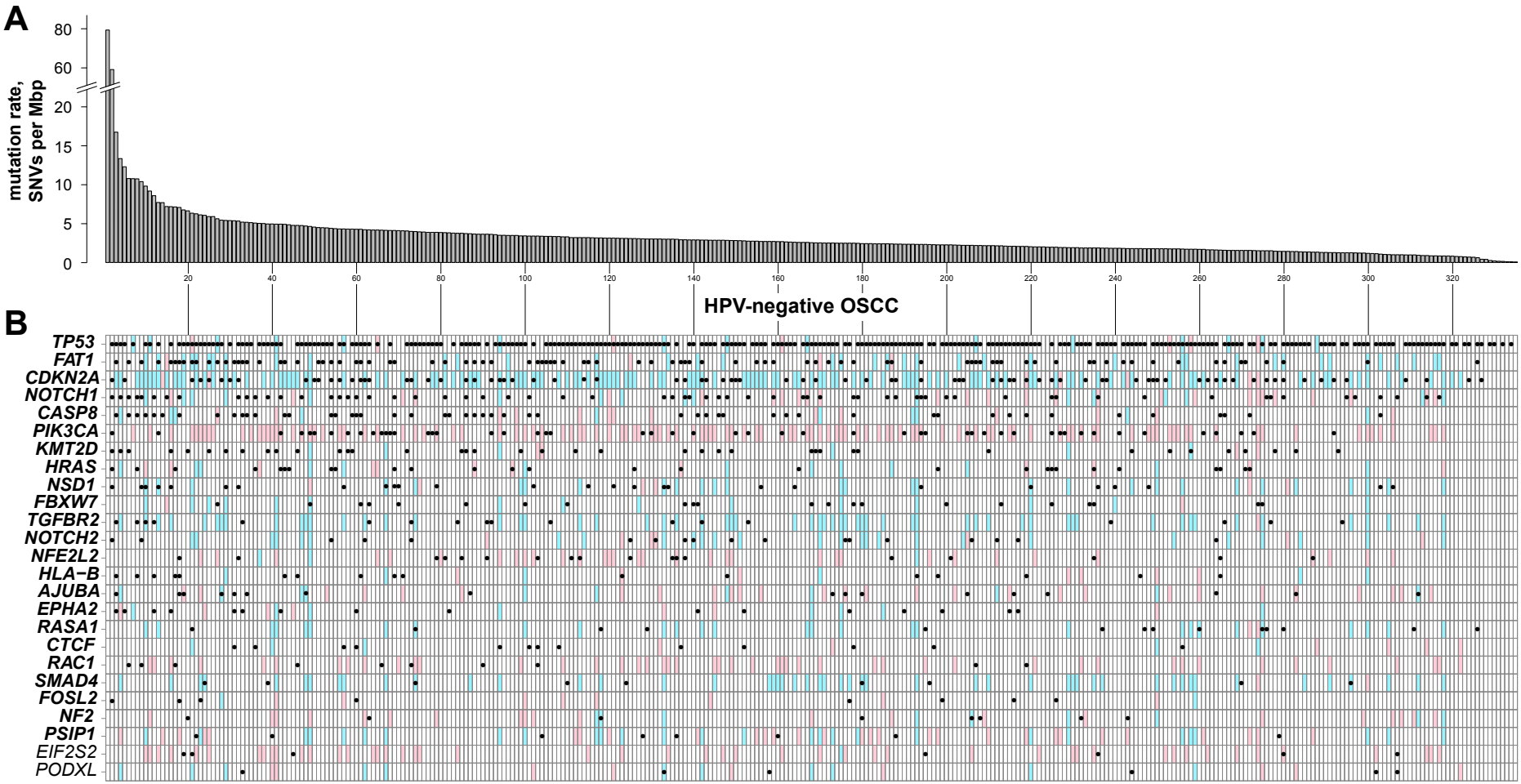
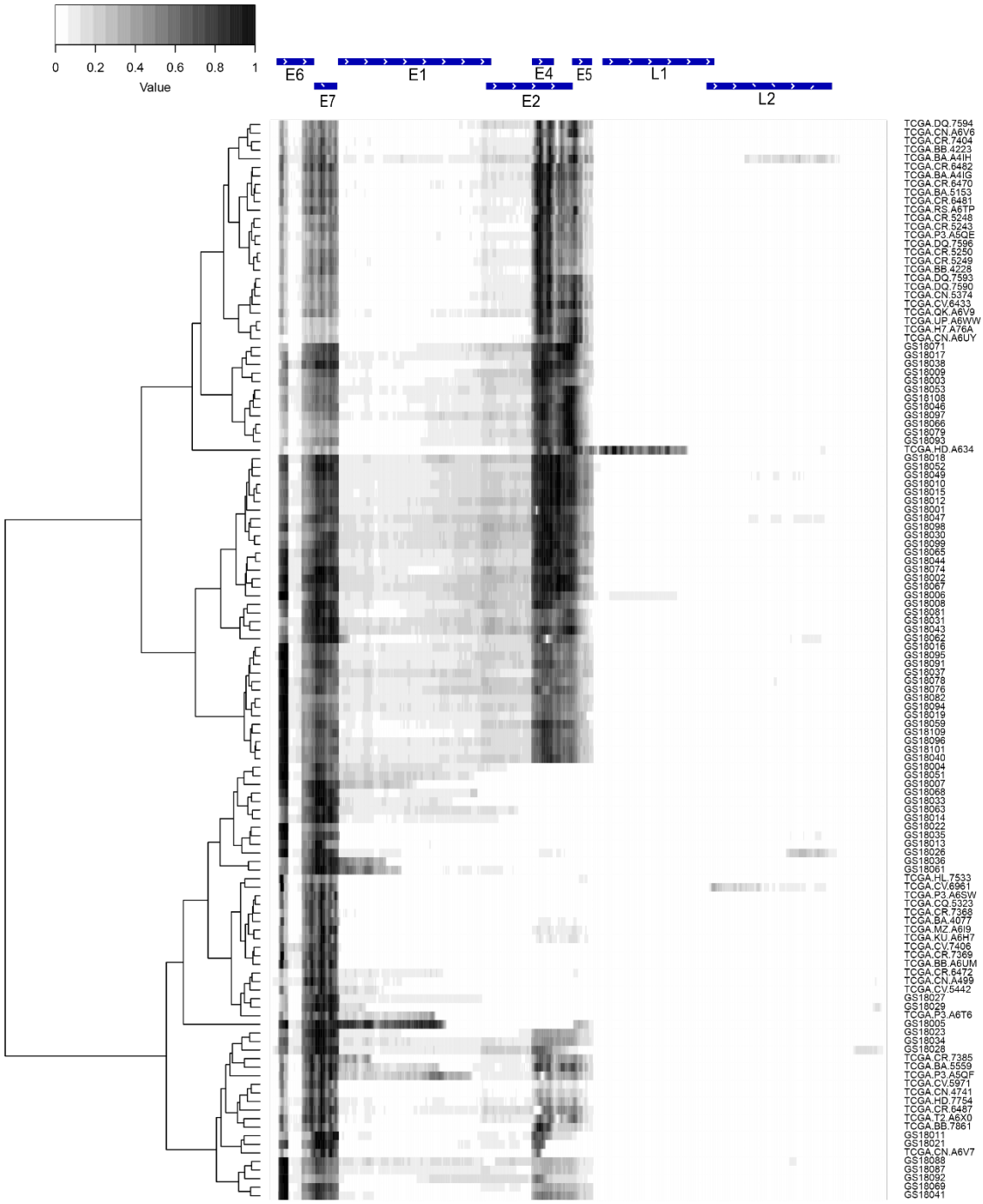
Gillison et al.

SUPPLEMENTAL FIGURES

TABLE OF CONTENTS

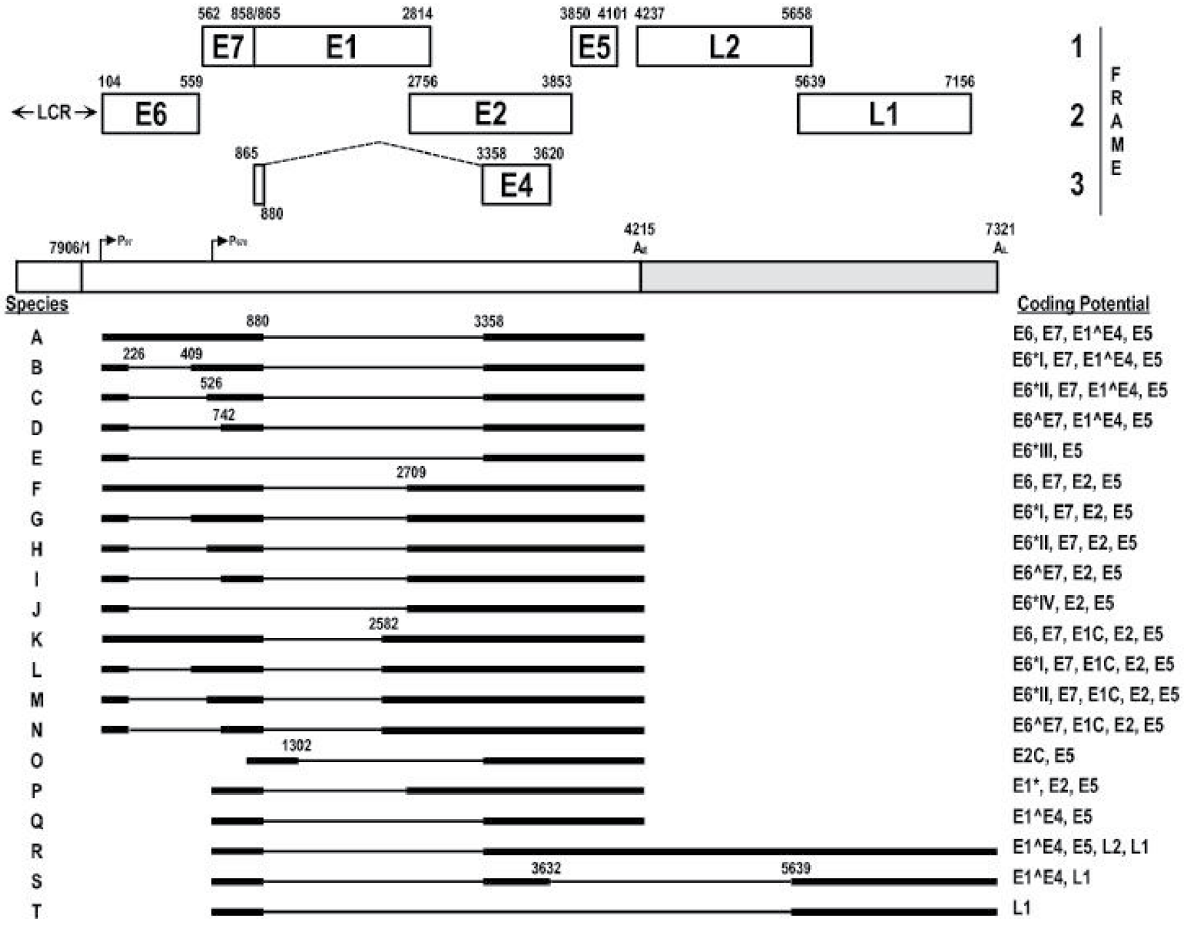


Supplemental Figure S1A. Somatic variants in 335 HPV-negative OSCC exomes. (A) The bar graph shows rates of somatic SNVs per Mbp per tumor in exomes of 335 HPV-negative OSCC. (B) SNVs and CNVs in 25 highly mutated genes in the cancers are shown. The order of samples (*x-axis*) was based on SNV rates (*y-axis, top*). Variants in 25 significantly mutated genes (*left*) were identified by MutSig (adjusted p-value < 0.2; (Lawrence, MS, et al., Nature 499: 214, 2013)). *Red*, copy number gain involving gene; *blue*, copy number loss; *black dots*, somatic variants including missense, nonsense, and splicing site SNVs, and frameshift or in-frame insertions or deletions (indels). *Bold text*, genes with mutations at q-value < 0.1. Copy number alterations were determined from 50 WGS data and 285 microarray-based CNV data (see also **Figure 1**; **Supplemental Table S3C**).



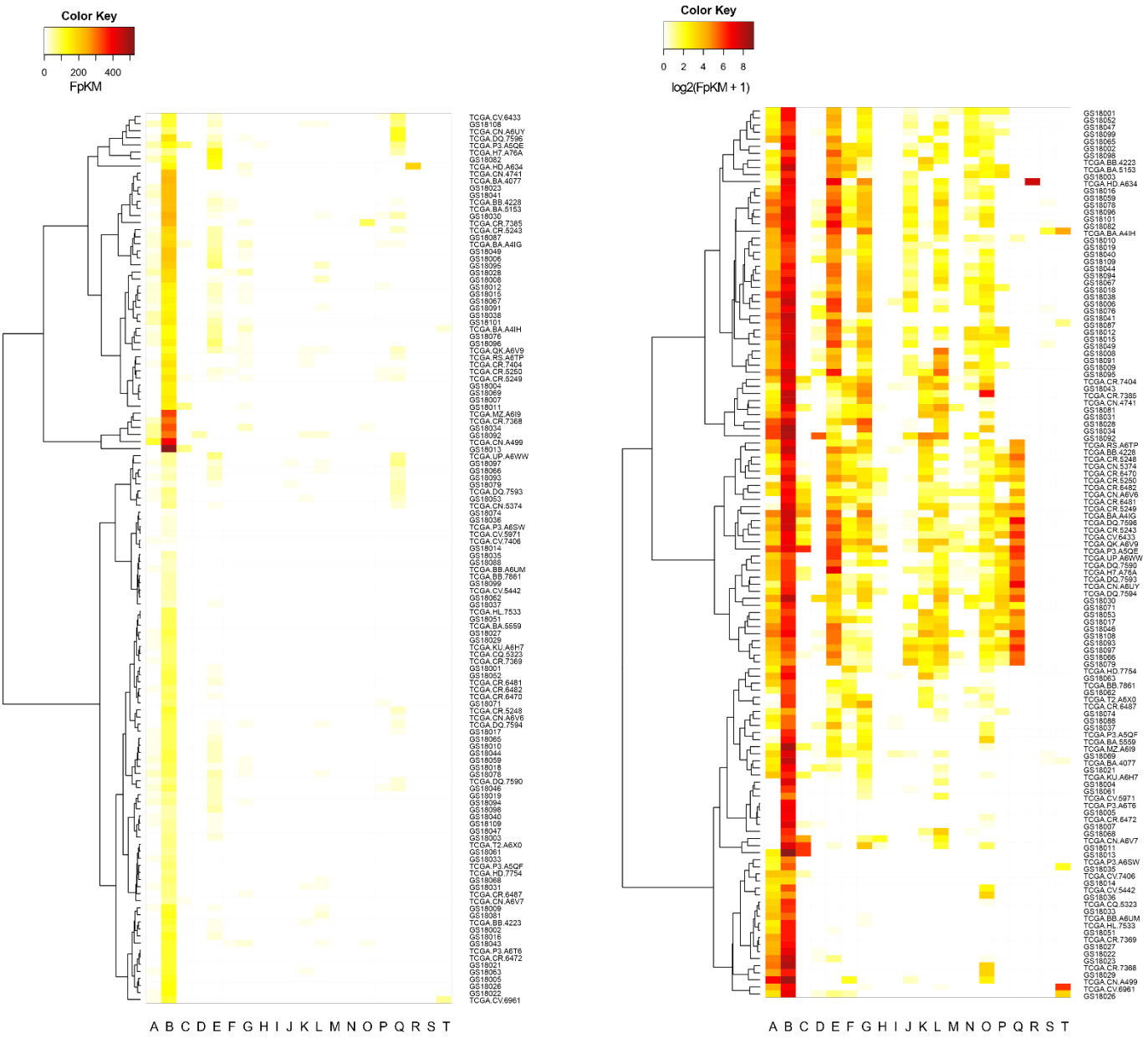
Supplemental Figure S1B. Viral transcripts expressed in HPV16-positive OSCC

Heatmap shows the normalized counts of RNA-seq reads expressed in 126 HPV16-positive tumors (*y-axis*) with available RNA-seq data (74 Ohio cohort and 52 TCGA), mapped to the linearized HPV16 genome (*x-axis*, *top schematic*). RNA-seq data were aligned and counted for each 10-bp bin of the virus genome. Intensities were normalized based on the maximum depth of sequencing coverage for each sample. *Gray scale*, the normalized depth of sequencing coverage ranging from 0 to 1. *Left,* hierarchical clustering of samples was performed using Ward D2 method (see also **Supplemental Table S1L**). Heatmaps were similar for non-16 types (data not shown).



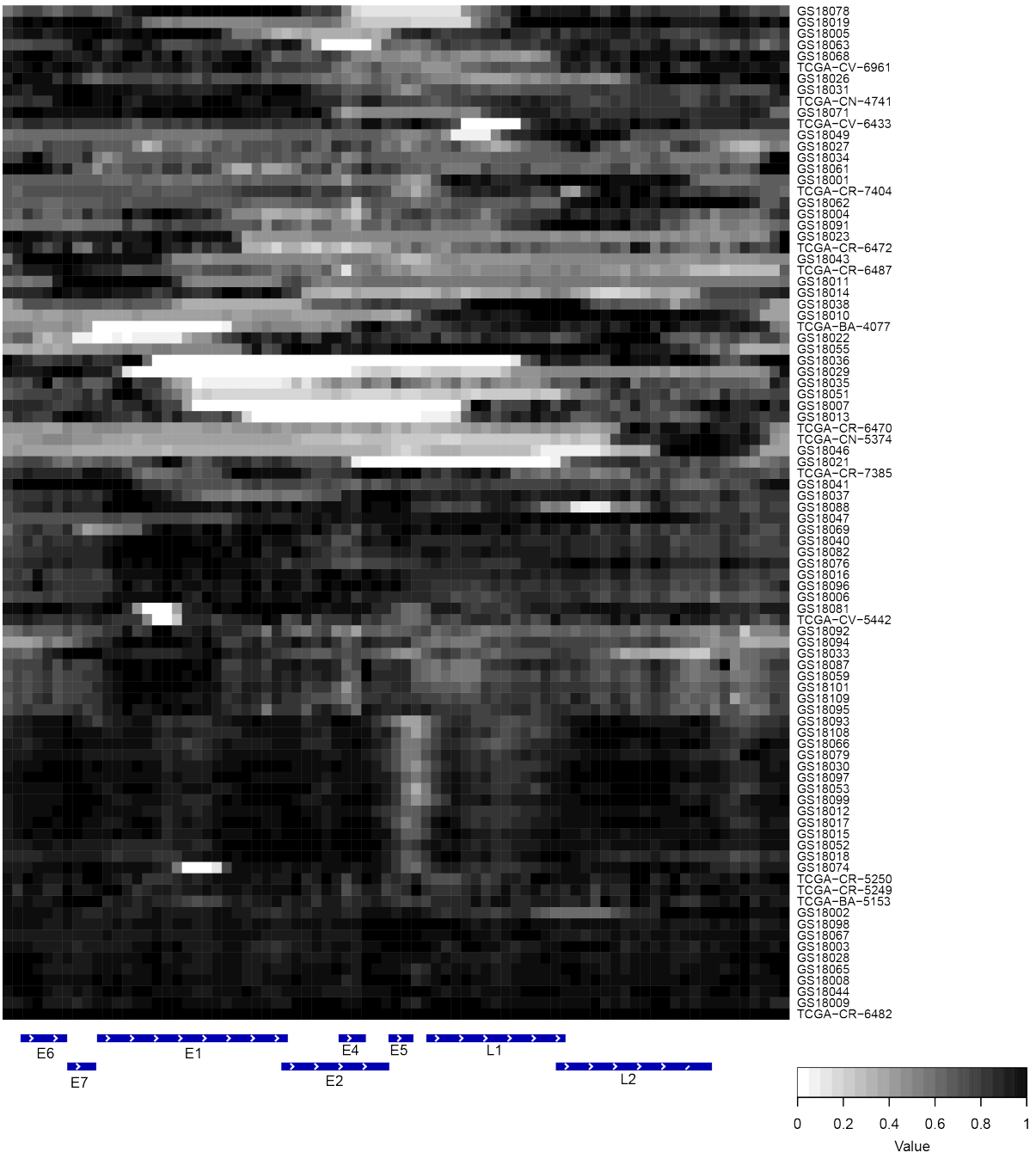
Supplemental Figure S1C. Schematic of HPV viral transcript splicing isoforms.

Reproduced from *Frontiers in Bioscience* **11**, 2286-2302, September 1, 2006.

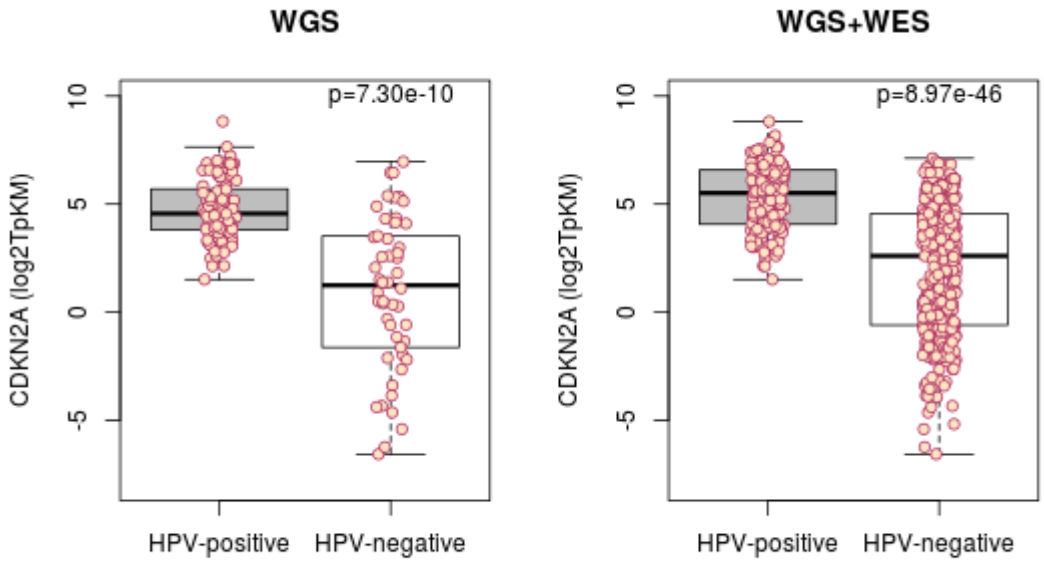


Supplemental Figure S1D. Viral splicing isoforms expressed in HPV16-positive OSCC.

Heatmaps show counts of HPV16 transcript isoforms identified in HPV16-positive OSCC RNA-seq data. *Left panel*, linear scale of transcript isoform expression (fragments per kilobase of transcript per million mapped reads, FPKM). *Right panel*, log2 transformation of transcript expression, shown as Log2 (FPKM + 1). *X-axis, bottom*, A-T, 20 HPV16 transcript structures, cf. Zheng et al. (*Frontiers in Bioscience* 11, 2286-2302, September 1, 2006), and **Supplemental Figure S1C**. *Top, color scale keys*, intensity of expression levels in FPKM for each isofom in *right,* 126 HPV16-positive OSCC. *Left,* hierarchical clustering was performed using Ward D2 method for samples (*IDs, right*).

