

Supplementary Data

Supplementary Table S1. Patient demographic differences between heavy smokers and never smokers in HPV-positive and HPV-negative HNSCC cohorts.

Supplementary Table S2. Frequency of *NSD1*, *CASP8* and *COL11A1* mutations in never smokers, light smokers (≤ 20 pack year history) and heavy smokers (>20 pack year history) in the TCGA and Broad Institute HPV-negative HNSCC cohorts.

Supplementary Table S3. List of the top 100 gene mutational differences between never smokers and heavy smokers in HPV-positive patients.

Supplementary Table S4. List of the top 100 gene mutational differences between never smokers and heavy smokers in HPV-negative patients.

Supplementary Table S5. Top 10 differentially deleted genes in HPV-positive and negative HNSCC tumours.

Supplementary Table S6. Univariate and multivariate models of overall survival for HPV-positive and negative TCGA HNSCC cohorts ($n=67$ and 430 , respectively). For multivariate analysis, samples with unreported clinical variables were excluded, and stepwise analysis was done in backward and forward steps.

Supplementary Table S7. Univariate and multivariate analysis of overall survival in locally advanced HPV-negative OSCC. The cumulative multivariate model was constructed after stepwise analysis.

Supplementary Table S8. Patient characteristics of *NSD1* wildtype compared with *NSD1* mutants in the HPV-negative cohort. (Gender, T-stage, N-stage, Overall stage and radiation compared with Fisher's exact test, Anatomical subsite and smoking history using a chi squared test, and age using Welch's t-test.)

Supplementary Table S9. List of the top 100 gene mutational differences between *NSD1* mutant vs. wild-type samples in the HPV-negative patients.

Supplementary Table S10. Top 20 pathway over-representations from the genes mutated at significantly different frequencies between *NSD1* mutant to *NSD1* wild-type HPV-negative HNSCC tumours.

Supplementary Table S11 – Association of *NSD1* mutations with mutations in other top 25 mutated genes in the HPV-negative HNSCC cohort (Fisher's Exact test was used to assess for mutually exclusive or co-occurrence of genes, and FDR correction using the Benjamin Hochberg method).

Supplementary Figure S1. A) Total mutation load differed significantly between HPV-positive and HPV-negative samples. Box plots show minimum, 25th percentile, median, 75th percentile, and maximum values (*** Mann-Whitney U test $p=0.0001$). **B)** The correlation between total mutation load and number of smoking pack-years in HPV-positive (top) and negative (bottom) patients.

Supplementary Figure S2. Top 20 differentially mutated genes between never smokers and heavy smokers in the HPV-positive TCGA HNSCC patients. No genes had significantly different mutation frequencies after FDR analysis. Fisher's exact test was used for comparison, and FDR correction through Benjamin-Hochberg method.

Supplementary Figure S3. The 5-year overall survival of never smokers and heavy smokers compared in **(A)** HPV-positive and **(B)** HPV-negative TCGA HNSCC patients.

Supplementary Figure S4. Comparison of hazard ratios describing overall survival differences obtained from stratification patients based on *NSD1* mutation status or HPV status in all TCGA HNSCC tumors.

Supplementary Figure S5. The 5-year overall survival differences between **A)** *CASP8* **B)** *COL11A1* wild-type and mutant HPV-negative TCGA HNSCC patients are not significant.

Supplementary Figure S6. Five-year overall survival comparison between H3K36M mutant and wild-type samples in the HPV-negative TCGA HNSCC cohort.

Supplementary Table S1. Patient demographic differences between heavy smokers and never smokers in HPV-positive and HPV-negative HNSCC cohorts.

Variable	HPV Positive Samples, No. (%) (n=44)			HPV Negative Samples, No. (%) (n=295)		
	Never Smokers (n=23)	Heavy Smokers (n=21)	p value	Never Smokers (n=90)	Heavy Smokers (n=205)	p value
Gender						
Male	21 (91)	18 (86)	0.44	44 (49)	158 (77)	<0.0001
Female	2 (9)	3 (14)		46 (51)	47 (23)	
Age median (range), years	53 (35-71)	57 (45-82)	0.05	62 (19-87)	61 (30-88)	0.68
Anatomical Sub-site						
Oral cavity	4 (17)	7 (33)	0.11	79 (88)	118 (58)	<0.0001
Oropharynx	19 (83)	13 (62)		3 (3)	13 (6)	
Hypopharynx	0	1 (5)		2 (2)	5 (2)	
Larynx	0	0		6 (7)	69 (34)	
Smoking						
Current	N/A	8	N/A	N/A	109	N/A
Former	N/A	13		N/A	96	
Never	23	N/A		90	N/A	
T stage						
T0-T2, TX	16 (70)	9 (43)	0.01	44 (49)	66 (32)	0.0007
T3-T4	7 (30)	12 (57)		46 (51)	137 (68)	
N stage						
N0-N2a, NX	12 (52)	11 (52)	0.98	74 (82)	147 (72)	0.03
N2b-N3	11 (48)	10 (48)		16 (18)	56 (28)	
Overall stage						
I-III	6 (26)	6 (29)	0.79	51 (57)	92 (45)	0.01
IV	17 (74)	15 (71)		37 (43)	106 (55)	
HPV Type						
HPV-16	20 (87)	19 (90)	0.56	N/A		
HPV-Others	3 (13)	2 (10)				

Supplementary Table S2. Frequency of NSD1, CASP8 and COL11A1 mutations in never smokers, light smokers (≤ 20 pack year history) and heavy smokers (>20 pack year history) in the TCGA and Broad Institute HPV-negative HNSCC cohorts.

TCGA HPV-negative cohort						
Gene	Never smokers (n=90)		Light smokers (≤ 20 pack-year) (n=44)		Heavy smokers (>20 pack-year) (n=205)	
	Count	Percentage	Count	Percentage	Count	Percentage
NSD1	2	2.22	2	4.55	38	18.5
CASP8	20	22.2	9	20.4	13	6.34
COL11A1	1	1.11	3	6.82	27	13.2
Broad Institute HPV-negative cohort						
Gene	Never smokers (n=11)		Light smokers (≤ 20 pack-year) (n=16)		Heavy smokers (>20 pack-year) (n=49)	
	Count	Percentage	Count	Percentage	Count	Percentage
NSD1	0	0	2	13	5	10
CASP8	3	27	2	13	1	2
COL11A1	0	0	1	6	3	6

Supplementary Table S3. List of the top 100 gene mutational differences between never smokers and heavy smokers in HPV-positive patients.

