	100	67.11%	145	43.28%	1.55	1.28E-06	2.44E-05	**
TAF5	22	14.77%	11	3.28%	4.50	1.48E-05	2.66E-04	**
CASZ1	27	18.12%	18	5.37%	3.37	2.61E-05	4.44E-04	**
HIST1H2AE	20	13.42%	14	4.18%	3.21	7.42E-04	1.19E-02	*
IFNGR1	19	12.75%	14	4.18%	3.05	1.32E-03	1.98E-02	*
RBL1	33	22.15%	43	12.84%	1.73	1.43E-02	2.00E-01	
TRAF3	31	20.81%	42	12.54%	1.66	2.70E-02	3.51E-01	
SLTM	10	6.71%	10	2.99%	2.25	8.05E-02	9.66E-01	
DDX3X	19	12.75%	25	7.46%	1.71	8.53E-02	9.66E-01	
FGFR3	26	17.45%	39	11.64%	1.50	1.11E-01	1.00E+00	
FBXW7	22	14.77%	34	10.15%	1.45	1.66E-01	1.00E+00	
ASAP1	43	28.86%	115	34.33%	0.84	2.50E-01	1.00E+00	
EP300	24	16.11%	42	12.54%	1.28	3.16E-01	1.00E+00	
RIPK4	18	12.08%	31	9.25%	1.31	3.33E-01	1.00E+00	
NSD1	20	13.42%	56	16.72%	0.80	4.17E-01	1.00E+00	
STAT1	8	5.37%	25	7.46%	0.72	4.42E-01	1.00E+00	
KMT2D	24	16.11%	47	14.03%	1.15	5.79E-01	1.00E+00	
NFKBIA	16	10.74%	32	9.55%	1.12	7.42E-01	1.00E+00	
TGFBR2	28	18.79%	68	20.30%	0.93	8.05E-01	1.00E+00	

Supplemental Table S5C. Comparison of mutation and copy number variation frequencies in 24 HPV-positive MutSig genes (HPV-positive vs. HPV-negative OSCC patients).
The 24 most significantly mutated genes in 149 HPV-positive OSCC tumors were identified using MutSig. Their mutation frequencies (including both mutations and copy number variants) were compared in 149 HPV-positive vs. 335 HPV-negative tumors. We counted the number of patients with somatic SNVs/INDELs having consequences with coding change (such as missense, stop gain, splice sites, frameshift, in-frame mutations) and copy number alterations for HPV-positive vs. HPV-negative tumors. The significance of the comparison was assessed using Fisher's Exact test. We applied the multiple testing correction on p-values using FDR method (see also **Supplemental Fig. S5B**).

gene	HPV-positive no. patients	HPV-positive % patients	HPV-negative no. patients	HPV-negative % patients	fold change in HPV-negative patients	p-value	FDR adjusted p-value	
TP53	11	7.38%	269	80.30%	10.88	2.64E-55	6.61E-54	**
CDKN2A	9	6.04%	205	61.19%	10.13	1.03E-33	2.48E-32	**
FAT1	10	6.71%	118	35.22%	5.25	2.64E-12	6.07E-11	**
PIK3CA	100	67.11%	145	43.28%	0.64	1.28E-06	2.82E-05	**
RAC1	4	2.68%	50	14.93%	5.56	2.98E-05	6.27E-04	**
CTCF	31	20.81%	24	7.16%	0.34	3.58E-05	7.16E-04	**
CASP8	10	6.71%	70	20.90%	3.11	5.69E-05	1.08E-03	**
PSIP1	10	6.71%	58	17.31%	2.58	1.67E-03	3.00E-02	*
NOTCH1	24	16.11%	96	28.66%	1.78	3.00E-03	5.10E-02	
EIF2S2	36	24.16%	47	14.03%	0.58	8.72E-03	1.39E-01	
NOTCH2	9	6.04%	44	13.13%	2.17	2.63E-02	3.94E-01	
HRAS	10	6.71%	45	13.43%	2.00	3.07E-02	4.30E-01	
NF2	6	4.03%	32	9.55%	2.37	4.32E-02	5.61E-01	
SMAD4	12	8.05%	49	14.63%	1.82	5.31E-02	6.38E-01	
RASA1	9	6.04%	37	11.04%	1.83	9.41E-02	1.00E+00	
HLA-B	19	12.75%	28	8.36%	0.66	1.37E-01	1.00E+00	
FBXW7	22	14.77%	34	10.15%	0.69	1.66E-01	1.00E+00	
NFE2L2	14	9.40%	47	14.03%	1.49	1.83E-01	1.00E+00	
NSD1	20	13.42%	56	16.72%	1.25	4.17E-01	1.00E+00	
KMT2D	24	16.11%	47	14.03%	0.87	5.79E-01	1.00E+00	
EPHA2	15	10.07%	29	8.66%	0.86	6.11E-01	1.00E+00	
PODXL	12	8.05%	24	7.16%	0.89	7.11E-01	1.00E+00	
AJUBA	17	11.41%	43	12.84%	1.13	7.65E-01	1.00E+00	
TGFBR2	28	18.79%	68	20.30%	1.08	8.05E-01	1.00E+00	
FOSL2	6	4.03%	16	4.78%	1.19	8.17E-01	1.00E+00	

Supplemental Table S5D. Comparison of mutation and copy number variation frequencies in 25 HPV-negative OSCC MutSig genes (HPV-positive vs. HPV-negative OSCC patients).

The 25 most significantly mutated genes in 335 HPV-negative OSCC were identified using MutSig. Their mutation and copy number variation frequencies were compared in 149 HPV-positive vs. 335 HPV-negative tumors. We counted the numbers of instances of somatic SNVs/INDELs having coding change consequences (such as missense, stop gain, splice sites, frameshift, in-frame mutations) and copy number alteration in the HPV-positive vs. HPV-negative tumors. The significance of the comparison was assessed using Fisher's Exact test. We applied multiple testing correction on p-values using the FDR method (see also **Supplemental Fig. S5B**).

Sample ID	virus type	position	reference allele	alternative allele	gene	consequence	amino acid change
GS18035	HPV16	712	C	A	E7	missense	51H>N
GS18081	HPV16	749	C	G	E7	missense	63S>C
GS18047	HPV16	784	G	A	E7	missense	75D>N
GS18014	HPV16	822	A	G	E7	missense	87L>M

Supplemental Table S5E. List of coding change mutations in E7 gene in HPV16-positive OSCC. Four HPV-positive OSCC samples with amino acid changes in E7 gene of HPV16 genome were detected. Shown here are the sample ID, position of mutation in HPV16 genome, reference allele, alternative allele, and amino acid change. Each of these mutations were observed only once among 90 HPV16-positive OSCC.

	RB1 WT	RB1 mutation
E7-WT	83	3
E7-mutation	4	0

	RB1 normal copy number	RB1 copy number loss
E7-WT	63	23
E7-mutation	2	2

	RB1 WT and normal copy number	RB1 mutation and/or copy number loss
E7-WT	60	26
E7-mutation	2	2

Supplemental Table S5F. Potential associations between HPV16 E7 mutation status and *RB1* alteration status.

Two-by-two tables show counts of HPV16-positive OSCC with or without mutations in HPV16 E7 and genetic disruptions present or absent in *RB1*. (*Top*) Number of samples with or without mutations (SNVs and small Indels with coding-change consequence) in E7 gene of HPV16 and number of samples with or without mutations in *RB1,* revealing 4 samples with rare missense coding-change variants in E7. (*Middle*) Number of samples with or without mutations (SNVs and small Indels with coding-change consequence) in E7 and number of samples with copy number loss of *RB1*. (*Bottom*) Number of samples with or without mutations (SNVs and small Indels with coding-change consequence) in E7 and number of samples withcopy number loss and/or mutation in *RB1*. Neither co-occurrence nor avoidance between E7 and *RB1* alterations were observed at significant levels. This analysis was limited by low numbers of E7 mutations in tumors, consistent with the cervical cancer literature.

Chr. arm	frequency mean (%sample)	frequency range (%sample)	gained length (Mbp)	% chr. arm gained	adj. p-value (mean)	adj. p-value (range)
1q	12.9%	12.6-14.6%	40	32.2%	1.83E-03	2.76E-04-2.24E-03
3q	55.0%	31.1-68.0%	104.5	97.6%	0.00E+00	0.00E+00-0.00E+00
5p	16.6%	13.6-20.4%	46	95.0%	1.34E-04	0.00E+00-7.81E-04
8p	13.6%	13.6-13.6%	0.5	1.1%	7.81E-04	7.81E-04-7.81E-04
8q	20.2%	14.6-27.2%	98	97.3%	1.69E-05	0.00E+00-2.76E-04
19q	14.5%	13.6-17.5%	13.5	41.4%	5.01E-04	6.93E-06-7.81E-04
20p	15.0%	13.6-17.5%	25.5	92.7%	2.99E-04	6.93E-06-7.81E-04
20q	20.5%	13.6-25.2%	7.5	21.1%	9.09E-05	0.00E+00-7.81E-04

Chr. arm	frequency mean (%sample)	frequency range (%sample)	lost length (Mbp)	% chr. arm lost	adj. p-value (mean)	adj. p-value (range)
1p	31.1%	28.2-33.0%	1.5	1.2%	0.00E+00	0.00E+00-0.00E+00
1q	22.8%	18.4-27.2%	1	0.8%	4.33E-07	0.00E+00-8.66E-07
2p	18.4%	16.5-20.4%	1	1.1%	4.33E-06	8.66E-07-7.79E-06
2q	13.5%	11.7-23.3%	14.5	9.7%	1.02E-03	0.00E+00-3.15E-03
3p	15.2%	11.7-19.4%	78	85.7%	4.60E-04	8.66E-07-3.15E-03
4p	11.7%	11.7-11.7%	0.5	1.0%	3.15E-03	3.15E-03-3.15E-03
5q	64.4%	60.2-68.9%	1.5	1.1%	0.00E+00	0.00E+00-0.00E+00
7q	15.5%	11.7-42.7%	4.5	4.5%	2.23E-03	0.00E+00-3.15E-03
8p	20.7%	13.6-32.0%	1.5	3.3%	1.08E-04	0.00E+00-3.15E-04
9p	37.0%	15.5-55.3%	8.5	17.3%	1.37E-06	0.00E+00-2.34E-05
9q	41.3%	12.6-43.7%	19	20.6%	2.81E-05	0.00E+00-1.04E-03
10p	15.5%	15.5-15.5%	0.5	1.2%	2.34E-05	2.34E-05-2.34E-05
10q	13.5%	11.7-59.2%	18	18.9%	2.68E-03	0.00E+00-3.15E-03
11q	38.2%	11.7-49.5%	67.5	83.0%	2.43E-05	0.00E+00-3.15E-03
13q	19.3%	11.7-27.2%	67.5	69.4%	1.85E-04	0.00E+00-3.15E-03
14q	18.1%	12.6-23.3%	45.5	50.7%	9.12E-05	0.00E+00-1.04E-03
15q	24.9%	12.6-53.4%	2.5	3.0%	4.17E-04	0.00E+00-1.04E-03
16p	16.5%	15.5-17.5%	1.5	4.1%	1.13E-05	2.60E-06-2.34E-05
16q	18.5%	15.5-21.4%	44	81.9%	2.68E-06	0.00E+00-2.34E-05
17q	16.5%	16.5-16.5%	0.5	0.9%	7.79E-06	7.79E-06-7.79E-06
21p	17.2%	15.5-18.4%	1.5	11.4%	8.95E-06	8.66E-07-2.34E-05
22q	18.4%	18.4-18.4%	0.5	1.4%	8.66E-07	8.66E-07-8.66E-07

Supplemental Table S5G. Recurrent chromosomal gains and losses in HPV-positive OSCC. Shown here are chromosome arms with significantly gained (*top*) or lost (*bottom*) segments in 103 HPV-positive tumors with WGS data. Copy number changes in tumor samples vs. matched normal samples were identified using CNAnorm as described in Methods. Recurrently gained or lost 500 kb bins containing any segmental copy number changes were identified. Gains were defined as estimated ploidy N >= 2.5, while losses were defined as N < 1.5. Mean and range of affected samples were determined as percentages of samples with copy number changes in 500 kb chromosomal bins (cf. **Fig. 5**). Statistical significance was assessed using the binomial distribution. Adjusted p-value, multiple testing correction with FDR method.

Chr. arm	frequency mean (%sample)	frequency range (%sample)	gained length (Mbp)	% chr. arm gained	adj. p-value (mean)	adj. p-value (range)
3q	34.2%	20.0-48.0%	66.5	62.1%	1.94E-04	0.00E+00-1.18E-03
5p	33.1%	30.0-40.0%	46.5	96.1%	1.12E-07	0.00E+00-8.66E-07
7p	22.6%	20.0-30.0%	50.5	84.3%	4.72E-04	8.66E-07-1.18E-03
8p	23.0%	20.0-24.0%	2	4.4%	3.47E-04	7.01E-05-1.18E-03
8q	40.5%	24.0-54.0%	100	99.2%	4.85E-07	0.00E+00-7.01E-05
9p	20.3%	20.0-22.0%	3	6.1%	1.03E-03	2.89E-04-1.18E-03
11q	29.3%	20.0-48.0%	11	13.5%	2.65E-04	0.00E+00-1.18E-03
14q	24.5%	20.0-32.0%	76.5	85.2%	2.29E-04	0.00E+00-1.18E-03
20p	20.4%	20.0-22.0%	9.5	34.5%	9.92E-04	2.89E-04-1.18E-03
20q	22.4%	20.0-26.0%	25	70.4%	3.52E-04	1.39E-05-1.18E-03

Chr. arm	frequency mean (%sample)	frequency range (%sample)	lost length (Mbp)	% chr. arm lost	adj. p-value (mean)	adj. p-value (range)
1p	32.0%	28.0-34.0%	1.5	1.2%	0.00E+00	0.00E+00-0.00E+00
1q	27.0%	24.0-30.0%	1	0.8%	0.00E+00	0.00E+00-0.00E+00
2q	18.0%	16.0-20.0%	1	0.7%	6.84E-05	2.60E-06-1.34E-04
3p	23.0%	16.0-34.0%	90.5	99.5%	2.48E-05	0.00E+00-1.34E-04
4p	22.0%	22.0-22.0%	0.5	1.0%	0.00E+00	0.00E+00-0.00E+00
4q	18.0%	18.0-18.0%	0.5	0.4%	1.90E-05	1.90E-05-1.90E-05
5q	68.7%	62.0-72.0%	1.5	1.1%	0.00E+00	0.00E+00-0.00E+00
7q	40.0%	40.0-40.0%	0.5	0.5%	0.00E+00	0.00E+00-0.00E+00
8p	20.0%	16.0-44.0%	42.5	93.2%	3.61E-05	0.00E+00-1.34E-04
9p	31.9%	16.0-58.0%	12.5	25.5%	1.72E-05	0.00E+00-1.34E-04
9q	40.4%	16.0-44.0%	21	22.8%	3.65E-06	0.00E+00-1.34E-04
10q	32.0%	18.0-46.0%	1	1.0%	9.52E-06	0.00E+00-1.90E-05
11q	16.9%	16.0-20.0%	10	12.3%	9.22E-05	2.60E-06-1.34E-04
14q	18.0%	18.0-18.0%	0.5	0.6%	1.90E-05	1.90E-05-1.90E-05
15q	29.3%	22.0-40.0%	1.5	1.8%	0.00E+00	0.00E+00-0.00E+00
17q	24.0%	24.0-24.0%	0.5	0.9%	0.00E+00	0.00E+00-0.00E+00
18q	20.9%	16.0-28.0%	26.5	43.5%	3.42E-05	0.00E+00-1.34E-04
21p	28.7%	26.0-32.0%	1.5	11.4%	0.00E+00	0.00E+00-0.00E+00
21q	16.2%	16.0-18.0%	5	14.3%	1.23E-04	1.90E-05-1.34E-04

Supplemental Table S5H. Recurrent chromosomal gains and losses in HPV-positive OSCC. Shown here are chromosome arms with significantly gained (*top*) or lost (*bottom*) segments in 50 HPV-negative tumors with WGS data. Copy number changes in tumor samples vs. matched normal samples were identified using CNAnorm as described in Methods. Recurrently gained or lost 500 kb bins containing any segmental copy number changes were identified. Gains were defined as estimated ploidy N >= 2.5, while losses were defined as N < 1.5. Mean and range of affected samples were determined as percentages of samples with copy number changes in 500 kb chromosomal bins (cf. **Fig. 5**). Statistical significance was assessed using the binomial distribution. Adjusted p-value, multiple testing correction with FDR method.

arm	no. genes	no. genes with expression (variance > 0.3)	no. genes tested (gain vs. normal)	no. genes differentially expressed (gain vs. normal)	%genes (gain vs. normal)	no. genes tested (loss vs. normal)	no. genes differentially expressed (loss vs. normal)	%genes (loss vs. normal)
1p	1,142	801	798	20	2.5%	327	7	2.1%
1q	917	668	665	48	7.2%	29	1	3.4%
2p	565	383	76	2	2.6%	86	2	2.3%
2q	776	584	156	2	1.3%	528	36	6.8%
3p	509	375	168	2	1.2%	375	79	21.1%
3q	604	570	570	301	52.8%	3	0	NA
4p	214	158	88	0	0.0%	158	11	7.0%
4q	486	376	173	4	2.3%	223	9	4.0%
5p	164	130	130	32	24.6%	4	0	NA
5q	725	517	71	1	1.4%	82	1	1.2%
6p	583	457	457	6	1.3%	422	22	5.2%
6q	441	334	78	1	1.3%	334	8	2.4%
7p	330	240	129	4	3.1%	5	0	NA
7q	645	489	187	2	1.1%	299	19	6.4%
8p	250	204	204	17	8.3%	182	2	1.1%
8q	458	366	366	81	22.1%	7	0	NA
9p	209	153	138	8	5.8%	17	0	NA
9q	565	407	398	5	1.3%	27	0	0.0%
10p	170	125	124	4	3.2%	125	6	4.8%
10q	550	410	18	1	NA	410	54	13.2%
11p	346	262	29	1	3.4%	235	9	3.8%
11q	676	492	259	8	3.1%	456	86	18.9%
12p	295	260	260	28	10.8%	224	4	1.8%
12q	770	518	178	3	1.7%	3	0	NA
13q	328	256	135	5	3.7%	256	43	16.8%
14q	829	648	255	3	1.2%	648	74	11.4%
15q	647	466	8	0	NA	437	8	1.8%
16p	489	345	43	1	2.3%	345	7	2.0%
16q	386	291	1	0	NA	291	90	30.9%
17p	350	264	20	0	0.0%	264	20	7.6%
17q	820	584	391	10	2.6%	125	9	7.2%
18p	88	66	66	5	7.6%	2	0	NA
18q	208	155	155	9	5.8%	16	0	NA
19p	584	422	422	1	0.2%	408	15	3.7%
19q	761	605	605	55	9.1%	2	0	NA
20p	165	118	118	16	13.6%	4	0	NA
20q	329	246	246	59	24.0%	182	8	4.4%
21p	1	1	0	0	NA	1	0	NA
21q	210	159	132	1	0.8%	102	2	2.0%
22q	517	381	12	0	NA	381	10	2.6%
Total	19,102	14,286	8,329	746	9.0%	8,025	642	8.0%