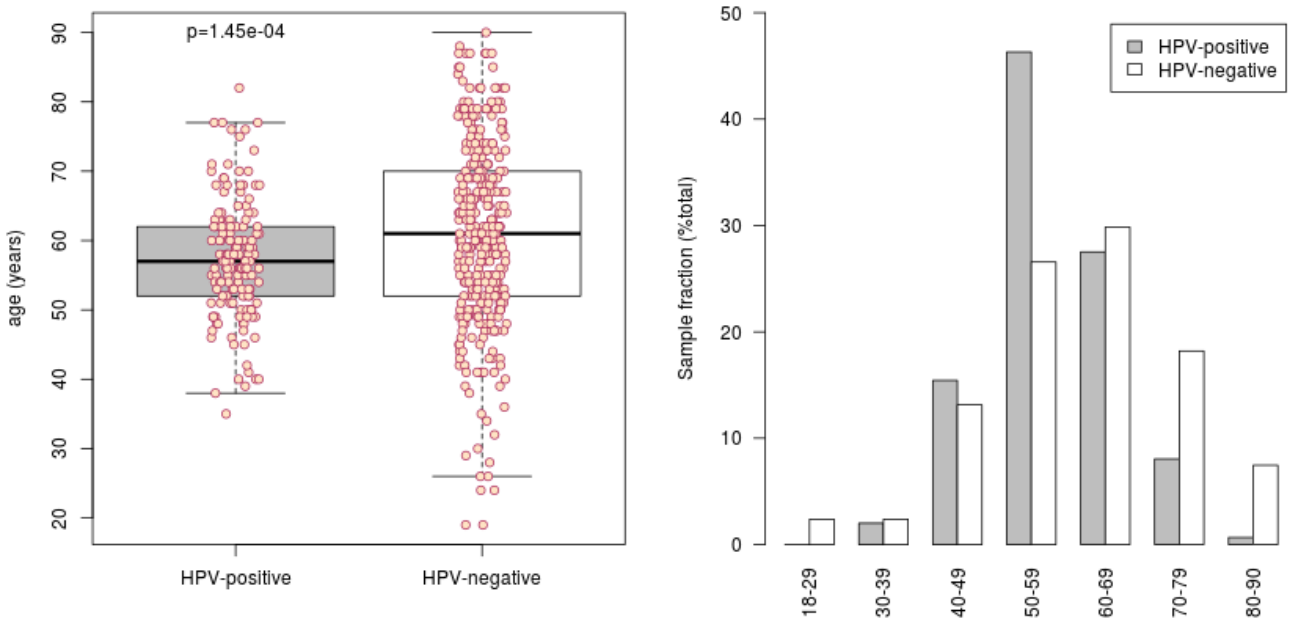


Supplemental Figure S1E. HPV genomic DNA detection in HPV16-positive OSCC. Heatmap shows the depth of coverage of WGS sequence reads mapping to the HPV genome for 90 HPV16-positive tumors (74 Ohio cohort and 16 TCGA). For display, the HPV genome (*bottom, x-axis*) was divided into 100 bp bins. The mean depth of coverage for each OSCC case (*right*, *y-axis*) was calculated for each bin. To normalize the data (*bottom right*, *heatmap grayscale*), the ratio of the depth of coverage in each bin divided by the maximum depth of coverage across bins in each sample was calculated. To reduce bias due to various nucleotide compositions across bin, the ratios for each bin were divided by the maximum ratio for each bin across all samples sequenced using data from the corresponding platform (CGI or Illumina for Ohio cohort samples or Illumina for TCGA samples). Hierarchical clustering of samples was performed. Heatmaps for non-16 types were similar (data not shown).

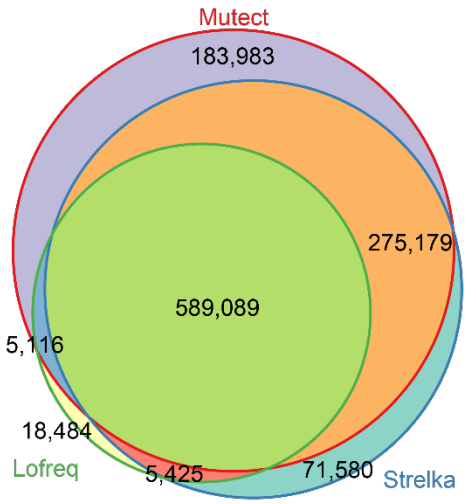


Supplemental Figure S1F. Comparison of *CDKN2A* transcript levels in HPV-positive vs. HPV-negative OSCC.

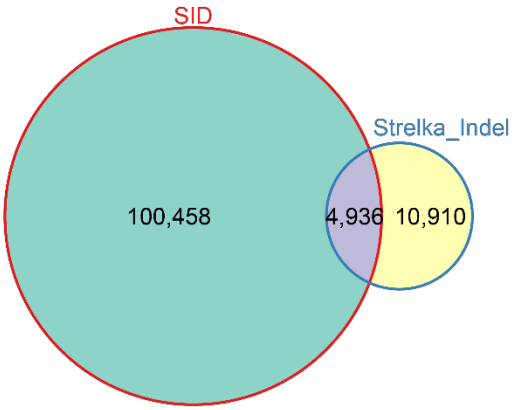
RNA-seq data were analyzed to compare *CDKN2A* transcript expression levels in HPV-positive vs. HPV-negative OSCC. *Left*, *CDKN2A* transcript levels in (*pink dots*) 101 HPV-positive tumors vs. 50 HPV-negative tumors with WGS data (i.e. 110 Ohio cohort and 41 TCGA samples; p = 7.30×10^{-10}, t-test). *Right, CDKN2A* transcript levels in tumors with WES or WGS data (i.e. 147 HPV-positive and 335 HPV-negative OSCC including 151 WGS and 331 WES samples; p = 8.97×10^{-46}, t-test). *Y-axis*, log2 transformed gene expression level (TPM, transcripts per million).



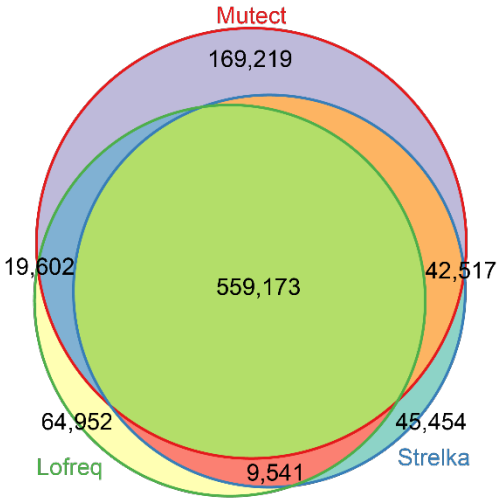
Supplemental Figure S1G. Age distribution of patients with HPV-positive vs. HPV-negative OSCC. (*Left*) Boxplot shows the age distributions of patients with HPV-positive (n = 149) vs. HPV-negative (n = 335) cancers. Mean age of HPV-positive patients is 57.3 years and of HPV-negative is 61.2 years ($p = 1.45 \times 10^{-4}$, t-test). (*Right*) Bar graph shows the fractions of all OSCC patients categorized in 10-year age increments. *X-axis*, age ranges; y-axis, fraction of patients; *gray fill*, HPV-positive; *white fill*, HPV-negative (see also **Supplemental Table S1A)**.



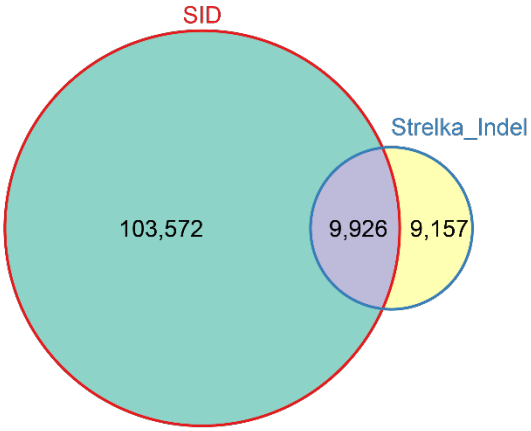
Ohio cohort - SNVs



Ohio cohort - Indels

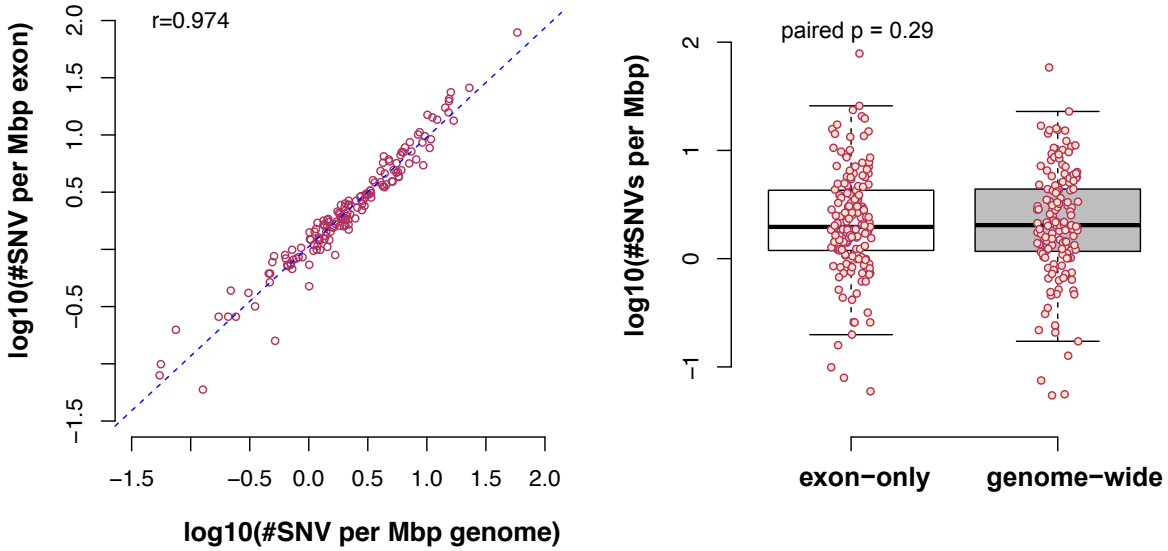


TCGA - SNVs



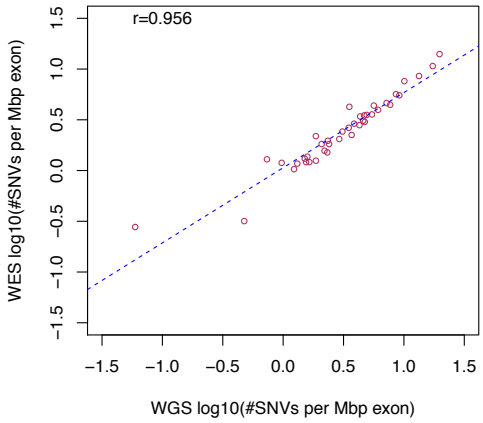
TCGA - Indels

Supplemental Figure S1H. Comparison of somatic variant detection using various bioinformatics tools.
Several variant detection software programs, including Mutect, Strelka, Lofreq, SomaticIndelDetector, and Strelka, were used to analyze WGS data from 53 tumors generated with Illumina platform for Ohio cohort samples (52 HPV-positive and 1 HPV-negative tumors), and 41 with Illumina platform for TCGA samples (17 HPV-positive and 24 HPV-negative tumors). Somatic variants (*left,* single nucleotide variants (SNVs) and *right*, insertion/deletion variants (indels)) as identified by the various packages were counted.



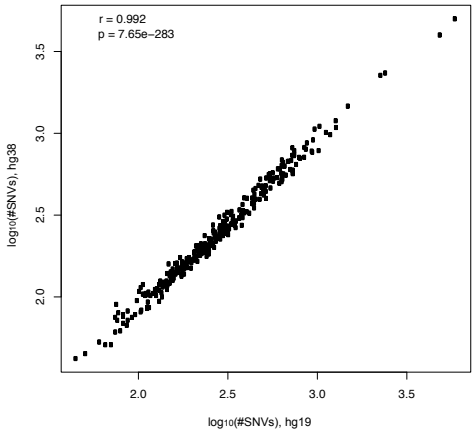
Supplemental Figure S1I. Comparison of SNV frequencies in WGS vs. "exonized" data from OSCC samples, and in WES vs. "exonized" WGS data.

Top (*Left*) Scatterplot of log-transformed, somatic SNV rates for (*y-axis*) exome extracted from WGS data vs. (*x-axis*) genome-wide, as measured for (*circles*) 153 individual WGS samples (i.e. 50 HPV-negative samples including 25 CGI and 25 Illumina WGS samples [24 TCGA and 1 Ohio cohort sample], and 103 HPV-positive samples including 34 CGI and 69 Illumina WGS [17 TCGA and 52 Ohio cohort samples). The exome was defined by as annotated by VEP v.79. Pearson's correlation coefficient was calculated (r = 0.974). (*Right*) Box-and-whisker plots of the same data (paired t-test, p=0.288).



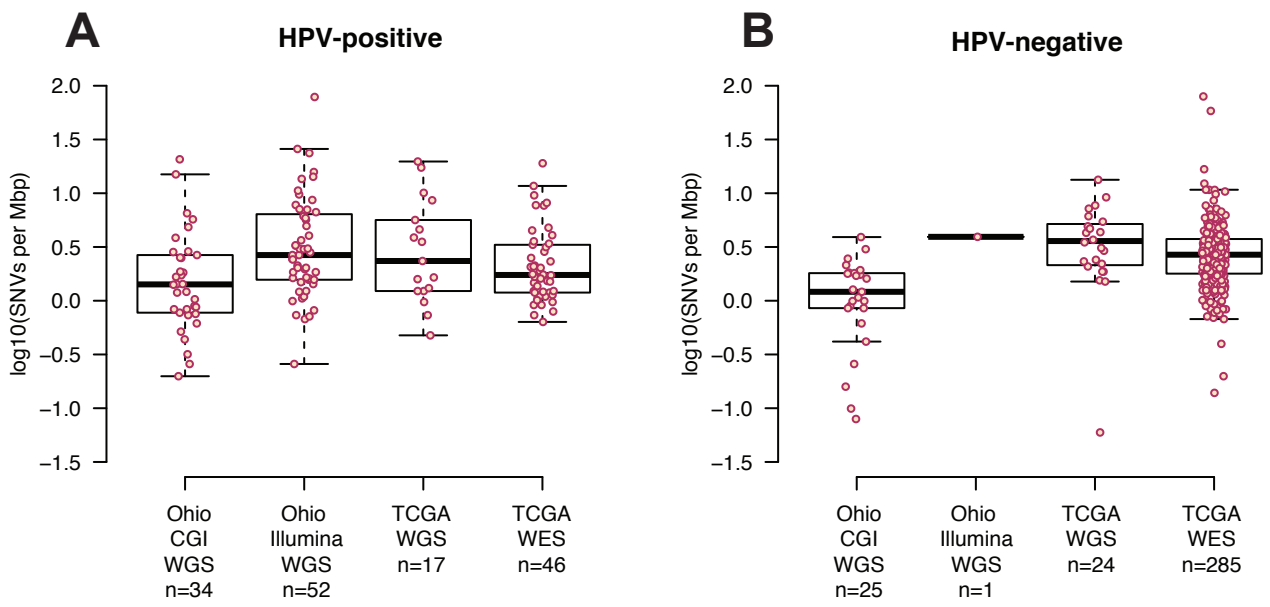
Supplemental Figure S1J. Comparison of SNV frequencies in WES v. "exonized" WGS data.

Scatterplot of log_{10}-transformed, somatic SNV rates for (*y-axis*) WES data vs. (*x-axis*) the exome extracted from WGS data from the same 40 OSCC samples, sequenced in the two ways by TCGA. *Circles,* individual data points from 16 HPV-positive and 24 HPV-negative OSCC. *X-axis* log_{10} transformed SNV rate in exome from WGS data; *y-axis*, log_{10} transformed SNV rate from WES data. The exome was defined by the target region from Agilent SureSelect Human All Exon v5 panel (**Fig. 1A**). Pearson's correlation coefficient was calculated (r = 0.956).



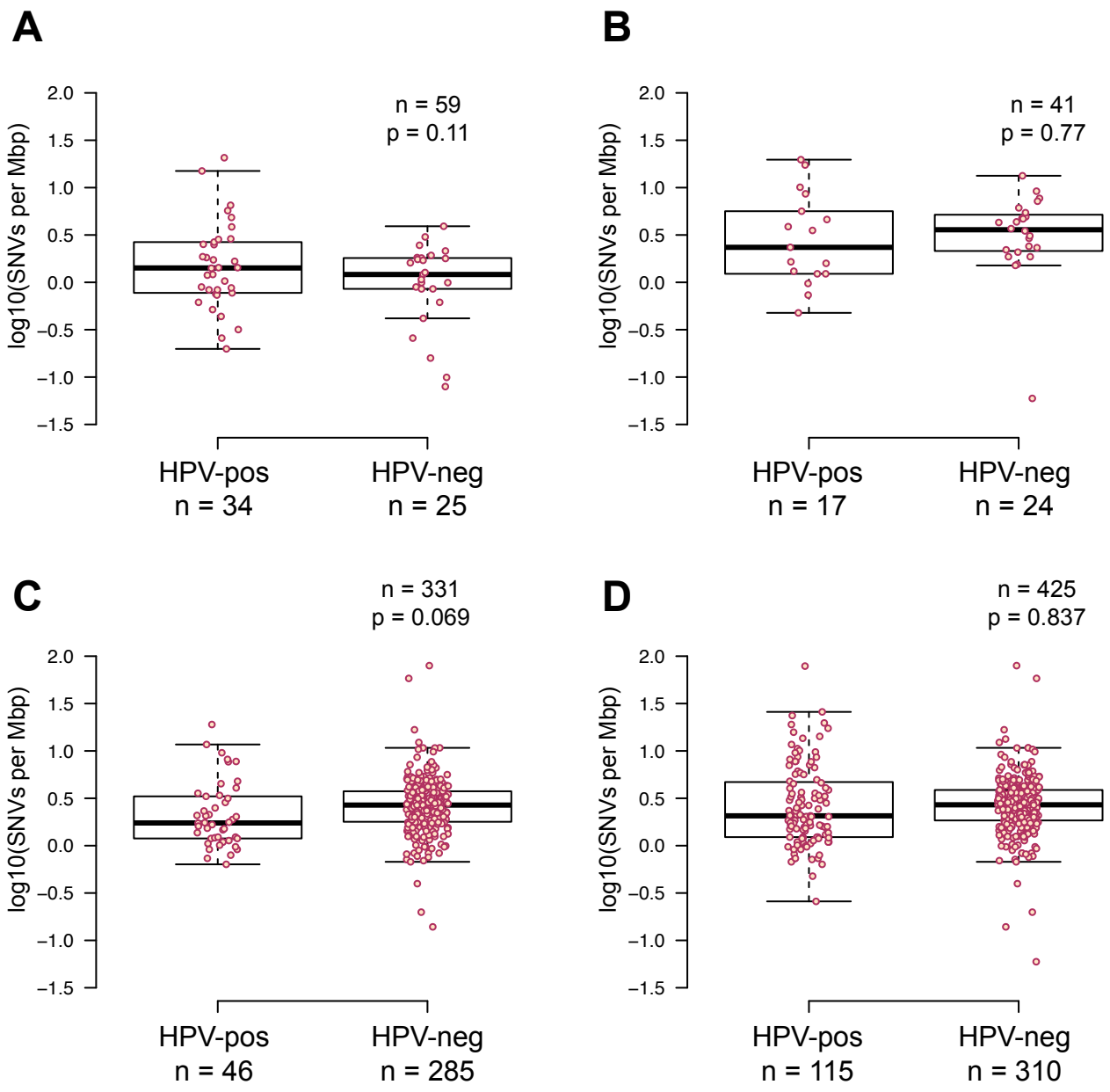
Supplemental Figure S1K. Comparison of SNV frequencies in OSCC WES data mapped to hg19 vs. hg38 reference genome assemblies.

Scatterplot of log_{10}-transformed, somatic SNV rates for WES data from HPV-positive or HPV-negative OSCC (n = 311; downloaded from TCGA) called by Mutect2 upon alignment to the (*y-axis*) hg38 human reference genome assembly vs. (*x-axis*) hg19 human reference genome assembly. *Symbols,* individual OSCC. Pearson's correlation coefficient was calculated (r = 0.992; p = 7.65×10^{-283}).



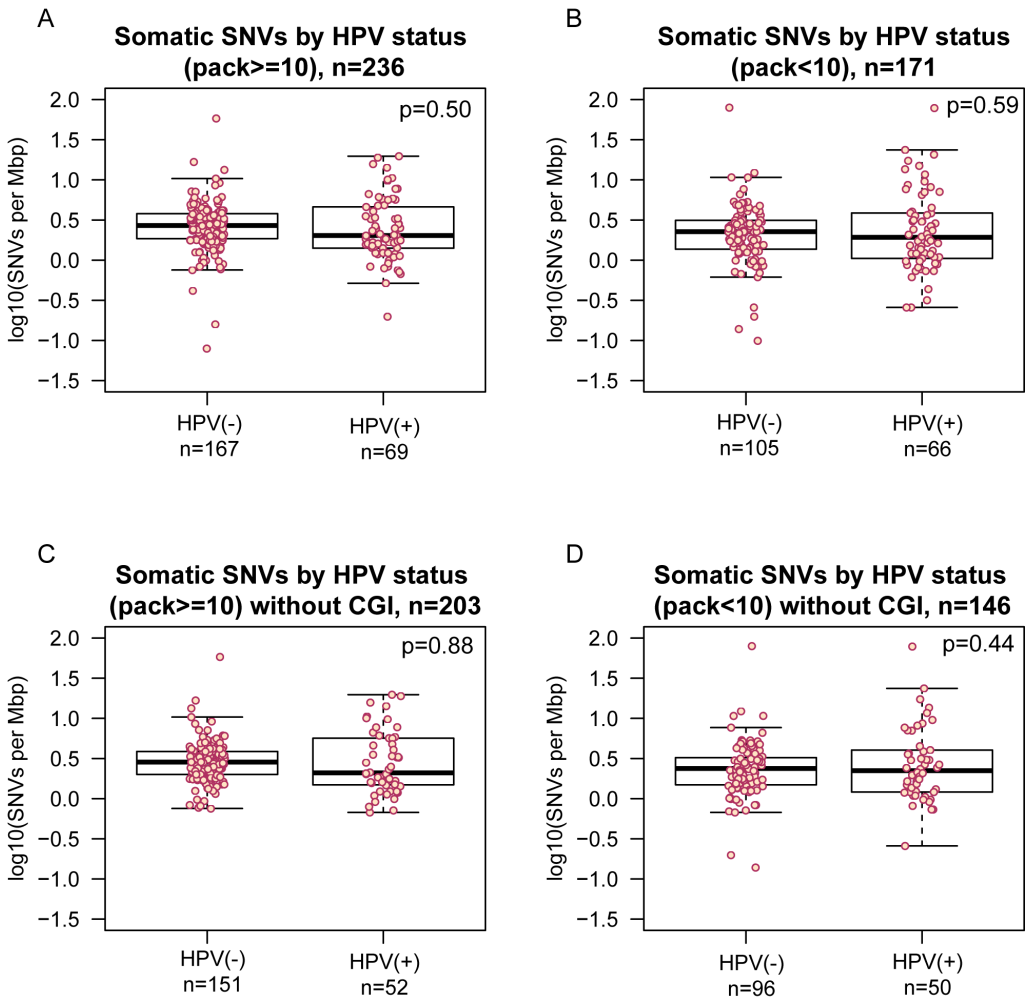
Supplemental Figure S1L. Somatic SNV rates by sequencing platform and HPV status.

Compared here in box and whisker plots are the $\log_{10}$ transformed somatic SNV rates for exonized WGS data generated at CGI or using Illumina platform for Ohio cohort or TCGA samples, and WES data from TCGA, for (A) HPV-positive and (B) HPV-negative OSCC. *Red dots*, individual OSCC samples. (A) HPV-positive tumors analyzed using various sequencing platforms (*bottom*). Log transformed SNV rates were compared between four platforms using ANOVA ($p = 0.00759$). Pairwise comparisons were performed using Tukey's test. SNV rates of Ohio Illumina WGS vs. Ohio CGI WGS platforms were significantly different ($p = 0.0055$). (B) HPV-negative tumors were analyzed using various sequencing platforms (*bottom*). Log-transformed SNV rates were compared between four platforms using ANOVA ($p = 1.1 \times 10^{-9}$). Rates of SNVs are significantly different between CGI and TCGA-WGS ($p = 2.0 \times 10^{-7}$), and between CGI and TCGA-WES ($p < 1 \times 10^{-7}$).