	Hugo_Symbol	Heavy-Smoker count	Never-Smoker count	p Val	OR	CI UP	CI DOWN	FDR
1	KALRN	5	1	0.0883	6.6028	0.6485	338.9480	0.6298
2	DNAH9	3	0	0.1004	Inf	0.4679	Inf	0.6298
3	DNMT3B	3	0	0.1004	Inf	0.4679	Inf	0.6298
4	DPP10	3	0	0.1004	Inf	0.4679	Inf	0.6298
5	FNDC1	3	0	0.1004	Inf	0.4679	Inf	0.6298
6	KCNT2	3	0	0.1004	Inf	0.4679	Inf	0.6298
7	KIAA0556	3	0	0.1004	Inf	0.4679	Inf	0.6298
8	KIAA1551	3	0	0.1004	Inf	0.4679	Inf	0.6298
9	KPRP	3	0	0.1004	Inf	0.4679	Inf	0.6298
10	LRBA	3	0	0.1004	Inf	0.4679	Inf	0.6298
11	LRRTM1	3	0	0.1004	Inf	0.4679	Inf	0.6298
12	MFSD2A	3	0	0.1004	Inf	0.4679	Inf	0.6298
13	NBPF10	3	0	0.1004	Inf	0.4679	Inf	0.6298
14	NLRC5	3	0	0.1004	Inf	0.4679	Inf	0.6298
15	PLCL1	3	0	0.1004	Inf	0.4679	Inf	0.6298
16	PLG	3	0	0.1004	Inf	0.4679	Inf	0.6298
17	PPP1R9A	3	0	0.1004	Inf	0.4679	Inf	0.6298
18	SLTM	3	0	0.1004	Inf	0.4679	Inf	0.6298
19	THSD7A	3	0	0.1004	Inf	0.4679	Inf	0.6298
20	TICRR	3	0	0.1004	Inf	0.4679	Inf	0.6298
21	TNKS1BP1	3	0	0.1004	Inf	0.4679	Inf	0.6298
22	ZNF543	3	0	0.1004	Inf	0.4679	Inf	0.6298
23	DNMT1	0	4	0.1093	0.0000	0.0000	1.5724	0.6298
24	FGFR3	0	4	0.1093	0.0000	0.0000	1.5724	0.6298
25	HFM1	0	4	0.1093	0.0000	0.0000	1.5724	0.6298
26	RYSR2	0	4	0.1093	0.0000	0.0000	1.5724	0.6298
27	LRP1B	4	1	0.1765	4.9997	0.4414	265.8529	0.6298
28	PDE4DIP	4	1	0.1765	4.9997	0.4414	265.8529	0.6298
29	RGPD3	1	5	0.1884	0.1863	0.0036	1.8947	0.6298
30	ACSS1	2	0	0.2220	Inf	0.2077	Inf	0.6298
31	ACTG2	2	0	0.2220	Inf	0.2077	Inf	0.6298
32	ADAM20	2	0	0.2220	Inf	0.2077	Inf	0.6298
33	AHSG	2	0	0.2220	Inf	0.2077	Inf	0.6298

	Hugo_Symbol	Heavy-Smoker count	Never-Smoker count	p Val	OR	CI UP	CI DOWN	FDR
34	AKAP4	2	0	0.2220	Inf	0.2077	Inf	0.6298
35	APOB	2	0	0.2220	Inf	0.2077	Inf	0.6298
36	AR	2	0	0.2220	Inf	0.2077	Inf	0.6298
37	ARHGAP15	2	0	0.2220	Inf	0.2077	Inf	0.6298
38	ATF7IP	2	0	0.2220	Inf	0.2077	Inf	0.6298
39	ATG2B	2	0	0.2220	Inf	0.2077	Inf	0.6298
40	ATXN1	2	0	0.2220	Inf	0.2077	Inf	0.6298
41	BCAN	2	0	0.2220	Inf	0.2077	Inf	0.6298
42	BIN3	2	0	0.2220	Inf	0.2077	Inf	0.6298
43	BOC	2	0	0.2220	Inf	0.2077	Inf	0.6298
44	BRF1	2	0	0.2220	Inf	0.2077	Inf	0.6298
45	BRINP1	2	0	0.2220	Inf	0.2077	Inf	0.6298
46	BTBD8	2	0	0.2220	Inf	0.2077	Inf	0.6298
47	C4BPA	2	0	0.2220	Inf	0.2077	Inf	0.6298
48	CCDC125	2	0	0.2220	Inf	0.2077	Inf	0.6298
49	CCDC40	2	0	0.2220	Inf	0.2077	Inf	0.6298
50	CCDC73	2	0	0.2220	Inf	0.2077	Inf	0.6298
51	CD226	2	0	0.2220	Inf	0.2077	Inf	0.6298
52	CDC40	2	0	0.2220	Inf	0.2077	Inf	0.6298
53	CLASP1	2	0	0.2220	Inf	0.2077	Inf	0.6298
54	CMYA5	2	0	0.2220	Inf	0.2077	Inf	0.6298
55	CNR1	2	0	0.2220	Inf	0.2077	Inf	0.6298
56	CSRNP3	2	0	0.2220	Inf	0.2077	Inf	0.6298
57	DAGLA	2	0	0.2220	Inf	0.2077	Inf	0.6298
58	DCHS1	2	0	0.2220	Inf	0.2077	Inf	0.6298
59	DEFB114	2	0	0.2220	Inf	0.2077	Inf	0.6298
60	DIS3L2	2	0	0.2220	Inf	0.2077	Inf	0.6298
61	DPYSL4	2	0	0.2220	Inf	0.2077	Inf	0.6298
62	DSG2	2	0	0.2220	Inf	0.2077	Inf	0.6298
63	EEA1	2	0	0.2220	Inf	0.2077	Inf	0.6298
64	EED	2	0	0.2220	Inf	0.2077	Inf	0.6298
65	EFHB	2	0	0.2220	Inf	0.2077	Inf	0.6298
66	EGFLAM	2	0	0.2220	Inf	0.2077	Inf	0.6298
67	EGFR	2	0	0.2220	Inf	0.2077	Inf	0.6298
68	EIF2AK4	2	0	0.2220	Inf	0.2077	Inf	0.6298
69	EIF2S1	2	0	0.2220	Inf	0.2077	Inf	0.6298

	Hugo_Symbol	Heavy-Smoker count	Never-Smoker count	p Val	OR	CI UP	CI DOWN	FDR
70	ENPP2	2	0	0.2220	Inf	0.2077	Inf	0.6298
71	EPAS1	2	0	0.2220	Inf	0.2077	Inf	0.6298
72	EPPK1	2	0	0.2220	Inf	0.2077	Inf	0.6298
73	ETS2	2	0	0.2220	Inf	0.2077	Inf	0.6298
74	FAM83B	2	0	0.2220	Inf	0.2077	Inf	0.6298
75	FAM83G	2	0	0.2220	Inf	0.2077	Inf	0.6298
76	FBXL13	2	0	0.2220	Inf	0.2077	Inf	0.6298
77	FCRLA	2	0	0.2220	Inf	0.2077	Inf	0.6298
78	FGGY	2	0	0.2220	Inf	0.2077	Inf	0.6298
79	FIG4	2	0	0.2220	Inf	0.2077	Inf	0.6298
80	FMO2	2	0	0.2220	Inf	0.2077	Inf	0.6298
81	FNDCC7	2	0	0.2220	Inf	0.2077	Inf	0.6298
82	FSIP2	2	0	0.2220	Inf	0.2077	Inf	0.6298
83	GCKR	2	0	0.2220	Inf	0.2077	Inf	0.6298
84	GRIA2	2	0	0.2220	Inf	0.2077	Inf	0.6298
85	GRIK3	2	0	0.2220	Inf	0.2077	Inf	0.6298
86	HIVEP2	2	0	0.2220	Inf	0.2077	Inf	0.6298
87	HOXD8	2	0	0.2220	Inf	0.2077	Inf	0.6298
88	HSPD1	2	0	0.2220	Inf	0.2077	Inf	0.6298
89	HSPG2	2	0	0.2220	Inf	0.2077	Inf	0.6298
90	HYDIN	2	0	0.2220	Inf	0.2077	Inf	0.6298
91	IGSF1	2	0	0.2220	Inf	0.2077	Inf	0.6298
92	IGSF9	2	0	0.2220	Inf	0.2077	Inf	0.6298
93	IRGQ	2	0	0.2220	Inf	0.2077	Inf	0.6298
94	ITPR2	2	0	0.2220	Inf	0.2077	Inf	0.6298
95	KIAA0226	2	0	0.2220	Inf	0.2077	Inf	0.6298
96	KIAA0430	2	0	0.2220	Inf	0.2077	Inf	0.6298
97	KIAA1429	2	0	0.2220	Inf	0.2077	Inf	0.6298
98	KIAA1549	2	0	0.2220	Inf	0.2077	Inf	0.6298
99	KRBA2	2	0	0.2220	Inf	0.2077	Inf	0.6298
100	KRT5	2	0	0.2220	Inf	0.2077	Inf	0.6298

p Val – p Value of Fisher’s exact test comparison between mutation count in heavy smokers vs. never smokers

OR – Odd’s ratio

CI UP – Upper level of 95% confidence interval

CI Down – Lower level of 95% confidence interval
FDR – p Value after adjustment for False Discovery Rate

Supplementary Table S4. List of the top 100 gene mutational differences between never smokers and heavy smokers in HPV-negative patients.