Supplemental Table S5I. Number of differentially expressed genes associated with subchromosomal copy number gains and losses in HPV-positive OSCC.
Presented in the table from left to right are: the number of coding genes in each chromosome arm; the number of genes in that arm with expression variance >0.3; the counts of genes in regions with copy number gain (*left, pink*) or loss (*right, light blue*), limited to regions with 3 or more samples with either local gain or loss; and the counts of genes in that region with significant differential expression (p<0.01) in association with the copy number changes. These data were derived from analysis of WGS data from 101 of 103 HPV-positive tumors with available RNA-seq data. See Methods for additional information.

arm	no. genes	no. genes with expression (variance > 0.3)	no. genes tested (gain vs. normal)	no. genes differentially expressed (gain vs. normal)	%genes (gain vs. normal)	no. genes tested (loss vs. normal)	no. genes differentially expressed (loss vs. normal)	%genes (loss vs. normal)
1p	1,033	838	133	4	3.0%	222	4	1.8%
1q	812	656	618	26	4.2%	12	0	NA
2p	514	400	8	0	NA	42	2	4.8%
2q	681	594	262	22	8.4%	224	1	0.4%
3p	442	354	2	0	NA	354	1	0.3%
3q	523	463	463	39	8.4%	0	0	NA
4p	187	168	168	4	2.4%	39	0	0.0%
4q	433	364	13	0	NA	103	0	0.0%
5p	148	146	146	44	30.1%	0	0	NA
5q	631	532	35	2	5.7%	492	15	3.0%
6p	500	435	64	2	3.1%	6	0	NA
6q	390	325	256	7	2.7%	93	4	4.3%
7p	288	259	259	45	17.4%	2	0	NA
7q	526	456	235	9	3.8%	32	0	0.0%
8p	206	202	145	17	11.7%	202	2	1.0%
8q	407	392	392	155	39.5%	11	0	NA
9p	182	182	177	50	28.2%	182	10	5.5%
9q	510	456	378	10	2.6%	84	2	2.4%
10p	148	120	2	0	NA	25	0	0.0%
10q	485	402	67	4	6.0%	20	0	0.0%
11p	319	280	40	1	2.5%	4	0	NA
11q	651	601	592	146	24.7%	319	5	1.6%
12p	240	225	204	6	2.9%	4	0	NA
12q	690	561	73	0	0.0%	3	0	NA
13q	289	263	158	12	7.6%	114	2	1.8%
14q	691	611	611	79	12.9%	142	2	1.4%
15q	541	457	57	1	1.8%	12	0	NA
16p	409	303	12	1	NA	12	0	NA
16q	359	293	2	0	NA	5	0	NA
17p	317	277	101	2	2.0%	228	2	0.9%
17q	722	596	329	12	3.6%	92	1	1.1%
18p	77	68	68	0	0.0%	8	1	NA
18q	181	173	61	8	13.1%	158	10	6.3%
19p	506	389	4	0	NA	159	9	5.7%
19q	665	534	19	0	NA	2	0	NA
20p	153	136	136	9	6.6%	5	0	NA
20q	305	252	252	39	15.5%	0	0	NA
21p	1	1	0	0	NA	1	0	NA
21q	179	161	3	0	0.0%	161	1	0.6%
22q	462	399	330	5	1.5%	27	1	3.7%
Total	16,803	14,324	6,875	761	11.1%	3,601	75	2.1%

Supplemental Table S5J. Number of differentially expressed genes associated with subchromosomal copy number gains and losses in HPV-negative OSCC.
Presented in the table from left to right are: the number of coding genes in each chromosome arm; the number of genes in that arm with expression variance >0.3; the counts of genes in regions with copy number gain (*left, pink*) or loss (*right, light blue*), limited to regions with 3 or more samples with either local gain or loss; and the counts of genes in that region with significant differential expression (p<0.01) in association with the copy number changes. These data were derived from analysis of WGS data from 50 HPV-negative tumors with available RNA-seq data. See methods for additional information.

arm	count	census gene	gene
1p	20	*SPEN*	*RP11-206L10.2, CCNL2, SLC35E2, HES2, DFFA, VPS13D, SNORA59A, DDI2, SPEN, RAP1GAP, SNHG12, KCNQ4, ZFYVE9, COA7, MIR553, GPSM2, CHI3L2, snoU13, MIR548AC, RP5-1042I8.7*
1q	48	*ABL2*	*GPR89B, SNORA40, LYSMD1, TUFT1, RP11-74C1.2, THEM4, AL590431.1, EFNA4, MTX1P1, ASH1L, SNORA42, IQGAP3, IFI16, DUSP23, TAGLN2, RNU4-42P, VANGL2, F11R, PVRL4, PPOX, NUF2, DARS2, GAS5, SNORD78, RP4-593C16.3, ABL2, RP11-46A10.5, C1orf21, RP11-295K2.3, NEK7, C1orf106, PKP1, LAD1, TMEM81, SRGAP2, PFKFB2, C1orf74, IRF6, FLVCR1, CAPN2, ENAH, SDE2, LIN9, MLK4, SNORA14B, B3GALNT2, HEATR1, RP11-488L18.6*
2p	2	--	*GEN1, RPIA*
2q	2	*NFE2L2*	*ADRA2B, NFE2L2*
3p	2	--	*HACL1, RNU4-78P*
3q	301	*TFG, POLQ, RPN1, STAG1, PIK3CB, FOXL2, ATR, WWTR1, GMPS, MLF1, TBL1XR1, PIK3CA, SOX2, MAP3K13, EIF4A2, BCL6, LPP, TP63, MUC4, TFRC*	*ARL13B, DHFRL1, NSUN3, ARL6, CRYBG3, CRYBG3, MINA, CLDND1, CPOX, DCBLD2, CMSS1, TBC1D23, LNP1, TFG, SENP7, TRMT10C, ZBTB11-AS1, RPL24, CEP97, BBX, CD47, IFT57, RP11-381E24.1, KIAA1524, DZIP3, ABHD10, C3orf52, RP11-757F18.5, GTPBP8, WDR52, SPICE1, KIAA2018, NAA50, ATP6V1A, GRAMD1C, ZDHHC23, KIAA1407, QTRTD1, MIR568, IGSF11, B4GALT4, TIMMDC1, COX17, GSK3B, RP11-18H7.1, GPR156, LRRC58, NDUFB4, RABL3, POLQ, GOLGB1, IQCB1, CCDC58, FAM162A, WDR5B, KPNA1, PARP9, DTX3L, HSPBAP1, DIRC2, PDIA5, PTPLB, CCDC14, UMPS, ITGB5, ZNF148, SNX4, Y_RNA, OSBPL11, RP11-379B18.5, RP11-666A20.4, FAM86JP, ALG1L, RP11-666A20.3, SLC41A3, RP11-124N2.1, ZXDC, TXNRD3, CHCHD6, PLXNA1, MCM2, PODXL2, ABTB1, RUVBL1, EEFSEC, RPN1, RAB7A, RP11-723O4.3, ACAD9, ISY1, HMCES, H1FX, RP11-529F4.1, RPL32P3, IFT122, TMCC1, AC083799.1, RP11-93K22.13, PIK3R4, ATP2C1, ASTE1, NEK11, NUDT16P1, NUDT16, RP11-517B11.7, ACPP, DNAJC13, NPHP3, TOPBP1, RYK, PPP2R3A, MSL2, PCCB, STAG1, RP11-85F14.5, NCK1, IL20RB, IL20RB-AS1, DZIP1L, CEP70, FAIM, PIK3CB, FOXL2, MRPS22, RP11-319G6.1, RBP1, SLC25A36, RP11-231L11.3, RNF7, ATP1B3, TFDP2, GK5, XRN1, ATR, U2SURP, RP11-91G21.1, C3orf58, RP11-274H2.5, RP11-274H2.3, PLOD2, PLSCR4, PLSCR1, GYG1, HLTF, HPS3, TM4SF1, TM4SF1-AS1, WWTR1, COMMD2, TSC22D2, SIAH2, SIAH2-AS1, GPR87, MBNL1-AS1, P2RY1, RP11-38P22.2, RP11-529G21.2, RAP2B, DHX36, C3orf33, GMPS, TIPARP, LINC00886, PA2G4P4, CCNL1, RP11-555M1.3, RP11-550I24.2, SHOX2, RSRC1, MLF1, GFM1, MFSD1, SCHIP1, IFT80, SMC4, TRIM59, KPNA4, KRT8P12, B3GALNT1, NMD3, PDCD10, HMGN1P8, GOLIM4, ACTRT3, MYNN, PHC3, PRKCI, RPL22L1, PLD1, PP13439, TNFSF10, ECT2, NAALADL2, TBL1XR1, ZMAT3, PIK3CA, LRRFIP1P1, ZNF639, MFN1, RP11-145M9.4, AC007620.3, GNB4, ACTL6A, MRPL47, NDUFB5, USP13, FXR1, DNAJC19, SOX2-OT, SOX2, RP11-646E18.2, ATP11B, DCUN1D1, MCCC1, B3GNT5, KLHL24, YEATS2, MAP6D1, PARL, ABCC5, EIF2B5, DVL3, AP2M1, ABCF3, ALG3, ECE2, PSMD2, EIF4G1, FAM131A, CLCN2, POLR2H, EPHB3, MAGEF1, VPS8, C3orf70, EHHADH, MAP3K13, TMEM41A, SENP2, RP11-537I16.2, IGF2BP2, RP11-443P15.2, TBCCD1, DNAJB11, PSMD10P2, RP11-573D15.9, EIF4A2, RFC4, RPL39L, RTP4, RP11-567G11.1, BCL6, LPP, TPRG1, TP63, LEPREL1, IL1RAP, CCDC50, OPA1, ATP13A3, TMEM44-AS1, TMEM44, AC046143.3, LSG1, AC046143.2, FAM43A, XXYLT1, ACAP2, PPP1R2, LINC00969, MIR570, MUC4, TNK2, RP11-480A16.1, SDHAP1, TFRC, PCYT1A, TM4SF19-AS1, TM4SF19, UBXN7, RNF168, WDR53, FBXO45, PIGX, CEP19, PAK2, SENP5, AC127904.2, NCBP2-AS2, PIGZ, MFI2, MFI2-AS1, DLG1, DLG1-AS1, BDH1, AC024560.3, KIAA0226, FYTTD1, LRCH3, RPL35A, LMLN*
4q	4	--	*NMU, RASGEF1B, TSPAN5, RP11-402J6.3*
5p	32	--	*CCDC127, C5orf55, CTD-2228K2.5, CEP72, AC026740.1, TRIP13, MRPL36, NDUFS6, IRX2, C5orf38, MED10, FAM173B, CTD-2256P15.2, TRIO, FAM105B, MYO10, GUSBP1, MTMR12, AMACR, LMBRD2, SKP2, NADK2, C5orf42, CTD-2127H9.1, RICTOR, TTC33, RPL37, OXCT1, CTD-2201E18.3, CCL28, C5orf34, NNT*
5q	1	--	*CTC-338M12.5*
6p	6	--	*GFOD1, JARID2, ZNF204P, ZNF391, TCP11, GLTSCR1L*
6q	1	--	*XXyac-YX65C7_A.2*

7p	4	--	*PRKAR1B, HEATR2, AC073957.15, INTS1*
7q	2	--	*FAM185BP, RP11-467H10.1*
8p	17	*KAT6A*	*RP11-115C21.2, AGPAT5, CTA-398F10.2, PINX1, XKR6, AF131215.2, AF131215.8, TNFRSF10A, RP11-1149O23.3, GTF2E2, ERLIN2, DDHD2, KAT6A, AP3M2, RP11-503E24.2, SMIM19, HGSNAT*
8q	81	*CHCHD7, COX6C, PABPC1, RECQL4*	*SPIDR, PRKDC, MCM4, RP11-110G21.1, RPS20, CHCHD7, ASPH, MTFR1, RRS1, MCMDC2, SNHG6, LACTB2, RPL7, GDAP1, RP11-27N21.3, MRPS28, RP11-26J3.3, DECR1, RAD54B, ESRP1, CCNE2, PLEKHF2, UQCRB, RPL30, HRSP12, STK3, RP11-410L14.2, VPS13B, COX6C, ANKRD46, PABPC1, RRM2B, NUDCD1, MED30, DSCC1, MTBP, ATAD2, RNF139-AS1, NDUFB9, RP11-532M24.1, PVT1, AC083843.1, TRAPPC9, CASC7, AGO2, PTK2, TSNARE1, LY6K, THEM6, ZFP41, RP13-582O9.5, ZNF696, RP13-582O9.7, TOP1MT, ZC3H3, RP11-661A12.14, PYCRL,TSTA3, ZNF623, MAPK15, SCRIB, OPLAH, EXOSC4, CYC1, FAM203A, MROH1, BOP1, ADCK5, CPSF1, SLC39A4, TONSL, KIFC2, PPP1R16A, GPT, MFSD3, RECQL4, C8orf82, ARHGAP39, ZNF34, RPL8, ZNF250*
9p	8	--	*FOXD4, RP11-12D24.6, PLGRKT, ACER2, IFT74, TCEA1P4, SNORD121A, FAM214B*
9q	5	--	*SNORA84, PIP5KL1, PKN3, DOLPP1, NSMF*
10p	4	--	*SEC61A2, MCM10, ANKRD26, MASTL*
10q	1	--	*ELOVL3*
11p	1	--	*DEPDC7*
11q	8	--	*SCGB1A1, RIN1, ANAPC15, STARD10, KCTD14, KCTD21, NARS2, ANKRD42*
12p	28	--	*CCDC77, FBXL14, ITFG2, FOXM1, RHNO1, TULP3, RAD51AP1, RP11-1038A11.3, CD27-AS1, MRPL51, NCAPD2, ING4, CDCA3, FAM66C, RIMKLB, PHC1, A2M-AS1, RP11-599J14.2, RP11-705C15.2, RP11-705C15.3, MAGOHB, LRP6, LOH12CR2, HIST4H4, LDHB, STK38L, KLHL42, DDX11-AS1*
12q	3	--	*RPS11P6, RP11-214K3.20, RP11-214K3.19*
13q	5	--	*SHISA2, GPR180, MBNL2, RAP2A, IRS2*
14q	3	--	*SLC7A8, HMGN2P6, RP11-47I22.3*
16p	1	--	*RP11-22P6.3*
17q	10	--	*ANKRD13B, RP11-166P13.3, MIR4737, RP11-583F2.6, CASKIN2, TNRC6C, CBX4, RP11-1055B8.7, RP11-498C9.15, FN3KRP*
18p	5	--	*TYMS, YES1, SOGA2, RP11-888D10.3, SLMO1*
18q	9	*BCL2*	*B4GALT6, RP11-549B18.1, ZNF396, RNF165, IER3IP1, RP11-126O1.5, RP11-27G24.1, BCL2, CBLN2*
19p	1	--	*ARID3A*
19q	55	*CIC, POLD1*	*SNORA68, CTD-2085J24.4, C19orf40, HAUS5, C19orf55, ARHGAP33, LRFN3, WDR62, SIPA1L3, CTB-102L5.8, RN7SL663P, ACTN4, CAPN12, AC008982.2, SARS2, PAK4, ZNF780B, MAP3K10, EXOSC5, ATP5SL, CIC, MEGF8, AC006953.1, ZNF283, ZNF45, ZNF221, ZNF112, ZNF180, ARHGAP35, BBC3, CCDC9, KPTN, LIG1, C19orf68, MAMSTR, LIN7B, PPFIA3, TEAD2, PRR12, PRMT1, PNKP, POLD1, SPACA6P, ZNF841, TSEN34, TMEM86B, ZNF784, ZFP28, ZNF71, TRAPPC2P1, ZNF749, AC012313.1, CTD-2619J13.14, CTD-2619J13.16, MZF1*
20p	16	--	*SNORD119, MAVS, PCNA, MCM8, ANKEF1, SLX4IP, DZANK1, MIR3192, RBBP9, DTD1, RP1-122P22.2, PYGB, GINS1, NINL, ZNF337-AS1, ZNF337*
20q	59	--	*RP3-324O17.4, HM13-IT1, TPX2, PDRG1, TM9SF4, PLAGL2, POFUT1, DNMT3B, RP5-1085F17.3, CDK5RAP1, SNTA1, CBFA2T2, NECAB3, ACTL10, E2F1, PXMP4, CHMP4B, RALY, RP1-64K7.4, AHCY, DYNLRB1, PIGU, NCOA6, GGT7, EIF6, UQCC1, CEP250, CPNE1, PHF20, LINC00657, DSN1, SOGA1, RBL1, SNORA71B, SNORA71A, SNORA71C, SNHG11, RN7SKP173, LPIN3, CHD6, IFT52, MYBL2, FITM2, DBNDD2, SNX21, SPATA25, ARFGEF2, SNORD12, TMEM189, ZFP64, FAM210B, HMGB1P1, RP1-309F20.3, SYCP2, FAM217B, PPP1R3D, RP11-157P1.4, LAMA5, RPS21*
21q	1	--	*KCNJ15*

Supplemental Table S5K. Genes whose differential up-regulation was associated with subchromosomal copy number gains in HPV-positive OSCC.