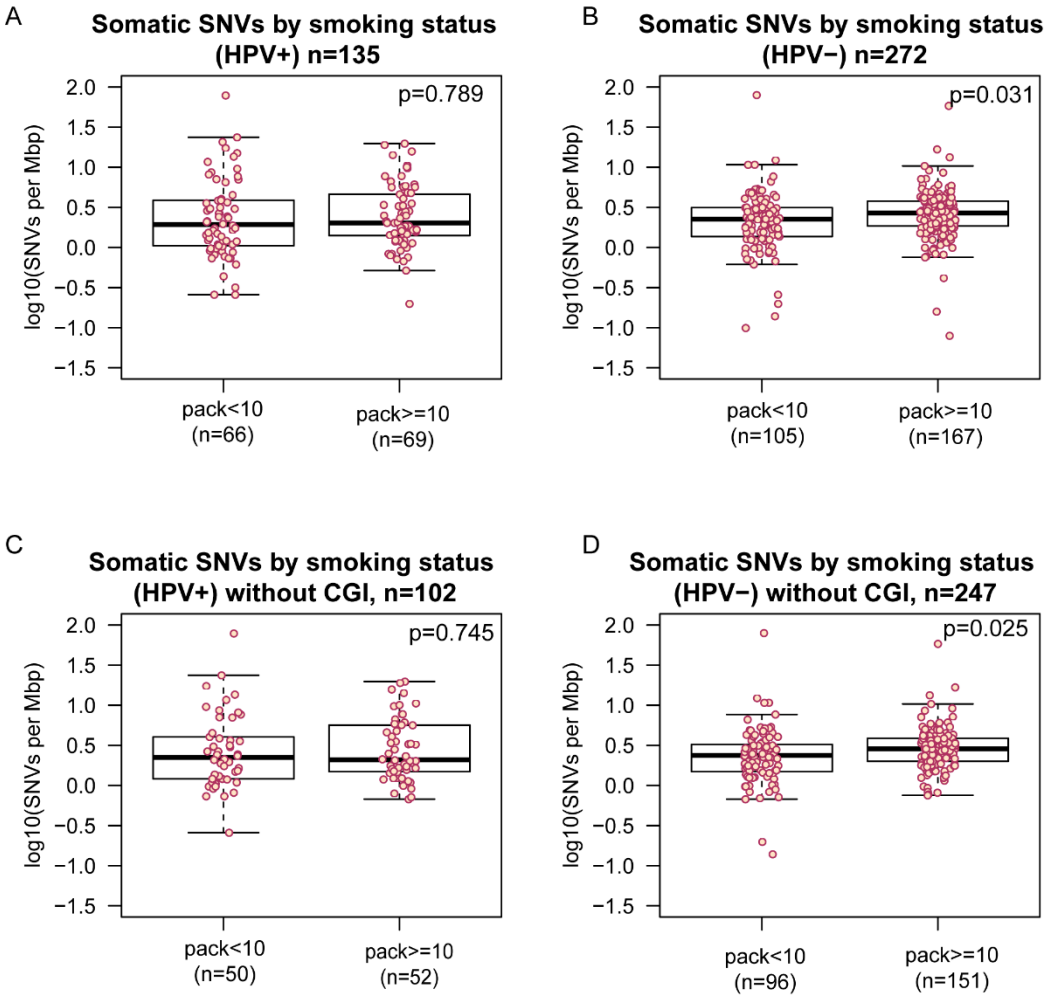
Supplemental Figure S1M. Comparison of somatic SNV rates by HPV status using various sequencing platforms

Shown here in box and whisker plots are the $\log_{10}$-transformed somatic SNV rates for HPV-positive vs. HPV-negative OSCC, using exonized WGS data (A) generated at CGI; and (B) using Illumina platform at TCGA; and (C) WES data from TCGA. (D) Data analyzed using the WES and Illumina WGS platforms were combined (i.e. including Ohio cohort-WGS, TCGA-WGS and TCGA-WES data, but excluding CGI WGS data). Statistical significance of comparisons was assessed using T statistics or ANOVA (see also **Supplemental Table S1M-N**).



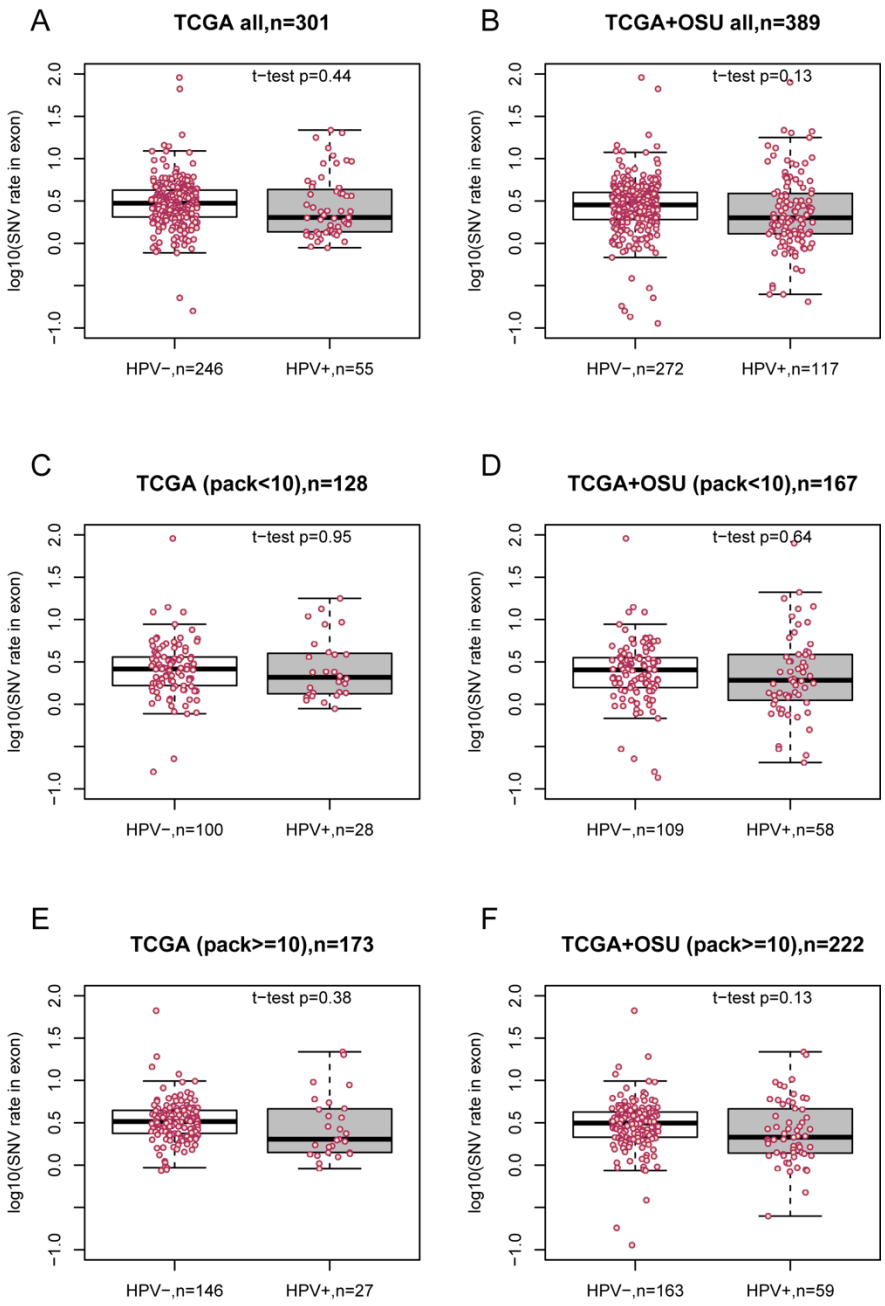
Supplemental Figure S1N. Comparison of somatic SNV rates by HPV status for various levels of cigarette smoking.

Shown here in box and whisker plots are the $\log_{10}$-transformed somatic SNV rates in exomes of HPV-positive vs. HPV-negative OSCC, displayed for various levels of cigarette smoking including (A) ≥10 pack-years; (B) < 10 pack-years; (C) ≥10 pack-years excluding CGI WGS data; and (D) < 10 pack-years excluding CGI WGS data. Statistical significance of comparisons was assessed using the t-statistic.



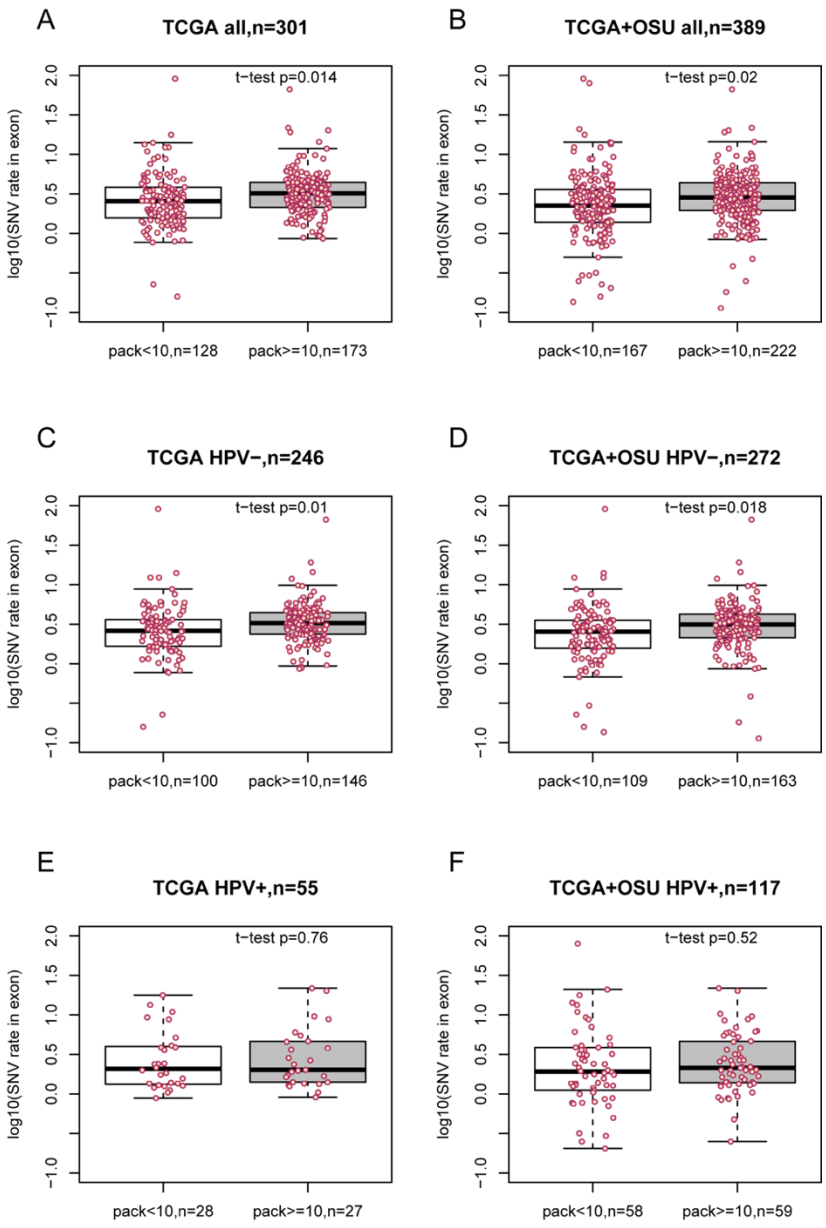
Supplemental Figure S1O. Comparison of somatic SNV rates by cigarette smoking for HPV-positive vs. HPV-negative OSCC with various sequencing platforms.

Shown here in box and whisker plots are the $\log_{10}$-transformed somatic SNV rates in exomes of OSCC patients with a cigarette smoking history of ≥10 pack-years, compared with <10 pack-years, for (A) HPV-positive OSCC; (B) HPV-negative OSCC; (C) HPV-positive OSCC excluding CGI WGS data; and (D) HPV-negative OSCC excluding WGS data. Statistical significance of comparisons was assessed using the t-statistic.



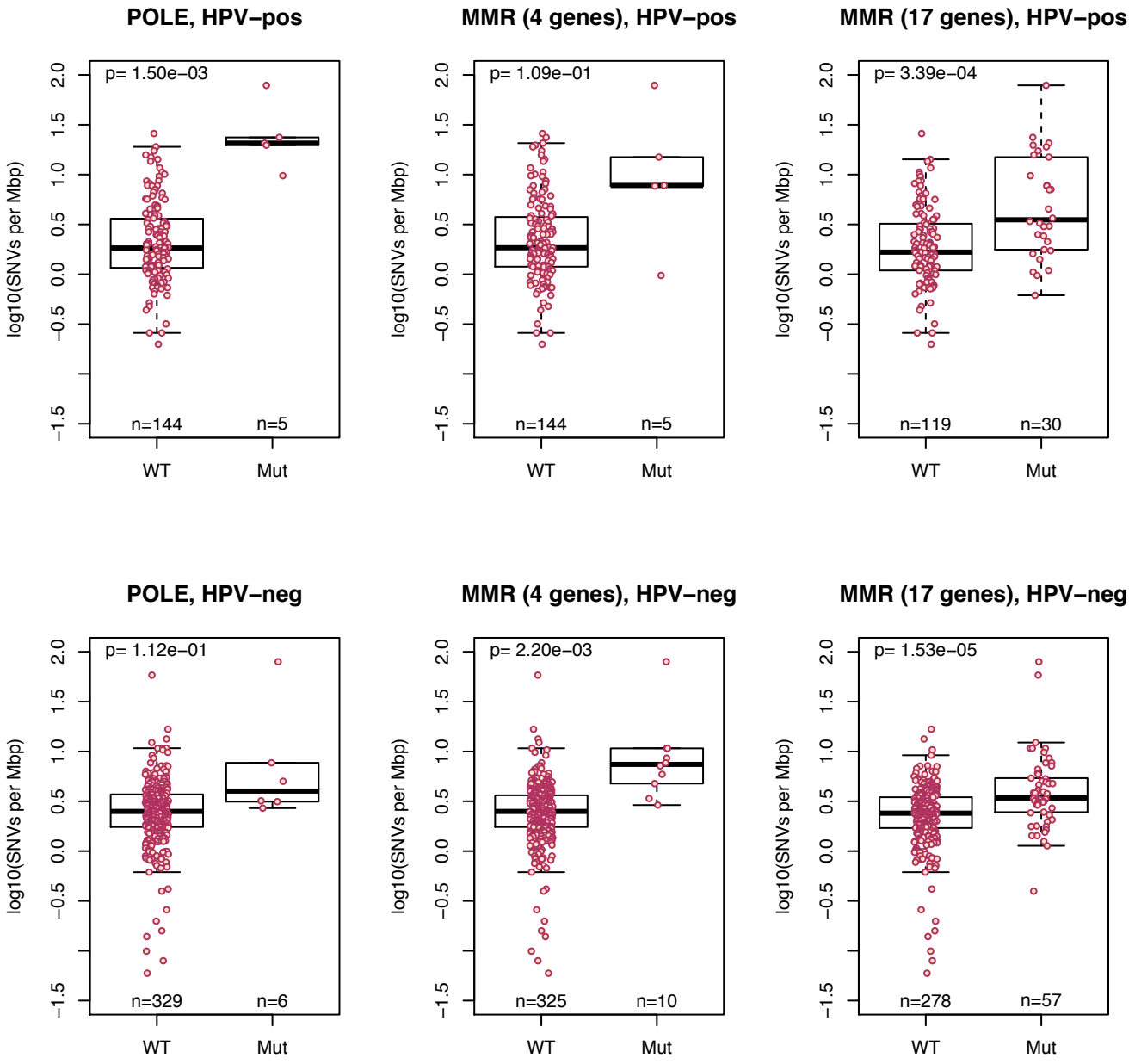
Supplemental Figure S1P. Comparison of somatic SNV rates by HPV status for various sequencing platforms and cigarette smoking.

Shown here in box and whisker plots are the $\log_{10}$-transformed somatic SNV rates in exomes of OSCC patients with HPV-positive vs. HPV-negative cancers, for (A) 301 cancers, regardless of smoking history, analyzed by WGS or WES at TCGA; (B) 389 cancers, regardless of smoking status, analyzed by WGS or WES at TCGA or in Ohio cohort; (C) 128 cancers analyzed by WGS or WES at TCGA with <10 pack-years cigarette smoking history in TCGA; (D) 167 cancers with <10 pack-years cigarette smoking history analyzed by WGS or WES at TCGA or in Ohio cohort; (E) 173 cancers with ≥10 pack-years smoking history in TCGA; and (F) 222 cancers with ≥10 pack-years smoking history at TCGA or in Ohio cohort. Statistical significance of comparisons was assessed using the t-statistic.



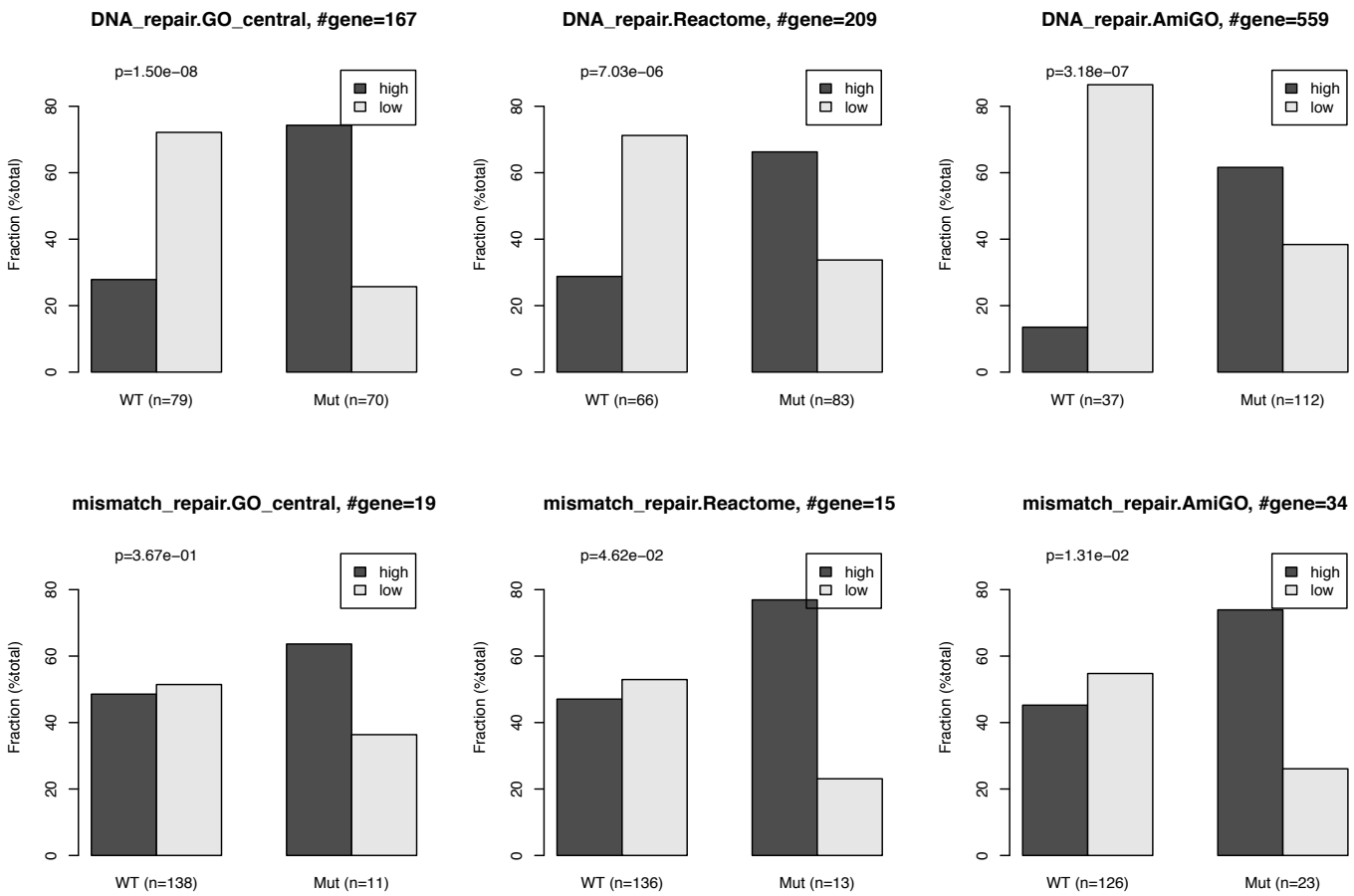
Supplemental Figure S1Q. Comparison of somatic SNV rates by HPV status for various sequencing platforms and cigarette smoking.

Shown here in box and whisker plots are the $\log_{10}$-transformed somatic SNV rates in exomes of 389 OSCC patients with available smoking history of <10 vs ≥pack-years of cigarette smoking, (A) 301 cancers studied at TCGA; (B) 389 cancers studied at TCGA or in the Ohio cohort; (C) HPV-negative analyzed by WGS or WES at TCGA; (D) HPV-negative analyzed by WGS or WES at TCGA or in Ohio cohort; (E) HPV-positive analyzed by WGS or WES at TCGA; and (F) HPV-positive analyzed by WGS or WES at TCGA or in Ohio cohort. Statistical significance of comparisons was assessed using the t-statistic.



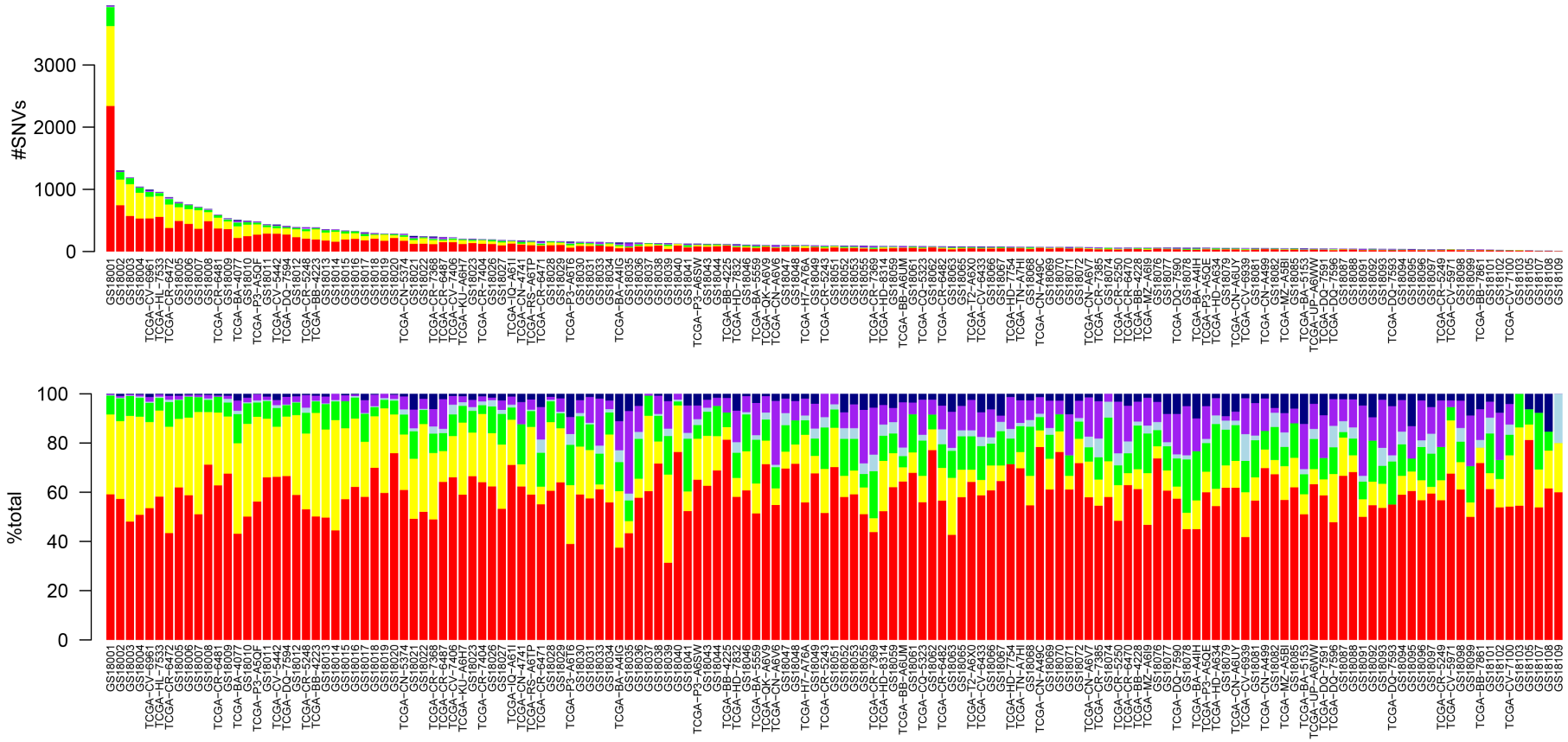
Supplemental Figure S1R. Comparison of somatic SNV rates by status of mutations in *POLE* or mismatch repair genes.