	Hugo_Symbol	Heavy-Smoker count	Never-Smoker count	p Val	OR	CI UP	CI DOWN	FDR
1	NSD1	38	2	0.0001	9.9640	2.4734	87.2895	0.0272
2	CASP8	13	20	0.0002	0.2384	0.1030	0.5339	0.0409
3	COL11A1	27	1	0.0004	13.4341	2.1412	557.4822	0.0610
4	ANK2	24	1	0.0012	11.7454	1.8533	489.6288	0.1361
5	CNTNAP2	17	0	0.0022	Inf	1.9157	Inf	0.1932
6	CSMD3	58	11	0.0027	2.8246	1.3687	6.3188	0.1942
7	MYH8	21	1	0.0036	10.1115	1.5748	423.9462	0.2254
8	NPAP1	25	2	0.0042	6.0868	1.4619	54.1488	0.2321
9	DNAH5	40	6	0.0049	3.3824	1.3510	10.1571	0.2349
10	ASTN2	14	0	0.0068	Inf	1.5174	Inf	0.2349
11	NRG3	14	0	0.0068	Inf	1.5174	Inf	0.2349
12	ZNF835	14	0	0.0068	Inf	1.5174	Inf	0.2349
13	ERBB4	15	0	0.0070	Inf	1.6488	Inf	0.2349
14	ERICH3	18	1	0.0101	8.5297	1.3053	360.3322	0.2547
15	RUNX1T1	18	1	0.0101	8.5297	1.3053	360.3322	0.2547
16	GPR98	22	2	0.0112	5.2682	1.2489	47.2287	0.2547
17	BCLAF1	13	0	0.0116	Inf	1.3875	Inf	0.2547
18	GRID1	13	0	0.0116	Inf	1.3875	Inf	0.2547
19	LTBP1	13	0	0.0116	Inf	1.3875	Inf	0.2547
20	SI	33	5	0.0135	3.2510	1.2012	11.0521	0.2547
21	FAM47C	17	1	0.0163	8.0136	1.2175	339.5706	0.2547
22	PAPPA2	25	3	0.0169	4.0134	1.1758	21.3274	0.2547
23	UNC13C	25	3	0.0169	4.0134	1.1758	21.3274	0.2547
24	CSMD1	21	2	0.0175	5.0019	1.1794	44.9636	0.2547
25	ARID1A	12	0	0.0207	Inf	1.2590	Inf	0.2547
26	ATP13A4	12	0	0.0207	Inf	1.2590	Inf	0.2547
27	DMBT1	12	0	0.0207	Inf	1.2590	Inf	0.2547
28	PDE10A	12	0	0.0207	Inf	1.2590	Inf	0.2547
29	PRB2	12	0	0.0207	Inf	1.2590	Inf	0.2547
30	ZNF479	12	0	0.0207	Inf	1.2590	Inf	0.2547
31	CDKN2A	44	31	0.0208	0.5214	0.2910	0.9383	0.2547
32	COL4A3	11	0	0.0209	Inf	1.1320	Inf	0.2547
33	PROX1	11	0	0.0209	Inf	1.1320	Inf	0.2547

	Hugo_Symbol	Heavy-Smoker count	Never-Smoker count	p Val	OR	CI UP	CI DOWN	FDR
34	REG1A	11	0	0.0209	Inf	1.1320	Inf	0.2547
35	RGS7	11	0	0.0209	Inf	1.1320	Inf	0.2547
36	TMEM200A	11	0	0.0209	Inf	1.1320	Inf	0.2547
37	RP1	23	3	0.0267	3.6525	1.0602	19.5100	0.2782
38	CNTNAP5	16	1	0.0271	7.5029	1.1306	319.0248	0.2782
39	VPS13C	16	1	0.0271	7.5029	1.1306	319.0248	0.2782
40	PEG3	20	2	0.0281	4.7385	1.1107	42.6709	0.2782
41	CTNND2	19	2	0.0285	4.4779	1.0428	40.5061	0.2782
42	DCHS2	19	2	0.0285	4.4779	1.0428	40.5061	0.2782
43	DNAH7	19	2	0.0285	4.4779	1.0428	40.5061	0.2782
44	LRP1	19	2	0.0285	4.4779	1.0428	40.5061	0.2782
45	RIMS2	19	2	0.0285	4.4779	1.0428	40.5061	0.2782
46	LRP1B	43	9	0.0301	2.3826	1.0776	5.8384	0.2792
47	PIK3CA	44	10	0.0347	2.1811	1.0147	5.1193	0.2792
48	BICC1	10	0	0.0350	Inf	1.0064	Inf	0.2792
49	DPP6	10	0	0.0350	Inf	1.0064	Inf	0.2792
50	GOLGB1	10	0	0.0350	Inf	1.0064	Inf	0.2792
51	LRRTM1	10	0	0.0350	Inf	1.0064	Inf	0.2792
52	MAP1B	10	0	0.0350	Inf	1.0064	Inf	0.2792
53	NLRP14	10	0	0.0350	Inf	1.0064	Inf	0.2792
54	OR5I1	10	0	0.0350	Inf	1.0064	Inf	0.2792
55	PHF3	10	0	0.0350	Inf	1.0064	Inf	0.2792
56	CUBN	22	3	0.0404	3.4749	1.0034	18.6163	0.2896
57	SPTA1	22	3	0.0404	3.4749	1.0034	18.6163	0.2896
58	ABCC8	14	1	0.0439	6.4976	0.9597	278.5671	0.2896
59	CHEK2	14	1	0.0439	6.4976	0.9597	278.5671	0.2896
60	NLRP8	14	1	0.0439	6.4976	0.9597	278.5671	0.2896
61	PTPRB	14	1	0.0439	6.4976	0.9597	278.5671	0.2896
62	GRM1	18	2	0.0439	4.2201	0.9756	38.3244	0.2896
63	KCNT2	18	2	0.0439	4.2201	0.9756	38.3244	0.2896
64	MROH2B	18	2	0.0439	4.2201	0.9756	38.3244	0.2896
65	MUC4	18	2	0.0439	4.2201	0.9756	38.3244	0.2896
66	SCN1A	18	2	0.0439	4.2201	0.9756	38.3244	0.2896
67	RYR2	29	5	0.0460	2.7930	1.0194	9.5666	0.2896
68	ANKRD30A	15	1	0.0462	6.9976	1.0447	298.6913	0.2896
69	ASXL3	15	1	0.0462	6.9976	1.0447	298.6913	0.2896

	Hugo_Symbol	Heavy-Smoker count	Never-Smoker count	p Val	OR	CI UP	CI DOWN	FDR
70	KLHL1	15	1	0.0462	6.9976	1.0447	298.6913	0.2896
71	FAT4	24	4	0.0538	2.8426	0.9352	11.6258	0.3026
72	DYNC1H1	8	9	0.0550	0.3669	0.1186	1.1132	0.3026
73	ACSM2B	9	0	0.0613	Inf	0.8824	Inf	0.3026
74	AFF3	9	0	0.0613	Inf	0.8824	Inf	0.3026
75	CDH11	9	0	0.0613	Inf	0.8824	Inf	0.3026
76	ELMO1	9	0	0.0613	Inf	0.8824	Inf	0.3026
77	IL1RAPL2	9	0	0.0613	Inf	0.8824	Inf	0.3026
78	KCNH7	9	0	0.0613	Inf	0.8824	Inf	0.3026
79	KIAA2022	9	0	0.0613	Inf	0.8824	Inf	0.3026
80	NACA	9	0	0.0613	Inf	0.8824	Inf	0.3026
81	NPIPA5	9	0	0.0613	Inf	0.8824	Inf	0.3026
82	PABPC5	9	0	0.0613	Inf	0.8824	Inf	0.3026
83	PLXNA4	9	0	0.0613	Inf	0.8824	Inf	0.3026
84	SELP	9	0	0.0613	Inf	0.8824	Inf	0.3026
85	STXBP5L	9	0	0.0613	Inf	0.8824	Inf	0.3026
86	TNIK	9	0	0.0613	Inf	0.8824	Inf	0.3026
87	TRPC4	9	0	0.0613	Inf	0.8824	Inf	0.3026
88	USP8	9	0	0.0613	Inf	0.8824	Inf	0.3026
89	ZIC4	9	0	0.0613	Inf	0.8824	Inf	0.3026
90	FMN2	20	3	0.0622	3.1255	0.8916	16.8580	0.3036
91	AHNAK	27	5	0.0663	2.5717	0.9315	8.8525	0.3172
92	EP300	17	2	0.0690	3.9650	0.9092	36.1388	0.3172
93	BAI3	16	2	0.0697	3.7125	0.8436	33.9858	0.3172
94	EPB41L3	16	2	0.0697	3.7125	0.8436	33.9858	0.3172
95	HCN1	16	2	0.0697	3.7125	0.8436	33.9858	0.3172
96	ABCA12	13	1	0.0715	6.0027	0.8758	258.6488	0.3172
97	ATP10A	13	1	0.0715	6.0027	0.8758	258.6488	0.3172
98	UNC5D	13	1	0.0715	6.0027	0.8758	258.6488	0.3172
99	ZNF99	13	1	0.0715	6.0027	0.8758	258.6488	0.3172
100	PRDM9	23	4	0.0789	2.7094	0.8869	11.1144	0.3434