Listed here are differentially expressed genes associated with local copy number gains (ploidy >=2.5 vs. 1.5 <= ploidy < 2.5), identified in 101 HPV-positive tumors with WGS and RNA-seq data. Transcript levels were compared for genes at sites with local copy number gains in at least 3 samples vs. normal copy numbers, using the one-tailed t-test (FDR adjusted p-value<0.01). The table shows the chromosome arm, number of genes that were significantly up-regulated by copy number gain, the subset of genes listed in the Cancer Gene Census database (https://cancer.sanger.ac.uk/census), and all genes whose expression was significantly up-regulated in association with local copy number gain.

arm	count	census gene	gene
1p	7	--	*SMIM1, CTNNBIP1, FBXO44, VPS13D, TMEM56, LINC01160, MIR548AC*
1q	1	--	*MIR4260*
2p	2	--	*RP11-521D12.5, SMC6*
2q	36	*ACSL3*	*CD8BP, SOWAHC, PSD4, RALB, CLASP1, CACNB4, BAZ2B, RBMS1, ITGA6, Y_RNA, INPP1, MFSD6, PIKFYVE, KANSL1L, AC007038.7, RPL37A, ZNF142, STK36, TTLL4, ABCB6, ACSL3, AGFG1, SLC16A14, SP140L, TIGD1, DGKD, USP40, AGAP1, PER2, TRAF3IP1, ASB1, HDAC4, ANKMY1, FARP2, AC133528.2, ING5*
3p	79	*VHL, PBRM1, FHIT, MITF*	*ITPR1, OGG1, VHL, IQSEC1, LSM3, FGD5-AS1, MRPS25, SH3BP5-AS1, HACL1, TBC1D5, SATB1, AC144521.1, NR1D2, CMC1, GPD1L, snoU13, EXOG, WDR48, ENTPD3-AS1, ZNF619, NKTR, ABHD5, ZNF852, RP11-348P10.2, TMEM42, LIMD1, LZTFL1, FYCO1, NBEAL2, NRADDP, SMARCC1, ZNF589, FCF1P2, TMA7, TREX1, UQCRC1, CELSR3, IP6K2, PRKAR2A, SLC25A20, P4HTM, WDR6, KLHDC8B, C3orf62, NICN1, MST1, TUSC2, RASSF1, HEMK1, VPRBP, ABHD14B, ABHD14A, ACY1, RPL29, POC1A, PPM1M, GLYCTK, DNAH1, PHF7, SMIM4, PBRM1, SNORD69, NEK4, SFMBT1, RP11-884K10.7, CCDC66, ARHGEF3, HESX1, APPL1, ABHD6, PXK, RP11-802O23.3, KCTD6, FHIT, ATXN7, KBTBD8, MITF, RP11-803B1.2, FAM86DP*
4p	11	--	*ZNF141, ZNF721, RP11-440L14.1, TMEM175, DGKQ, MAEA, AC016773.1, MXD4, HTT, BOD1L1, STIM2*
4q	9	--	*MAD2L1, SPATA5, JADE1, RPS3A, RP11-372K14.2, RP11-164P12.3, RP11-555K12.1, RP11-798M19.3, SAP30*
5q	1	--	*SMN2*
6p	22	*PIM1*	*SERPINB1, RP11-560J1.2, LRRC16A, ZNF165, ZKSCAN8, ZKSCAN3, HLA-G, C6orf47, MSH5, HSPA1A, PPT2, AGPAT1, PBX2, XXbac-BPG157A10.21, UHRF1BP1, ANKS1A, PIM1, TMEM217, PRICKLE4, PEX6, TMEM63B, RN7SKP116*
6q	8	*MYB, ARID1B*	*PTP4A1, KIAA1009, RRAGD, MYB, ARID1B, ZDHHC14, SNORA20, RP1-167A14.2*
7q	19	--	*AC004980.8, AC004980.9, IFRD1, TMEM168, RP11-274B21.10, EPHB6, PIP, CASP2, ZNF786, ZNF862, ATP6V0E2, ZBED6CL, REPIN1, ZNF775, NUB1, PRKAG2, GALNT11, PTPRN2, NCAPG2*
8p	2	--	*ASAH1, RP11-90P5.7*
10p	6	--	*SEC61A2, THNSL1, MASTL, RP11-305E6.4, MAP3K8, PARD3*
10q	54	*PTEN, KIAA1598*	*FAM21A, PBLD, TYSND1, RP11-152N13.5, GLUD1P3, VDAC2, COMTD1, RPS24, PPIF, TMEM254-AS1, TMEM254, FAM213A, GLUD1, MINPP1, ATAD1, KLLN, PTEN, LIPA, KIF20B, FGFBP3, KIF11, MYOF, CEP55, PDLIM1, ENTPD1-AS1, CCNJ, RP11-175O19.4, PGAM1, CNNM1, DNMBP, ERLIN1, SEMA4G, DPCD, NPM3, KCNIP2, ELOVL3, RP11-18I14.10, CNNM2, USMG5, SLK, SFR1, GSTO1, MXI1, KIAA1598, RGS10, PLEKHA1, ACADSB, LHPP, STK32C, LRRC27, RP11-122K13.12, FUOM, PAOX, MTG1*
11p	9	*DDB2*	*RRM1, SWAP70, AMPD3, GALNT18, PIK3C2A, NUCB2, SAA4, C11orf94, DDB2*
11q	86	*NUMA1, EED, ATM, FLI1*	*PPP1R32, PLCB3, CCDC88B, POLD4, CHKA, MRPL21, IGHMBP2, TPCN2, ORAOV1, FADD, PPFIA1, CTTN, DHCR7, NADSYN1, FAM86C1, NUMA1, LRTOMT, ANAPC15, INPPL1, CLPB, ATG16L2, FCHSD2, RELT, PAAF1, C2CD3, PPME1, POLD3, RP11-147I3.1, NEU3, RPS3, SNORD15B, UVRAG, PRKRIR, PAK1, CLNS1A, RSF1, AAMDC, NDUFC2, KCTD21-AS1, KCTD21, USP35, GAB2, NARS2, PRCP, C11orf82, RAB30, RAB30-AS1, RP11-727A23.4, RP11-727A23.5, RP11-727A23.11, ANKRD42, CREBZF, EED, RP11-320L11.2, TMEM135, SMCO4, TAF1D, SNORD6, C11orf54, MRE11A, RP11-712B9.2, SESN3, FAM76B, JRKL, DCUN1D5, RP11-693N9.2, CASP4, CASP1P2, CARD17, ACAT1, NPAT, ATM, IL18, C11orf71, SIDT2, IL10RA, AMICA1, RPS25, UBASH3B, CRTAM, GRAMD1B, SIAE, AP001007.1, FLI1, ZBTB44, SNX19*

12p	4	--	*TEAD4, RP11-726G1.1, DUSP16, RP11-153K16.1*
13q	43	--	*CRYL1, IFT88, RP11-124N19.3, MIPEP, PARP4, CENPJ, MTMR6, NUPL1, USP12, RPL21, RASL11A, UBL3, HSPH1, PDS5B, RFC3, RFXAP, ELF1, AKAP11, EPSTI1, SLC25A30, LRCH1, RCBTB2, FNDC3A, ARL11, EBPL, RNASEH2B, RP11-24B19.3, WDFY2, ALG11, NEK3, BORA, COMMD6, UCHL3, MYCBP2, NDFIP2, STK24, DOCK9, H2AFZP3, TEX30, BIVM, TUBGCP3, DCUN1D2, UPF3A*
14q	74	*ARHGAP5, KTN1, GPHN, RAD51B, DICER1, HSP90AA1*	*PNP, ZNF219, TRAV12-3, TRAJ5, HECTD1, NUBPL, ARHGAP5, MBIP, MIS18BP1, POLE2, L2HGDH, ATP5S, RP11-247L20.4, MAP4K5, PYGL, LINC00640, FRMD6, ERO1L, GNPNAT1, WDHD1, DLGAP5, KTN1-AS1, KTN1, CTD-2002H8.2, DAAM1, GPR135, L3HYPDH, DHRS7, SLC38A6, RP11-902B17.1, SGPP1, SYNE2, MTHFD1, GPHN, PLEKHH1, ARG2, RAD51B, MAP3K9, RP6-65G23.3, RP6-114E22.1, DCAF4, PNMA1, PTGR2, ENTPD5, FLVCR2, C14orf1, RP11-7F17.7, RP11-7F17.1, POMT2, GSTZ1, SNORA46, SPTLC2, GALC, KCNK10, FOXN3-AS1, TTC7B, RPS6KA5, NDUFB1, LGMN, ASB2, DICER1, CCDC85C, WARS, RP11-1029J19.4, DYNC1H1, HSP90AA1, ANKRD9, RCOR1, RP11-73M18.10, ZFYVE21, C14orf2, ADSSL1, CEP170B, AHNAK2*
15q	8	--	*RP11-540B6.6, ARHGAP11A, RNU6-353P, SQRDL, RP11-69G7.1, RPP25, PDE8A, MIR3174*
16p	7	--	*PRSS22, DNASE1, RP11-473I1.9, SYT17, IL4R, SULT1A1, PRSS8*
16q	90	*CBFB*	*SHCBP1, ORC6, GPT2, NETO2, SIAH1, RP11-21B23.2, RP11-467J12.4, CHD9, RBL2, LPCAT2, CAPNS2, BBS2, MT1H, MT1X, NLRC5, CCL17, GPR56, KATNB1, ZNF319, USB1, SETD6, SNORA50, SLC38A7, GOT2, CMTM4, DYNC1LI2, NAE1, PDP2, RP11-61A14.2, RP11-61A14.3, CDH16, FAM96B, CES2, CES3, CBFB, C16orf70, TRADD, ELMO3, PLEKHG4, HSD11B2, GFOD2, NUTF2, PSKH1, NFATC3, ESRP2, CDH3, HAS3, CIRH1A, SNTB2, NIP7, EXOSC6, MARVELD3, ATXN1L, DHODH, RP11-252A24.2, RP11-252A24.7, RFWD3, LDHD, BCAR1, CFDP1, RP11-252K23.2, ADAT1, MON1B, NUDT7, CMC2, CENPN, GCSH, GAN, CMIP, OSGIN1, TAF1C, RP11-486L19.2, RP11-517C16.2, TLDC1, KLHL36, C16orf74, ZCCHC14, FLJ00104, CYBA, MVD, PIEZO1, APRT, GALNS, RP11-46C24.7, ZNF778, VPS9D1, ZNF276, AFG3L1P, DBNDD1, GAS8*
17p	20	--	*GEMIN4, RNMTL1, SERPINF2, RAP1GAP2, GSG2, ANKFY1, MINK1, ZFP3, ZNF594, KIAA0753, GPS2, AC025335.1, LINC00324, CTC1, RANGRF, AC135178.1, USP32P2, TVP23B, FAM83G, B9D1*
17q	9	--	*ARL17B, TRIM47, FOXJ1, AC127496.1, C17orf89, MAFG, CCDC57, HEXDC, TBCD*
19p	15	--	*RNF126, ATP5D, CIRBP, NDUFS7, LSM7, DOHH, AC007292.6, AC027319.1, PPAN, QTRT1, GIPC1, CYP4F3, JUND, NR2C2AP, ZNF682*
20q	8	--	*C20orf24, RP4-564F22.5, SNHG17, TTPAL, Z97053.1, SNORD12C, FAM210B, CABLES2*
21q	2	--	*BACH1, AP001056.1*
22q	10	--	*VPREB3, DDT, SNORD125, LIF, DUSP18, MIR659, MICALL1, TSPO, PLXNB2, RABL2B*

Supplemental Table S5L. Genes whose differential down-regulation was associated with subchromosomal copy number losses in HPV-positive OSCC.

Listed here are differentially expressed genes associated with local copy number losses (ploidy < 1.5 vs. 1.5 <= ploidy < 2.5), identified in 101 HPV-positive tumors with WGS and RNA-seq data. Transcript levels were compared for genes at sites with local copy number losses in at least 3 samples vs. normal copy numbers, using the one-tailed t-test (FDR adjusted p-value<0.01). The table shows the chromosome arm, the number of genes that were significantly down-regulated by copy number loss, the subset of genes listed in the Cancer Gene Census database, (https://cancer.sanger.ac.uk/census) and all genes whose expression was significantly down-regulated in association with local copy number loss.

arm	count	census gene	gene
1p	4	--	*C1orf210, RP4-592A1.2, CTBS, ARHGAP29*
1q	26	--	*SETDB1, LYSMD1, SNX27, SPRR2D, SPRR2A, S100A14, ILF2, CKS1B, EFNA4, RAB25, RRNAD1, DUSP23, IGSF9, USP21, RP11-122G18.5, NUF2, C1orf112, LAD1, UBE2T, SLC41A1, PIGR, TMEM206, FLVCR1, C1orf131, MLK4, CHRM3*
2q	22	*SF3B1*	*ITPRIPL1, DAPL1, GRB14, SSB, CDCA7, SP3, AC010894.3, HNRNPA3, AC079305.10, PRKRA, RP11-65L3.4, SLC39A10, PGAP1, SF3B1, C2orf69, TYW5, ORC2, FZD7, NOP58, FAM117B, NBEAL1, ABI2*
3q	39	*TBL1XR1, MAP3K13*	*DHFRL1, PDCL3P4, MIR567, HSPBAP1, RP11-93K22.13, FAM86HP, SLC25A36, RNF7, TM4SF1-AS1, COMMD2, C3orf33, SLC33A1, CCNL1, RSRC1, IFT80, MYNN, PRKCI, TBL1XR1, ACTL6A, MRPL47, ATP11B, DCUN1D1, PARL, EIF2B5, DVL3, AP2M1, CLCN2, POLR2H, MAGEF1, MAP3K13, DNAJB11, FETUB, LINC01063, NCBP2, NCBP2-AS2, DLG1, AC128709.3, AC024560.3, RPL35A*
4p	4	--	*ATP5I, GRK4, PPP2R2C, CCDC96*
5p	44	*SDHA, DROSHA*	*CCDC127, SDHA, PDCD6, C5orf55, EXOC3, CTD-2228K2.5, CEP72, AC026740.1, CTD-2589H19.6, BRD9, TRIP13, CLPTM1L, LPCAT1, RP11-43F13.1, MRPL36, NDUFS6, MED10, NSUN2, PAPD7, FASTKD3, SNORD123, FAM173B, CCT5, CTD-2256P15.2, MARCH6, DAP, FAM105B, DROSHA, C5orf22, CTD-2186M15.3, RAI14, CTD-2517O10.6, RAD1, BRIX1, DNAJC21, WDR70, CTD-2127H9.1, RPL37, C5orf51, ZNF131, PAIP1, NNT-AS1, MRPS30, RP11-53O19.3*
5q	2	--	*ESM1, BOD1*
6p	2	--	*ABCF1, EFHC1*
6q	7	--	*RP3-355L5.5, VNN1, VNN3, PERP, SNORA2, RP11-288H12.3, SLC22A3*
7p	45	*PMS2, EGFR*	*COX19, INTS1, MAFK, PSMG3, MAD1L1, FTSJ2, SNX8, EIF3B, CHST12, BRAT1, RNF216P1, TNRC18, PMS2, USP42, DAGLB, ZNF12, PMS2CL, COL28A1, RPA3, PHF14, BZW2, TWISTNB, KLHL7, NUPL2, MPP6, OSBPL3, CYCS, HOXA1, SCRN1, PLEKHA8, RPS27P16, ZNRF2, AVL9, RP11-379H18.1, POU6F2, COA1, POLD2, PURB, SNHG15, SNORA5C, HUS1, C7orf57, FIGNL1, EGFR, CCT6A*
7q	9	--	*TMEM60, PCLO, ZKSCAN1, RP11-126L15.4, UFSP1, LINC01004, PUS7, CBLL1, PTPRZ1*
8p	17	*WRN, WHSC1L1*	*REEP4, PPP2R2A, PPP2CB, WRN, ZNF703, ERLIN2, PROSC, BRF2, RAB11FIP1, ASH2L, STAR, LSM1, BAG4, DDHD2, WHSC1L1, RP11-503E24.2, FNTA*
8q	155	*TCEA1, CHCHD7, NCOA2, NBN, PABPC1, UBR5, EIF3E, RAD21, EXT1, NDRG1, RECQL4*	*UBE2V2, PCMTD1, RP11-110G21.1, RB1CC1, TCEA1, LYPLA1, MRPL15, TMEM68, TGS1, CHCHD7, IMPAD1, GGH, YTHDF3, RP11-16E18.3, ARMC1, MTFR1, RRS1, MYBL1, VCPIP1, ARFGEF1, NCOA2, TRAM1, LACTB2, TERF1, RPL7, UBE2W, TMEM70, PEX2, MRPS28, IMPA1, ZFAND1, SNX16, C8orf59, WWP1, RMDN1, CPNE3, OSGIN2, NBN, DECR1, TMEM64, GS1-251I9.4, OTUD6B, TRIQK, TMEM67, RAD54B, KIAA1429, ESRP1, DPY19L4, INTS8, CCNE2, PLEKHF2, C8orf37, UQCRB, MTERFD1, PTDSS1, MTDH, RNU7-177P, LAPTM4B, HRSP12, STK3, VPS13B, RNF19A, ANKRD46, PABPC1, YWHAZ, GRHL2, UBR5, AZIN1, ATP6V1C1, FZD6, SLC25A32, DCAF13, LRP12, OXR1, RP11-649G15.2, EIF3E, NUDCD1, ENY2, EBAG9, UTP23, RAD21, MED30, EXT1, TAF2, DSCC1, MRPL13, MTBP, TBC1D31, ZHX1, ATAD2, FAM91A1, TMEM65, TRMT12, RNF139, TATDN1, RP11-532M24.1, RP11-1082L8.3, ZNF572, KIAA0196, TRIB1, FAM84B, PVT1, FAM49B, EFR3A, PHF20L1, NDRG1, AC083843.1, CTA-204B4.2, CHRAC1, CASC7, AGO2, PTK2, TSNARE1, LY6K, ZFP41, GLI4, RP13-582O9.5, ZNF696, RP13-582O9.7, ZC3H3, TIGD5, PYCRL, ZNF623, ZNF707, FAM83H, SCRIB, PUF60, PLEC, GRINA, OPLAH, EXOSC4, GPAA1, CYC1, MAF1, MROH1, HSF1, DGAT1, SLC52A2, CPSF1, TONSL, MFSD3, RECQL4, LRRC14, C8orf82, ARHGAP39, ZNF251, ZNF34, RPL8, ZNF517, ZNF7, COMMD5, ZNF250, ZNF16, ZNF252P, C8orf33*

9p	50	*FANCG*	*TMEM261, MPDZ, HAUS6, SCARNA8, RPS6, TUSC1, CAAP1, TOPORS-AS1, NDUFB6, GVQW1, APTX, TCEA1P4, DNAJA1, SMU1, B4GALT1, RP11-326F20.5, CHMP5, NOL6, UBE2R2, UBAP2, DCAF12, UBAP1, KIF24, NUDT2, KIAA1161, FAM219A, RPP25L, DCTN3, SIGMAR1, VCP, FANCG, PIGO, STOML2, FAM214B, TESK1, ARHGEF39, CREB3, GBA2, RGP1, HINT2, GLIPR2, CLTA, HMGB3P24, RNF38, ZCCHC7, GRHPR, POLR1E, TOMM5, TRMT10B, EXOSC3*
9q	10	--	*PTAR1, AGTPBP1, NAA35, ERCC6L2, TMEM246, SUSD1, POLE3, C9orf43, TRIM32, ZBTB34*
10q	4	*CCDC6*	*CCDC6, MCU, RP11-464F9.20, GLUD1P3*
11p	1	--	*LGR4*
11q	146	*MEN1, CCND1, NUMA1*	*RNU6-118P, SNHG1, C11orf84, DNAJC4, VEGFB, RP11-783K16.5, BAD, TRMT112, MAP4K2, MEN1, CDC42BPG, PPP2R5B, ARL2, SAC3D1, ZNHIT2, AP003068.18, CDC42EP2, DPF2, SSSCA1-AS1, SSSCA1, PCNXL3, CFL1, MUS81, FIBP, CCDC85B, FOSL1, C11orf68, DRAP1, SART1, EIF1AD, BANF1, RAB1B, YIF1A, BRMS1, MRPL11, CTD-3074O7.5, RBM4, RP11-658F2.8, C11orf80, RCE1, SYT12, KDM2A, ADRBK1, ANKRD13D, SSH3, RAD9A, PPP1CA, RPS6KB2, CORO1B, TMEM134, AIP, GSTP1, NDUFV1, FAM86C2P, NDUFS8, SUV420H1, C11orf24, LRP5, PPP6R3, CPT1A, MRPL21, IGHMBP2, TPCN2, MIR3164, MYEOV, CCND1, ORAOV1, ANO1, FADD, PPFIA1, AP000487.6, CTTN, DHCR7, RP11-660L16.2, NADSYN1, FAM86C1, RP11-849H4.2, RNF121, NUMA1, LRTOMT, LAMTOR1, ANAPC15, INPPL1, CLPB, ARAP1, STARD10, ATG16L2, FCHSD2, RP11-800A3.2, P2RY2, RP11-800A3.4, RAB6A, MRPL48, COA4, PAAF1, RP11-707G14.7, PPME1, AP001372.2, SPCS2, RPS3, UVRAG, PRKRIR, RP11-111M22.3, C11orf30, TSKU, RP11-21L23.3, RP11-21L23.2, PAK1, CLNS1A, RSF1, INTS4, NDUFC2, ALG8, KCTD21-AS1, KCTD21, USP35, NARS2, PRCP, RAB30-AS1, RP11-727A23.5, ANKRD42, CCDC90B, RP11-680H20.2, MRE11A, KIAA1377, YAP1, BIRC2, TMEM123, RP11-315O6.1, RP11-690D19.3, DCUN1D5, DYNC2H1, PTS, RP11-159N11.4, ZW10, TMPRSS13, MPZL3, IFT46, HINFP, NLRX1, PVRL1, TRIM29, AP001007.1, DCPS, ARHGAP32, RP11-679I18.4*
12p	6	*ETNK1*	*SCARNA10, WBP11, AEBP2, ETNK1, ARNTL2, KLHL42*
13q	12	*BRCA2*	*GJA3, LNX2, BRCA2, N4BP2L2, ELF1, CLN5, FBXL3, TGDS, GPR180, DZIP1, DOCK9, TMTC4*
14q	79	*ARHGAP5, KTN1, HIF1A*	*PARP2, RAB2B, MRPL52, PABPN1, NGDN, TM9SF1, AP4S1, HECTD1, ARHGAP5-AS1, ARHGAP5, BRMS1L, MBIP, FKBP3, POLE2, KLHDC2, NEMF, VCPKMT, SOS2, MAP4K5, SAV1, PYGL, RP11-218E20.3, LINC00640, FRMD6, ERO1L, GNPNAT1, GMFB, CGRRF1, WDHD1, FBXO34, ATG14, KTN1-AS1, KTN1, EXOC5, AP5M1, ARID4A, TIMM9, KIAA0586, DAAM1, PCNXL4, DHRS7, MNAT1, HIF1A, SNAPC1, HSPA2, PLEK2, RDH11, ZFYVE26, DCAF5, EXD2, ZFYVE1, PTGR2, POMT2, SPTLC2, CEP128, GTF2A1, STON2, FLRT2, ZC3H14, EML5, TTC8, FOXN3-AS1, NRDE2, TTC7B, BTBD7, IFI27, IFI27L2, ATG2B, CCNK, EML1, DYNC1H1, RCOR1, TRMT61A, APOPT1, RP11-894P9.1, LINC00638, AHNAK2, PACS2, MTA1*
15q	1	--	*MYO5C*
16p	1	--	*SNRNP25*
17p	2	--	*SMG6, FOXO3B*
17q	12	--	*UNC119, ABHD15, RP11-68I3.4, SUZ12P, CTD-2349P21.9, PEX12, GGNBP2, MRM1, RP11-156P1.3, TOM1L1, CTD-3010D24.3, ST6GALNAC1*
18q	8	--	*CABLES1, OSBPL1A, ATP5A1, HAUS1, C18orf25, PIAS2, HDHD2, IER3IP1*
20p	9	--	*FAM110A, PCED1A, UBOX5, MIR103A2, TMEM230, CRLS1, MKKS, ZNF133, ZNF337*
20q	39	*ASXL1, SS18L1*	*PDRG1, POFUT1, ASXL1, NECAB3, PXMP4, AHCY, ITCH, NCOA6, NFS1, EPB41L1, SOGA1, TTI1, RPRD1B, RALGAPB, FAM83D, DHX35, OSER1, UBE2C, ZSWIM1, CSE1L, SPATA2, UBE2V1, PARD6B, ZFP64, ZNF217, AURKA, PPP4R1L, VAPB, STX16, NELFCD, SLMO2, TAF4, SS18L1, RP11-157P1.4, MRGBP, DIDO1, ABHD16B, UCKL1, AL118506.1*