Box and whisker plots display somatic SNV rates for OSCC with and without mutations in *POLE* and with and without mutations in mismatch repair genes. For 149 HPV-positive cancers, $\log_{10}$-transformed SNV rates in exomes of cancers with (*top left*) wildtype (WT) or mutant (mut) *POLE*; (*top middle*) 4 WT or mut mismatch repair (MMR) genes, defined here as *MLH1, PMS2, MSH2, and MSH6*; or (*top right*) 17 WT or mut MMR genes defined here as *APC*, *BMPR1A*, *EPCAM*, *FBXO11*, *GALNT12*, *MLH1*, *MLH3*, *MSH2*, *MSH3*, *MSH6*, *MUTYH*, *PMS2*, *POLD1*, *POLE*, *PTEN*, *SMAD4*, *STK11.* For 335 HPV-negative OSCC tumors, similar plots show SNV rates in (*bottom left*) WT or mut *POLE*; (*bottom middle*) 4 WT or mut MMR genes; and (*bottom right*) 17 WT or mut MMR genes. Statistical significance of comparisons was assessed using t statistics.



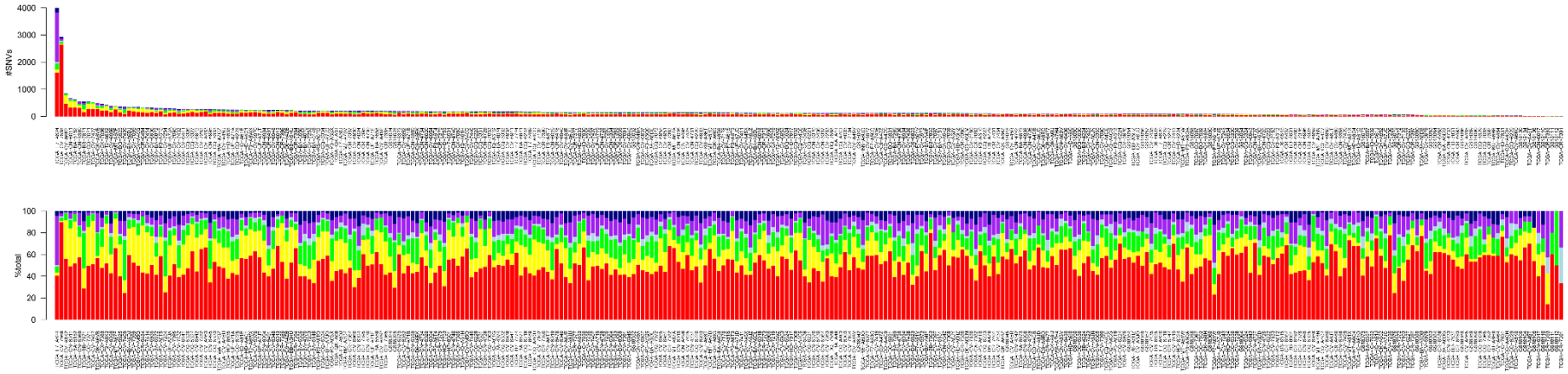
Supplemental Figure S1S. Fraction of tumors with deleterious mutations in DNA-repair and mismatch repair genes segregating into groups with overall mutation rates above and below the median.

Bar graphs display (y-axis) fractions of 149 HPV-positive OSCC samples harboring (*key, dark gray*, high [≥ the median mutation rate]; and *light gray*, low levels [<median] of SNVs) with and without mutations in DNA repair genes or in mismatch repair genes (*x-axis*, *counts of affected samples indicated*), as categorized by GO central, Reactome or AmiGO ontology databases (*top of each panel*, labels and counts of genes per category). *P-values*, statistical significance of comparisons was assessed using Fisher's exact test, without multiple testing correction.

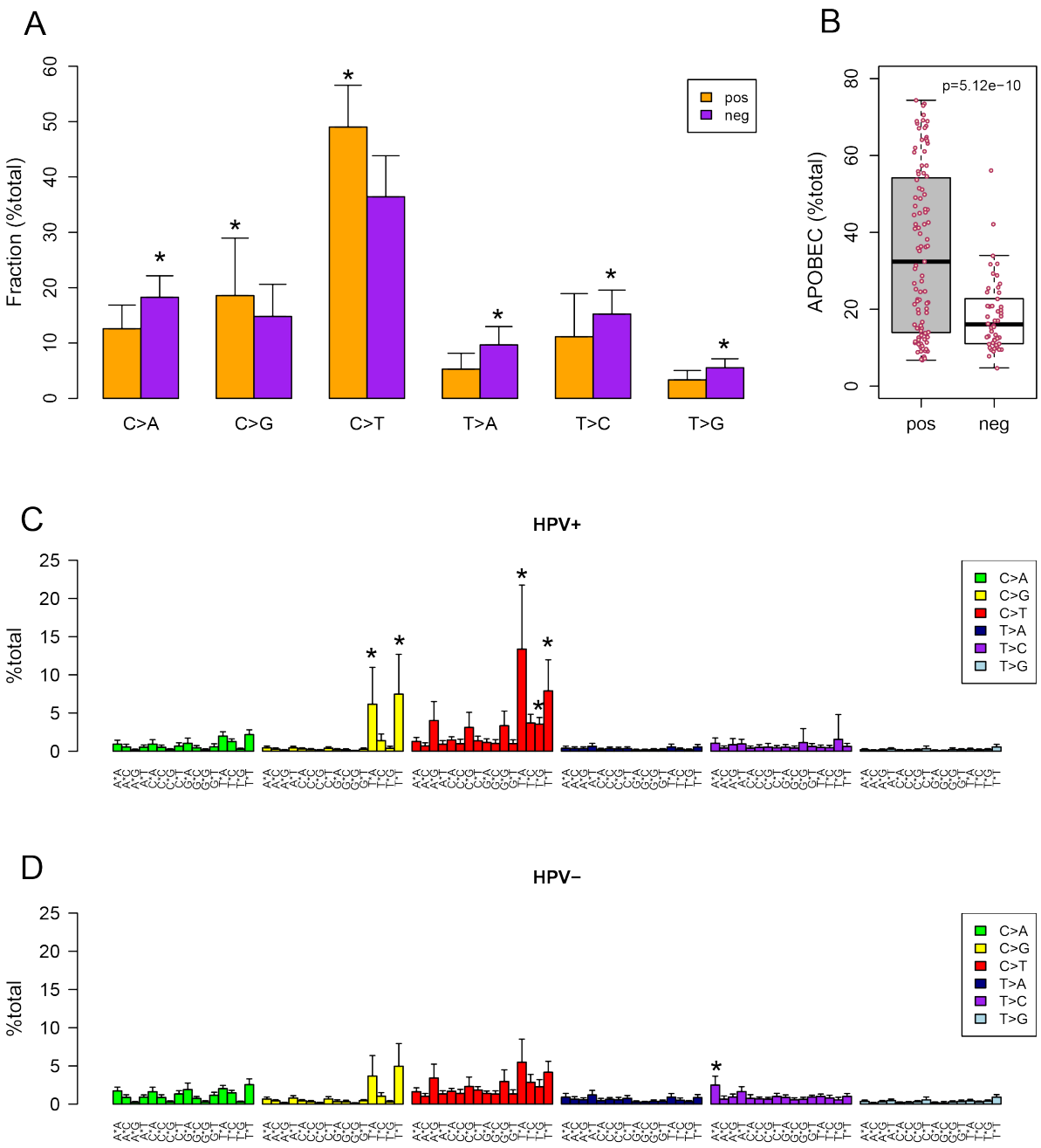


Supplemental Figure S2A. Somatic nucleotide substitutions in 149 HPV-positive OSCC exomes.

Bar graphs show (*top*) total counts of somatic SNVs per tumor and (*bottom*) fractions of nucleotide substitutions in exomes of 149 HPV-positive OSCC. Samples (*x-axis*) were ordered by the number of variants (*y-axis, top*). *Colors*: counts or fractions of variants: C>T, red; C>G, yellow; C>A, green; T>G, light blue; T>C, purple; T>A, navy blue (see also **Supplemental Tables S2A-B**).

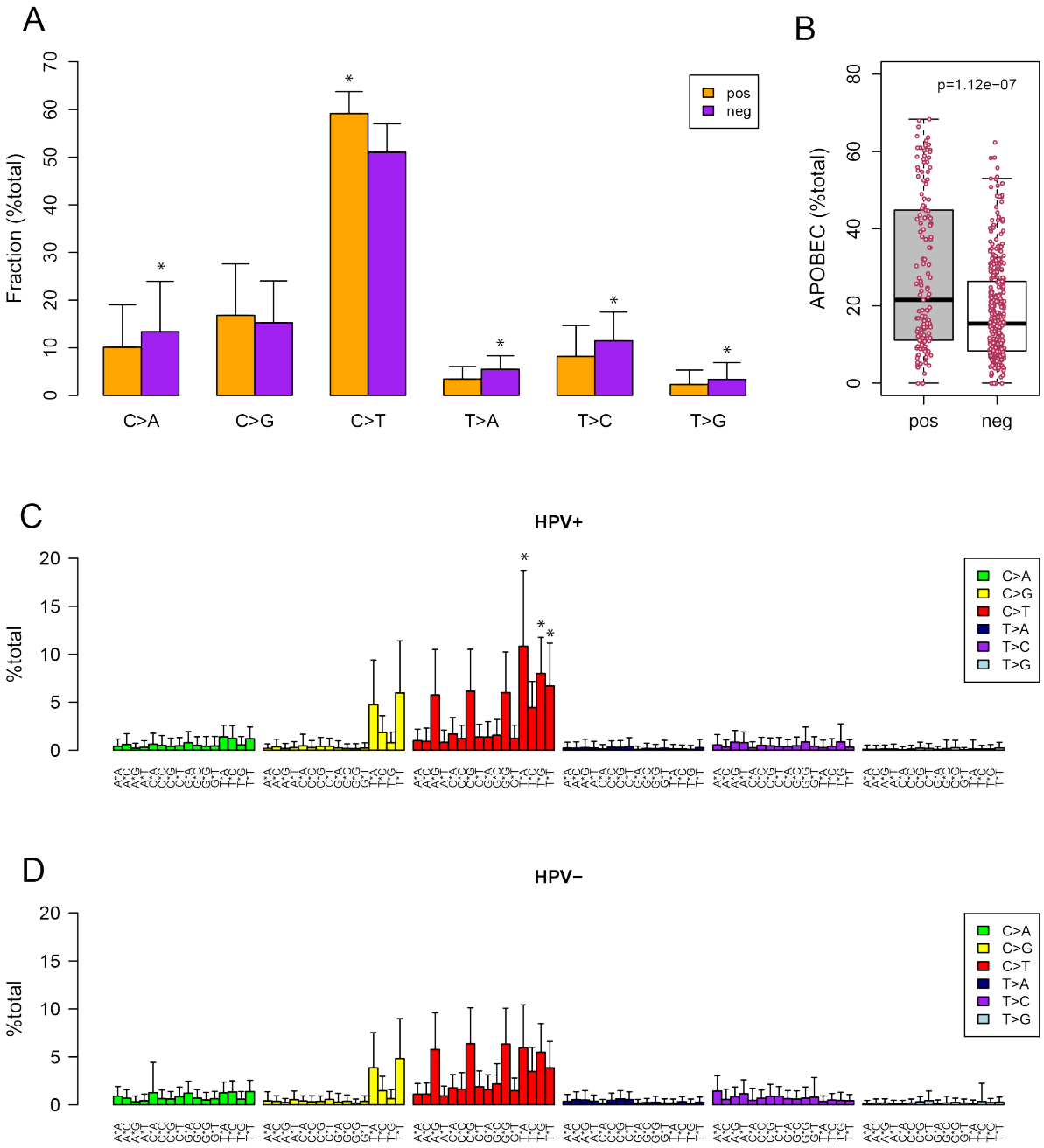


Supplemental Figure S2B. Somatic nucleotide substitutions in 335 HPV-negative OSCC exomes.
Bar graphs show (*top*) total counts of somatic SNVs per tumor and (*bottom*) fractions of nucleotide substitutions in exomes of 335 HPV-negative OSCC. Samples (*x-axis*) were ordered by the number of variants (*y-axis, top*). *Colors*: counts or fractions of variants: C>T, red; C>G, yellow; C>A, green; T>G, light blue; T>C, purple; T>A, navy blue. Samples are ordered based on rates of SNVs (see also **Supplemental Table S2A-B**).



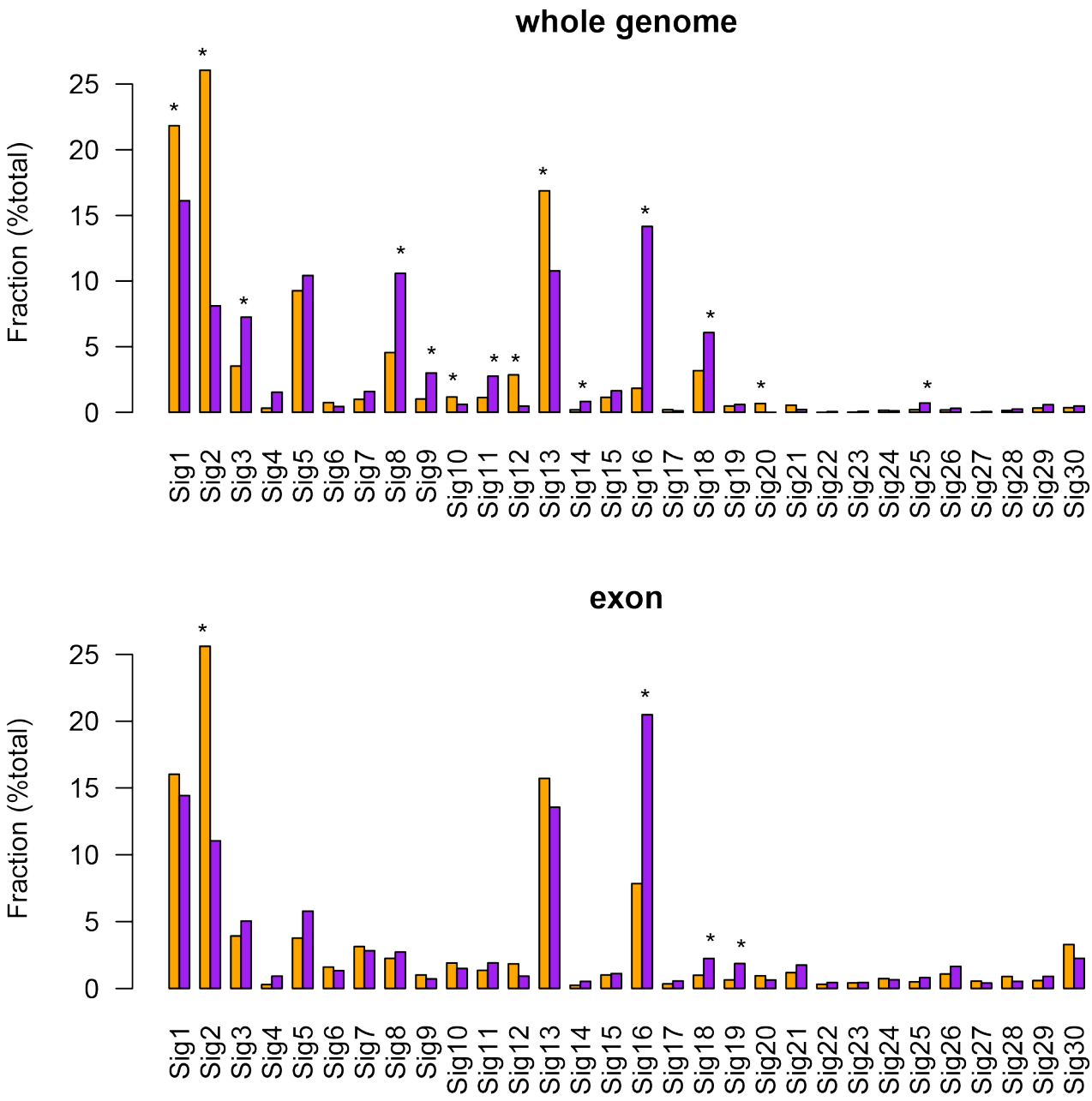
Supplemental Figure S2C. Nucleotide substitutions in WGS of HPV-positive vs. HPV-negative OSCC.

(**A**) Bar graph shows mean fractions per cancer of the 6 possible somatic single nucleotide substitutions in WGS data of *orange*, 103 HPV-positive OSCC; *purple*, 50 HPV-negative OSCC. *Error bars*, standard deviation; *asterisk*, significant difference between HPV-positive and HPV-negative samples by t-test ($p<0.01$). (**B**) Box and whiskers plot shows fraction of APOBEC signature mutation (C>T or C>G mutation in TCW trinucleotide context) for each sample ($p = 5.12 \times 10^{-10}$, t-test). (**C, D**) Bar graphs show mean fraction of SNVs per tumor (*y-axis*) in 96 possible trinucleotide contexts (*x-axis*) in (**C**) HPV-positive and (**D**) HPV-negative tumors. *Colors, key at right*, SNV at central nucleotide of trinucleotide sequences; *asterisk*, significant increase in HPV-positive samples compared to HPV-negative samples by t-test ($p<0.01$; change in mean fraction >1%). *Error bars*, standard deviation (see also **Supplemental Table S2A**).

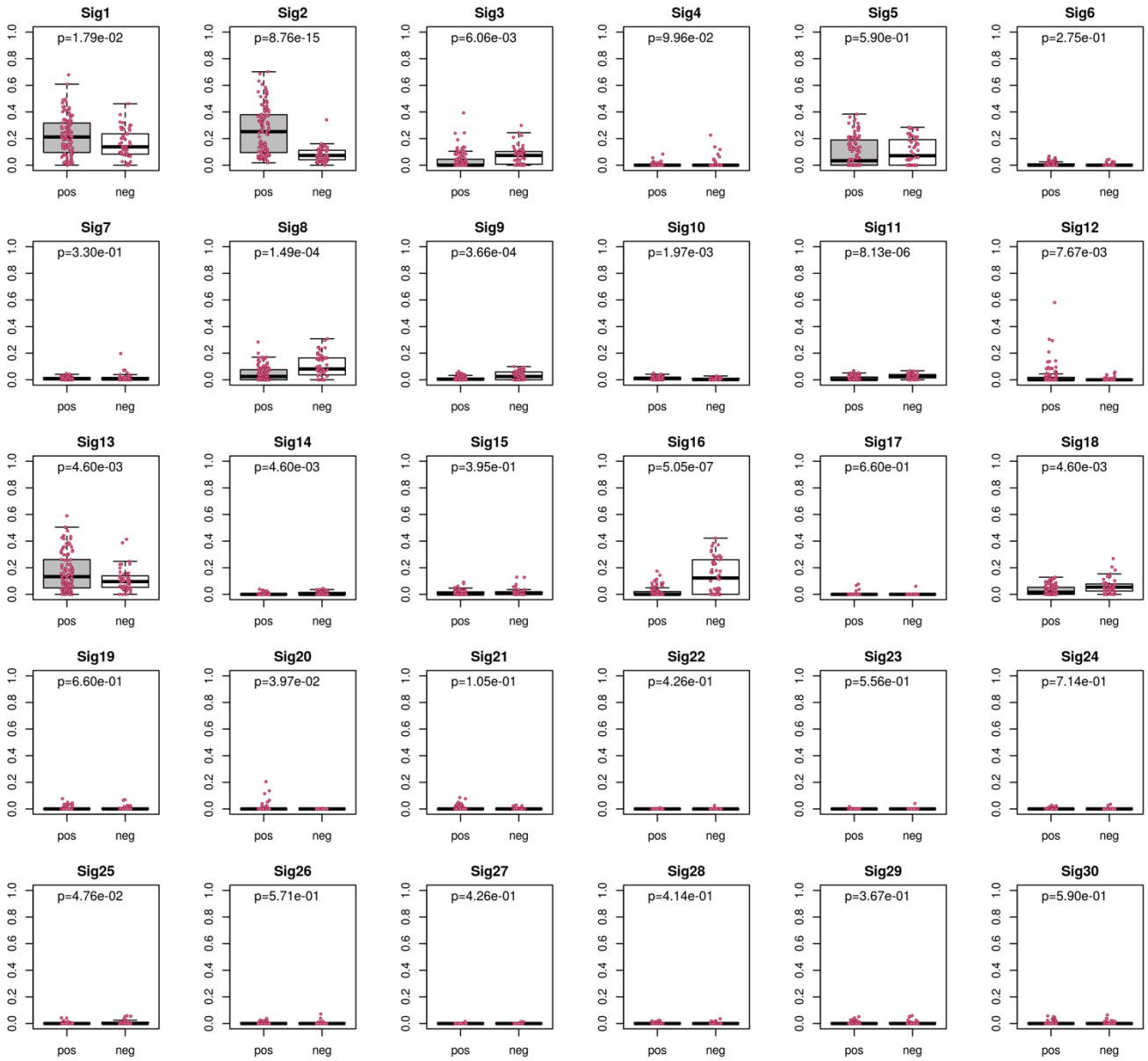


Supplemental Figure S2D. Nucleotide substitutions in exomes of HPV-positive vs. HPV-negative OSCC.