p Val – p Value of Fisher’s exact test comparison between mutation count in heavy smokers vs. never smokers

OR – Odd’s ratio

CI UP – Upper level of 95% confidence interval

CI Down – Lower level of 95% confidence interval

FDR – p Value after adjustment for False Discovery Rate

Supplementary Table S5. Top 10 differentially deleted genes in HPV-positive and negative HNSCC tumours.

HPV-Positive cohort deep deletions				
Gene	Never smokers (n=23)	Heavy smokers (n=21)	p value	Adjusted p value
PTEN	7	1	0.05	1
ATAD1	4	0	0.11	1
CFL1P1	4	0	0.11	1
RN7SL78P	4	0	0.11	1
KLLN	4	0	0.11	1
LRP1B	0	2	0.22	1
RBMS3	0	2	0.22	1
U3	0	2	0.22	1
TGFBR2	0	2	0.22	1
GADL1	0	2	0.22	1
HPV-Negative cohort deep deletions				
Gene	Never smokers (n=90)	Heavy smokers (n=205)	p value	Adjusted p value
NCKAP5	18	14	0.0019	1
DUX4L2	8	3	0.0045	1
DUX4L3	8	3	0.0045	1
DUX4L4	8	3	0.0045	1
DUX4L5	8	3	0.0045	1
DUX4L6	8	3	0.0045	1
DUX4L7	8	3	0.0045	1
DUX4	8	3	0.0045	1
FRG1	8	3	0.0045	1
FRG2	8	3	0.0045	1
NCKAP5	18	14	0.0019	1

Supplementary Table S6. Univariate and multivariate models of overall survival for HPV-positive and negative TCGA HNSCC cohorts (n=67 and 430, respectively). For multivariate analysis, samples with unreported clinical variables were excluded, and stepwise analysis was done in backward and forward steps.

		HPV-positive				HPV-negative			
		Univariate		Multivariate		Univariate		Multivariate	
Variables		HR (95% CI)	P value	HR (95% CI)	P value	HR (95% CI)	P value	HR (95% CI)	P value
Gender	Male vs. female	1.00 (0.223 - 4.51)	0.996			0.761 (0.559-1.04)	0.0833	0.748 (0.508-1.10)	0.142
Age		0.990 (0.940-1.04)	0.713	0.937(0.856-1.03)	0.160	1.02 (1.00-1.03)	0.028		
Subsite	Larynx vs. hypopharynx	0.905 (0-Inf)	1.00			0.449 (0.160-1.26)	0.127		
	Oral cavity vs. hypopharynx	1.88e08(0-Inf)	1.00	2.35e06 (0-Inf)	0.999	0.550 (0.203-1.49)	0.240		
	Oropharynx vs. hypopharynx	6.22e07(0-Inf)	1.00	2.72e07 (0-Inf)	0.999	0.804 (0.262-2.47)	0.704		
Smoking history	Light vs. never	0.858 (0.164-4.48)	0.856			1.322 (0.751-2.33)	0.333		
	Heavy vs. never	2.07 (0.560-7.65)	0.275			0.978 (0.663-1.44)	0.912		
T stage	T3-T4 vs. T0-T2, Tx	1.07 (0.376-3.03)	0.902			1.10 (0.804-1.51)	0.55		
N stage	N2b-N3 vs. N0-N2a, NX	0.362 (0.123-1.07)	0.0653			1.76 (1.29-2.41)	0.00042	2.02 (1.34-3.04)	0.000747
Overall stage	IV vs. I-III	0.792 (0.270 - 2.32)	0.672			1.16 (0.861-1.55)	0.336		
Radiation	Yes vs. No	0.290 (0.0871-0.963)	0.0432	0.108 (0.0235-0.484)	0.00373	0.654 (0.470-0.911)	0.012	0.628 (0.418-0.942)	0.0246
HPV type	HPV others vs. HPV 16	3.28 (1.12-9.62)	0.0306						
NSD1	Mutant vs. Wildtype					0.429 (0.244-0.755)	0.00331	0.365 (0.169-0.789)	0.0104

Supplementary Table S7. Univariate and multivariate analysis of overall survival in locally advanced HPV-negative OSCC. The cumulative multivariate model was constructed after stepwise analysis.

		Univariate		Multivariate	
Variable		HR (95% CI)	p value	HR (95% CI)	p value
Gender	Male <i>vs.</i> female	1.67 (0.724-3.87)	0.229	2.16 (0.817-5.74)	0.121
Age		0.999 (0.973-1.03)	0.953		
Smoking history	Light <i>vs.</i> never	1.02 (0.310-3.35)	0.974	0.874 (0.239-3.20)	0.839
	Heavy <i>vs.</i> never	2.30 (0.933-5.69)	0.0703	2.67 (0.903-7.91)	0.0757
Alcohol abuse	Positive <i>vs.</i> negative	1.96 (0.988-3.88)	0.054		
N stage	N2b-N3 <i>vs.</i> N0-N2a	3.83 (1.82-8.04)	0.000391	4.59 (2.06-10.2)	0.00020
Margins	Positive <i>vs.</i> Negative	2.99 (0.903-9.88)	0.0729	9.26 (2.19-39.2)	0.00249
<i>NSDI</i>	Mutated <i>vs.</i> wildtype	0.323 (0.0772-1.35)	0.122	0.230 (0.0530-0.998)	0.0496

Supplementary Table S8. Patient characteristics of *NSD1* wildtype compared with *NSD1* mutants in the HPV-negative cohort. (Gender, T-stage, N-stage, Overall stage and radiation compared with Fisher's exact test, Anatomical subsite and smoking history using a chi squared test, and age using Welch's t-test.)

Variable	HPV Negative Samples, No. (%) (n=430)		
	<i>NSD1</i> wildtype (n=375)	<i>NSD1</i> mutants (n=55)	p value
Gender			
Male	262	43	0.265
Female	113	12	
Age median (range), years	62 (19-90)	60 (38-80)	0.112
Anatomical Sub-site			
Oral cavity	267	21	< 0.0001
Oropharynx	22	3	
Hypopharynx	8	0	
Larynx	78	31	
Smoking History			
Heavy	166	38	0.00105
Light	42	2	
None	88	2	
Unknown	36	13	
T stage			
T0-T2, TX	131	13	0.0947
T3-T4	240	42	
Unknown	3	0	
N stage			
N0-N2a, NX	272	44	0.326
N2b-N3	99	11	
Unknown	4	0	
Overall stage			
I-III	168	27	0.662
IV	194	27	
Unknown	13	1	
Radiation			
Yes	206	32	1.00
No	119	19	
Unknown	50	4	

Supplementary Table S9. List of the top 100 gene mutational differences between *NSD1* mutant vs. wild-type samples in the HPV-negative patients.