22q	5	--	*CECR1, RANBP1, MORC2, TRIOBP, MIEF1*

Supplemental Table S5M. Genes whose differential up-regulation was associated with subchromosomal copy number gains in HPV-negative OSCC.

Listed here are differentially expressed genes associated with local copy number gains (ploidy >=2.5 vs. 1.5 <= ploidy < 2.5), identified in 50 HPV-negative tumors with WGS and RNA-seq data. Transcript levels were compared for genes at sites with local copy number gains in at least 3 samples vs. normal copy numbers, using the one-tailed t-test (FDR adjusted p-value<0.01). The table shows the chromosome arm, number of genes that were significantly up-regulated by copy number gain, the subset of genes listed in the Cancer Gene Census database (https://cancer.sanger.ac.uk/census), and all genes whose expression was significantly up-regulated in association with local copy number gain.

arm	count	census gene	gene
1p	4	*NRAS*	*snoU13, CPT2, USP1, NRAS*
2p	2	--	*REEP1, CH17-132F21.1*
2q	1	--	*CXCR2P1*
3p	1	--	*MANF*
5q	15	--	*NDUFS4, PPAP2A, ZSWIM6, POC5, ZBED3, BHMT2, ATG10, ST8SIA4, C5orf56, IRF1, APBB3, VTRNA1-3, SPRY4, COL23A1, C5orf45*
6q	4	--	*MB21D1, SLC22A1, MPC1, PSMB1*
8p	2	--	*ERICH1, UBXN8*
9p	10	*JAK2, CDKN2A*	*JAK2, CDKN2A, IFT74, GVQW1, KIAA1161, DCTN3, VCP, RUSC2, SLC25A51, SHB*
9q	2	--	*PPP1R26, AGPAT2*
11q	5	*PAFAH1B2*	*TENM4, TAF1D, PPP2R1B, PAFAH1B2, APLP2*
13q	2	--	*FAM207BP, UFM1*
14q	2	--	*LINC00641, SPTSSA*
17p	2	--	*TXNDC17, MYH10*
17q	1	--	*TUBG1*
18p	1	--	*PSMG2*
18q	10	--	*C18orf8, NPC1, TTC39C, RNF125, MOCOS, CXXC1, WDR7, VPS4B, SOCS6, RBFA*
19p	9	--	*ABCA7, HMHA1, SBNO2, NCLN, TMIGD2, C19orf10, DPP9, TICAM1, EVI5L*
21q	1	--	*UBE2G2*
22q	1	--	*CTA-250D10.23*

Supplemental Table S5N. Genes whose differential down-regulation was associated with subchromosomal copy number losses in HPV-negative OSCC.
Listed here are genes whose differential expression associated with local copy number losses (ploidy < 1.5 vs. 1.5 <= ploidy < 2.5), identified in HPV-negative tumors with WGS data. Transcript levels were compared for genes at sites with local copy number losses in at least 3 samples vs. normal copy numbers, using the one-tailed t-test (FDR adjusted p-value<0.01). The table shows the chromosome arm, the number of genes that were significantly down-regulated by copy number loss, the subset of genes listed in the Cancer Gene Census database (https://cancer.sanger.ac.uk/census), and all genes whose expression was significantly down-regulated in association with local copy number loss.

GO biological process complete	**no. genes queried (n = 21042)**	**observed (n = 543)**	**expected**	**fold enrich-ment**	**P-value**	**genes**
non-motile cilium assembly (GO:1905515)	41	9	1.06	8.51	1.46E-02	*ARL13B, CEP250, IFT122, IFT52, IFT57, IFT74, IFT80, NPHP3, VANGL2*
DNA-dependent DNA replication (GO:0006261)	115	16	2.97	5.39	8.03E-04	*CCNE2, GINS1, LIG1, MCM10, MCM2, MCM4, MCM8, MCMDC2, PCNA, PNKP, POLD1, POLQ, RFC4, RRM2B, TONSL, TOPBP1*
translation (GO:0006412)	383	27	9.88	2.73	3.46E-02	*ABTB1, AGO2, DARS2, EEFSEC, EIF2B5, EIF4A2, EIF4G1, EIF6, GFM1, MRPL36, MRPL47, MRPL51, MRPS22, MRPS28, NARS2, PABPC1, RPL22L1, RPL24, RPL30, RPL35A, RPL37, RPL39L, RPL7, RPL8, RPS20, RPS21, SARS2*
mitotic cell cycle process (GO:1903047)	653	43	16.85	2.55	3.22E-04	*ANAPC15, ARID3A, CCNE2, CEP250, CEP70, CEP72, CHMP4B, DLG1, DSCC1, DSN1, E2F1, ECT2, FOXM1, GINS1, GPSM2, HAUS5, IQGAP3, LIG1, MASTL, MCM10, MCM2, MCM4, MCM8, MYBL2, NAA50, NCAPD2, NEK11, NINL, PCNA, PINX1, PRKDC, PRMT1, PSMD2, RRS1, SMC4, SPICE1, TFDP2, TOPBP1, TP63, TPX2, TRIP13, TYMS, WDR62*
cellular response to DNA damage stimulus (GO:0006974)	741	46	19.12	2.41	5.83E-04	*ACTL6A, ARID3A, ASTE1, ATR, BBC3, BCL2, CHCHD6, DTX3L, E2F1, FAAP24, FBXO45, FOXM1, GEN1, HIST1H4A, IFI16, ING4, ISY1, LIG1, MASTL, MCM10, MCM8, MCMDC2, NCOA6, NEK11, PARP9, PCNA, PNKP, POLD1, POLQ, POLR2H, PRKDC, PRMT1, RAD51AP1, RAD54B, RFC4, RHNO1, RNF168, RRM2B, RUVBL1, SPIDR, TFDP2, TONSL, TOPBP1, TP63, TRIP13, ZMAT3*

Supplemental Table S5O. PANTHER biological process ontology terms over-represented among differentially upregulated genes in HPV-positive tumors with copy number gains (gain vs. normal) We tested the statistical significance of GO Biological Process terms for genes that were differentially expressed in samples with copy number gains compared to samples with normal copy number in HPV-positive tumors, using PANTHER Overrepresentation Test (Released 2017-12-27) (http://www.PANTHERdb.org). We used GO Biological Process complete terms and applied multiple testing correction (Bonferroni p<0.05). The PANTHER test gave results with gene ontology clusters; similar ontology terms were clustered together such as ‘cell cycle’ and ‘mitotic cell cycle’. The most significantly increased GO terms for each ontology cluster in PANTHER outputs are shown in this table. Shown here are the GO term, the number of genes queried, (*obs.*) the number of genes observed in the differentially expressed gene list, (*exp.*) the number of genes expected by chance, fold enrichment (observed/expected), p-values, and the list of genes that were significantly upregulated.

GO biological process complete	no. genes queried (n = 21042)	obs. (n =501)	exp.	fold enrich -ment	P-value	genes
organophosphate biosynthetic process (GO:0090407)	514	34	12.24	2.78	1.29E-03	*ABHD5, ACAT1, ADSSL1, AGPAT1, AMPD3, APRT, ATM, ATP5D, ATP5S, CHKA, CRYL1, DGKQ, DHODH, GPD1L, GPHN, INPPL1, IP6K2, LHPP, LPCAT2, MBIP, MTHFD1, MTMR6, MVD, NADSYN1, PIK3C2A, PIKFYVE, PLEKHA1, PNP, PRKAG2, PTEN, PYGL, RRM1, SPTLC2, VPS9D1*
nucleoside phosphate metabolic process (GO:0006753)	525	31	12.5	2.48	4.55E-02	*ABHD14B, ACAT1, ADSSL1, AMPD3, APRT, ATP5D, ATP5S, DHODH, ENTPD5, FHIT, GPD1L, GPHN, HSPA1A, LHPP, MBIP, MTHFD1, NADSYN1, NDUFB1, NDUFC2, NDUFS7, NUDT7, OGG1, PDE8A, PGAM1, PNP, PRKAG2, RNASEH2B, RRM1, SULT1A1, UQCRC1, VPS9D1*
cell cycle (GO:0007049)	1361	68	32.4	2.1	7.31E-05	*ANAPC15, APPL1, ATM, BACH1, BORA, C2CD3, CABLES2, CASP2, CENPJ, CENPN, CEP55, CLASP1, DDIAS, DLGAP5, DYNC1H1, DYNC1LI2, GPS2, GSG2, HSP90AA1, HTT, KATNB1, KIAA0753, KIF11, KIF20B, KLHDC8B, MAD2L1, MAEA, MAP3K8, MASTL, MIS18BP1, MRE11, MSH5, MYB, NAE1, NCAPG2, NEK3, NEK4, NPAT, NUMA1, NUP58, ORAOV1, ORC6, PARD3, PBRM1, PDS5B, PIM1, POC1A, POLD3, POLD4, POLE2, PPME1, PRKAG2, PTP4A1, RAD51B, RALB, RASSF1, RBL2, RFC3, RFWD3, RPS3, RRAGD, RRM1, SIAH1, TBCD, TUBGCP3, TUSC2, UVRAG, WDR6*
organelle organization (GO:0006996)	3182	114	75.76	1.5	3.63E-02	*AGFG1, ARID1B, ATG16L2, ATM, ATP5D, ATP5S, ATP6V0E2, ATXN7, B9D1, BBS2, BCAR1, C2CD3, CARMIL1, CCDC88B, CELSR3, CENPJ, CENPN, CEP162, CHD9, CLASP1, CTC1, CTTN, DAAM1, DDB2, DLGAP5, DNAH1, DYNC1H1, DYNC1LI2, EED, FARP2, FOXJ1, FRMD6, GALNT11, GAN, GAS8, GLUD1, GSG2, HDAC4, HSP90AA1, HSPA1A, HTT, IFT88, IGHMBP2, ING5, INPPL1, JADE1, KATNB1, KIAA0753, KIF11, LIMD1, LSM3, MAD2L1, MAEA, MASTL, MBIP, MINK1, MIPEP, MIS18BP1, MRE11, MSH5, MYB, NCAPG2, NDUFAF8, NDUFB1, NDUFC2, NDUFS7, NIP7, NUBPL, NUMA1, NUP58, NUTF2, PAK1, PBRM1, PDS5B, PER2, PEX6, POC1A, POLD3, POLD4, POLE2, PPAN, PPIF, PTEN, RAB30, RABL2B, RAD51B, RBL2, RCOR1, RFC3, RPS3, RPS6KA5, RSF1, SAP30, SATB1, SETD6, SFMBT1, SHTN1, SLK, SMARCC1, SMC6, SNX19, STK36, SWAP70, SYT17, TMEM135, TMEM175, TPCN2, TRAF3IP1, TSPO, TUBGCP3, UVRAG, VPRBP, WDHD1, ZKSCAN3*

Supplemental Table S5P. PANTHER biological process ontology terms over-represented among differentially downregulated genes in HPV-positive tumors with copy number loss (loss vs. normal) We tested the statistical significance of GO Biological Process terms for genes that were differentially expressed in samples with copy number loss compared to samples with normal copy number in HPV-positive tumors, using PANTHER Overrepresentation Test (Released 2017-12-27). We used GO Biological Process complete terms and applied multiple testing correction (Bonferroni p<0.05). The PANTHER test gave results with gene ontology clusters. The most significantly increased GO terms for each ontology cluster in PANTHER outputs are shown in this table. Shown here are the GO term, the number of genes in reference human genome that were queried, (*obs.*) the number of genes observed in the differentially expressed gene list, (*exp.*) the number of genes expected by chance, fold enrichment (observed/expected), p-values, and the list of genes that were significantly downregulated.

GO biological process complete	no. genes queried (n = 21042)	obs. (n = 610)	exp.	fold enrich-ment	P-value	genes
double-strand break repair (GO:0006302)	162	20	4.7	4.26	9.44E-04	*APTX, BRCA2, EXD2, HUS1, KDM2A, MRE11, MTA1, MUS81, NBN, ORAOV1, PAPD7, RAD21, RAD54B, RPA3, TONSL, TRIP13, UVRAG, VCP, WRN, ZFYVE26*
meiotic cell cycle process (GO:1903046)	164	19	4.75	4	5.00E-03	*AURKA, BRCA2, CCNE2, FIGNL1, HSF1, HSPA2, HUS1, MRE11, MUS81, MYBL1, NSUN2, ORAOV1, RAD1, RAD21, RAD54B, TERF1, TOP6BL, TRIP13, WASHC5*
nucleic acid phosphodiester bond hydrolysis (GO:0090305)	290	26	8.41	3.09	5.91E-03	*AGO2, APTX, CPSF1, DCPS, DROSHA, EXD2, EXOSC3, EXOSC4, HRSP12, LACTB2, MRE11, MUS81, NCBP2, PGAP1, PMS2, POLD2, RAD1, RAD9A, RPA3, RRS1, SMG6, TATDN1, TOP6BL, UTP23, WRN, ZC3H3*
ncRNA processing (GO:0034470)	405	33	11.74	2.81	1.66E-03	*AGO2, CPSF1, DCAF13, DROSHA, EXOSC3, EXOSC4, INTS1, INTS4, INTS8, MRM2, NGDN, NOL6, NOP58, NSUN2, PRKRA, PUS7, RPL35A, RPL37, RPL7, RPL8, RPS3, RPS6, RRNAD1, RRS1, SART1, SSB, TRMT10B, TRMT112, TRMT12, TRMT61A, TYW5, UTP23, WBP11*
translation (GO:0006412)	383	31	11.1	2.79	4.46E-03	*ABCF1, AGO2, COA1, EGFR, EIF1AD, EIF2B5, EIF3B, EIF3E, IGHMBP2, MRPL11, MRPL13, MRPL15, MRPL21, MRPL36, MRPL47, MRPL48, MRPL52, MRPS28, MRPS30, NARS2, PABPC1, PAIP1, RPL35A, RPL37, RPL7, RPL8, RPS3, RPS6, RPS6KB2, TCEA1, TRMT112*
cell division (GO:0051301)	490	38	14.2	2.68	6.78E-04	*ANAPC15, AURKA, BOD1, BRCA2, CABLES1, CCND1, CCNE2, CCNK, CFL1, CHMP5, CKS1B, CLTA, DCTN3, DYNC1H1, EFHC1, FAM83D, FZD7, HAUS1, HAUS6, KLHL42, MAD1L1, NSUN2, NUF2, NUMA1, PARD6B, PPP1CA, RAD21, REEP4, RPS3, SAC3D1, SSSCA1, TERF1, TRIOBP, UNC119, WASHC5, ZFYVE26, ZNF16, ZW10*
mitotic cell cycle process (GO:1903047)	653	47	18.93	2.48	1.80E-04	*ANAPC15, AURKA, BANF1, BOD1, BRCA2, CCND1, CCNE2, CEP126, CEP72, CFL1, CHMP5, CKS1B, DCTN3, DLG1, DSCC1, DYNC1H1, EFHC1, EML1, HAUS1, HAUS6, HINFP, HUS1, MAD1L1, MNAT1, MRE11, MUS81, NBN, NUMA1, NUPL2, ORAOV1, ORC2, PAPD7, POLE2, POLE3, PPME1, PPP2R2A, RAD9A, RANBP1, REEP4, RPA3, RPS6, RRS1, TAF2, TERF1, TRIP13, UNC119, ZW10*
organelle localization (GO:0051640)	577	38	16.73	2.27	3.29E-02	*AP2M1, ATG14, AURKA, BOD1, BRAT1, CEP72, CHMP5, DCTN3, DLG1, DYNC1H1, EXOC5, FAM83D, HAUS1, HAUS6, HIF1A, KIF24, LAMTOR1, MAD1L1, MAP4K2, MKKS, NUMA1, PCLO, PDCD6, PPP6R3, PTK2, RAB1B, RAB6A, RRS1, SCRIB, STX16, SYT12, TERF1, TMEM230, TMEM67, TSNARE1, UVRAG, YWHAZ, ZW10*

intracellular transport (GO:0046907)	1265	72	36.67	1.96	4.71E-04	*AIP, AP2M1, AP4S1, AP5M1, ATG14, ATP5A1, ATP5I, BANF1, CHMP5, CLN5, CLTA, CPSF1, CPT1A, CTTN, CYC1, DCTN3, DDHD2, DYNC1H1, DYNC2H1, ENY2, HIF1A, IER3IP1, IFT46, IFT80, LAPTM4B, MKKS, NCBP2, NOL6, NUPL2, PCLO, PDCD6, PEX12, PLEKHA8, PPP6R3, PRKCI, PTK2, RAB1B, RAB6A, RANBP1, RGP1, RPL35A, RPL37, RPL7, RPL8, RPS3, RPS6, RRS1, SCRIB, SMG6, SNX16, SNX27, SNX8, SPCS2, SSB, STAR, STOML2, STX16, SYT12, TGS1, TMEM230, TOM1L1, TOMM5, TRAM1, TSNARE1, UBAP1, VAPB, VCP, WASHC5, YIF1A, YWHAZ, ZC3H3, ZW10*
regulation of cell cycle (GO:0051726)	1164	62	33.74	1.84	3.57E-02	*ANAPC15, AURKA, BAD, BIRC2, BRCA2, CABLES1, CCND1, CCNE2, CCNK, CCNL1, CEP72, CGRRF1, CHMP5, CKS1B, CREB3, DCTN3, DLG1, DYNC1H1, EGFR, FAM83D, FANCG, FIGNL1, FOSL1, HAUS1, HAUS6, HINFP, HSF1, HSPA2, HUS1, LAMTOR1, MAD1L1, MAP4K2, MEN1, MNAT1, MRE11, MTBP, MUS81, NBN, NDRG1, NSUN2, NUMA1, ORAOV1, PAK1, PPP2R5B, RAD1, RAD21, RAD9A, RANBP1, RPA3, RPRD1B, RPS6, SART1, SCRIB, STK3, TERF1, TMEM67, TOM1L1, TRIP13, UVRAG, ZNF16, ZNF703, ZW10*
cellular protein localization (GO:0034613)	1332	68	38.61	1.76	4.84E-02	*AIP, AP2M1, AP4S1, ARL2, AURKA, BAD, BAG4, BOD1, BRCA2, CEP72, CLTA, CORO1B, CPSF1, CSE1L, CTTN, DLG1, DNAJA1, DYNC2H1, EFR3A, EGFR, ENY2, EXOC3, FAM83D, FAM83H, FNTA, FRMD6, GPAA1, IFT46, IFT80, LAMTOR1, MTBP, NCBP2, NOL6, NUPL2, PACS2, PDCD6, PEX12, PIGR, PRKCI, RAB6A, RAD21, RANBP1, RPL35A, RPL37, RPL7, RPL8, RPS3, RPS6, RRS1, SCRIB, SMG6, SNX16, SNX27, SNX8, SPCS2, SSB, STX16, TGS1, TOM1L1, TOMM5, TRAM1, TSNARE1, TTC7B, TTC8, VCP, YWHAZ, ZC3H3, ZW10*