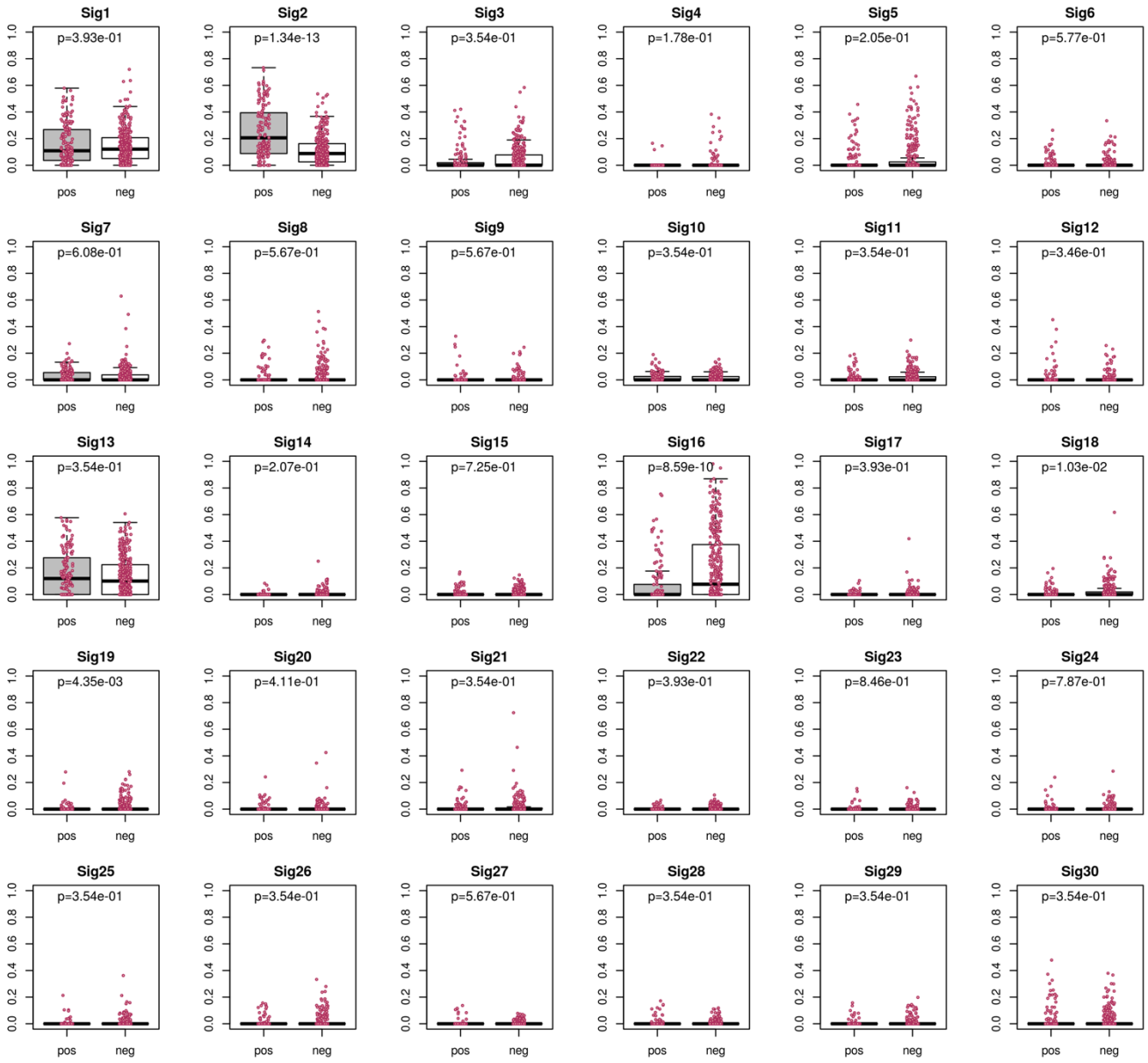
(**A**) Bar graph shows mean fractions per cancer of the 6 possible somatic single nucleotide substitutions in exomes of *orange*, 149 HPV-positive OSCC; *purple*, 335 HPV-negative OSCC. *Error bars*, standard deviation; *asterisk*, significant difference between HPV-positive and HPV-negative samples by T-test ($p<0.01$). (**B**) Box and whiskers plot shows fraction of APOBEC signature mutation (C>T or C>G mutation in TCW trinucleotide context) for each sample ($p = 1.12 \times 10^{-7}$, t-test). (**C, D**) Bar graphs show mean fraction of SNVs per tumor (*y-axis*) in 96 possible trinucleotide contexts (*x-axis*) in (**C**) HPV-positive and (**D**) HPV-negative tumors. *Colors, key at right*, SNV in central nucleotide of trinucleotide sequences; *asterisk*, significant increase in HPV-positive samples compared to HPV-negative samples by t-test ($p < 0.01$; change in mean fraction >1%). *Error bars*, standard deviation (see also **Supplemental Table S2B**).



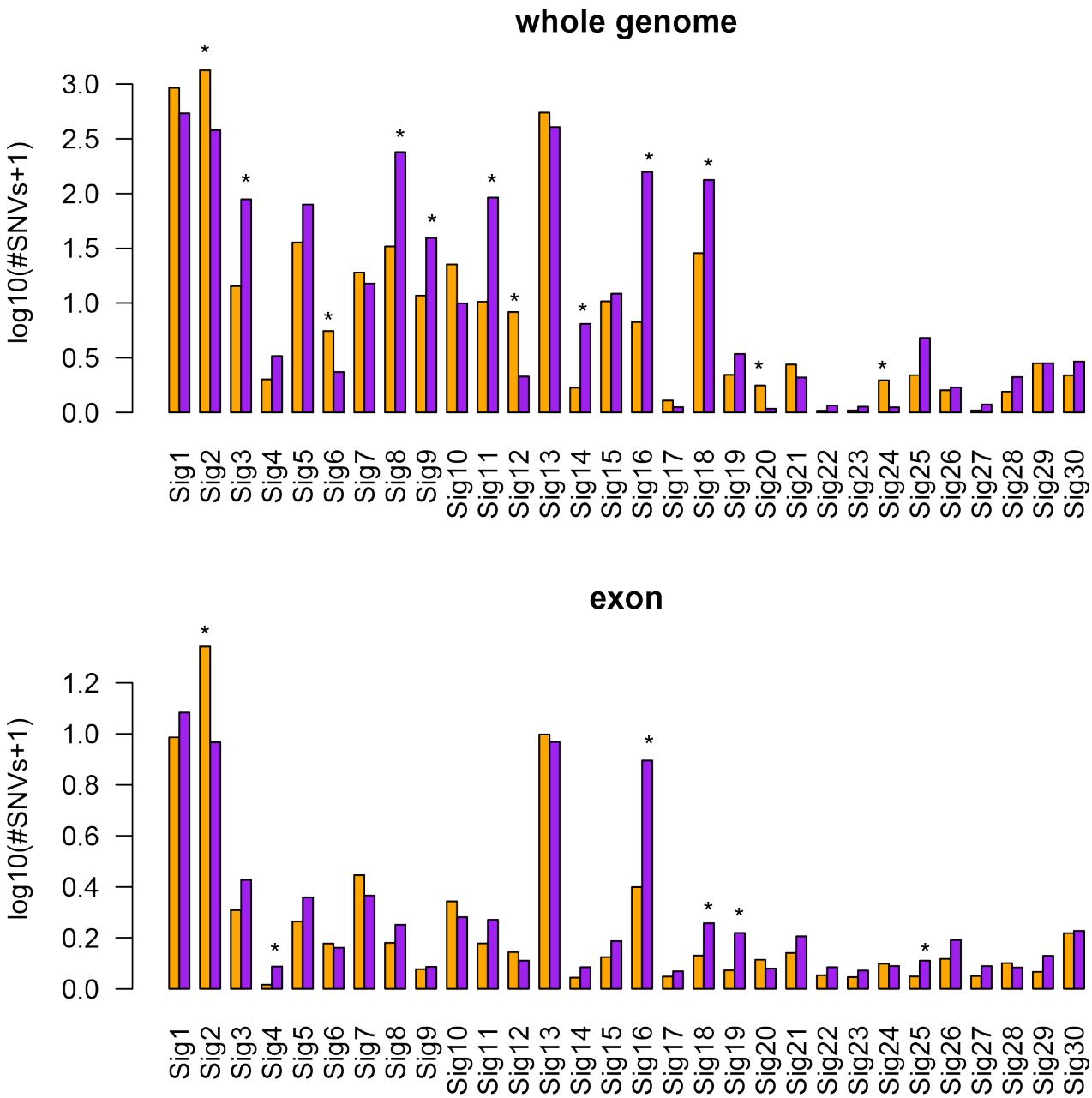
Supplemental Figure S2E. Fractions of mutational signatures in HPV-positive vs. HPV-negative OSCC.
Bar plots display the mean fractions (*y-axis*) of 30 mutational signatures (*x-axis*) in (*orange*) HPV-positive and (*purple*) HPV-negative OSCC. SNVs were identified in (*top*) WGS data from 103 HPV-positive and 50 HPV-negative tumors (purple); and (*bottom*) in exomes of 149 HPV-positive (orange) and 335 HPV-negative tumors (purple). *Asterisks*, significant difference between HPV-positive and HPV-negative tumors (t-test, FDR adjusted p<0.05). See also **Supplemental Table S2E**.



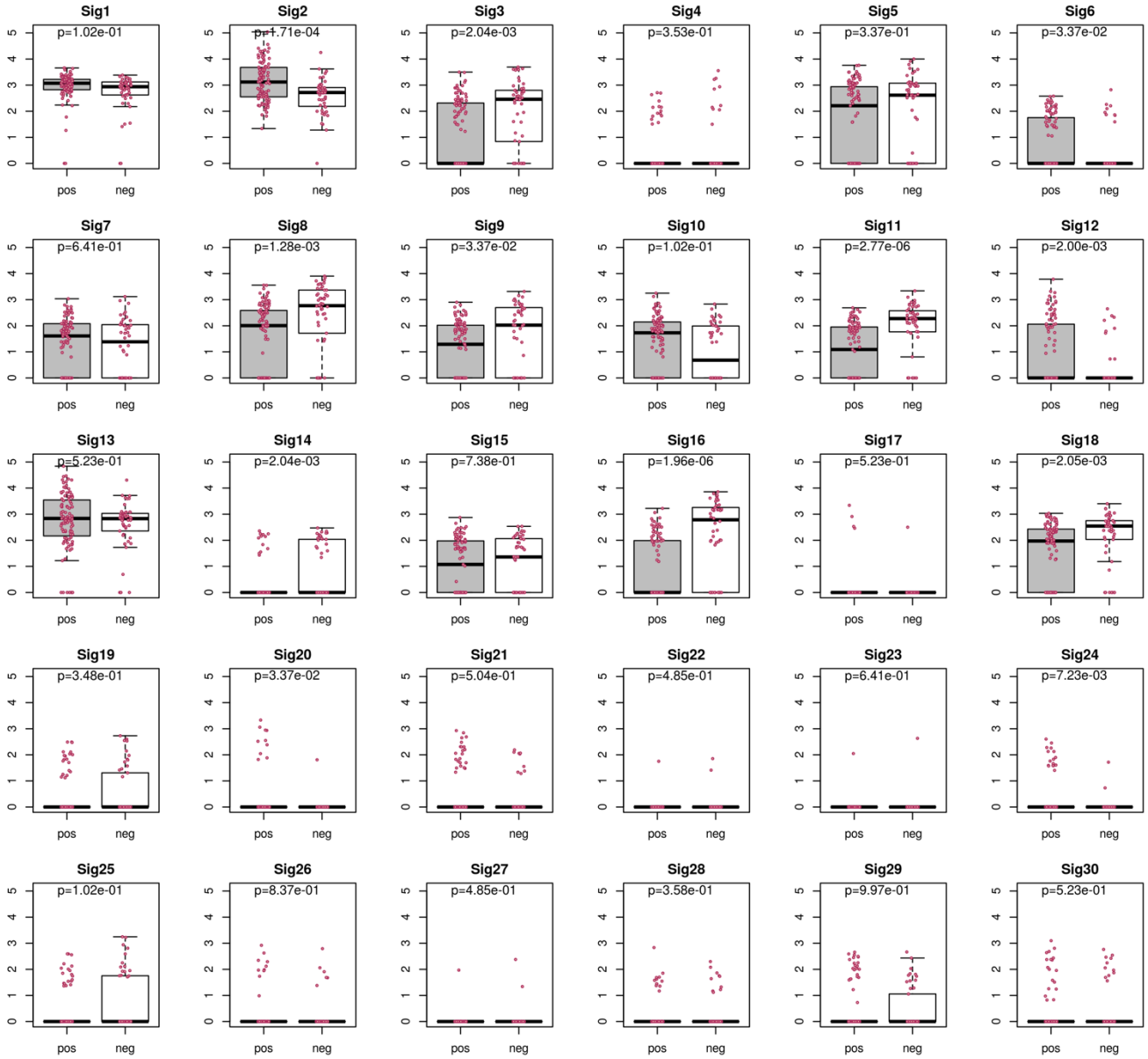
Supplemental Figure S2F. Fractions of 30 mutational signatures found in WGS data of HPV-positive vs. HPV-negative OSCC.
Shown here in box and whisker plots are the fractions of mutational signatures per tumor (*red dots*) in 103 HPV-positive vs. 50 HPV-negative OSCC. P-values were calculated by t-test with false discovery rate multiple testing correction (see also **Supplemental Table S2E**).



Supplemental Figure S2G. Fractions of 30 mutational signatures found in exomes of HPV-positive vs. HPV-negative OSCC.
Shown here in box and whisker plots are the fractions of mutational signatures per tumor (*red dots*) in exomes of 149 HPV-positive vs. 335 HPV-negative OSCC. P-values were calculated by t-test with false discovery rate multiple testing correction (see also **Supplemental Table S2F**).

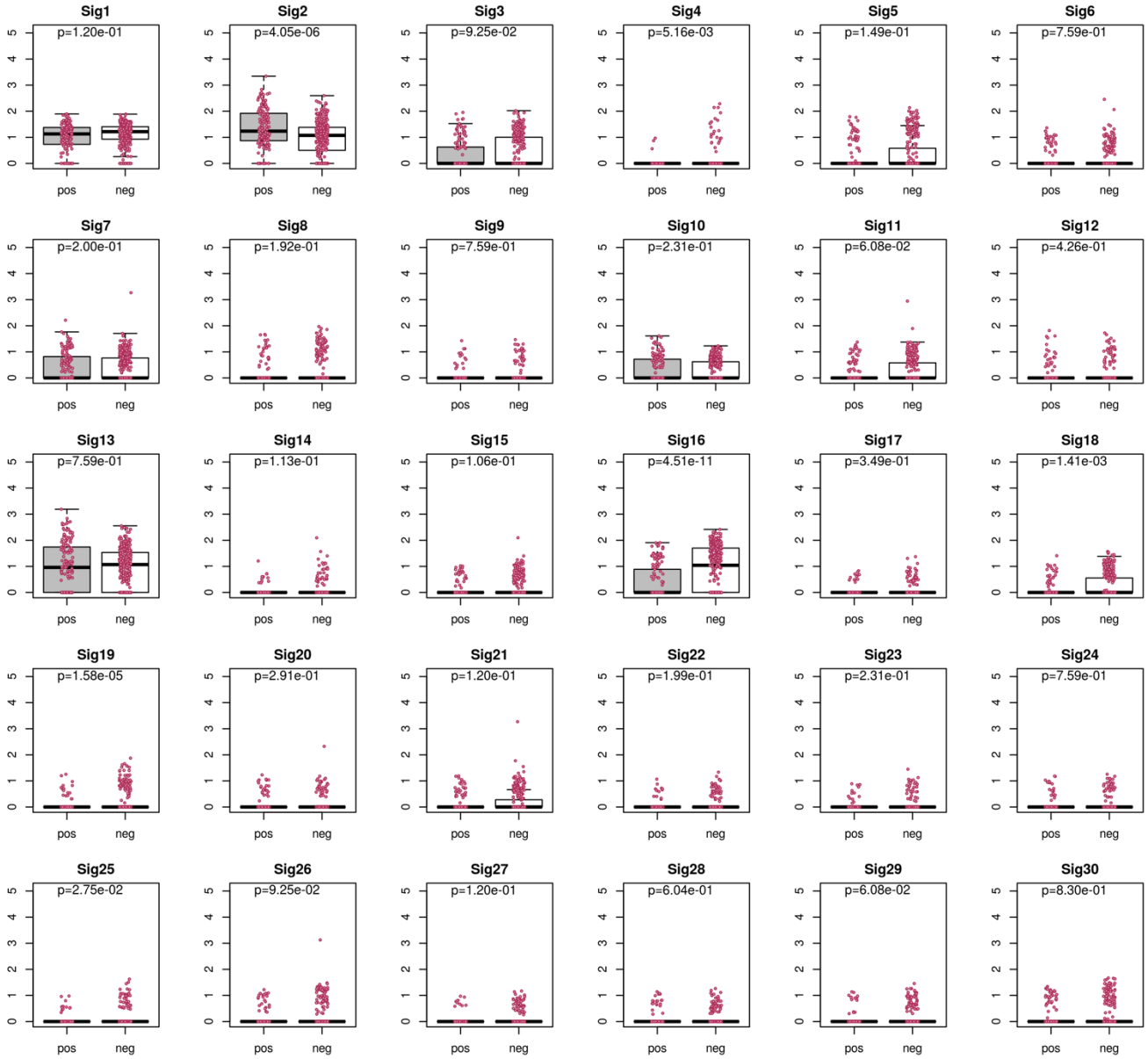


Supplemental Figure S2H. Counts of mutational signatures in HPV-positive vs. HPV-negative OSCC. Bar plots display the log transformation of mean counts (*y-axis*) of 30 mutational signatures (*x-axis*) in (*orange*) HPV-positive and (*purple*) HPV-negative OSCC. SNVs were identified in (*top*) WGS data from 103 HPV-positive and 50 HPV-negative tumors (purple); and (*bottom*) in exomes extracted from WES and WGS data of 149 HPV-positive (*orange*) and 335 HPV-negative tumors (*purple*). *Asterisks*, significant difference between HPV-positive and HPV-negative tumors (t-test, FDR adjusted p<0.05). See also **Supplemental Tables S2E-F**.



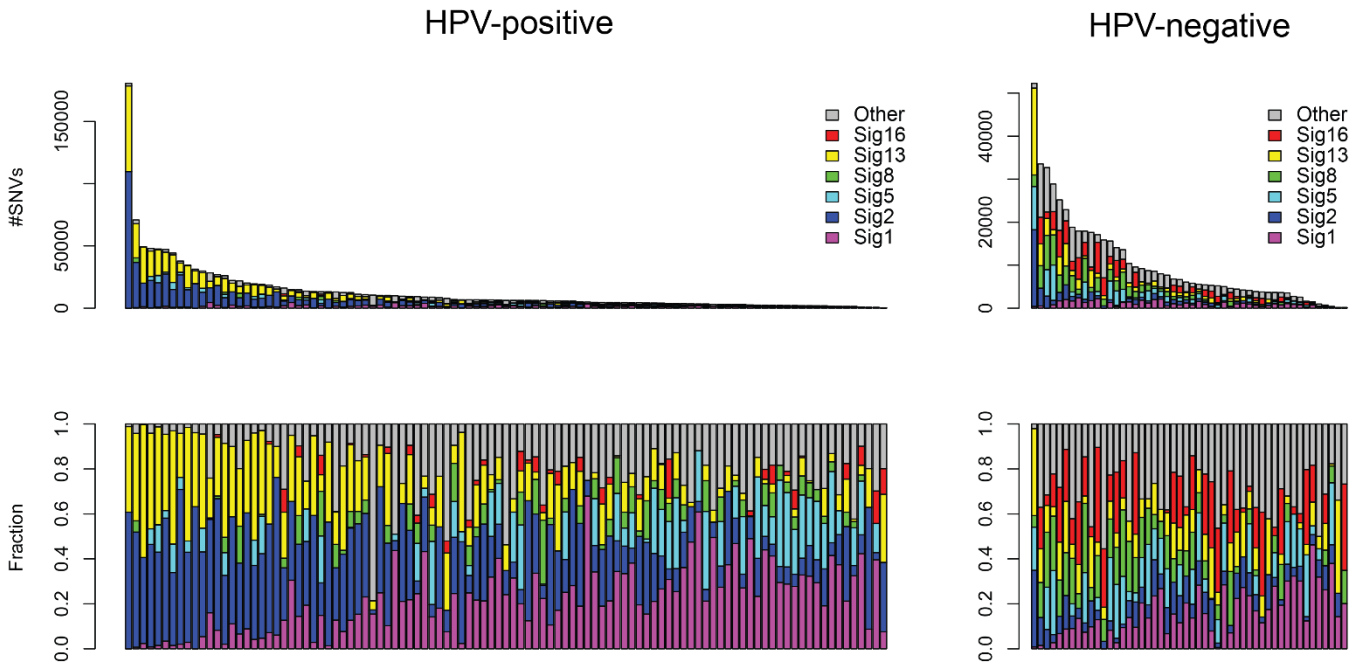
Supplemental Figure S2I. Counts of 30 mutational signatures found in WGS data of HPV-positive vs. HPV-negative OSCC.

Shown here in box and whisker plots are the $\log_{10}$ transformed counts of mutational signatures per tumor (*red dots*) in 103 HPV-positive vs. 50 HPV-negative OSCC. P-values were calculated by t-test with false discovery rate multiple testing correction (see also **Supplemental Table S2E**).



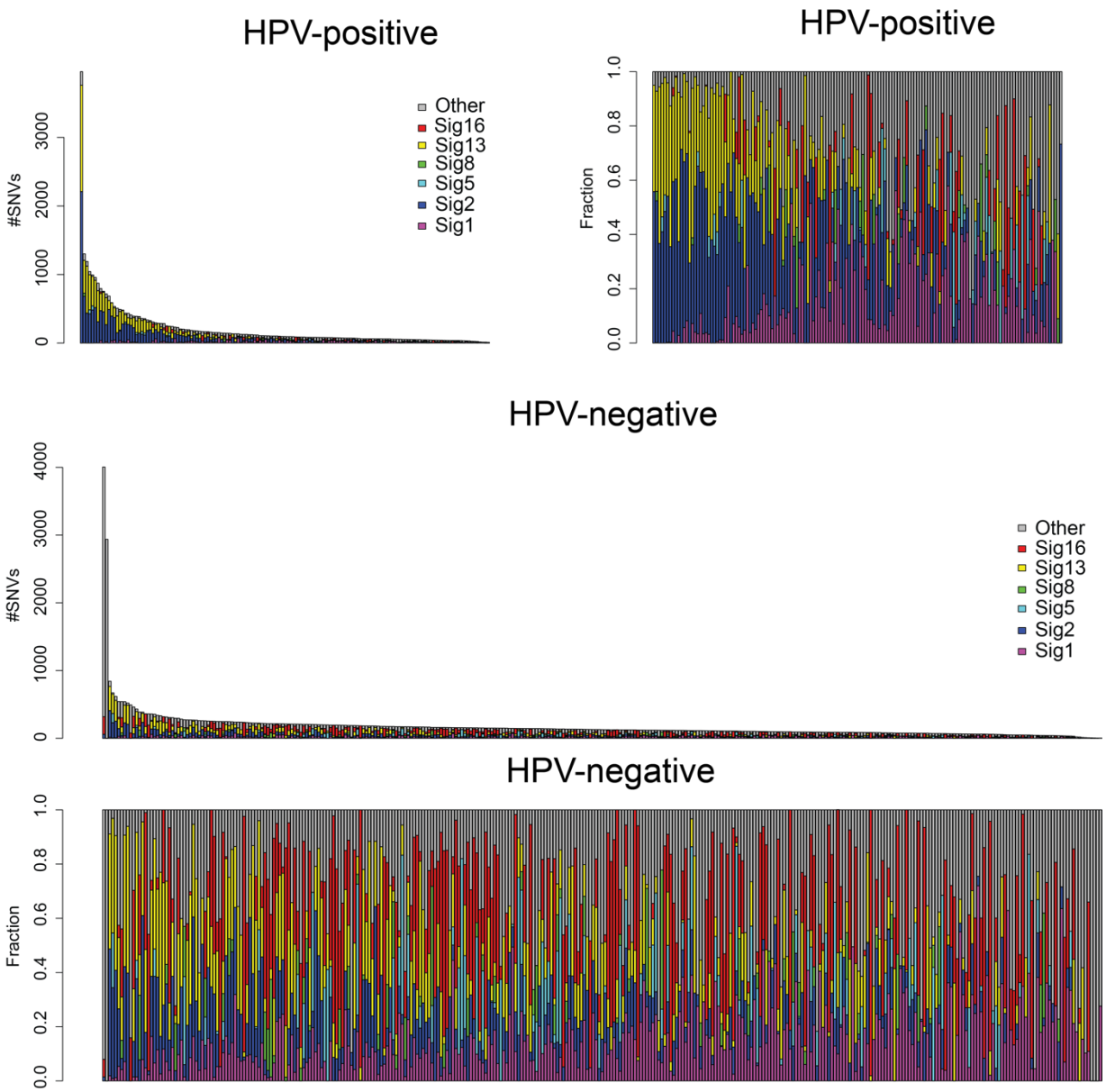
Supplemental Figure S2J. Counts of 30 mutational signatures found in exomes of HPV-positive vs. HPV-negative OSCC.

Shown here in box and whisker plots are the $\log_{10}$ transformed counts of mutational signatures per tumor (*red dots*) in exomes of 149 HPV-positive vs. 335 HPV-negative OSCC. P-values were calculated by t-test with false discovery rate multiple testing correction (see also **Supplemental Table S2F**).