	Hugo_Symbol	NSD1 mutant	NSD1 wild-type	p Val	OR	CI UP	CI DOWN	FDR
1	WDR17	11	5	0.0000	18.3000	5.5558	70.5449	0.0002
2	CSMD3	28	68	0.0000	4.6751	2.4839	8.8352	0.0008
3	PTPRD	12	10	0.0000	10.1121	3.7577	27.8644	0.0009
4	SYNE1	25	58	0.0000	4.5472	2.3811	8.6579	0.0009
5	COL11A1	17	26	0.0000	5.9829	2.7830	12.6795	0.0011
6	SLITRK4	9	5	0.0000	14.3435	4.1116	56.8331	0.0026
7	FSTL5	10	8	0.0000	10.1221	3.4016	31.2107	0.0036
8	ZNF676	10	8	0.0000	10.1221	3.4016	31.2107	0.0036
9	DNAH8	15	24	0.0000	5.4657	2.4553	11.9188	0.0047
10	LPHN3	10	9	0.0000	8.9758	3.0975	26.4620	0.0056
11	OR5M3	6	1	0.0000	45.1553	5.3134	2090.2435	0.0056
12	PKHD1L1	17	34	0.0000	4.4770	2.1365	9.1816	0.0074
13	AMPH	8	5	0.0000	12.4898	3.4400	50.7012	0.0074
14	SLIT2	10	10	0.0000	8.0624	2.8430	22.9121	0.0074
15	RP1	13	20	0.0000	5.4746	2.3247	12.5644	0.0095
16	COL24A1	9	8	0.0001	8.9155	2.8966	28.0269	0.0097
17	LAMA1	11	14	0.0001	6.4163	2.4732	16.3089	0.0107
18	COL4A3	8	6	0.0001	10.3903	3.0148	38.0337	0.0110
19	ZIC1	8	6	0.0001	10.3903	3.0148	38.0337	0.0110
20	GTF2A1L	6	2	0.0001	22.5766	3.9014	233.8789	0.0110
21	NYAP2	6	2	0.0001	22.5766	3.9014	233.8789	0.0110
22	SCN2A	7	4	0.0001	13.4065	3.2694	64.8458	0.0116
23	CNTNAP5	10	12	0.0001	6.6903	2.4397	18.0064	0.0132
24	APOB	14	26	0.0001	4.5721	2.0377	9.9569	0.0140
25	DCHS2	11	16	0.0001	5.5884	2.1974	13.7873	0.0165
26	SI	16	35	0.0001	3.9789	1.8795	8.1990	0.0165
27	SELE	5	1	0.0002	36.9037	4.0163	1758.7061	0.0165
28	MYH2	9	10	0.0002	7.1042	2.4181	20.6126	0.0165
29	PXDNL	9	10	0.0002	7.1042	2.4181	20.6126	0.0165
30	CUX2	7	5	0.0002	10.7098	2.8023	44.5816	0.0165
31	EGFLAM	7	5	0.0002	10.7098	2.8023	44.5816	0.0165
32	FER1L6	7	5	0.0002	10.7098	2.8023	44.5816	0.0165
33	IQUB	7	5	0.0002	10.7098	2.8023	44.5816	0.0165
34	OR4M1	7	5	0.0002	10.7098	2.8023	44.5816	0.0165

	Hugo_Symbol	NSD1 mutant	NSD1 wild-type	p Val	OR	CI UP	CI DOWN	FDR
35	C8B	6	3	0.0002	15.0410	3.0965	95.9606	0.0178
36	OR5H14	6	3	0.0002	15.0410	3.0965	95.9606	0.0178
37	COL19A1	8	8	0.0002	7.7635	2.4167	24.9854	0.0205
38	LAMA2	13	25	0.0003	4.3237	1.8830	9.5819	0.0218
39	PAPPA2	14	29	0.0003	4.0665	1.8316	8.7303	0.0218
40	CNTN3	7	6	0.0004	8.9092	2.4523	33.5567	0.0269
41	MME	7	6	0.0004	8.9092	2.4523	33.5567	0.0269
42	MAGEC1	10	15	0.0004	5.3146	2.0083	13.5407	0.0272
43	DNAH5	18	47	0.0004	3.3925	1.6767	6.7068	0.0272
44	DSP	9	12	0.0004	5.8942	2.0735	16.2086	0.0279
45	SPTA1	11	19	0.0004	4.6713	1.8789	11.1477	0.0279
46	C7	6	4	0.0005	11.2695	2.5708	56.2653	0.0279
47	CAPRIN2	6	4	0.0005	11.2695	2.5708	56.2653	0.0279
48	CCDC40	6	4	0.0005	11.2695	2.5708	56.2653	0.0279
49	PABPC5	6	4	0.0005	11.2695	2.5708	56.2653	0.0279
50	UNC13C	12	23	0.0005	4.2616	1.7989	9.6952	0.0279
51	CD93	5	2	0.0005	18.4552	2.9288	198.2901	0.0279
52	LHCGR	5	2	0.0005	18.4552	2.9288	198.2901	0.0279
53	MPEG1	5	2	0.0005	18.4552	2.9288	198.2901	0.0279
54	MRGPRX2	5	2	0.0005	18.4552	2.9288	198.2901	0.0279
55	DNAH9	13	27	0.0005	3.9823	1.7475	8.7337	0.0279
56	CDH9	10	16	0.0005	4.9705	1.8955	12.4934	0.0297
57	GRM1	10	16	0.0005	4.9705	1.8955	12.4934	0.0297
58	DAAM2	7	7	0.0006	7.6233	2.1798	26.7017	0.0321
59	GRIA2	7	7	0.0006	7.6233	2.1798	26.7017	0.0321
60	SLC39A12	7	7	0.0006	7.6233	2.1798	26.7017	0.0321
61	COL22A1	12	24	0.0007	4.0734	1.7272	9.2100	0.0321
62	LRP2	12	24	0.0007	4.0734	1.7272	9.2100	0.0321
63	AFF3	8	10	0.0007	6.1848	2.0148	18.4034	0.0321
64	DYSF	8	10	0.0007	6.1848	2.0148	18.4034	0.0321
65	NBAS	8	10	0.0007	6.1848	2.0148	18.4034	0.0321
66	PDZD2	8	10	0.0007	6.1848	2.0148	18.4034	0.0321
67	NRXN1	10	17	0.0008	4.6667	1.7943	11.5876	0.0357
68	RIMS2	10	17	0.0008	4.6667	1.7943	11.5876	0.0357
69	USH2A	17	46	0.0008	3.1979	1.5607	6.3761	0.0357
70	FSHR	6	5	0.0009	9.0007	2.1988	38.7711	0.0400
71	TJP1	6	5	0.0009	9.0007	2.1988	38.7711	0.0400