Supplemental Table S5Q. PANTHER biological process ontology terms over-represented among differentially expressed genes in HPV-negative tumors with copy number gain (gain vs. normal). We tested the enrichment of GO Biological Process terms for genes differentially expressed between samples with cop number gain and samples with normal copy number in HPV-negative tumors using PANTHER Overrepresentation Test (Released 20171205). We tested the GO Biological Process complete terms and applied multiple testing correction (Bonferroni p<0.05). The PANTHER test shows the result with gene ontology clusters; similar ontology terms are clustered together. The most enriched GO term for each ontology clusters in PANTHER output are shown in this table. Shown here are the GO term, the number of genes in reference human genome, (*obs.*) the number of genes observed in the differentially expressed gene list, (*exp.*) the number of genes expected by chance, fold enrichment (observed/expected), p-values, and the list of genes. We did not observe significantly enriched Biological Process terms for genes differentially expressed between samples with copy number loss and samples with normal copy number.

Supplemental Table S5R. PANTHER biological process ontology terms over-represented among differentially expressed genes in HPV-negative tumors with copy number loss (loss vs. normal). No significantly enriched Biological Process ontology terms were observed in differentially expressed genes (loss vs. normal) in HPV-negative tumors.

Chr.	start	stop	length (Mbp)	cytobands	fraction of altered samples in HPV-pos (%total)	fraction of altered samples in HPV-neg (%total)	FDR-adjusted p-value
2	156,000,001	156,500,000	0.50	q24.1	0.0%	8.0%	3.05E-02
2	173,000,001	180,000,000	7.00	q31.1, q31.2	1.0%	12.0%	2.12E-02
3	93,500,001	148,000,000	54.50	q11.1, q11.2, q12.1, q12.2, q12.3, q13.11, q13.12, q13.13, q13.2, q13.31, q13.32, q13.33, q21.1, q21.2, q21.3, q22.1, q22.2, q22.3, q23, q24	52.4%	18.0%	3.46E-03
3	153,500,001	198,022,430	44.52	q25.2, q25.31, q25.32, q25.33, q26.1, q26.2, q26.31, q26.32, q26.33, q27.1, q27.2, q27.3, q28, q29	66.0%	40.0%	1.59E-02
5	29,000,001	29,500,000	0.50	p13.3	15.5%	36.0%	3.50E-02
5	37,500,001	49,500,000	12.00	p13.2, p13.1, p12, p11, q11.1	13.6%	34.0%	2.82E-02
7	1	62,500,000	62.50	p22.3, p22.2, p22.1, p21.3, p21.2, p21.1, p15.3, p15.2, p15.1, p14.3, p14.2, p14.1, p13, p12.3, p12.2, p12.1, p11.2, p11.1, q11.1, q11.21	1.9%	22.0%	6.86E-04
7	92,000,001	92,500,000	0.50	q21.2	2.9%	16.0%	2.85E-02
8	44,000,001	49,500,000	5.50	p11.1, q11.1, q11.21	7.2%	25.0%	2.59E-02
8	83,500,001	84,000,000	0.50	q21.13	17.5%	38.0%	4.31E-02
8	89,500,001	98,500,000	9.00	q21.3, q22.1	20.9%	44.0%	2.94E-02
8	104,500,001	140,500,000	36.00	q22.3, q23.1, q23.2, q23.3, q24.11, q24.12, q24.13, q24.21, q24.22, q24.23, q24.3	22.3%	47.0%	2.01E-02
9	35,000,001	36,500,000	1.50	p13.3, p13.2	3.9%	20.0%	1.30E-02
9	43,500,001	44,000,000	0.50	p12, p11.2	1.9%	14.0%	2.67E-02
11	54,500,001	55,000,000	0.50	q11	1.9%	14.0%	2.67E-02
11	64,000,001	103,500,000	39.50	q13.1, q13.2, q13.3, q13.4, q13.5, q14.1, q14.2, q14.3, q21, q22.1, q22.2, q22.3	1.9%	14.0%	4.90E-03

14	19,000,001	107,349,540	88.35	q11.1, q11.2, q12, q13.1, q13.2, q13.3, q21.1, q21.2, q21.3, q22.1, q22.2, q22.3, q23.1, q23.2, q23.3, q24.1, q24.2, q24.3, q31.1, q31.2, q31.3, q32.11, q32.12, q32.13, q32.2, q32.31, q32.32, q32.33	1.9%	24.0%	2.91E-04
16	35,000,001	46,500,000	11.50	p11.1, q11.1, q11.2	0.0%	12.8%	1.06E-02
18	15,500,001	18,500,000	3.00	p11.1, q11.1	2.0%	14.3%	3.27E-02
19	24,500,001	27,500,000	3.00	p11, q11	1.0%	12.2%	2.55E-02
22	16,000,001	17,000,000	1.00	q11.1	0.0%	8.0%	3.05E-02
22	23,000,001	40,500,000	17.50	q11.22, q11.23, q12.1, q12.2, q12.3, q13.1	0.0%	10.0%	1.24E-02
5	103,000,001	103,500,000	0.50	q21.2	1.0%	12.0%	4.46E-02
8	4,000,001	4,500,000	0.50	p23.2	7.8%	26.0%	4.46E-02
8	38,500,001	43,000,000	4.50	p11.22, p11.21	1.9%	16.0%	1.05E-02
9	7,000,001	17,500,000	10.50	p24.1, p23, p22.3, p22.2	1.0%	14.0%	2.19E-02
9	38,500,001	39,000,000	0.50	p13.1	0.0%	10.0%	3.15E-02
11	69,000,001	71,000,000	2.00	q13.3, q13.4	17.5%	0.0%	3.64E-02
11	78,500,001	134,000,000	55.50	q14.1, q14.2, q14.3, q21, q22.1, q22.2, q22.3, q23.1, q23.2, q23.3, q24.1, q24.2, q24.3, q25	42.7%	10.0%	3.58E-03
13	49,000,001	49,500,000	0.50	q14.2	27.2%	6.0%	4.52E-02
14	78,000,001	107,349,540	29.35	q24.3, q31.1, q31.2, q31.3, q32.11, q32.12, q32.13, q32.2, q32.31, q32.32, q32.33	21.4%	2.0%	3.64E-02
16	34,500,001	35,500,000	1.00	p11.2, p11.1	17.0%	0.0%	4.08E-02
16	46,000,001	56,000,000	10.00	q11.2, q12.1, q12.2	21.4%	2.0%	3.64E-02
18	35,000,001	78,077,248	43.08	q12.2, q12.3, q21.1, q21.2, q21.31, q21.32, q21.33, q22.1, q22.2, q22.3, q23	1.9%	20.0%	4.62E-03

Supplemental Table S5S. Copy number changes in HPV-positive vs. HPV-negative OSCC.
The table shows genomic regions that are differentially gained (*pink*) or lost (*light blue*) in comparing HPV-positive vs. HPV-negative tumors. We divided the reference genome into 500 kb bins and compared the number of samples with copy number gains between 103 HPV-positive and 50 HPV-negative OSCC using Fisher’s Exact test. We applied the multiple testing correction for p-values using FDR ($p<0.05$). Shown in the table (*left to right*) are: the affected chromosome; the coordinates of merged regions with significant gain or loss; the length of the region in Mbp; the chromosomal arm and cytobands; frequencies of affected HPV-positive and HPV-negative OSCC (averaged over the affected genomic length); and the FDR adjusted p-value.

Chr.	start	stop	fraction of altered samples in HPV-pos (%total)	fraction of altered samples in HPV-neg (%total)	FDR-adjusted p-value	no. genes	no. genes with CNV	no. genes with altered expression	no. census genes	cancer census genes
2	156,000,001	156,500,000	0.0%	8.0%	3.05E-02	8	8	0	0	---
2	173,000,001	180,000,000	1.0%	12.0%	2.12E-02	159	150	18	0	---
3	93,500,001	148,000,000	52.4%	18.0%	3.46E-03	873	872	201	9	*TFG, CBLB, POLQ, RPN1, CNBP, STAG1, PIK3CB, FOXL2, ATR*
3	153,500,001	198,022,430	66.0%	40.0%	1.59E-02	763	727	180	13	*GMPS, MLF1, TBL1XR1, PIK3CA, SOX2, MAP3K13, EIF4A2, BCL6, LPP, TP63, MB21D2, MUC4, TFRC*
5	29,000,001	29,500,000	15.5%	36.0%	3.50E-02	6	6	0	0	---
5	37,500,001	49,500,000	13.6%	34.0%	2.82E-02	125	125	24	0	---
7	1	62,500,000	1.9%	22.0%	6.86E-04	1013	1002	133	3	*PMS2, RAC1, EGFR*
7	92,000,001	92,500,000	2.9%	16.0%	2.85E-02	12	12	6	1	*CDK6*
8	44,000,001	49,500,000	7.2%	25.0%	2.59E-02	47	47	4	0	---
8	83,500,001	84,000,000	17.5%	38.0%	4.31E-02	4	4	0	0	---
8	89,500,001	98,500,000	20.9%	44.0%	2.94E-02	138	138	28	1	*NBN*
8	104,500,001	140,500,000	22.3%	47.0%	2.01E-02	384	384	63	5	*EIF3E, RAD21, EXT1, MYC, NDRG1*
9	35,000,001	36,500,000	3.9%	20.0%	1.30E-02	66	66	21	1	*FANCG*
9	43,500,001	44,000,000	1.9%	14.0%	2.67E-02	7	6	0	0	---
11	54,500,001	55,000,000	1.9%	14.0%	2.67E-02	0	0	0	0	---
11	64,000,001	103,500,000	1.9%	14.0%	4.90E-03	945	842	198	7	*MEN1, MALAT1, CCND1, NUMA1, PICALM, EED, BIRC3*
14	19,000,001	107,349,540	1.9%	24.0%	2.91E-04	2244	2168	318	11	*ARHGAP5, NIN, KTN1, HIF1A, RAD51B, TRIP11, GOLGA5, DICER1, BCL11B, HSP90AA1, AKT1*
16	35,000,001	46,500,000	0.0%	12.8%	1.06E-02	6	0	0	0	---
18	15,500,001	18,500,000	2.0%	14.3%	3.27E-02	0	0	0	0	---
19	24,500,001	27,500,000	1.0%	12.2%	2.55E-02	0	0	0	0	---
22	16,000,001	17,000,000	0.0%	8.0%	3.05E-02	41	35	0	0	---
22	23,000,001	40,500,000	0.0%	10.0%	1.24E-02	603	537	21	2	*BCR, EWSR1*
5	103,000,001	103,500,000	1.0%	12.0%	4.46E-02	2	2	0	0	---
8	4,000,001	4,500,000	7.8%	26.0%	4.46E-02	2	2	0	0	---
8	38,500,001	43,000,000	1.9%	16.0%	1.05E-02	98	98	10	1	*IKBKB*

9	7,000,001	17,500,000	1.0%	14.0%	2.19E-02	86	86	2	0	---
9	38,500,001	39,000,000	0.0%	10.0%	3.15E-02	15	14	0	0	---
11	69,000,001	71,000,000	17.5%	0.0%	3.64E-02	41	41	5	1	*CCND1*
11	78,500,001	134,000,000	42.7%	10.0%	3.58E-03	1001	1001	133	9	*PICALM, EED, MAML2, ATM, DDX10, POU2AF1, SDHD, CBL, FLI1*
13	49,000,001	49,500,000	27.2%	6.0%	4.52E-02	8	8	3	1	*RB1*
14	78,000,001	107,349,540	21.4%	2.0%	3.64E-02	829	825	113	6	*TRIP11, GOLGA5, DICER1, TCL1A, HSP90AA1, AKT1*
16	34,500,001	35,500,000	17.0%	0.0%	4.08E-02	52	52	0	0	---
16	46,000,001	56,000,000	21.4%	2.0%	3.64E-02	200	199	36	1	*CYLD*
18	35,000,001	78,077,248	1.9%	20.0%	4.62E-03	511	511	52	2	*MALT1, KDSR*

Supplemental Table S5T. Differentially expressed genes in CNV regions of HPV-positive vs. HPV-negative OSCC.
Shown here are genes that are differentially expressed, identified from chromosomal regions with copy number gain or loss in HPV-positive vs. negative OSCC (identified in **Supplemental Table S5S**). For both types of OSCC, genes with significant changes in copy numbers with associated changes in expression were identified by comparison with samples having normal copy number using the one tailed t-test with FDR correction (number of genes with altered gene expression). Shown here is the subset of differentially expressed genes annotated in the Sanger Institute Cancer Census gene list.

GO biological process complete	no. genes queried (n = 21042)	observed (n = 68)	expected	fold enrich-ment	P-value	genes
complement activation, classical pathway (GO:0006958)	142	29	0.46	63.2	1.03E-40	*C7, CR2, IGHA2, IGHD, IGHG4, IGHM, IGHV1-2, IGHV1-3, IGHV1-58, IGHV2-26, IGHV3-13, IGHV3-15, IGHV3-20, IGHV3-43, IGHV3-53, IGHV3-64, IGHV3-66, IGHV3-9, IGKV1-16, IGKV1-33, IGKV1D-16, IGKV2-28, IGKV2-29, IGKV2-30, IGKV2D-40, IGLC7, IGLV2-11, IGLV2-8, IGLV3-27*
immunoglobulin production (GO:0002377)	115	23	0.37	61.89	5.06E-31	*AICDA, IGKV1-27, IGKV1D-13, IGKV1D-16, IGKV1D-42, IGKV1D-8, IGKV2-24, IGKV2-28, IGKV2D-24, IGKV2D-29, IGKV6-21, IGKV6D-21, IGLV10-54, IGLV1-36, IGLV2-18, IGLV3-10, IGLV3-9, IGLV4-60, IGLV5-37, IGLV5-45, IGLV7-46, IGLV8-61, IGLV9-49*
phagocytosis, recognition (GO:0006910)	73	13	0.24	55.11	2.4E-15	*IGHA2, IGHD, IGHG4, IGHM, IGHV1-3, IGHV1-58, IGHV2-26, IGHV3-15, IGHV3-20, IGHV3-43, IGHV3-64, IGHV3-66, IGLC7*
regulation of complement activation (GO:0030449)	112	18	0.36	49.73	1.01E-21	*C7, CR2, IGHG4, IGHV1-2, IGHV3-13, IGHV3-53, IGHV3-9, IGKV1-16, IGKV1-33, IGKV1D-16, IGKV2-28, IGKV2-29, IGKV2-30, IGKV2D-40, IGLC7, IGLV2-11, IGLV2-8, IGLV3-27*
phagocytosis, engulfment (GO:0006911)	86	13	0.28	46.78	1.96E-14	*IGHA2, IGHD, IGHG4, IGHM, IGHV1-3, IGHV1-58, IGHV2-26, IGHV3-15, IGHV3-20, IGHV3-43, IGHV3-64, IGHV3-66, IGLC7*
B cell receptor signaling pathway (GO:0050853)	92	13	0.3	43.73	4.64E-14	*IGHA2, IGHD, IGHG4, IGHM, IGHV1-3, IGHV1-58, IGHV2-26, IGHV3-15, IGHV3-20, IGHV3-43, IGHV3-64, IGHV3-66, IGLC7*
Fc-gamma receptor signaling pathway involved in phagocytosis (GO:0038096)	133	16	0.43	37.23	6.06E-17	*IGHG4, IGHV1-2, IGHV3-13, IGHV3-53, IGHV3-9, IGKV1-16, IGKV1-33, IGKV1D-16, IGKV2-28, IGKV2-29, IGKV2-30, IGKV2D-40, IGLC7, IGLV2-11, IGLV2-8, IGLV3-27*
positive regulation of B cell activation (GO:0050871)	132	13	0.43	30.48	4.59E-12	*IGHA2, IGHD, IGHG4, IGHM, IGHV1-3, IGHV1-58, IGHV2-26, IGHV3-15, IGHV3-20, IGHV3-43, IGHV3-64, IGHV3-66, IGLC7*
receptor-mediated endocytosis (GO:0006898)	266	16	0.86	18.61	2.91E-12	*IGHA2, IGHV1-2, IGHV3-13, IGHV3-53, IGHV3-9, IGKV1-16, IGKV1-33, IGKV1D-16, IGKV2-28, IGKV2-29, IGKV2-30, IGKV2D-40, IGLC7, IGLV2-11, IGLV2-8, IGLV3-27*

defense response to bacterium (GO:0042742)	280	14	0.9	15.47	2.97E-09	*AICDA, IGHA2, IGHD, IGHG4, IGHM, IGHV1-3, IGHV1-58, IGHV2-26, IGHV3-15, IGHV3-20, IGHV3-43, IGHV3-64, IGHV3-66, IGLC7*
leukocyte migration (GO:0050900)	360	17	1.16	14.61	1.57E-11	*IGHA2, IGHM, IGHV1-2, IGHV3-13, IGHV3-53, IGHV3-9, IGKV1-16, IGKV1-33, IGKV1D-16, IGKV2-28, IGKV2-29, IGKV2-30, IGKV2D-40, IGLC7, IGLV2-11, IGLV2-8, IGLV3-27*
innate immune response (GO:0045087)	707	15	2.28	6.57	0.0000557	*C7, CR2, IGHA2, IGHD, IGHG4, IGHM, IGHV1-3, IGHV1-58, IGHV2-26, IGHV3-15, IGHV3-20, IGHV3-43, IGHV3-64, IGHV3-66, IGLC7*
proteolysis (GO:0006508)	1331	17	4.3	3.95	0.00712	*AC027319, IGHG4, IGHV1-2, IGHV3-13, IGHV3-53, IGHV3-9, IGKV1-16, IGKV1-33, IGKV1D-16, IGKV2-28, IGKV2-29, IGKV2-30, IGKV2D-40, IGLC7, IGLV2-11, IGLV2-8, IGLV3-27*

Supplemental Table S5U1. Biological process terms enriched in cluster 1 from the gene expression profile of HPV-positive tumors. We tested the enrichment of GO Biological Process terms for gene cluster 1 identified by hierarchical clustering of RNA-seq data from 147 HPV-positive OSCC. We used PANTHER Overrepresentation Test (Released 20171205), and tested the GO Biological Process complete terms and applied multiple testing correction (Bonferroni $p<0.05$). The PANTHER test shows the result with gene ontology clusters; similar ontology terms are clustered together. The most enriched GO term for each ontology clusters in PANTHER output are shown in this table. Shown here are the GO term, the number of genes in reference human genome, the number of genes observed in the gene cluster, the number of genes expected by chance, fold enrichment (observed/expected), p-values, and the list of genes (see also **Supplemental Fig. S5G**).