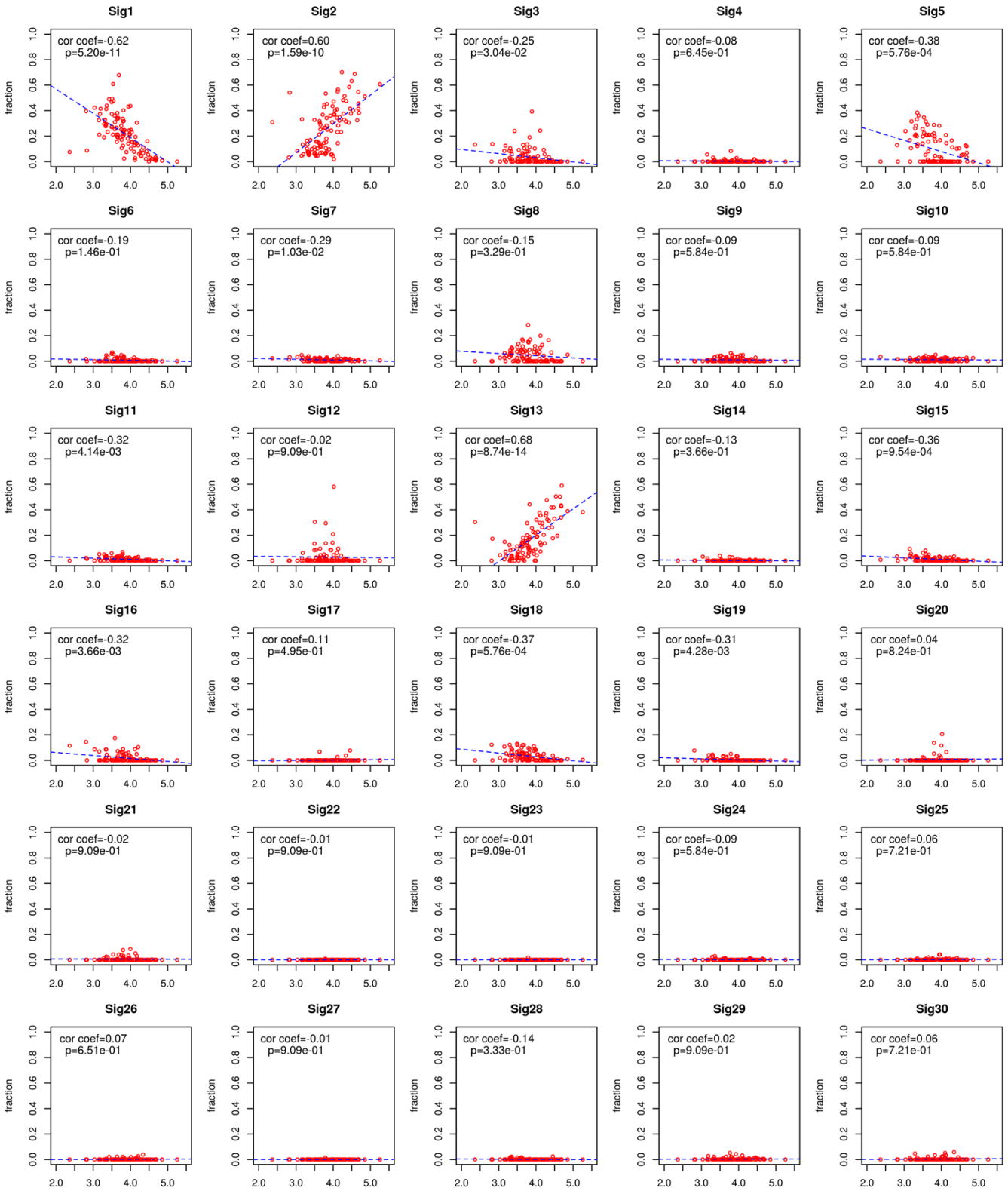
Supplemental Figure S2K. Counts and fractions of mutational signatures per tumor found in WGS data of HPV-positive and HPV-negative OSCC.

Bar plots display the (*top*) counts and (*bottom*) fractions per tumor (*x-axis*) found in WGS data from (*left*) 103 HPV-positive and (*right*) 50 HPV-negative OSCC. *Colors, key*: *pink*, signature 1; *navy*, signature 2; *light blue*, signature 5; *green,* signature 8; *yellow*, signature 13; *red,* signature 16; and *gray*, others (see also **Supplemental Table S2E**).

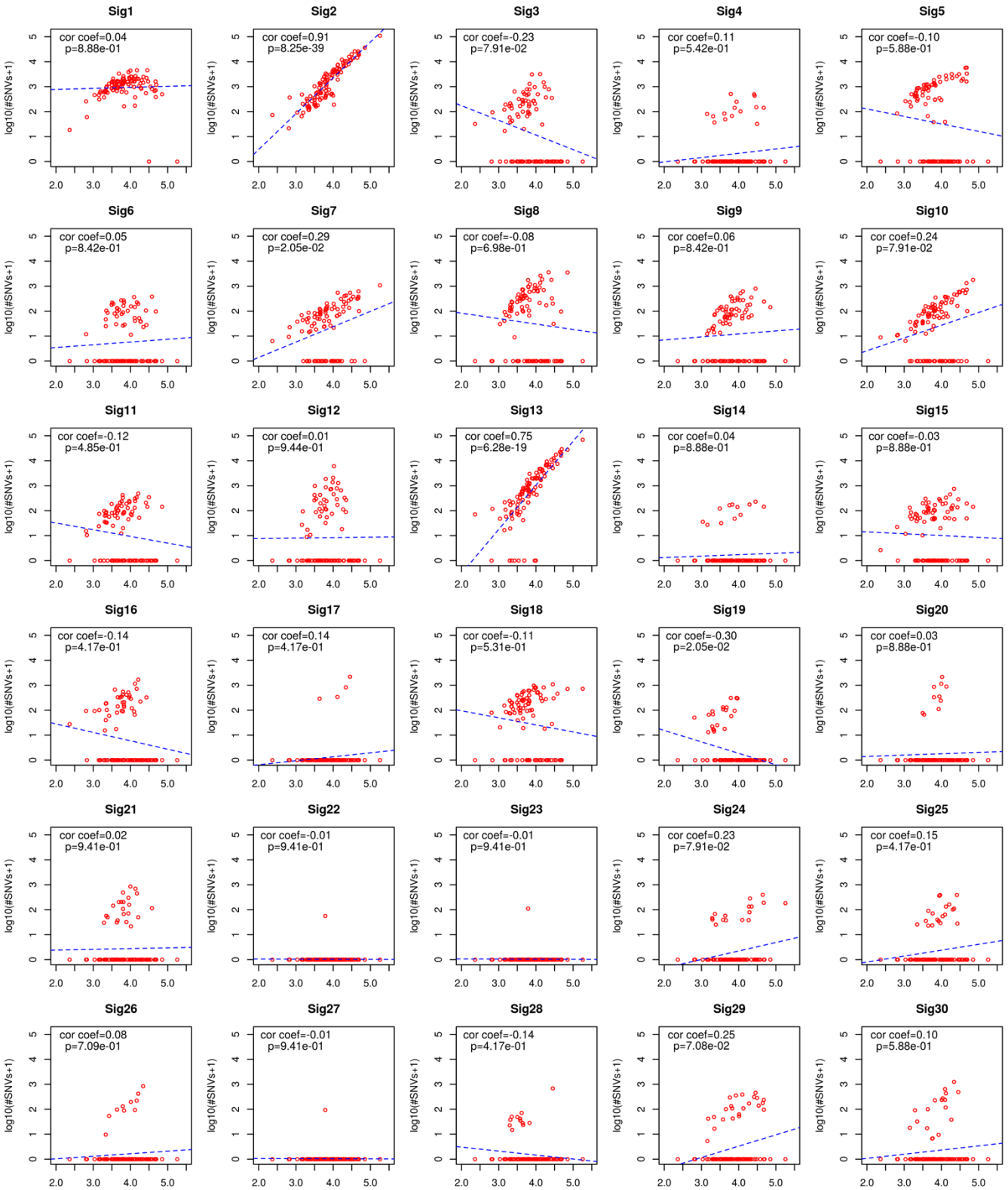


Supplemental Figure S2L. Counts and fractions of mutational signatures per tumor found in exomes of HPV-positive and HPV-negative OSCC.

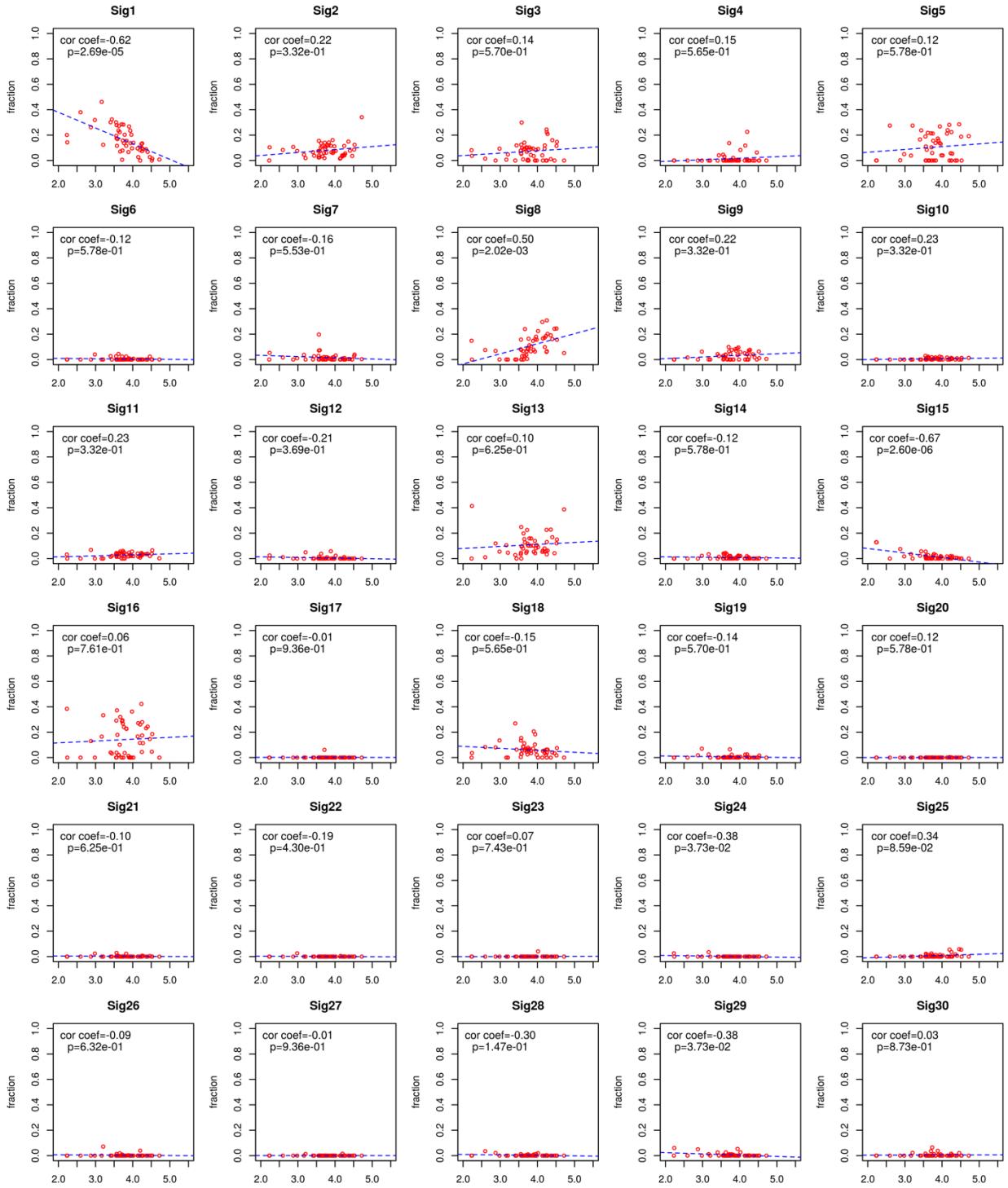
Bar plots display the (*top*) counts and (*bottom*) fractions per tumor (*x-axis*) found in exomes derived from WGS and WES data from (*left*) 149 HPV-positive and (*right*) 335 HPV-negative OSCC. *Colors, key*: *pink*, signature 1; *navy*, signature 2; *light blue*, signature 5; *green,* signature 8; *yellow*, signature 13; *red,* signature 16; and *gray*, others (see also **Supplemental Table S2F**).



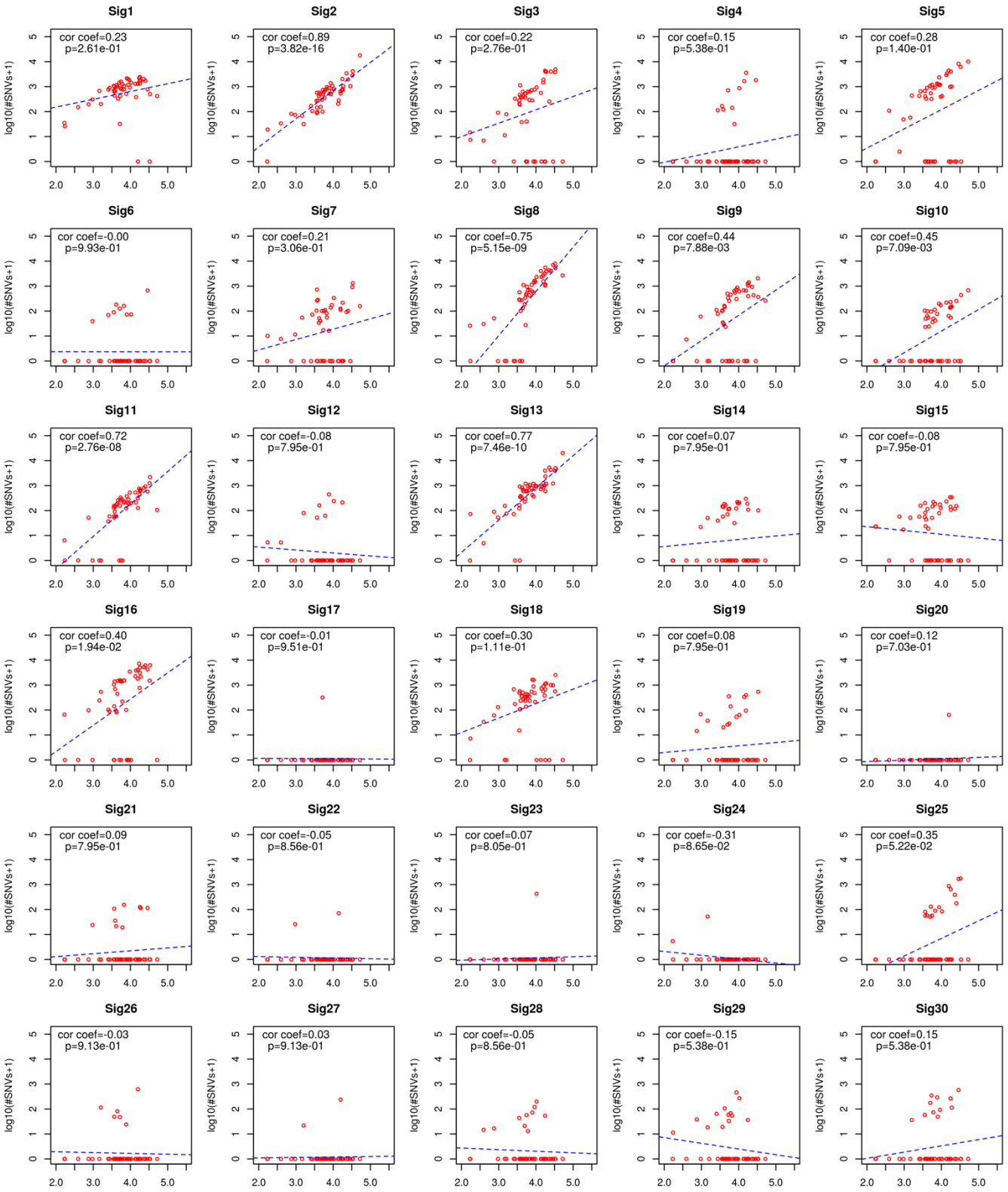
Supplemental Figure S2M1. Associations between total number of SNVs and fractions of mutational signatures in HPV-positive OSCC, using WGS data. Scatter plots show associations in HPV-positive OSCC (n = 103 WGS). *X-axis*, $\log_{10}$-transformed total number of SNVs, $\log_{10}$(no.SNVs+1); *y*-axis, fraction of mutational signature per tumor (red dots). Pearson's correlation coefficient and adjusted p-values using FDR multiple testing correction were calculated for each signature (see also **Supplemental Table S2G**).



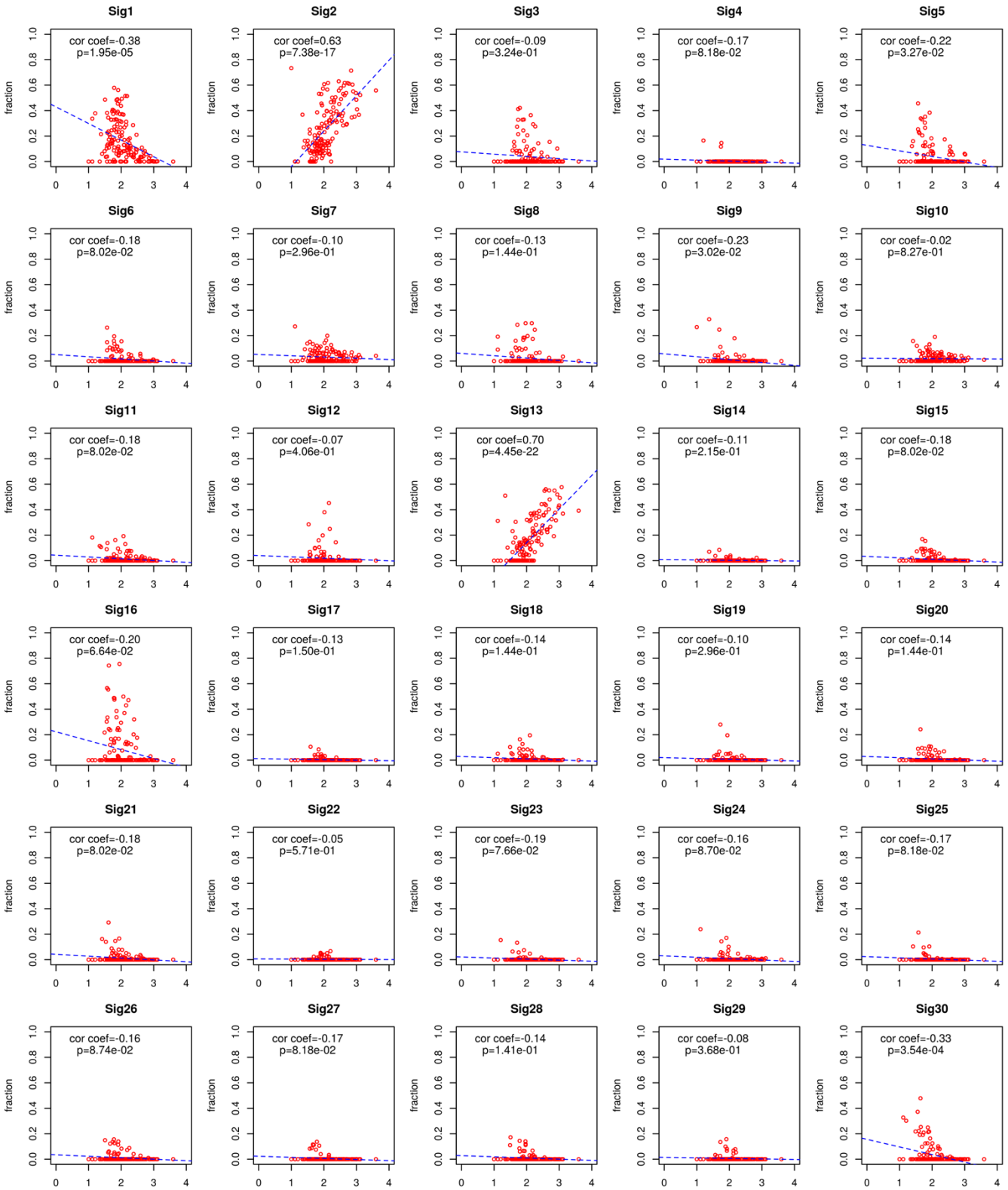
Supplemental Figure S2M2. Associations between total number of SNVs and counts of mutational signatures in HPV-positive OSCC, using WGS data. Scatter plots show associations in HPV-positive OSCC (n = 103 WGS). *X-axis*, $\log_{10}$-transformed total number of SNVs, $\log_{10}$(no. SNVs+1); *y*-axis, log transformed count of mutational signature per tumor (*red dots*). Pearson's correlation coefficient and adjusted p-values using multiple testing correction with FDR were calculated for each signature (see also **Supplemental Table S2G**).



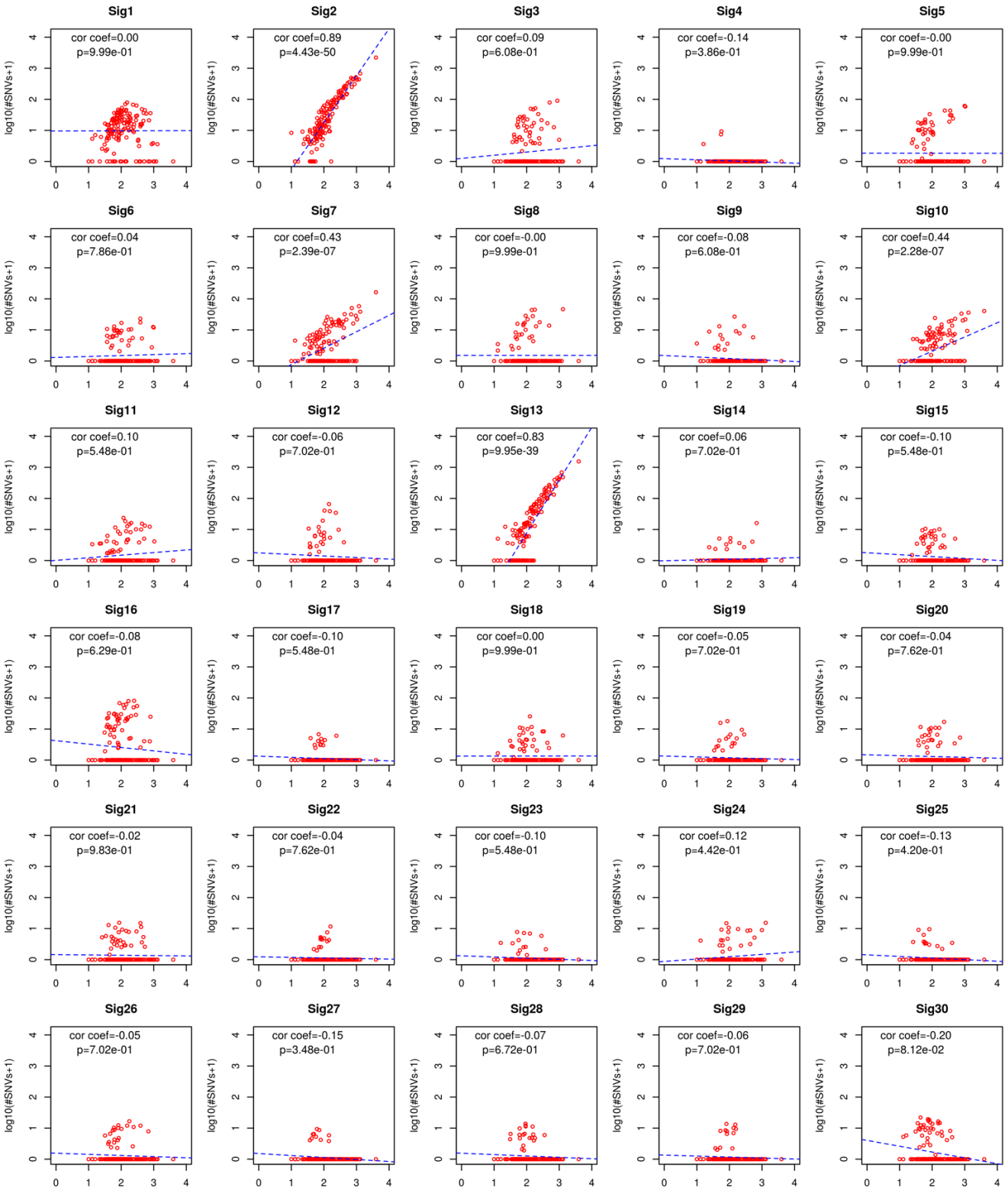
Supplemental Figure S2M3. Associations between total number of SNVs and fractions of mutational signatures in HPV-negative OSCC, using WGS data. Scatter plots show associations in HPV-negative OSCC (n = 50 WGS). *X-axis*, $\log_{10}$-transformed total number of SNVs, $\log_{10}$(no. SNVs+1); *y-axis*, fraction of mutational signature per tumor (*red dots*). Pearson's correlation coefficient and adjusted p-values using multiple testing correction with FDR were calculated for each signature (see also **Supplemental Table S2I**).