	Hugo_Symbol	NSD1 mutant	NSD1 wild-type	p Val	OR	CI UP	CI DOWN	FDR
72	ERICH3	9	14	0.0009	5.0287	1.8128	13.3125	0.0400
73	ZNF804A	9	14	0.0009	5.0287	1.8128	13.3125	0.0400
74	CD163	8	11	0.0011	5.6106	1.8593	16.2222	0.0436
75	EYS	8	11	0.0011	5.6106	1.8593	16.2222	0.0436
76	OR5D13	7	8	0.0011	6.6572	1.9614	22.0769	0.0436
77	RP1L1	7	8	0.0011	6.6572	1.9614	22.0769	0.0436
78	PKHD1	10	18	0.0011	4.3966	1.7029	10.7990	0.0436
79	AVPR1A	5	3	0.0012	12.2940	2.3136	81.5660	0.0436
80	C2CD5	5	3	0.0012	12.2940	2.3136	81.5660	0.0436
81	CCDC85A	5	3	0.0012	12.2940	2.3136	81.5660	0.0436
82	CNDP1	5	3	0.0012	12.2940	2.3136	81.5660	0.0436
83	OR10A7	5	3	0.0012	12.2940	2.3136	81.5660	0.0436
84	OR2M3	5	3	0.0012	12.2940	2.3136	81.5660	0.0436
85	TRPC1	5	3	0.0012	12.2940	2.3136	81.5660	0.0436
86	RYSR2	15	38	0.0013	3.3226	1.5567	6.8555	0.0491
87	THSD7A	10	19	0.0015	4.1548	1.6199	10.1065	0.0536
88	ZFHX4	12	27	0.0015	3.5922	1.5410	7.9935	0.0536
89	ITGA8	8	12	0.0016	5.1315	1.7257	14.4929	0.0545
90	ADAMTSL1	6	6	0.0017	7.4887	1.9209	29.2308	0.0545
91	ASTN1	6	6	0.0017	7.4887	1.9209	29.2308	0.0545
92	KDR	6	6	0.0017	7.4887	1.9209	29.2308	0.0545
93	OR4Q3	6	6	0.0017	7.4887	1.9209	29.2308	0.0545
94	REG1A	6	6	0.0017	7.4887	1.9209	29.2308	0.0545
95	SPATA31E1	6	6	0.0017	7.4887	1.9209	29.2308	0.0545
96	ADCY1	7	9	0.0017	5.9056	1.7823	18.7652	0.0545
97	FAM155A	7	9	0.0017	5.9056	1.7823	18.7652	0.0545
98	PLCL1	7	9	0.0017	5.9056	1.7823	18.7652	0.0545
99	ROBO2	7	9	0.0017	5.9056	1.7823	18.7652	0.0545
100	SLITRK2	7	9	0.0017	5.9056	1.7823	18.7652	0.0545

p Val – p Value of Fisher’s exact test comparison between mutation count in heavy-smokers vs. never smokers.

OR – Odd’s ratio

CI UP – Upper level of 95% confidence interval

CI Down – Lower level of 95% confidence interval

FDR – p Value after adjustment for False Discovery Rate

Supplementary Table S10. Top 20 pathway over-representations from the genes mutated at significantly different frequencies between *NSDI* mutant to *NSDI* wild-type HPV-negative HNSCC tumours.

1	Pathway identifier	R-HSA-8874081		
	Description	#Entities total: 32 MET activates PTK2 signaling		
	#Entities mapped in Reactome	5	Submitted entities found	LAMA2, COL24A1, LAMA1, COL11A1
	p-value	6.38E-06	FDR q-value	9.30E-04
2	Pathway identifier	R-HSA-3000171		
	Description	#Entities total: 61 Non-integrin membrane-ECM interactions		
	#Entities mapped in Reactome	6	Submitted entities found	LAMA2, LAMA1, NRXN1, COL11A1, COL4A3
	p-value	1.02E-05	FDR q-value	9.30E-04
3	Pathway identifier	R-HSA-1474244		
	Description	#Entities total: 329 Extracellular matrix organization		
	#Entities mapped in Reactome	12	Submitted entities found	LAMA2, COL24A1, MME, LAMA1, COL11A1, COL22A1, NRXN1, COL4A3, KDR, ITGA8, COL19A1
	p-value	1.22E-05	FDR q-value	9.30E-04
4	Pathway identifier	R-HSA-8948216		
	Description	#Entities total: 44 Collagen chain trimerization		
	#Entities mapped in Reactome	5	Submitted entities found	COL24A1, COL11A1, COL22A1, COL4A3, COL19A1
	p-value	2.91E-05	FDR q-value	1.46E-03
5	Pathway identifier	R-HSA-8875878		
	Description	#Entities total: 45 MET promotes cell motility		
	#Entities mapped in Reactome	5	Submitted entities found	LAMA2, COL24A1, LAMA1, COL11A1
	p-value	3.24E-05	FDR q-value	1.46E-03
6	Pathway identifier	R-HSA-3000157		
	Description	#Entities total: 31 Laminin interactions		
	#Entities mapped in Reactome	4	Submitted entities found	LAMA2, LAMA1, COL4A3
	p-value	1.17E-04	FDR q-value	4.45E-03
7	Pathway identifier	R-HSA-1474290		
	Description	#Entities total: 104		

		Collagen formation		
	#Entities mapped in Reactome	6	Submitted entities found	COL24A1, LAMA2, COL11A1, COL22A1, COL4A3, COL19A1
	p-value	1.90E-04	FDR q-value	6.09E-03
8	Pathway identifier		R-HSA-1650814	
	Description	#Entities total: 76 Collagen biosynthesis and modifying enzymes		
	#Entities mapped in Reactome	5	Submitted entities found	COL24A1, COL11A1, COL22A1, COL4A3, COL19A1
	p-value	3.67E-04	FDR q-value	1.03E-02
9	Pathway identifier		R-HSA-6806834	
	Description	#Entities total: 87 Signaling by MET		
	#Entities mapped in Reactome	5	Submitted entities found	LAMA2, COL24A1, LAMA1, COL11A1
	p-value	6.73E-04	FDR q-value	1.68E-02
10	Pathway identifier		R-HSA-6794362	
	Description	#Entities total: 93 Protein-protein interactions at synapses		
	#Entities mapped in Reactome	5	Submitted entities found	PTPRD, SLITRK4, NRXN1, GRM1
	p-value	9.05E-04	FDR q-value	1.99E-02
11	Pathway identifier		R-HSA-166665	
	Description	#Entities total: 8 Terminal pathway of complement		
	#Entities mapped in Reactome	2	Submitted entities found	C7, C8B
	p-value	1.90E-03	FDR q-value	3.50E-02
12	Pathway identifier		R-HSA-8935964	
	Description	#Entities total: 8 RUNX1 regulates expression of components of tight junctions		
	#Entities mapped in Reactome	2	Submitted entities found	TJP1
	p-value	1.90E-03	FDR q-value	3.50E-02
13	Pathway identifier		R-HSA-2022090	
	Description	#Entities total: 67 Assembly of collagen fibrils and other multimeric structures		
	#Entities mapped in Reactome	4	Submitted entities found	COL24A1, LAMA2, COL11A1, COL4A3
	p-value	2.06E-03	FDR q-value	3.50E-02
14	Pathway identifier		R-HSA-1442490	
	Description	#Entities total: 69 Collagen degradation		
	#Entities mapped in Reactome	4	Submitted entities found	MME, COL11A1, COL4A3, COL19A1