GO biological process complete	no. genes queried (n = 21042)	observed (n = 100)	exp.	fold enrichment	P-value	genes
skeletal system morphogenesis (GO:0048705)	213	9	1.01	8.89	0.00803	*ALX3, COL11A1, COMP, EYA1, HOXC11, LTF, MMP13, PAX1, SP5*

Supplemental Table S5U2. Biological process terms enriched in cluster 2 from the gene expression profile of HPV-positive tumors. We tested the enrichment of GO Biological Process terms for gene cluster 2 identified by hierarchical clustering of RNA-seq data from 147 HPV-positive OSCC. We used PANTHER Overrepresentation Test (Released 20171205), and tested the GO Biological Process complete terms and applied multiple testing correction (Bonferroni p<0.05). The PANTHER test shows the result with gene ontology clusters; similar ontology terms are clustered together. The most enriched GO term for each ontology clusters in PANTHER output are shown in this table. The table shows the GO term, the number of genes in reference human genome, the number of genes observed in the gene cluster, (*exp.*) the number of genes expected by chance, fold enrichment (observed/expected), p-values, and the list of genes (see also **Supplemental Fig. S5G**).

GO biological process complete	no. genes queried (n = 21042)	obs. (n = 27)	exp.	fold enrich-ment	P-value	genes
muscle filament sliding (GO:0030049)	38	12	0.05	> 100	1.77E-22	*ACTA1, ACTC1, ACTN2, DES, MYBPC1, MYBPC2, MYL1, MYL2, TNNC1, TNNC1, TNNI1, TNNT3*
striated muscle myosin thick filament assembly (GO:0071688)	13	3	0.02	> 100	5.91E-03	*MYBPC1, MYBPC2, MYBPH*
sarcomere organization (GO:0045214)	39	6	0.05	> 100	1.01E-07	*ACTN2, ANKRD1, KLHL41, MYBPC1, MYBPC2, MYBPH*
skeletal muscle contraction (GO:0003009)	28	4	0.04	> 100	4.65E-04	*TNNC1, TNNC1, TNNI1, TNNT3*
ventricular cardiac muscle tissue morphogenesis (GO:0055010)	48	4	0.06	64.94	3.95E-03	*MYL2, TNNC1, TNNC1, TNNI1*
cardiac muscle contraction (GO:0060048)	74	6	0.09	63.19	4.55E-06	*ACTC1, MYL1, MYL2, TNNC1, TNNC1, TNNI1*
regulation of striated muscle contraction (GO:0006942)	88	5	0.11	44.28	8.29E-04	*MYBPH, MYL2, TNNC1, TNNI1, TNNT3*
regulation of ATPase activity (GO:0043462)	76	4	0.1	41.02	2.42E-02	*SLN, TNNC1, TNNC1, TNNT3*

Supplemental Table S5U3. Biological process terms enriched in cluster 3 from the gene expression profile of HPV-positive tumors. We tested the enrichment of GO Biological Process terms for gene cluster 3 identified by hierarchical clustering of RNA-seq data from 147 HPV-positive OSCC. We used PANTHER Overrepresentation Test (Released 20171205), and tested the GO Biological Process complete terms and applied multiple testing correction (Bonferroni p<0.05). The PANTHER test shows the result with gene ontology clusters; similar ontology terms are clustered together. The most enriched GO term for each ontology clusters in PANTHER output are shown in this table. The table shows the GO term, the number of genes in reference human genome, the number of genes observed in the gene cluster, (*exp.*) the number of genes expected by chance, fold enrichment (observed/expected), p-values, and the list of genes (see also **Supplemental Fig. S5G**).

GO biological process complete	no. genes queried (n = 21042)	observed (n = 135)	exp.	fold enrich-ment	P-value	genes
gamete generation (GO:0007276)	658	16	4.22	3.79	4.90E-02	*C14orf39, CCDC155, CFTR, FOXJ1, FOXL2, HORMAD1, LEP, MEIOB, NOS2, PRSS21, SHCBP1L, SOHLH1, TAF7L, TDRD9, TEX15, ZPBP2*
epithelium development (GO:0060429)	1074	21	6.89	3.05	4.62E-02	*AC010524, AC022596, ACTL8, AKR1C2, AL137224, ELF5, ALX1, CALB1, CASP14, EPHA7, FLRT3, FOXJ1, FOXL2, GSTA1, HOXC13, HOXD11, HOXD13, KRT14, PITX2, SLITRK6, SPINK6, UPK1B*
cell differentiation (GO:0030154)	3525	48	22.6	2.12	9.24E-04	*AC010524, AC010677, AC022596, ACTL8, AKR1C2, AL137224, ELF5, ALX1, BMP3, C14orf39, CASP14, CCDC155, CFTR, CHL1, EPHA7, FABP4, FLRT3, FOXE3, FOXJ1, FOXL2, GABRB2, GBX1, GNGT1, GSTA1, HORMAD1, KRT14, LEP, MGST1, NEFH, NEFL, NHLH2, NOS2, NPPC, PITX2, PRAME, RANBP3L, SERPINB12, SHCBP1L, SLITRK5, SLITRK6, SOHLH1, SPINK6, TAF7L, TDRD9, TENM2, TEX15, UPK1B, ZIC1, ZPBP2*
multicellular organism development (GO:0007275)	4795	55	30.76	1.79	2.13E-02	*AC010524, AC010677, AC022596, AL137224, ELF5, AL161626, ALX1, BMP3, C14orf39, CALB1, CASP14, CDH16, CHL1, CHL1, CHRM3, CLDN4, DMRT2, EPHA7, FLRT3, FOXE3, FOXJ1, FOXL2, GABRB2, GBX1, GDA, GNGT1, HORMAD1, HOXC13, HOXD11, HOXD13, IRX4, KRT14, LEP, MGST1, MSLN, NEFH, NEFL, NHLH2, NKX2-4, NOS2, NPPC, PITX2, RANBP3L, SCGB1A1, SERPINB12, SLC6A11, SLITRK5, SLITRK6, SOHLH1, SPINK6, TAF7L, TDRD9, TENM2, TEX15, WDR72, ZIC1*

Supplemental Table S5U4. Biological process terms enriched in cluster 4 from the gene expression profile of HPV-positive tumors. We tested the enrichment of GO Biological Process terms for gene cluster 4 identified by hierarchical clustering of RNA-seq data from 147 HPV-positive OSCC. We used PANTHER Overrepresentation Test (Released 20171205), and tested the GO Biological Process complete terms and applied multiple testing correction (Bonferroni p<0.05). The PANTHER test shows the result with gene ontology clusters; similar ontology terms are clustered together. The most enriched GO term for each ontology clusters in PANTHER output are shown in this table. Shown here are the GO term, the number of genes in reference human genome, the number of genes observed in the gene cluster, (*exp.*) the number of genes expected by chance, fold enrichment (observed/expected), p-values, and the list of genes (see also **Supplemental Fig. S5G**).

GO biological process complete	no. genes queried (n = 21042)	obs. (n = 74)	exp.	fold enrich-ment	P-value	genes
peptide cross-linking (GO:0018149)	59	15	0.21	72.29	7.04E-20	*FLG, KRT1, LCE3A, LCE3D, LCE3E, PI3, SPRR1A, SPRR2A, SPRR2B, SPRR2D, SPRR2E, SPRR2F, SPRR3, SPRR4, TGM3*
cornification (GO:0070268)	113	23	0.4	57.88	3.39E-30	*DSG1, FLG, KLK12, KLK13, KLK5, KLK8, KRT1, KRT13, KRT24, KRT4, KRT6C, KRT78, LCE3D, PI3, RPTN, SPRR1A, SPRR2A, SPRR2B, SPRR2D, SPRR2E, SPRR2F, SPRR2G, SPRR3*
keratinization (GO:0031424)	227	27	0.8	33.82	4.83E-30	*DSG1, FLG, KLK12, KLK13, KLK5, KLK8, KRT1, KRT13, KRT24, KRT4, KRT6C, KRT78, LCE3A, LCE3D, LCE3E, PI3, RPTN, SPRR1A, SPRR2A, SPRR2B, SPRR2D, SPRR2E, SPRR2F, SPRR2G, SPRR3, SPRR4, TGM3*

Supplemental Table S5U5. Biological process enriched in cluster 5 from the gene expression profile of HPV-positive tumors. We tested the enrichment of GO Biological Process terms for gene cluster 5 identified by hierarchical clustering of RNA-seq data from 147 HPV-positive OSCC. We used PANTHER Overrepresentation Test (Released 20171205), and tested the GO Biological Process complete terms and applied multiple testing correction (Bonferroni p<0.05). The PANTHER test shows the result with gene ontology clusters; similar ontology terms are clustered together. The most enriched GO term for each ontology clusters in PANTHER output are shown in this table. Shown here are the GO term, the number of genes in reference human genome, the number of genes observed in the gene cluster, the number of genes expected by chance, fold enrichment (observed/expected), p-values, and the list of genes (see also **Supplemental Fig. S5G**).

Supplemental Table S5U6. **Biological process enriched in cluster 6 from the gene expression profile of HPV-positive tumors**
No significantly enriched GO terms were found for genes in cluster 6 (see also **Supplemental Fig. S5G**).

GO biological process complete	no. genes queried (n = 21042)	observed (n = 140)	expected	fold enrich-ment	P-value	genes
fat-soluble vitamin metabolic process (GO:0006775)	33	5	0.22	22.77	2.87E-02	*AC005336, AC005336, CYP26A1, LRAT, RPE65*
cellular hormone metabolic process (GO:0034754)	110	8	0.73	10.93	7.82E-03	*ADH7, AL031005, CYP26A1, LRAT, RBP4, RPE65, SULT1E1, UGT1A7*

Supplemental Table S5V1. Biological process enriched in cluster 1 from the gene expression profile of HPV-negative tumors. We tested the enrichment of GO Biological Process terms for gene cluster 1 identified by hierarchical clustering of RNA-seq data from 335 HPV-negative OSCC. We used PANTHER Overrepresentation Test (Released 20171205), and tested the GO Biological Process complete terms and applied multiple testing correction (Bonferroni p<0.05). The PANTHER test shows the result with gene ontology clusters; similar ontology terms are clustered together. The most enriched GO term for each ontology clusters in PANTHER output are shown in this table. Shown here are the GO term, the number of genes queried, the number of genes observed in the gene cluster, the number of genes expected by chance, fold enrichment (observed/expected), p-values, and the list of genes (see **Supplemental Fig. S5H**).

GO biological process complete	no. genes queried (n = 21042)	obs. (n = 96)	exp.	fold enrichment	P-value	genes
detection of muscle stretch (GO:0035995)	4	4	0.02	> 100	3.71E-05	*CAV3, CSRP3, TCAP, TTN*
skeletal muscle thin filament assembly (GO:0030240)	6	5	0.03	> 100	9.77E-07	*ACTA1, ACTC1, LMOD3, TCAP, TTN*
positive regulation of skeletal muscle fiber development (GO:0048743)	9	5	0.04	> 100	7.34E-06	*LMOD3, MYF5, MYF6, MYOD1, MYOG*
muscle filament sliding (GO:0030049)	38	21	0.17	> 100	1.46E-33	*ACTA1, ACTC1, ACTN2, DES, MYBPC1, MYBPC2, MYH2, MYH7, MYH8, MYL1, MYL2, MYL3, NEB, TCAP, TNNC1, TNNC1, TNNC2, TNNI1, TNNI2, TNNT3, TTN*
cardiac muscle fiber development (GO:0048739)	9	4	0.04	97.42	9.33E-04	*MYO18B, MYPN, TCAP, TTN*
skeletal muscle contraction (GO:0003009)	28	12	0.13	93.94	1.50E-16	*CHRNA1, CHRND, MB, MYH7, MYH8, TCAP, TNNC1, TNNC1, TNNC2, TNNI1, TNNI2, TNNT3*
transition between fast and slow fiber (GO:0014883)	7	3	0.03	93.94	4.45E-02	*MYH7, TNNC1, TNNI1*
striated muscle myosin thick filament assembly (GO:0071688)	13	5	0.06	84.3	4.55E-05	*MYBPC1, MYBPC2, MYBPH, TCAP, TTN*
skeletal muscle adaptation (GO:0043501)	11	4	0.05	79.7	2.07E-03	*ACTA1, MYOD1, MYOG, TRIM63*
muscle cell fate commitment (GO:0042693)	15	5	0.07	73.06	9.24E-05	*MYF5, MYF6, MYL2, MYOD1, MYOG*
sarcomere organization (GO:0045214)	39	13	0.18	73.06	9.23E-17	*ACTN2, ANKRD1, CASQ1, CASQ2, KLHL41, LDB3, LMOD2, MYBPC1, MYBPC2, MYBPH, MYPN, TCAP, TTN*
cardiac myofibril assembly (GO:0055003)	16	5	0.07	68.5	1.27E-04	*ACTC1, CSRP3, MYL2, TCAP, TTN*
positive regulation of myoblast differentiation (GO:0045663)	22	6	0.1	59.78	9.68E-06	*CSRP3, MYF5, MYF6, MYOD1, MYOG, SMYD1*
skeletal muscle fiber development (GO:0048741)	24	6	0.11	54.8	1.62E-05	*ACTA1, KLHL40, KLHL41, LMOD3, MYOD1, MYOG*
response to stimulus involved in regulation of muscle adaptation (GO:0014874)	16	4	0.07	54.8	9.10E-03	*CASQ1, MYOG, SGCA, TRIM63*
positive regulation of myoblast fusion (GO:1901741)	18	4	0.08	48.71	1.45E-02	*MYF5, MYF6, MYOD1, MYOG*

cardiac muscle contraction (GO:0060048)	74	15	0.34	44.43	1.35E-16	*ACTC1, ATP1A2, CASQ2, CAV3, CSRP3, MYH7, MYL1, MYL2, MYL3, TCAP, TNNC1, TNNC1, TNNI1, TNNI2, TTN*
regulation of release of sequestered calcium ion into cytosol by sarcoplasmic reticulum (GO:0010880)	25	5	0.11	43.84	1.15E-03	*ATP1A2, CASQ1, CASQ2, HRC, TRDN*
regulation of the force of heart contraction (GO:0002026)	27	5	0.12	40.59	1.67E-03	*ATP1A2, CSRP3, MYH7, MYL2, MYL3*
regulation of cardiac muscle contraction by calcium ion signaling (GO:0010882)	22	4	0.1	39.85	3.18E-02	*ATP1A2, CASQ2, HRC, TNNC1*
skeletal muscle tissue regeneration (GO:0043403)	29	5	0.13	37.79	2.37E-03	*MYF6, MYOD1, MYOG, SGCA, TMEM8C*
skeletal muscle cell differentiation (GO:0035914)	54	8	0.25	32.47	1.77E-06	*ANKRD1, KLHL40, KLHL41, MYF5, MYF6, MYOD1, MYOG, SMYD1*
ventricular cardiac muscle tissue morphogenesis (GO:0055010)	48	6	0.22	27.4	9.49E-04	*MYH7, MYL2, MYL3, TNNC1, TNNC1, TNNI1*
negative regulation of ion transmembrane transporter activity (GO:0032413)	66	6	0.3	19.93	6.01E-03	*ACTN2, ATP1A2, CASQ2, CAV3, SLN, TRDN*
regulation of cardiac conduction (GO:1903779)	69	6	0.31	19.06	7.76E-03	*ATP1A2, CASQ1, CASQ2, HRC, TNNC1, TRDN*
negative regulation of cation transmembrane transport (GO:1904063)	73	6	0.33	18.02	1.07E-02	*ACTN2, ATP1A2, CASQ2, CAV3, SLN, TRDN*
regulation of calcium ion transmembrane transporter activity (GO:1901019)	87	7	0.4	17.64	1.55E-03	*ATP1A2, CASQ1, CASQ2, CAV3, HRC, SLN, TRDN*
cardiac conduction (GO:0061337)	87	6	0.4	15.12	2.92E-02	*ATP1A2, CACNA1S, CACNG1, CACNG6, CACNG6, CASQ2*

Supplemental Table S5V2. Biological process enriched in cluster 2 from the gene expression profile of HPV-negative tumors.
We tested the enrichment of GO Biological Process terms for gene cluster 2 identified by hierarchical clustering of RNA-seq data from 335 HPV-negative OSCC. We used PANTHER Overrepresentation Test (Released 20171205), and tested the GO Biological Process complete terms and applied multiple testing correction (Bonferroni p<0.05). The PANTHER test shows the result with gene ontology clusters; similar ontology terms are clustered together. The most enriched GO term for each ontology clusters in PANTHER output are shown in this table. The tables shows GO term, the number of genes queried in reference human genome, (*obs.*) the number of genes observed in the gene cluster, (*exp.*) the number of genes expected by chance, fold enrichment (observed/expected), p-values, and the list of genes (see **Supplemental Fig. S5H**).

Supplemental Table S5V3. Biological process enriched in cluster 3 from the gene expression profile of HPV-negative tumors.
No significantly enriched GO terms were found for genes in cluster 3 (see **Fig. Supplemental S5H**).