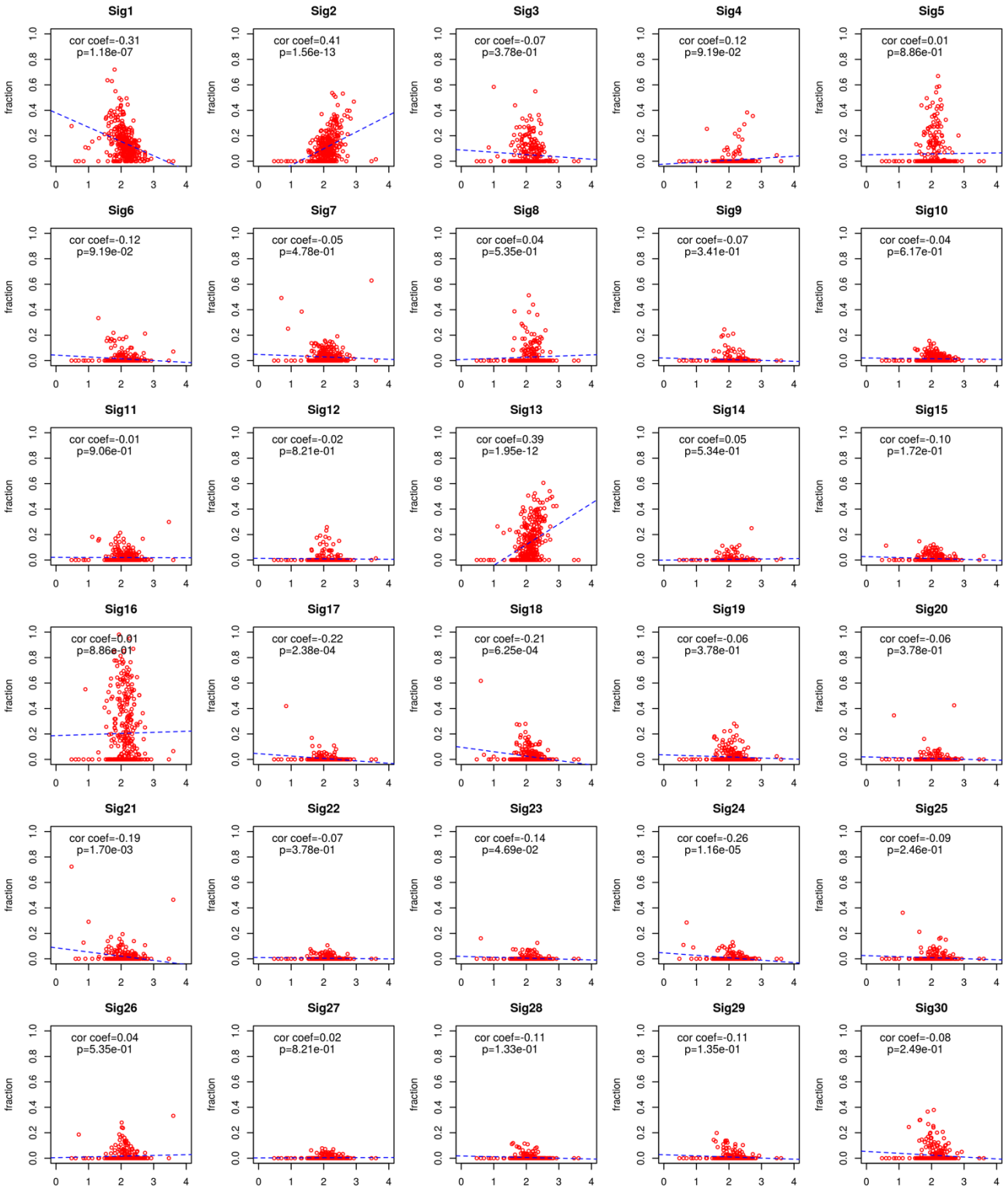
Supplemental Figure S2M4. Associations between total number of SNVs and counts of mutational signatures in HPV-negative OSCC, WGS data. Scatter plots show associations in HPV-negative OSCC (n = 50 WGS). *X-axis*, $\log_{10}$-transformed total number of SNVs; *y*-axis, log-transformed count of mutational signature per tumor (*red dots*). Pearson's correlation coefficient and adjusted p-values using multiple testing correction with FDR were calculated for each signature (see also **Supplemental Table S2I**).



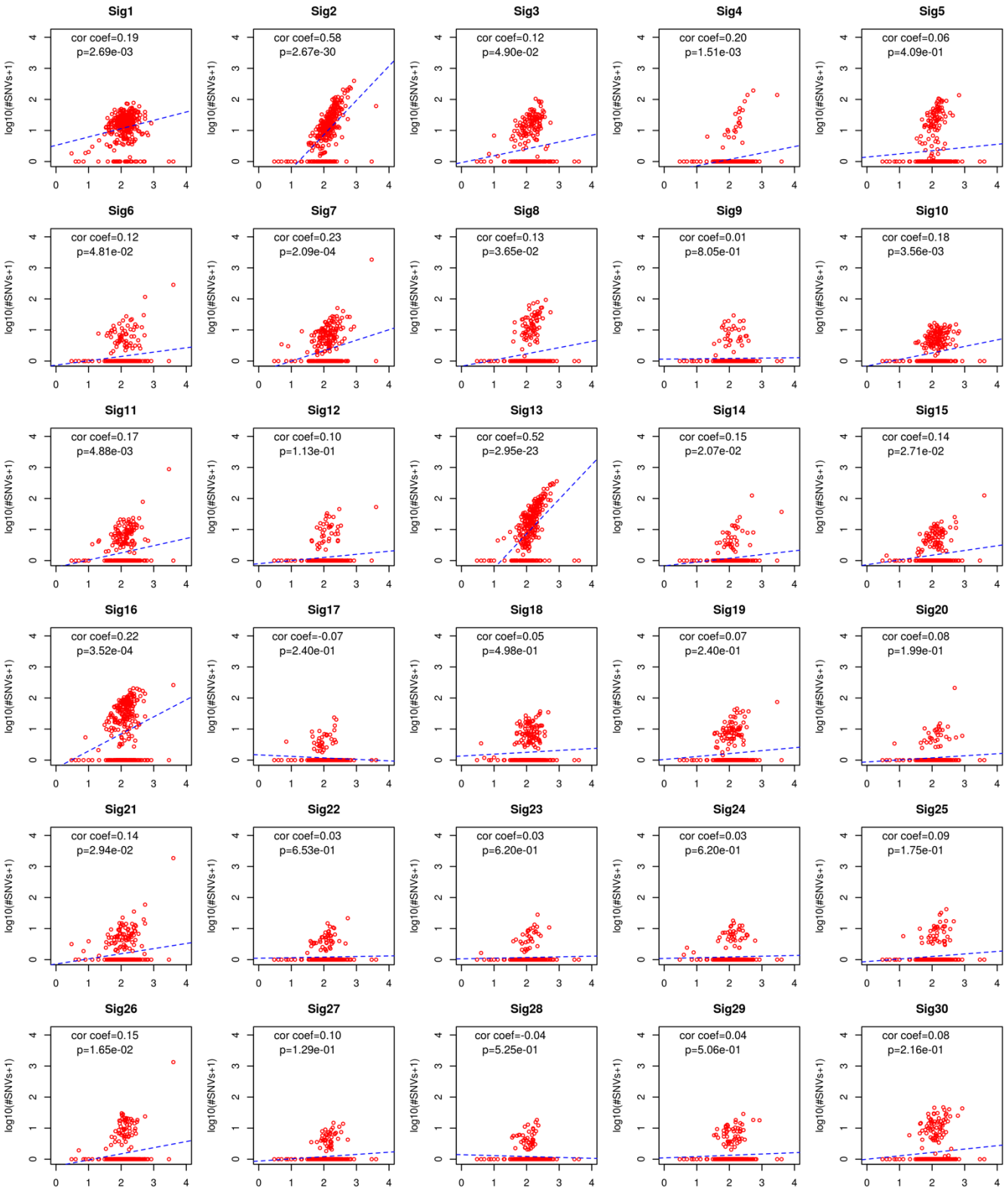
Supplemental Figure S2M5. Associations between total number of SNVs and fraction of mutational signatures in HPV-positive OSCC, using exome data. Scatter plots show associations in HPV-positive OSCC (n = 149 exomes). *X-axis*, $\log_{10}$-transformed total number of SNVs, $\log_{10}$(no. SNVs+1); *y*-axis, fraction of mutational signature per tumor (*red dots*). Pearson's correlation coefficient and adjusted p-values using FDR multiple testing correction were calculated for each signature (see also **Supplemental Table S2H**).



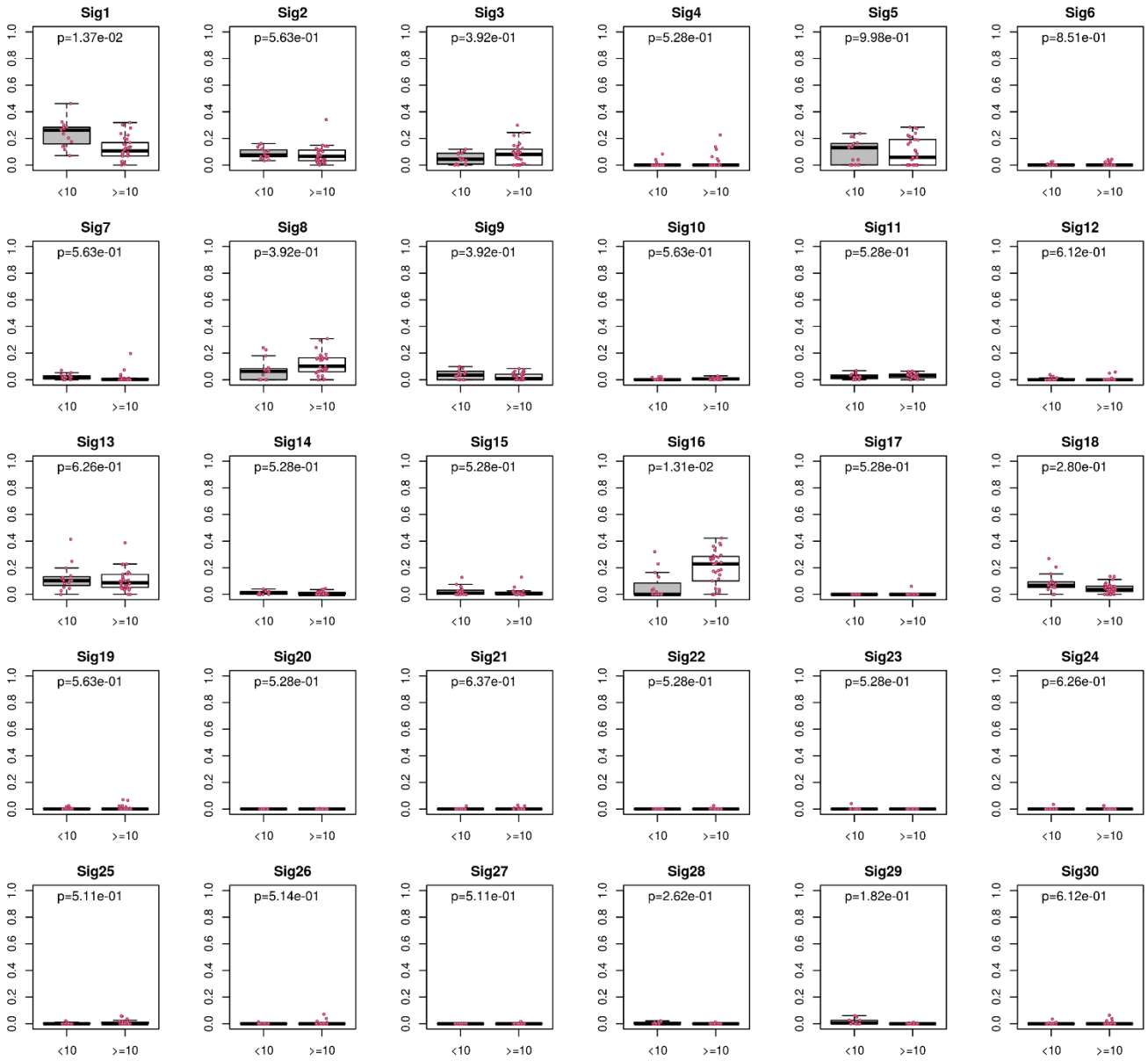
Supplemental Figure S2M6. Associations between total number of SNVs and counts of mutational signatures in HPV-positive OSCC, using exome data. Scatter plots show associations in HPV-positive OSCC (n = 149 exomes). *X-axis*, $\log_{10}$-transformed total number of SNVs, $\log_{10}$(no. SNVs+1); *y*-axis, log transformed count of mutational signatures per tumor (*red dots*). Pearson's correlation coefficient and adjusted p-values using FDR multiple testing correction were calculated for each signature (see also **Supplemental Table S2H**).



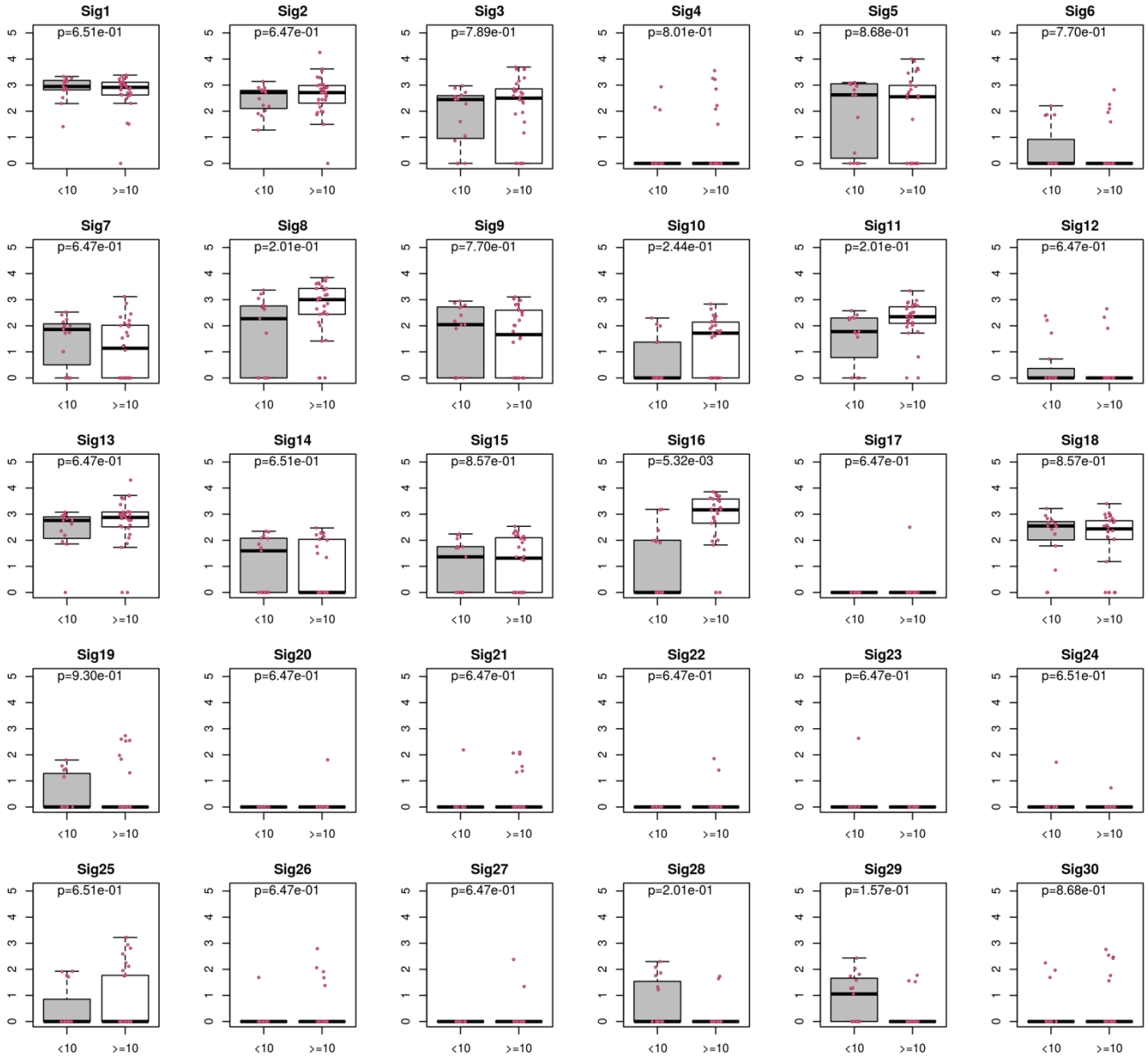
Supplemental Figure S2M7. Associations between total number of SNVs and fraction of mutational signatures in HPV-negative OSCC, using exome data. Scatter plots show associations in HPV-negative OSCC (n = 335 exomes). *X-axis*, $\log_{10}$-transformed total number of SNVs, $\log_{10}$(no.SNVs+1); *y*-axis, fraction of mutational signature per tumor (*red dots*). Pearson's correlation coefficient and adjusted p-values using multiple testing correction with FDR were calculated for each signature (see also **Supplemental Table S2J**).



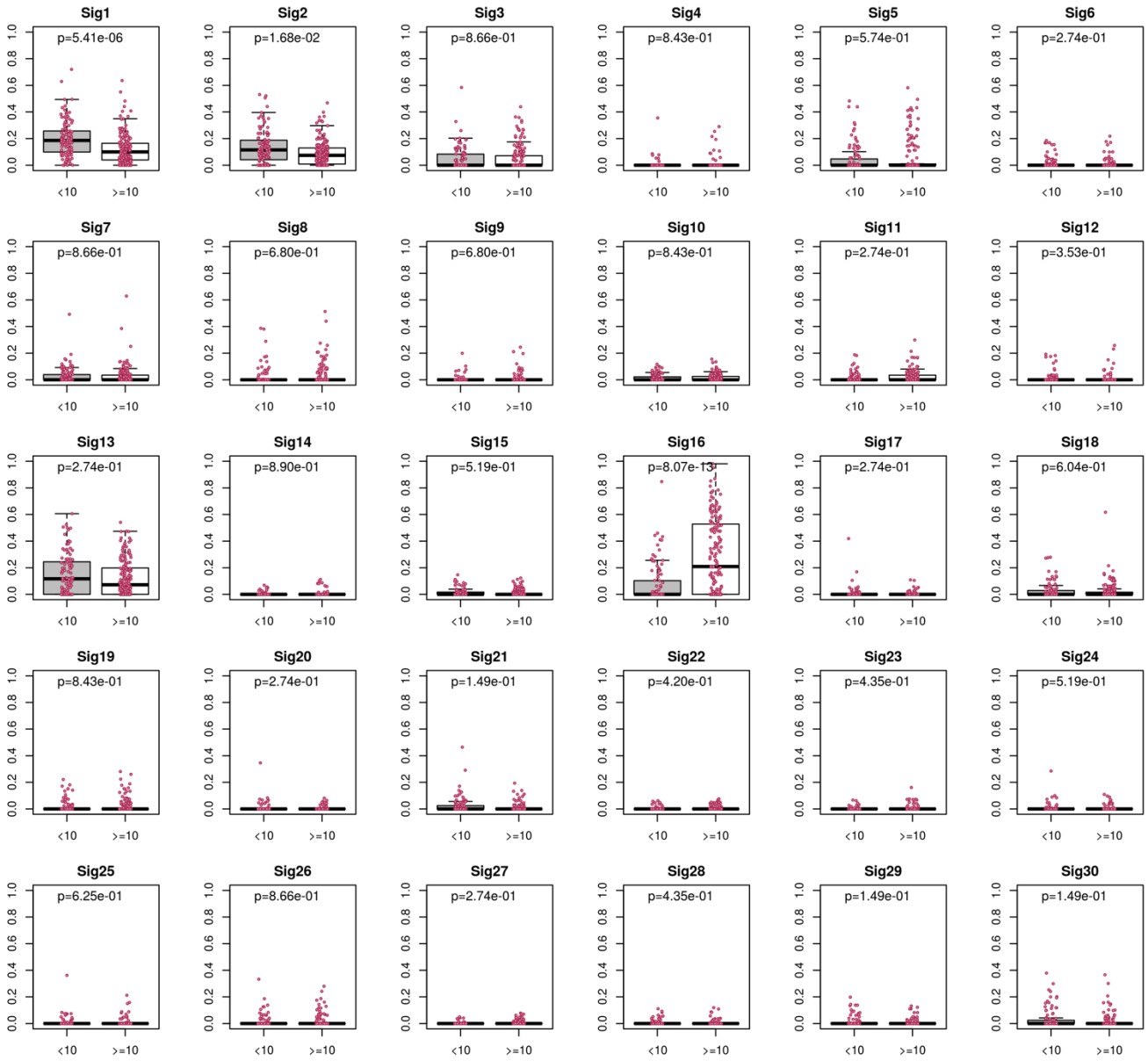
Supplemental Figure S2M8. Associations between total number of SNVs and counts of mutational signatures in HPV-negative OSCC, using exome data. Scatter plots show associations in HPV-negative OSCC (n = 335 exomes). *X-axis*, $\log_{10}$-transformed total number of SNVs, $\log_{10}$(no.SNVs+1); *y*-axis, log transformed count of mutational signatures per tumor (*red dots*). Pearson's correlation coefficient and adjusted p-values using multiple testing correction with FDR were calculated for each signature (see also **Supplemental Table S2J**).