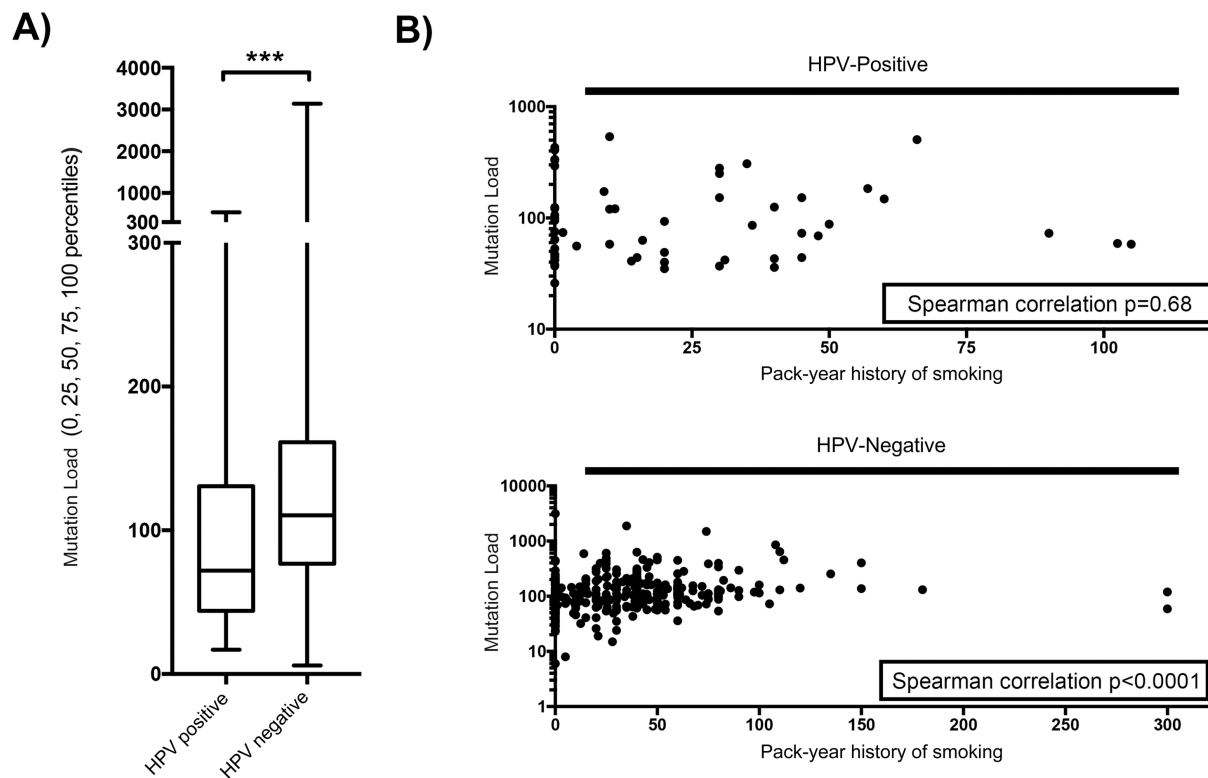
	p-value	2.29E-03	FDR q-value	3.66E-02
15	Pathway identifier	R-HSA-351906		
	Description	#Entities total: 11 Apoptotic cleavage of cell adhesion proteins		
	#Entities mapped in Reactome	2	Submitted entities found	TJP1, DSP
	p-value	3.53E-03	FDR q-value	5.18E-02
16	Pathway identifier	R-HSA-3000178		
	Description	#Entities total: 79 ECM proteoglycans		
	#Entities mapped in Reactome	4	Submitted entities found	LAMA2, LAMA1, COL4A3, ITGA8
	p-value	3.70E-03	FDR q-value	5.18E-02
17	Pathway identifier	R-HSA-428542		
	Description	#Entities total: 12 Regulation of commissural axon pathfinding by SLIT and ROBO		
	#Entities mapped in Reactome	2	Submitted entities found	ROBO2, SLIT2
	p-value	4.18E-03	FDR q-value	5.43E-02
18	Pathway identifier	R-HSA-375281		
	Description	#Entities total: 13 Hormone ligand-binding receptors		
	#Entities mapped in Reactome	2	Submitted entities found	LHCGR, FSHR
	p-value	4.88E-03	FDR q-value	5.86E-02
19	Pathway identifier	R-HSA-2214320		
	Description	#Entities total: 15 Anchoring fibril formation		
	#Entities mapped in Reactome	2	Submitted entities found	LAMA2, COL4A3
	p-value	6.43E-03	FDR q-value	7.24E-02
20	Pathway identifier	R-HSA-1474228		
	Description	#Entities total: 5 Degradation of the extracellular matrix		
	#Entities mapped in Reactome	148	Submitted entities found	MME, LAMA2, COL11A1, COL4A3, COL19A1
	p-value	6.58E-03	FDR q-value	7.24E-02

Supplementary Table S11 – Association of *NSD1* mutations with mutations in other top 25

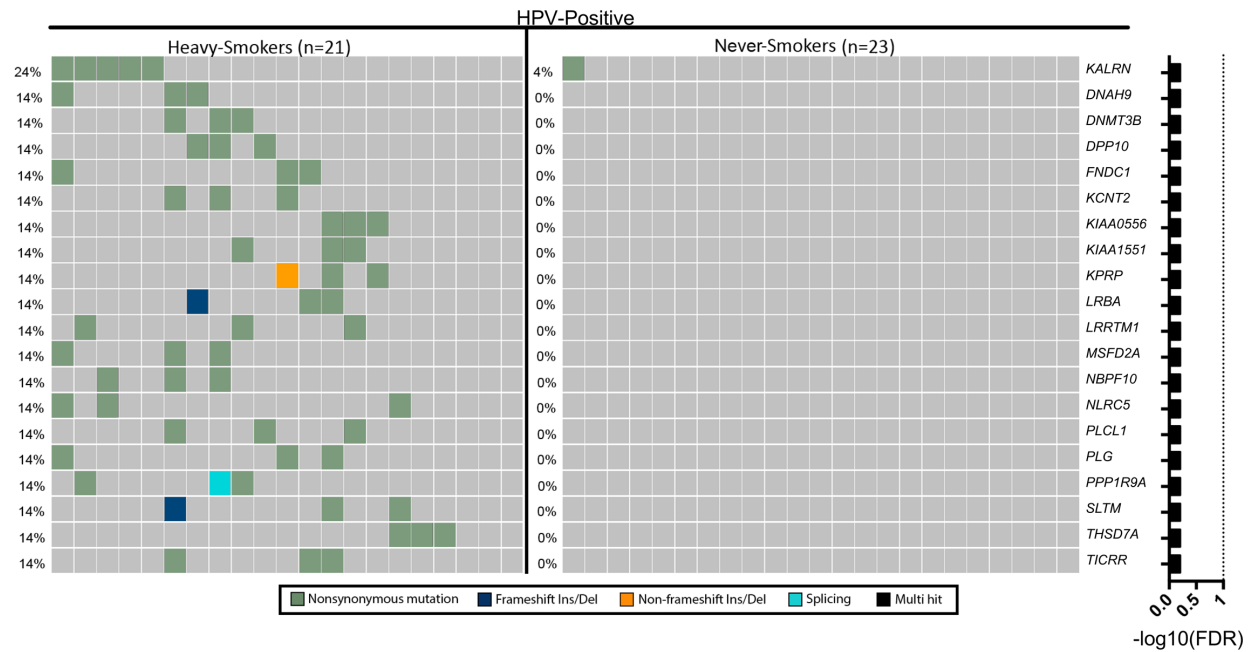
mutated genes in the HPV-negative HNSCC cohort (Fisher's Exact test was used to assess for mutually exclusive or co-occurrence of genes, and FDR correction using the Benjamin Hochberg method).

Gene 1	Gene 2	Wildtype-Wildtype	Mutated-mutated	Wildtype-mutated	Mutated-wildtype	Odds Ratio	p Value	FDR
NSD1	CSMD3	307	28	68	27	2.8	0.000176	0.00264
	SYNE1	317	25	58	30	2.93	0.000211	0.00264
	PKHD1L1	341	17	34	38	3.4	0.000533	0.00444
	SI	340	16	35	39	3.11	0.00126	0.00785
	DNAH5	328	18	47	37	2.6	0.00404	0.0202
	USH2A	329	17	46	38	2.51	0.00641	0.0229
	RYR2	337	15	38	40	2.68	0.00589	0.0229
	XIRP2	333	15	42	40	2.43	0.014	0.0398
	DMD	340	13	35	42	2.53	0.0143	0.0398
	FLG	325	16	50	39	2.18	0.0218	0.0496
	AHNAK	343	12	32	43	2.55	0.0201	0.0496
	PCLO	318	17	57	38	2.03	0.0296	0.0617
	LRP1B	314	18	61	37	2.01	0.0344	0.0661
	DST	337	12	38	43	2.15	0.0489	0.0873
	MUC16	307	18	68	37	1.8	0.0611	0.102
	FRG1B	302	18	73	37	1.68	0.0957	0.15
	FAM135B	340	10	35	45	1.94	0.107	0.157
	TTN	227	30	148	25	1.38	0.2	0.277
	CDKN2A	274	10	101	45	0.675	0.328	0.431
	CASP8	325	4	50	51	0.546	0.376	0.47
	PIK3CA	312	11	63	44	1.19	0.58	0.691
	FAT1	275	13	100	42	0.887	0.873	0.949
	KMT2D	313	8	47	62	0.88	0.848	0.949
	TP53	65	46	310	9	1.01	1	1
	NOTCH1	301	11	74	44	1.01	1	1