GO biological process complete	no. genes queried (n = 21042)	obs. (n = 84)	exp.	fold enrich-ment	P-value	genes
peptide cross-linking (GO:0018149)	59	22	0.24	93.41	4.87E-33	*KRT1, KRT2, LCE1A, LCE1B, LCE1F, LCE2A, LCE2B, LCE2C, LCE2D, LCE3A, LCE3C, LCE3D, LCE3E, LOR, PRR9, SPRR1A, SPRR2B, SPRR2E, SPRR2F, SPRR3, SPRR4, TGM3*
cornification (GO:0070268)	113	23	0.45	50.99	9.85E-29	*CASP14, DSC1, DSG1, KLK12, KLK13, KLK14, KRT1, KRT13, KRT2, KRT24, KRT4, KRT76, LCE1A, LCE3D, LOR, RPTN, SPINK6, SPRR1A, SPRR2B, SPRR2E, SPRR2F, SPRR2G, SPRR3*
keratinization (GO:0031424)	227	35	0.91	38.62	3.88E-42	*CASP14, DSC1, DSG1, KLK12, KLK13, KLK14, KRT1, KRT13, KRT2, KRT24, KRT4, KRT76, LCE1A, LCE1B, LCE1F, LCE2A, LCE2B, LCE2C, LCE2D, LCE3A, LCE3C, LCE3D, LCE3E, LCE6A, LOR, RPTN, SPINK6, SPRR1A, SPRR2B, SPRR2E, SPRR2F, SPRR2G, SPRR3, SPRR4, TGM3*

Supplemental Table S5V4. Biological process enriched in cluster 4 from the gene expression profile of HPV-negative tumors. We tested the enrichment of GO Biological Process terms for gene cluster 4 identified by hierarchical clustering of RNA-seq data from 335 HPV-negative OSCC. We used PANTHER Overrepresentation Test (Released 20171205), and tested the GO Biological Process complete terms and applied multiple testing correction (Bonferroni p<0.05). The PANTHER test shows the result with gene ontology clusters; similar ontology terms are clustered together. The most enriched GO term for each ontology clusters in PANTHER output are shown in this table. Shown here are the GO term, the number of genes queried in reference human genome, (*obs.*) the number of genes observed in the gene cluster, (*exp.*) the number of genes expected by chance, fold enrichment (observed/expected), p-values, and the list of genes (see **Supplemental Fig. S5H**).

GO biological process complete	no. genes queried (n = 21042)	obs. (n = 107)	exp.	fold enrich-ment	P-value	genes
complement activation, classical pathway (GO:0006958)	142	71	0.72	98.33	1.97E-122	*IGHA1, IGHA2, IGHD, IGHG1, IGHG2, IGHG3, IGHG4, IGHM, IGHV1-18, IGHV1-2, IGHV1-24, IGHV1-3, IGHV1-46, IGHV1-69, IGHV2-26, IGHV2-5, IGHV2-70, IGHV3-11, IGHV3-13, IGHV3-15, IGHV3-20, IGHV3-21, IGHV3-23, IGHV3-30, IGHV3-33, IGHV3-43, IGHV3-48, IGHV3-49, IGHV3-53, IGHV3-66, IGHV3-7, IGHV3-73, IGHV3-74, IGHV3-9, IGHV4-34, IGHV4-39, IGHV4-4, IGHV4-59, IGHV4-61, IGHV5-51, IGKV1-12, IGKV1-16, IGKV1-17, IGKV1-5, IGKV1D-12, IGKV1D-16, IGKV2-29, IGKV2-30, IGKV2D-40, IGKV3-11, IGKV3-15, IGKV3-20, IGKV3D-20, IGKV4-1, IGKV5-2, IGLC7, IGLV1-40, IGLV1-44, IGLV1-47, IGLV1-51, IGLV2-11, IGLV2-14, IGLV2-23, IGLV2-8, IGLV3-1, IGLV3-19, IGLV3-21, IGLV3-25, IGLV3-27, IGLV6-57, IGLV7-43*
regulation of complement activation (GO:0030449)	112	52	0.57	91.3	4.46E-84	*IGHG1, IGHG2, IGHG3, IGHG4, IGHV1-2, IGHV1-46, IGHV1-69, IGHV2-5, IGHV2-70, IGHV3-11, IGHV3-13, IGHV3-23, IGHV3-30, IGHV3-33, IGHV3-48, IGHV3-53, IGHV3-7, IGHV3-9, IGHV4-34, IGHV4-39, IGHV4-59, IGKV1-12, IGKV1-16, IGKV1-17, IGKV1-5, IGKV1D-12, IGKV1D-16, IGKV2-29, IGKV2-30, IGKV2D-40, IGKV3-11, IGKV3-15, IGKV3-20, IGKV3D-20, IGKV4-1, IGKV5-2, IGLC7, IGLV1-40, IGLV1-44, IGLV1-47, IGLV1-51, IGLV2-11, IGLV2-14, IGLV2-23, IGLV2-8, IGLV3-1, IGLV3-19, IGLV3-21, IGLV3-25, IGLV3-27, IGLV6-57, IGLV7-43*
Fc-gamma receptor signaling pathway involved in phagocytosis (GO:0038096)	133	52	0.68	76.89	3.22E-80	*IGHG1, IGHG2, IGHG3, IGHG4, IGHV1-2, IGHV1-46, IGHV1-69, IGHV2-5, IGHV2-70, IGHV3-11, IGHV3-13, IGHV3-23, IGHV3-30, IGHV3-33, IGHV3-48, IGHV3-53, IGHV3-7, IGHV3-9, IGHV4-34, IGHV4-39, IGHV4-59, IGKV1-12, IGKV1-16, IGKV1-17, IGKV1-5, IGKV1D-12, IGKV1D-16, IGKV2-29, IGKV2-30, IGKV2D-40, IGKV3-11, IGKV3-15, IGKV3-20, IGKV3D-20, IGKV4-1, IGKV5-2, IGLC7, IGLV1-40, IGLV1-44, IGLV1-47, IGLV1-51, IGLV2-11, IGLV2-14, IGLV2-23, IGLV2-8, IGLV3-1, IGLV3-19, IGLV3-21, IGLV3-25, IGLV3-27, IGLV6-57, IGLV7-43*

phagocytosis, recognition (GO:0006910)	73	25	0.37	67.35	3.51E-34	*IGHA1, IGHA2, IGHD, IGHG1, IGHG2, IGHG3, IGHG4, IGHM, IGHV1-18, IGHV1-24, IGHV1-3, IGHV2-26, IGHV3-15, IGHV3-20, IGHV3-21, IGHV3-23, IGHV3-43, IGHV3-49, IGHV3-66, IGHV3-73, IGHV3-74, IGHV4-4, IGHV4-61, IGHV5-51, IGLC7*
phagocytosis, engulfment (GO:0006911)	86	25	0.44	57.17	2.01E-32	*IGHA1, IGHA2, IGHD, IGHG1, IGHG2, IGHG3, IGHG4, IGHM, IGHV1-18, IGHV1-24, IGHV1-3, IGHV2-26, IGHV3-15, IGHV3-20, IGHV3-21, IGHV3-23, IGHV3-43, IGHV3-49, IGHV3-66, IGHV3-73, IGHV3-74, IGHV4-4, IGHV4-61, IGHV5-51, IGLC7*
B cell receptor signaling pathway (GO:0050853)	92	25	0.47	53.44	1.06E-31	*IGHA1, IGHA2, IGHD, IGHG1, IGHG2, IGHG3, IGHG4, IGHM, IGHV1-18, IGHV1-24, IGHV1-3, IGHV2-26, IGHV3-15, IGHV3-20, IGHV3-21, IGHV3-23, IGHV3-43, IGHV3-49, IGHV3-66, IGHV3-73, IGHV3-74, IGHV4-4, IGHV4-61, IGHV5-51, IGLC7*
Fc-epsilon receptor signaling pathway (GO:0038095)	183	48	0.93	51.58	4.58E-65	*IGHV1-2, IGHV1-46, IGHV1-69, IGHV2-5, IGHV2-70, IGHV3-11, IGHV3-13, IGHV3-23, IGHV3-30, IGHV3-33, IGHV3-48, IGHV3-53, IGHV3-7, IGHV3-9, IGHV4-34, IGHV4-39, IGHV4-59, IGKV1-12, IGKV1-16, IGKV1-17, IGKV1-5, IGKV1D-12, IGKV1D-16, IGKV2-29, IGKV2-30, IGKV2D-40, IGKV3-11, IGKV3-15, IGKV3-20, IGKV3D-20, IGKV4-1, IGKV5-2, IGLC7, IGLV1-40, IGLV1-44, IGLV1-47, IGLV1-51, IGLV2-11, IGLV2-14, IGLV2-23, IGLV2-8, IGLV3-1, IGLV3-19, IGLV3-21, IGLV3-25, IGLV3-27, IGLV6-57, IGLV7-43*
immunoglobulin production (GO:0002377)	115	30	0.58	51.3	2.47E-38	*IGKV1-12, IGKV1-27, IGKV1-5, IGKV1-6, IGKV1-9, IGKV1D-13, IGKV1D-16, IGKV1D-17, IGKV1D-8, IGKV2-24, IGKV2D-29, IGKV3-7, IGKV3D-11, IGKV3D-15, IGKV3D-20, IGKV4-1, IGKV6-21, IGKV6D-21, IGLV10-54, IGLV1-36, IGLV2-18, IGLV3-10, IGLV3-16, IGLV3-9, IGLV4-60, IGLV4-69, IGLV5-45, IGLV7-46, IGLV8-61, IGLV9-49*
positive regulation of B cell activation (GO:0050871)	132	25	0.67	37.25	7.59E-28	*IGHA1, IGHA2, IGHD, IGHG1, IGHG2, IGHG3, IGHG4, IGHM, IGHV1-18, IGHV1-24, IGHV1-3, IGHV2-26, IGHV3-15, IGHV3-20, IGHV3-21, IGHV3-23, IGHV3-43, IGHV3-49, IGHV3-66, IGHV3-73, IGHV3-74, IGHV4-4, IGHV4-61, IGHV5-51, IGLC7*

receptor-mediated endocytosis (GO:0006898)	266	50	1.35	36.97	5.21E-61	*IGHA1, IGHA2, IGHV1-2, IGHV1-46, IGHV1-69, IGHV2-5, IGHV2-70, IGHV3-11, IGHV3-13, IGHV3-23, IGHV3-30, IGHV3-33, IGHV3-48, IGHV3-53, IGHV3-7, IGHV3-9, IGHV4-34, IGHV4-39, IGHV4-59, IGKV1-12, IGKV1-16, IGKV1-17, IGKV1-5, IGKV1D-12, IGKV1D-16, IGKV2-29, IGKV2-30, IGKV2D-40, IGKV3-11, IGKV3-15, IGKV3-20, IGKV3D-20, IGKV4-1, IGKV5-2, IGLC7, IGLV1-40, IGLV1-44, IGLV1-47, IGLV1-51, IGLV2-11, IGLV2-14, IGLV2-23, IGLV2-8, IGLV3-1, IGLV3-19, IGLV3-21, IGLV3-25, IGLV3-27, IGLV6-57, IGLV7-43*
leukocyte migration (GO:0050900)	360	51	1.83	27.86	2.93E-56	*IGHA1, IGHA2, IGHM, IGHV1-2, IGHV1-46, IGHV1-69, IGHV2-5, IGHV2-70, IGHV3-11, IGHV3-13, IGHV3-23, IGHV3-30, IGHV3-33, IGHV3-48, IGHV3-53, IGHV3-7, IGHV3-9, IGHV4-34, IGHV4-39, IGHV4-59, IGKV1-12, IGKV1-16, IGKV1-17, IGKV1-5, IGKV1D-12, IGKV1D-16, IGKV2-29, IGKV2-30, IGKV2D-40, IGKV3-11, IGKV3-15, IGKV3-20, IGKV3D-20, IGKV4-1, IGKV5-2, IGLC7, IGLV1-40, IGLV1-44, IGLV1-47, IGLV1-51, IGLV2-11, IGLV2-14, IGLV2-23, IGLV2-8, IGLV3-1, IGLV3-19, IGLV3-21, IGLV3-25, IGLV3-27, IGLV6-57, IGLV7-43*
defense response to bacterium (GO:0042742)	280	26	1.42	18.26	2.69E-21	*IGHA1, IGHA2, IGHD, IGHG1, IGHG2, IGHG3, IGHG4, IGHM, IGHV1-18, IGHV1-24, IGHV1-3, IGHV2-26, IGHV3-15, IGHV3-20, IGHV3-21, IGHV3-23, IGHV3-43, IGHV3-49, IGHV3-66, IGHV3-73, IGHV3-74, IGHV4-4, IGHV4-61, IGHV5-51, IGKV3-20, IGLC7*
proteolysis (GO:0006508)	1331	52	6.77	7.68	1.39E-29	*IGHG1, IGHG2, IGHG3, IGHG4, IGHV1-2, IGHV1-46, IGHV1-69, IGHV2-5, IGHV2-70, IGHV3-11, IGHV3-13, IGHV3-23, IGHV3-30, IGHV3-33, IGHV3-48, IGHV3-53, IGHV3-7, IGHV3-9, IGHV4-34, IGHV4-39, IGHV4-59, IGKV1-12, IGKV1-16, IGKV1-17, IGKV1-5, IGKV1D-12, IGKV1D-16, IGKV2-29, IGKV2-30, IGKV2D-40, IGKV3-11, IGKV3-15, IGKV3-20, IGKV3D-20, IGKV4-1, IGKV5-2, IGLC7, IGLV1-40, IGLV1-44, IGLV1-47, IGLV1-51, IGLV2-11, IGLV2-14, IGLV2-23, IGLV2-8, IGLV3-1, IGLV3-19, IGLV3-21, IGLV3-25, IGLV3-27, IGLV6-57, IGLV7-43*
innate immune response (GO:0045087)	707	25	3.6	6.95	1.41E-10	*IGHA1, IGHA2, IGHD, IGHG1, IGHG2, IGHG3, IGHG4, IGHM, IGHV1-18, IGHV1-24, IGHV1-3, IGHV2-26, IGHV3-15, IGHV3-20, IGHV3-21, IGHV3-23, IGHV3-43, IGHV3-49, IGHV3-66, IGHV3-73, IGHV3-74, IGHV4-4, IGHV4-61, IGHV5-51, IGLC7*

Supplemental Table S5V5. Biological process enriched in cluster 5 from the gene expression profile of HPV-negative tumors.
We tested the enrichment of GO Biological Process terms for gene cluster 5 identified by hierarchical clustering of RNA-seq data from 335 HPV-negative OSCC. We used PANTHER Overrepresentation Test (Released 20171205), and tested the GO Biological Process complete terms and applied multiple testing correction (Bonferroni p<0.05). The PANTHER test shows the result with gene ontology clusters; similar ontology terms are clustered together. The most enriched GO term for each ontology clusters in PANTHER output are shown in this table. Shown here are the GO term, the number of genes queried in reference human genome, the number of genes observed in the gene cluster, the number of genes expected by chance, fold enrichment (observed/expected), p-values, and the list of genes (see **Supplemental Fig. S5H**).

Supplemental Table S5W1. Biological process enriched in cluster 1 from the gene expression profile of combined HPV-positive and HPV-negative tumors.
No significantly enriched GO terms were found for genes in cluster 1 (see also **Supplemental Fig. S5I**).

GO biological process complete	no. genes queried (n = 21042)	obs. (n = 80)	exp.	fold enrichment	P-value	genes
detection of muscle stretch (GO:0035995)	4	3	0.02	> 100	4.84E-03	*CAV3, CSRP3, TCAP*
skeletal muscle thin filament assembly (GO:0030240)	6	4	0.02	> 100	8.90E-05	*ACTA1, ACTC1, LMOD3, TCAP*
muscle filament sliding (GO:0030049)	38	19	0.14	> 100	6.82E-31	*ACTA1, ACTC1, ACTN2, DES, MYBPC1, MYBPC2, MYH2, MYH7, MYH8, MYL1, MYL2, MYL3, NEB, TCAP, TNNC1, TNNC2, TNNI1, TNNT3*
transition between fast and slow fiber (GO:0014883)	7	3	0.03	> 100	2.57E-02	*MYH7, TNNC1, TNNI1*
skeletal muscle contraction (GO:0003009)	28	11	0.11	> 100	1.93E-15	*CHRNA1, CHRND, MB, MYH7, MYH8, TCAP, TNNC1, TNNC2, TNNI1, TNNT3*
striated muscle myosin thick filament assembly (GO:0071688)	13	4	0.05	80.93	1.92E-03	*MYBPC1, MYBPC2, MYBPH, TCAP*
sarcomere organization (GO:0045214)	39	11	0.15	74.19	7.16E-14	*ACTN2, ANKRD1, CASQ1, CASQ2, KLHL41, LMOD2, MYBPC1, MYBPC2, MYBPH, MYPN, TCAP*
cardiac myofibril assembly (GO:0055003)	16	4	0.06	65.76	4.37E-03	*ACTC1, CSRP3, MYL2, TCAP*
skeletal muscle fiber development (GO:0048741)	24	5	0.09	54.8	3.74E-04	*ACTA1, KLHL40, KLHL41, LMOD3, MYOG*
regulation of the force of heart contraction (GO:0002026)	27	5	0.1	48.71	6.69E-04	*ATP1A2, CSRP3, MYH7, MYL2, MYL3*
striated muscle adaptation (GO:0014888)	27	5	0.1	48.71	6.69E-04	*ACTA1, MYH7, MYOG, TCAP, TRIM63*
positive regulation of myoblast differentiation (GO:0045663)	22	4	0.08	47.82	1.54E-02	*CSRP3, MYF6, MYOG, SMYD1*
cardiac muscle contraction (GO:0060048)	74	13	0.28	46.21	2.76E-14	*ACTC1, ATP1A2, CASQ2, CAV3, CSRP3, MYH7, MYL1, MYL2, MYL3, TCAP, TNNC1, TNNI1*
regulation of release of sequestered calcium ion into cytosol by sarcoplasmic reticulum (GO:0010880)	25	4	0.1	42.08	2.54E-02	*ATP1A2, CASQ1, CASQ2, TRDN*
positive regulation of myotube differentiation (GO:0010831)	32	5	0.12	41.1	1.54E-03	*CAV3, LMOD3, MYF6, MYOG, SMYD1*
regulation of striated muscle contraction (GO:0006942)	88	11	0.33	32.88	4.77E-10	*ATP1A2, CASQ1, CASQ2, CAV3, MYBPH, MYH7, MYL2, MYL3, TNNC1, TNNI1, TNNT3*

ventricular cardiac muscle tissue morphogenesis (GO:0055010)	48	6	0.18	32.88	3.17E-04	*MYH7, MYL2, MYL3, TNNC1, TNNC1, TNNI1*
skeletal muscle cell differentiation (GO:0035914)	54	6	0.21	29.23	6.32E-04	*ANKRD1, KLHL40, KLHL41, MYF6, MYOG, SMYD1*
negative regulation of ion transmembrane transporter activity (GO:0032413)	66	6	0.25	23.91	2.03E-03	*ACTN2, ATP1A2, CASQ2, CAV3, SLN, TRDN*
negative regulation of calcium ion transport (GO:0051926)	59	5	0.22	22.29	3.03E-02	*ATP1A2, CASQ2, CAV3, SLN, TRDN*
regulation of calcium ion transmembrane transporter activity (GO:1901019)	87	6	0.33	18.14	1.00E-02	*ATP1A2, CASQ1, CASQ2, CAV3, SLN, TRDN*
cardiac conduction (GO:0061337)	87	6	0.33	18.14	1.00E-02	*ATP1A2, CACNA1S, CACNG1, CACNG6, CASQ2*
regulation of membrane potential (GO:0042391)	400	10	1.52	6.58	2.61E-02	*ACTN2, ATP1A2, CACNA1S, CASQ2, CAV3, CHRNA1, CHRND, CHRNG, SCN4A, TRDN*

Supplemental Table S5W2. Biological process enriched in cluster 2 from the gene expression profile of combined HPV-positive and HPV-negative tumors.
We tested the enrichment of GO Biological Process terms for gene cluster 2 as identified by hierarchical clustering of RNA-seq data from combined HPV-positive (n=147) and HPV-negative (n=335) OSCC. We used PANTHER Overrepresentation Test (Released 20171205), and tested the GO Biological Process complete terms and applied multiple testing correction (Bonferroni p<0.05). The PANTHER test shows the result with gene ontology clusters; similar ontology terms are clustered together. The most enriched GO term for each ontology clusters in PANTHER output are shown in this table. Shown here are the GO term, the number of genes queried in reference human genome, the number of genes observed in the gene cluster, the number of genes expected by chance, fold enrichment (observed/expected), p-values, and the list of genes (see also **Supplemental Fig. S5I**).

Supplemental Table S5W3. Biological process enriched in cluster 3 from the gene expression profile of combined HPV-positive and HPV-negative tumors No significantly enriched GO terms were found for genes in cluster 3 (see also **Supplemental Fig. S5I**).

GO biological process complete	no. genes queried (n = 21042)	obs. (n = 94)	exp.	fold enrich-ment	P-value	genes
retinoic acid metabolic process (GO:0042573)	22	4	0.1	40.7	2.93E-02	*ADH7, CYP26A1, LRAT, UGT1A7*
neuron fate specification (GO:0048665)	34	5	0.15	32.92	4.65E-03	*DMRTA2, HOXC10, ISL1, POU4F1, TLX3*

Supplemental Table S5W4. Biological process enriched in cluster 4 from the gene expression profile of combined HPV-positive and HPV-negative tumors.
We tested the enrichment of GO Biological Process terms for gene cluster 4 as identified by hierarchical clustering of RNA-seq data from combined HPV-positive (n=147) and HPV-negative (n=335) OSCC. We used PANTHER Overrepresentation Test (Released 20171205), and tested the GO Biological Process complete terms and applied multiple testing correction (Bonferroni p<0.05). The PANTHER test shows the result with gene ontology clusters; similar ontology terms are clustered together. The most enriched GO term for each ontology clusters in PANTHER output are shown in this table. Shown here are the GO term, the number of genes queried in reference human genome, (*obs.*) the number of genes observed in the gene cluster, (*exp.*) the number of genes expected by chance, fold enrichment (observed/expected), p-values, and the list of genes **(**see also **Supplemental Fig. S5I)**.

GO biological process complete	no. genes queried (n = 21042)	obs. (n = 78)	exp.	fold enrichment	P-value	genes
positive regulation of defense response to bacterium (GO:1900426)	8	3	0.03	> 100	0.0355	*KLK5, KLK7, NLRP10*
peptide cross-linking (GO:0018149)	59	21	0.22	96.02	1.03E-31	*KRT1, KRT2, LCE1A, LCE1F, LCE2A, LCE2B, LCE2C, LCE2D, LCE3A, LCE3C, LCE3D, LCE3E, LOR, PI3, PRR9, SPRR2B, SPRR2E, SPRR2F, SPRR3, SPRR4, TGM3*
cornification (GO:0070268)	113	23	0.42	54.91	1.39E-29	*DSC1, DSG1, KLK12, KLK13, KLK14, KLK5, KLK8, KRT1, KRT13, KRT2, KRT24, KRT4, KRT6C, LCE1A, LCE3D, LOR, PI3, RPTN, SPRR2B, SPRR2E, SPRR2F, SPRR2G, SPRR3*
keratinization (GO:0031424)	227	34	0.84	40.41	1.04E-41	*DSC1, DSG1, KLK12, KLK13, KLK14, KLK5, KLK8, KRT1, KRT13, KRT2, KRT24, KRT4, KRT6C, LCE1A, LCE1F, LCE2A, LCE2B, LCE2C, LCE2D, LCE3A, LCE3C, LCE3D, LCE3E, LCE6A, LOR, PI3, RPTN, SPRR2B, SPRR2E, SPRR2F, SPRR2G, SPRR3, SPRR4, TGM3*

Supplemental Table S5W5. Biological process enriched in cluster 5 from the gene expression profile of combined HPV-positive and HPV-negative tumors.

We tested the enrichment of GO Biological Process terms for gene cluster 5 as identified by hierarchical clustering of RNA-seq data from combined HPV-positive (n=147) and HPV-negative (n=335) OSCC. We used PANTHER Overrepresentation Test (Released 20171205), and tested the GO Biological Process complete terms and applied multiple testing correction (Bonferroni p<0.05). The PANTHER test shows the result with gene ontology clusters; similar ontology terms are clustered together. The most enriched GO term for each ontology clusters in PANTHER output are shown in this table. Shown here are the GO term, the number of genes queried in reference human genome, (*obs.*) the number of genes observed in the gene cluster, (*exp.*) the number of genes expected by chance, fold enrichment (observed/expected), p-values, and the list of genes (see also **Supplemental Fig. S5I**).

GO biological process complete	no. genes queried (n = 21042)	obs. (n = 110)	exp.	fold enrich-ment	P-value	genes
complement activation, classical pathway (GO:0006958)	142	72	0.74	96.99	1.68E-123	*CR2, IGHA1, IGHA2, IGHD, IGHG1, IGHG2, IGHG3, IGHG4, IGHM, IGHV1-18, IGHV1-2, IGHV1-24, IGHV1-3, IGHV1-46, IGHV1-58, IGHV1-69, IGHV2-26, IGHV2-5, IGHV2-70, IGHV3-11, IGHV3-13, IGHV3-15, IGHV3-20, IGHV3-21, IGHV3-23, IGHV3-30, IGHV3-33, IGHV3-43, IGHV3-48, IGHV3-49, IGHV3-53, IGHV3-64, IGHV3-66, IGHV3-7, IGHV3-73, IGHV3-74, IGHV3-9, IGHV4-34, IGHV4-39, IGHV4-4, IGHV4-59, IGHV5-51, IGKV1-12, IGKV1-16, IGKV1-17, IGKV1-5, IGKV1D-12, IGKV1D-16, IGKV2-29, IGKV2-30, IGKV2D-40, IGKV3-11, IGKV3-15, IGKV3-20, IGKV3D-20, IGKV4-1, IGKV5-2, IGLC7, IGLV1-40, IGLV1-44, IGLV1-47, IGLV2-11, IGLV2-14, IGLV2-23, IGLV2-8, IGLV3-1, IGLV3-19, IGLV3-21, IGLV3-25, IGLV3-27, IGLV6-57, IGLV7-43*
regulation of complement activation (GO:0030449)	112	52	0.59	88.81	3.07E-83	*CR2, IGHG1, IGHG2, IGHG3, IGHG4, IGHV1-2, IGHV1-46, IGHV1-69, IGHV2-5, IGHV2-70, IGHV3-11, IGHV3-13, IGHV3-23, IGHV3-30, IGHV3-33, IGHV3-48, IGHV3-53, IGHV3-7, IGHV3-9, IGHV4-34, IGHV4-39, IGHV4-59, IGKV1-12, IGKV1-16, IGKV1-17, IGKV1-5, IGKV1D-12, IGKV1D-16, IGKV2-29, IGKV2-30, IGKV2D-40, IGKV3-11, IGKV3-15, IGKV3-20, IGKV3D-20, IGKV4-1, IGKV5-2, IGLC7, IGLV1-40, IGLV1-44, IGLV1-47, IGLV2-11, IGLV2-14, IGLV2-23, IGLV2-8, IGLV3-1, IGLV3-19, IGLV3-21, IGLV3-25, IGLV3-27, IGLV6-57, IGLV7-43*
Fc-gamma receptor signaling pathway involved in phagocytosis (GO:0038096)	133	51	0.7	73.35	3.06E-77	*IGHG1, IGHG2, IGHG3, IGHG4, IGHV1-2, IGHV1-46, IGHV1-69, IGHV2-5, IGHV2-70, IGHV3-11, IGHV3-13, IGHV3-23, IGHV3-30, IGHV3-33, IGHV3-48, IGHV3-53, IGHV3-7, IGHV3-9, IGHV4-34, IGHV4-39, IGHV4-59, IGKV1-12, IGKV1-16, IGKV1-17, IGKV1-5, IGKV1D-12, IGKV1D-16, IGKV2-29, IGKV2-30, IGKV2D-40, IGKV3-11, IGKV3-15, IGKV3-20, IGKV3D-20, IGKV4-1, IGKV5-2, IGLC7, IGLV1-40, IGLV1-44, IGLV1-47, IGLV2-11, IGLV2-14, IGLV2-23, IGLV2-8, IGLV3-1, IGLV3-19, IGLV3-21, IGLV3-25, IGLV3-27, IGLV6-57, IGLV7-43*

agocytosis, recognition (GO:0006910)	73	26	0.38	68.13	8.64E-36	*IGHA1, IGHA2, IGHD, IGHG1, IGHG2, IGHG3, IGHG4, IGHM, IGHV1-18, IGHV1-24, IGHV1-3, IGHV1-58, IGHV2-26, IGHV3-15, IGHV3-20, IGHV3-21, IGHV3-23, IGHV3-43, IGHV3-49, IGHV3-64, IGHV3-66, IGHV3-73, IGHV3-74, IGHV4-4, IGHV5-51, IGLC7*
phagocytosis, engulfment (GO:0006911)	86	26	0.45	57.83	5.83E-34	*IGHA1, IGHA2, IGHD, IGHG1, IGHG2, IGHG3, IGHG4, IGHM, IGHV1-18, IGHV1-24, IGHV1-3, IGHV1-58, IGHV2-26, IGHV3-15, IGHV3-20, IGHV3-21, IGHV3-23, IGHV3-43, IGHV3-49, IGHV3-64, IGHV3-66, IGHV3-73, IGHV3-74, IGHV4-4, IGHV5-51, IGLC7*
B cell receptor signaling pathway (GO:0050853)	92	26	0.48	54.06	3.29E-33	*IGHA1, IGHA2, IGHD, IGHG1, IGHG2, IGHG3, IGHG4, IGHM, IGHV1-18, IGHV1-24, IGHV1-3, IGHV1-58, IGHV2-26, IGHV3-15, IGHV3-20, IGHV3-21, IGHV3-23, IGHV3-43, IGHV3-49, IGHV3-64, IGHV3-66, IGHV3-73, IGHV3-74, IGHV4-4, IGHV5-51, IGLC7*
immunoglobulin production (GO:0002377)	115	32	0.6	53.23	1.22E-41	*IGKV1-12, IGKV1-27, IGKV1-5, IGKV1-6, IGKV1-9, IGKV1D-13, IGKV1D-16, IGKV1D-17, IGKV1D-8, IGKV2-24, IGKV2D-24, IGKV2D-29, IGKV3-7, IGKV3D-11, IGKV3D-15, IGKV3D-20, IGKV4-1, IGKV6-21, IGKV6D-21, IGLV10-54, IGLV1-36, IGLV2-18, IGLV3-10, IGLV3-16, IGLV3-9, IGLV4-60, IGLV4-69, IGLV5-37, IGLV5-45, IGLV7-46, IGLV8-61, IGLV9-49*
Fc-epsilon receptor signaling pathway (GO:0038095)	183	47	0.96	49.13	2.21E-62	*IGHV1-2, IGHV1-46, IGHV1-69, IGHV2-5, IGHV2-70, IGHV3-11, IGHV3-13, IGHV3-23, IGHV3-30, IGHV3-33, IGHV3-48, IGHV3-53, IGHV3-7, IGHV3-9, IGHV4-34, IGHV4-39, IGHV4-59, IGKV1-12, IGKV1-16, IGKV1-17, IGKV1-5, IGKV1D-12, IGKV1D-16, IGKV2-29, IGKV2-30, IGKV2D-40, IGKV3-11, IGKV3-15, IGKV3-20, IGKV3D-20, IGKV4-1, IGKV5-2, IGLC7, IGLV1-40, IGLV1-44, IGLV1-47, IGLV2-11, IGLV2-14, IGLV2-23, IGLV2-8, IGLV3-1, IGLV3-19, IGLV3-21, IGLV3-25, IGLV3-27, IGLV6-57, IGLV7-43*
positive regulation of B cell activation (GO:0050871)	132	26	0.69	37.68	3.36E-29	*IGHA1, IGHA2, IGHD, IGHG1, IGHG2, IGHG3, IGHG4, IGHM, IGHV1-18, IGHV1-24, IGHV1-3, IGHV1-58, IGHV2-26, IGHV3-15, IGHV3-20, IGHV3-21, IGHV3-23, IGHV3-43, IGHV3-49, IGHV3-64, IGHV3-66, IGHV3-73, IGHV3-74, IGHV4-4, IGHV5-51, IGLC7*

receptor-mediated endocytosis (GO:0006898)	266	49	1.39	35.24	2.03E-58	*IGHA1, IGHA2, IGHV1-2, IGHV1-46, IGHV1-69, IGHV2-5, IGHV2-70, IGHV3-11, IGHV3-13, IGHV3-23, IGHV3-30, IGHV3-33, IGHV3-48, IGHV3-53, IGHV3-7, IGHV3-9, IGHV4-34, IGHV4-39, IGHV4-59, IGKV1-12, IGKV1-16, IGKV1-17, IGKV1-5, IGKV1D-12, IGKV1D-16, IGKV2-29, IGKV2-30, IGKV2D-40, IGKV3-11, IGKV3-15, IGKV3-20, IGKV3D-20, IGKV4-1, IGKV5-2, IGLC7, IGLV1-40, IGLV1-44, IGLV1-47, IGLV2-11, IGLV2-14, IGLV2-23, IGLV2-8, IGLV3-1, IGLV3-19, IGLV3-21, IGLV3-25, IGLV3-27, IGLV6-57, IGLV7-43*
leukocyte migration (GO:0050900)	360	50	1.88	26.57	9.03E-54	*IGHA1, IGHA2, IGHM, IGHV1-2, IGHV1-46, IGHV1-69, IGHV2-5, IGHV2-70, IGHV3-11, IGHV3-13, IGHV3-23, IGHV3-30, IGHV3-33, IGHV3-48, IGHV3-53, IGHV3-7, IGHV3-9, IGHV4-34, IGHV4-39, IGHV4-59, IGKV1-12, IGKV1-16, IGKV1-17, IGKV1-5, IGKV1D-12, IGKV1D-16, IGKV2-29, IGKV2-30, IGKV2D-40, IGKV3-11, IGKV3-15, IGKV3-20, IGKV3D-20, IGKV4-1, IGKV5-2, IGLC7, IGLV1-40, IGLV1-44, IGLV1-47, IGLV2-11, IGLV2-14, IGLV2-23, IGLV2-8, IGLV3-1, IGLV3-19, IGLV3-21, IGLV3-25, IGLV3-27, IGLV6-57, IGLV7-43*
defense response to bacterium (GO:0042742)	280	27	1.46	18.45	2.46E-22	*IGHA1, IGHA2, IGHD, IGHG1, IGHG2, IGHG3, IGHG4, IGHM, IGHV1-18, IGHV1-24, IGHV1-3, IGHV1-58, IGHV2-26, IGHV3-15, IGHV3-20, IGHV3-21, IGHV3-23, IGHV3-43, IGHV3-49, IGHV3-64, IGHV3-66, IGHV3-73, IGHV3-74, IGHV4-4, IGHV5-51, IGKV3-20, IGLC7*
proteolysis (GO:0006508)	1331	51	6.96	7.33	1.05E-27	*IGHG1, IGHG2, IGHG3, IGHG4, IGHV1-2, IGHV1-46, IGHV1-69, IGHV2-5, IGHV2-70, IGHV3-11, IGHV3-13, IGHV3-23, IGHV3-30, IGHV3-33, IGHV3-48, IGHV3-53, IGHV3-7, IGHV3-9, IGHV4-34, IGHV4-39, IGHV4-59, IGKV1-12, IGKV1-16, IGKV1-17, IGKV1-5, IGKV1D-12, IGKV1D-16, IGKV2-29, IGKV2-30, IGKV2D-40, IGKV3-11, IGKV3-15, IGKV3-20, IGKV3D-20, IGKV4-1, IGKV5-2, IGLC7, IGLV1-40, IGLV1-44, IGLV1-47, IGLV2-11, IGLV2-14, IGLV2-23, IGLV2-8, IGLV3-1, IGLV3-19, IGLV3-21, IGLV3-25, IGLV3-27, IGLV6-57, IGLV7-43*

Supplemental Table S5W6. Biological process enriched in cluster 6 from the gene expression profile of combined HPV-positive and HPV-negative tumors. We tested the enrichment of GO Biological Process terms for gene cluster 2 as identified by hierarchical clustering of RNA-seq data from combined HPV-positive (n=147) and HPV-negative (n=335) OSCC. We used PANTHER Overrepresentation Test (Released 20171205), and tested the GO Biological Process complete terms and applied multiple testing correction (Bonferroni p<0.05). The PANTHER test shows the result with gene ontology clusters; similar ontology terms are clustered together. The most enriched GO term for each ontology clusters in PANTHER output are shown in this table. Shown here are the GO term, the number of genes queried in reference human genome, (*obs.*) the number of genes observed in the gene cluster, (*exp.*) the number of genes expected by chance, fold enrichment (observed/expected), p-values, and the list of genes (see also **Supplemental Fig. S5I**).

gene	no. variants	no. variants in top 5%	fraction of variants in top 5%	p-value	FDR adjusted p-value
TP53	251	104	41.4%	1.36E-78	8.90E-75
CDKN2A	64	36	56.3%	2.14E-34	7.01E-31
FAT1	94	22	23.4%	5.31E-12	1.16E-08
NOTCH1	60	17	28.3%	5.06E-11	8.31E-08
CASP8	50	13	26.0%	2.93E-08	3.85E-05
EPHA2	16	7	43.8%	9.11E-07	9.97E-04
HRAS	28	8	28.6%	6.39E-06	6.00E-03
NSD1	24	7	29.2%	2.11E-05	1.73E-02
OR4C11	6	4	66.7%	2.84E-05	2.07E-02
AJUBA	18	6	33.3%	3.58E-05	2.35E-02
PIK3CA	50	9	18.0%	9.40E-05	5.20E-02
FBXW7	21	6	28.6%	9.50E-05	5.20E-02

Supplemental Table S6A. Genes with somatic variants having high variant allelic fraction in HPV-negative tumors.
Genes with somatic variants with significant variant high allelic fraction (VAF) in HPV-negative tumors. We extracted variant allelic fraction (VAF) of somatic variants with depth of coverage 20x or greater. We ranked variants based on their VAF and counted the number of coding-change variants ranked at the top 5% highest variant allelic fraction. For genes with three or more somatic variants, we determined the enrichment of coding change variants in the top 5% highest VAF group. We calculated the significance of enrichment using binomial statistics and applied multiple testing correction on p-values using FDR method. The table shows the genes with FDR-adjusted p-value < 0.2 (see also **Supplemental Fig. S6**).

gene	no. variants	no. variants in top 5%	fraction of variants in top 5%	p-value	FDR adjusted p-value
ZNF750	18	14	77.8%	2.23E-19	7.56E-16
PIK3CA	41	10	24.4%	8.75E-08	1.48E-04
EP300	18	7	38.9%	2.21E-07	2.50E-04
CYLD	12	4	33.3%	2.03E-04	1.38E-01
NSD1	12	4	33.3%	2.03E-04	1.38E-01

Supplemental Table S6B. Genes with somatic variants having high variant allelic fraction in HPV-positive tumors.
Genes with somatic variants having significant high allelic fraction in HPV-positive tumors. We extracted variant allelic fraction (VAF) of variants with their depth of coverage 20x or higher. We ranked variants based on their variant allelic fraction (VAF) and counted the number of coding change variants ranked int top 5% highest allelic fraction. For genes with three or more somatic variants, we determined the enrichment of coding change variants in the top 5% highest VAF allelic fraction for each gene. We used binomial statistics to determine the significance level of enrichment and applied multiple testing correction on p-values using FDR method. The table shows the genes with FDR adjusted p-value < 0.2 See also **Supplemental Fig. S6**).