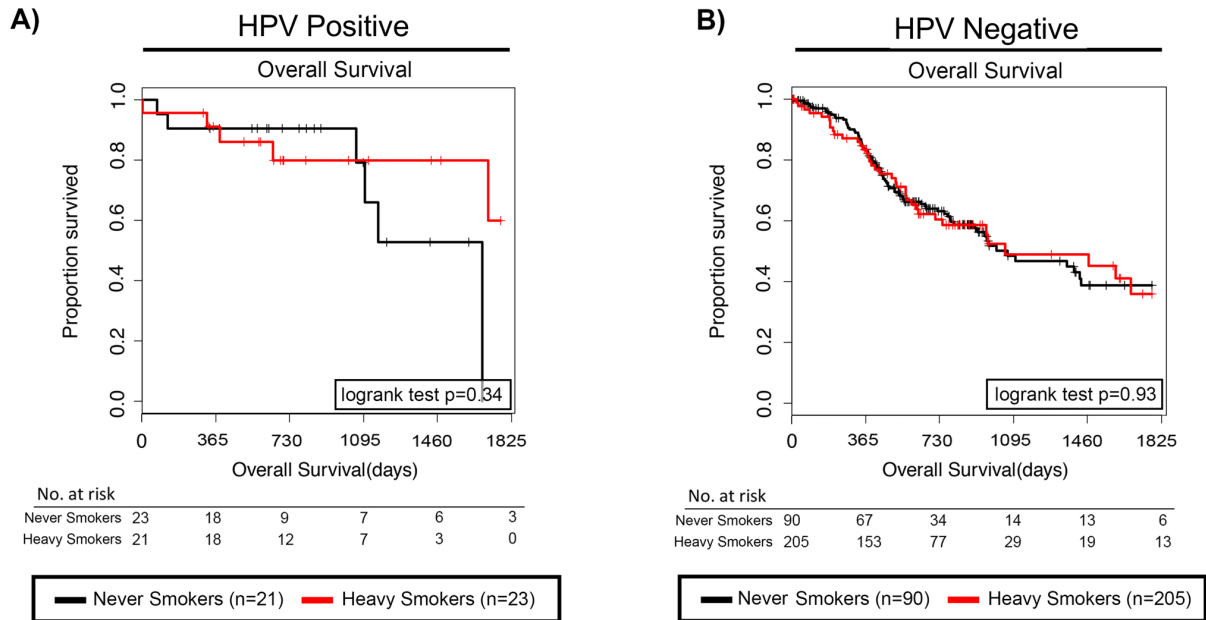
Supplementary Figure S1. A) Total mutation load differed significantly between HPV-positive and HPV-negative samples. Box plots show minimum, 25th percentile, median, 75th percentile, and maximum values (***) Mann-Whitney U test $p=0.0001$). **B)** The correlation between total mutation load and number of smoking pack-years in HPV-positive (top) and negative (bottom) patients.



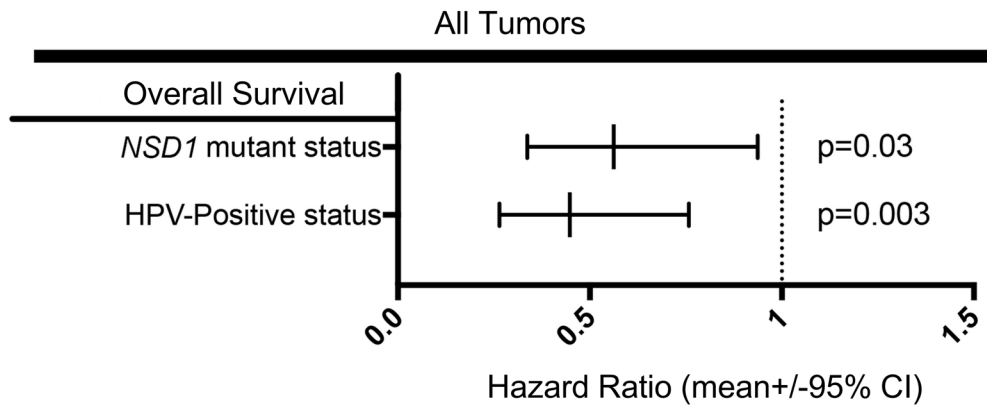
Supplementary Figure S2. Top 20 differentially mutated genes between never smokers and heavy smokers in the HPV-positive TCGA HNSCC patients. No genes had significantly different mutation frequencies after FDR analysis. Fisher's exact test was used for comparison, and FDR correction through Benjamin-Hochberg method.



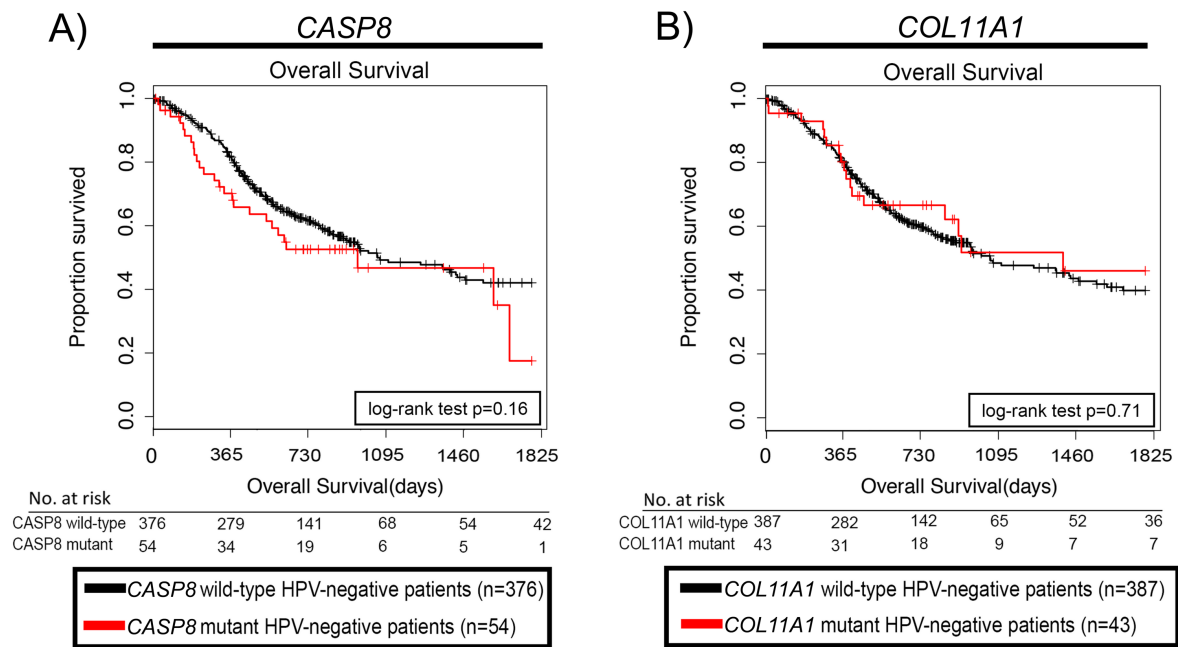
Supplementary Figure S3. The 5-year overall survival of never smokers and heavy smokers compared in **(A)** HPV-positive and **(B)** HPV-negative TCGA HNSCC patients.



Supplementary Figure S4. Comparison of hazard ratios describing overall survival differences obtained from stratification patients based on *NSD1* mutation status or HPV status in all TCGA HNSCC tumors.



Supplementary Figure S5. The 5-year overall survival differences between **A) *CASP8*** **B) *COL11A1*** wild-type and mutant HPV-negative TCGA HNSCC patients are not significant.



Supplementary Figure S6. Five-year overall survival comparison between H3K36M mutant and wild-type samples in the HPV-negative TCGA HNSCC cohort.