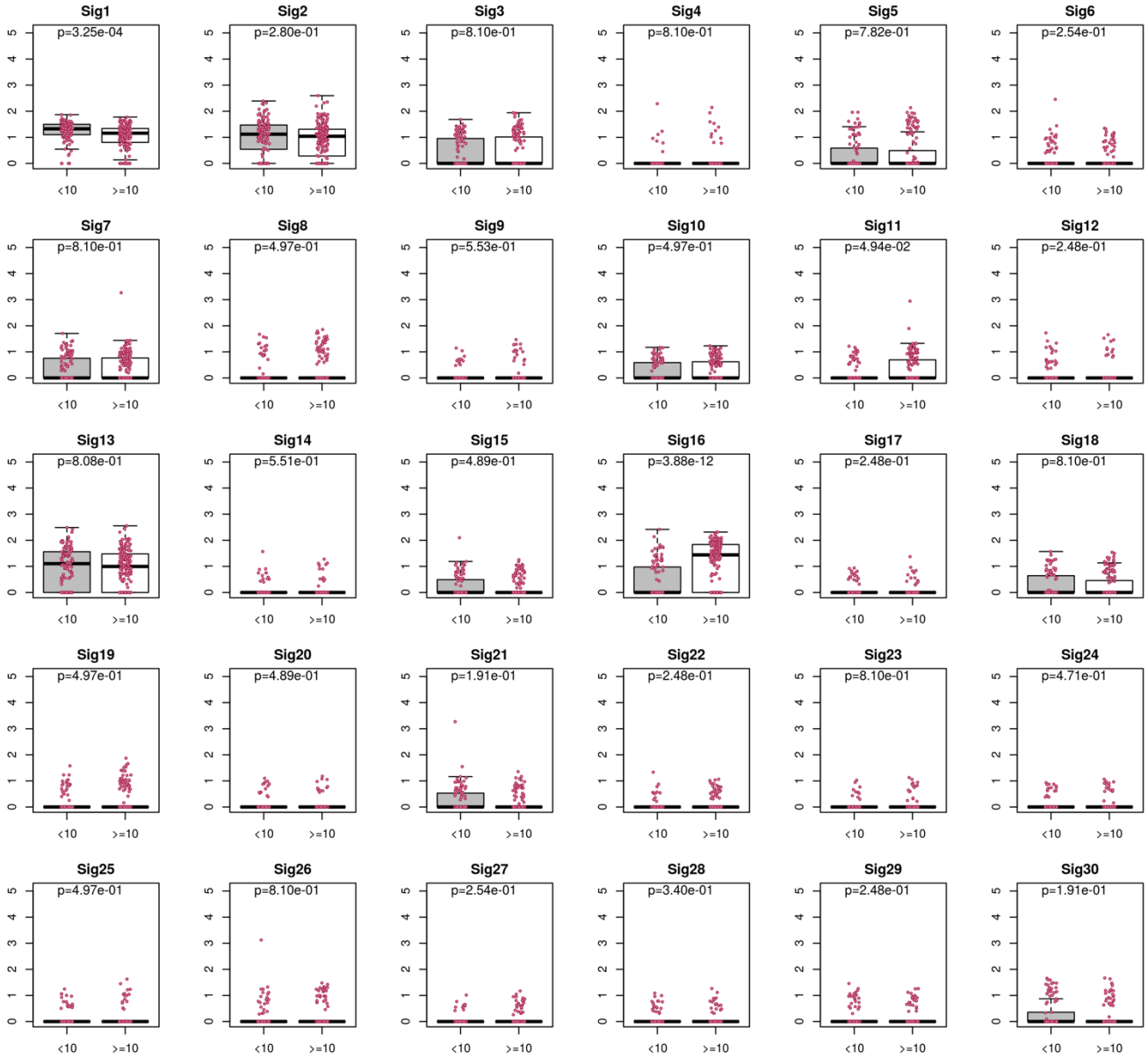
Supplemental Figure S2N1. Associations between cigarette smoking and fractions of mutational signature in HPV-negative OSCC, using WGS data.
Shown here in box and whisker plots are the fractions (*y-axis*) of mutational signatures per tumor (*red dots*), comparing (*x-axis*) non-/low-smokers (less than 10 pack-years, n=15) vs. heavy-smokers (greater than or equal to 10 pack-years, n=30). Adjusted P-values were calculated by t-test with false discovery rate multiple testing correction (see also **Supplemental Table S2K**).



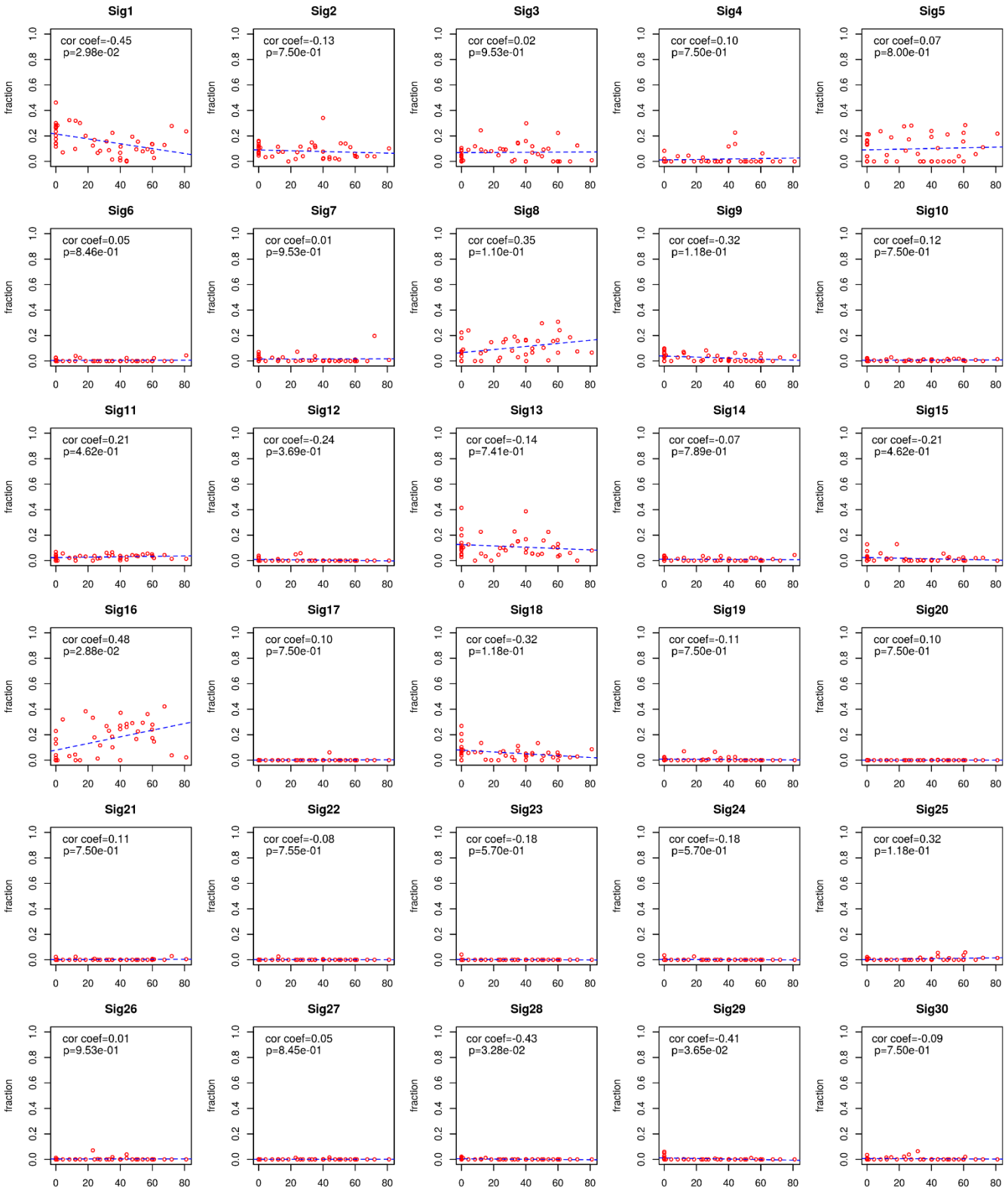
Supplemental Figure S2N2. Associations between cigarette smoking and counts of mutational signatures in HPV-negative OSCC, using WGS data.
Shown here in box and whisker plots are the $\log_{10}$-transformed counts of SNVs (*y-axis*) in mutational signatures per tumor (*red dots*), comparing (*x-axis*) non-/low-smokers (less than 10 pack-years, n=15) vs. heavy-smokers (greater than or equal to 10 pack-years, n=30). Adjusted P-values were calculated by t-test with false discovery rate multiple testing correction (see also **Supplemental Table S2K**).



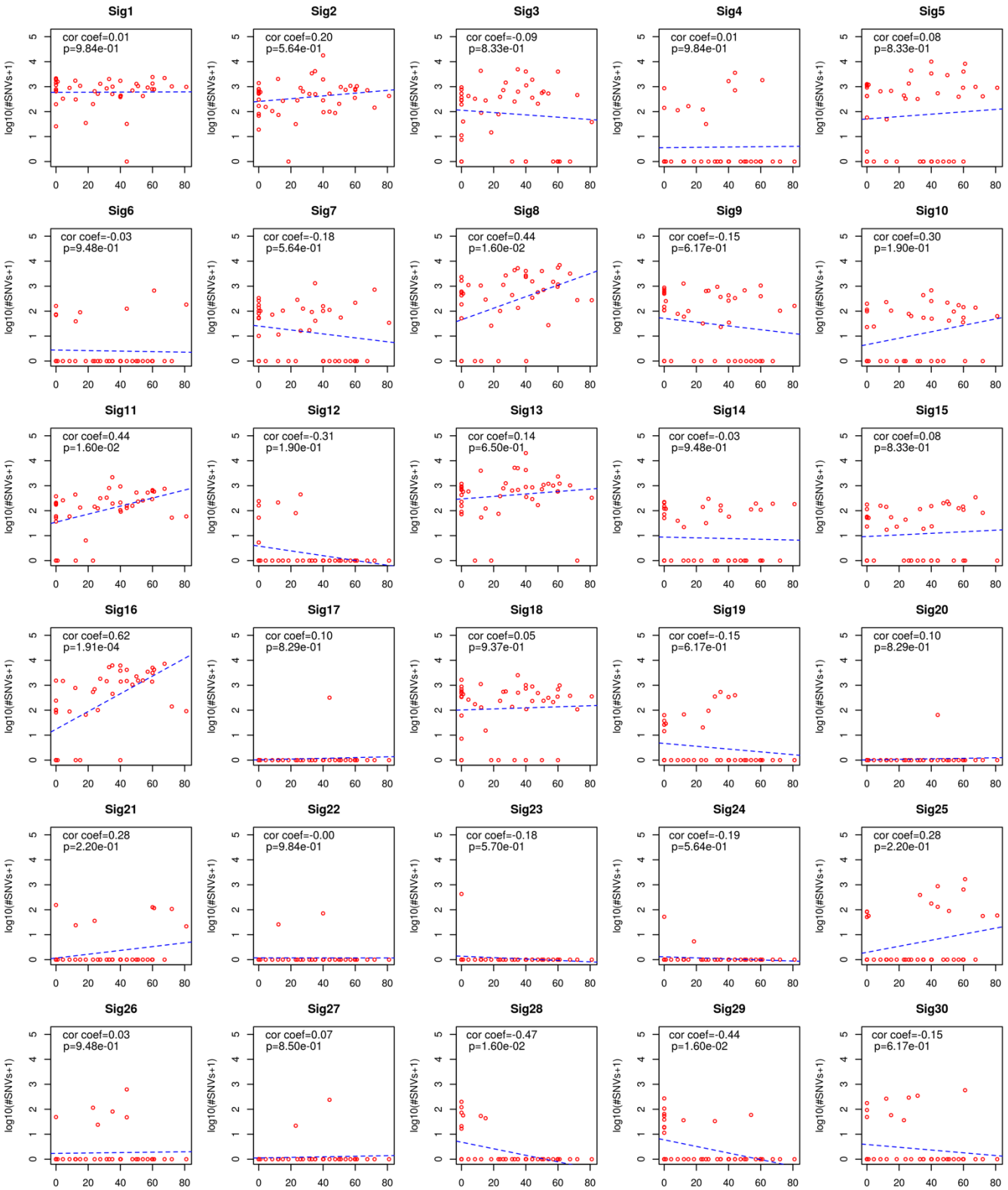
Supplemental Figure S2N3. Associations between cigarette smoking and fractions of mutational signatures in HPV-negative OSCC, using exome data.
Shown here in box and whisker plots are the fractions counts of SNVs (*y-axis*) in mutational signatures per tumor (*red dots*), comparing (*x-axis*) non-/low-smokers (less than 10 pack-years, n=105) vs. heavy-smokers (greater than or equal to 10 pack-years, n=167). Adjusted P-values were calculated by t-test with false discovery rate multiple testing correction (see also **Supplemental Table S2L**).



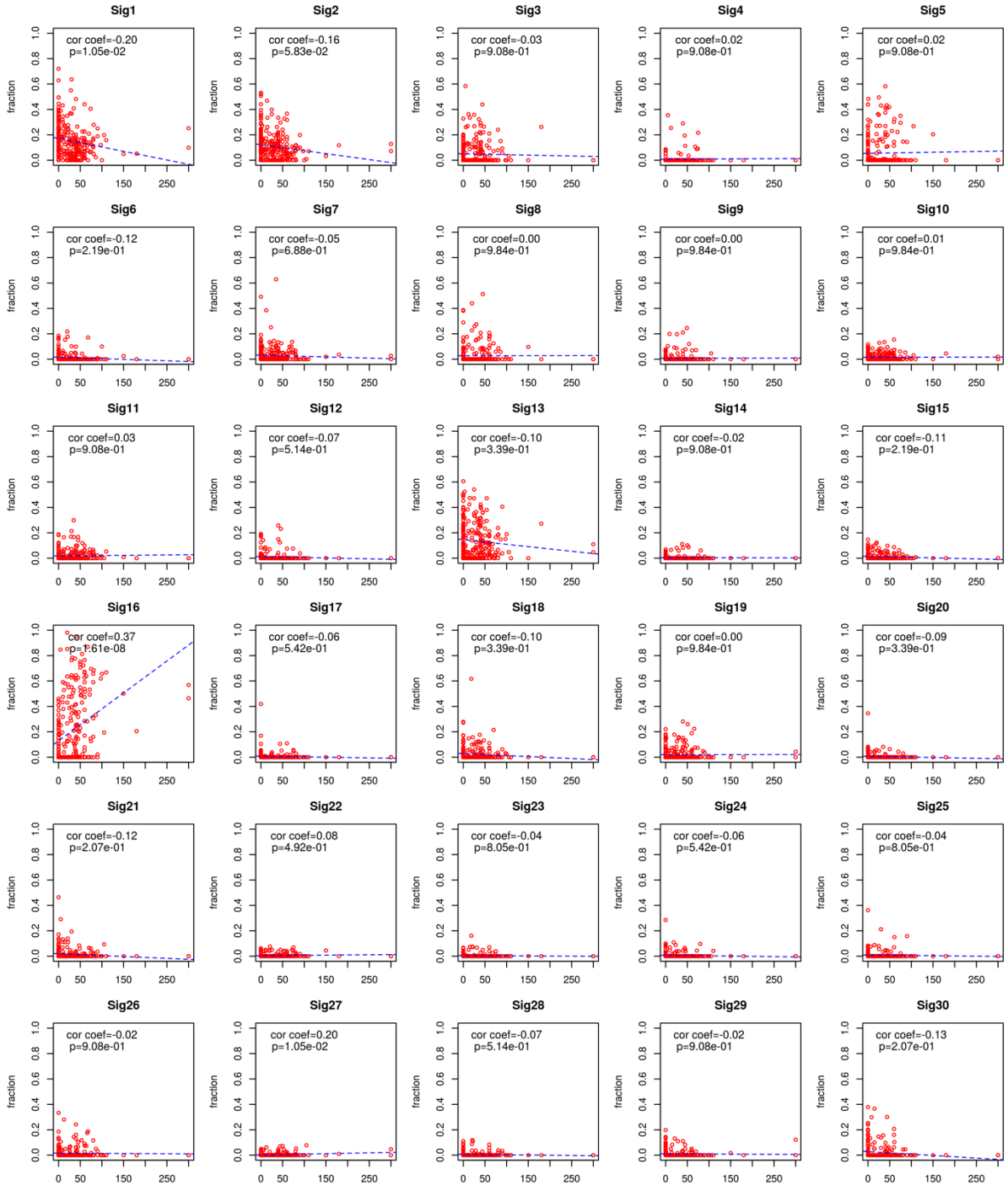
Supplemental Figure S2N4. Associations between cigarette smoking and counts of mutational signatures in HPV-negative OSCC, using exome data.
Shown here in box and whisker plots are the $\log_{10}$-transformed counts of SNVs (*y-axis*) in mutational signatures per tumor (*red dots*), comparing (*x-axis*) non-/low-smokers (<10 pack-years, n=105) vs. heavy-smokers (≥10 pack-years, n=167). Adjusted P-values were calculated by t-test with false discovery rate multiple testing correction (see also **Supplemental Table S2L**).



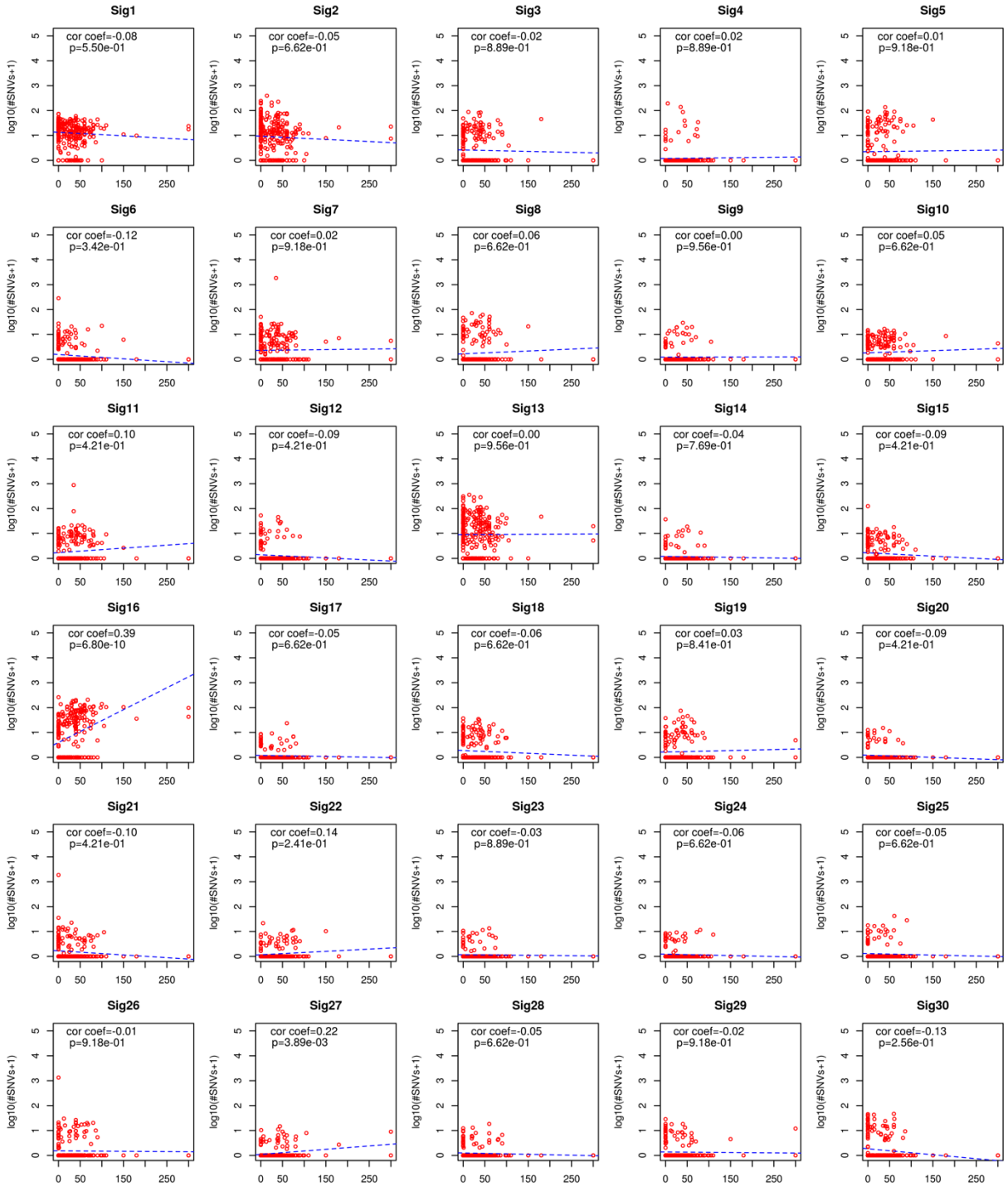
Supplemental Figure S2N5. Associations between cigarette smoking and fractions of mutational signatures in HPV-negative OSCC, using WGS data. Scatter plots show associations in HPV-negative OSCC (n = 45). *X-axis*, pack-years of cigarette smoking; *y-axis*, fraction of mutational signatures per tumor (*red dots*). Pearson's correlation coefficient and adjusted p-values using multiple testing correction with FDR were calculated for each signature (see also **Supplemental Table S2M**).



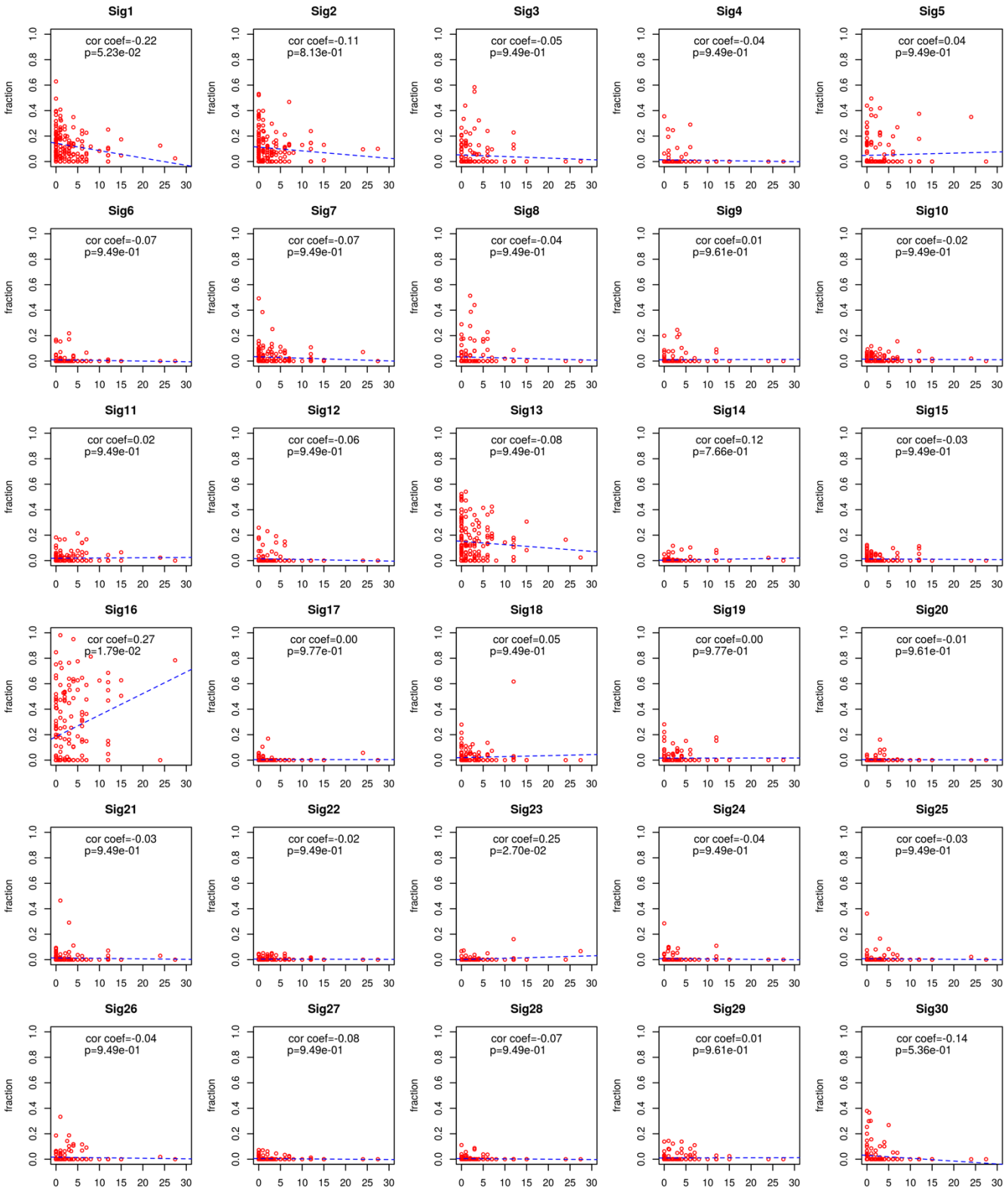
Supplemental Figure S2N6. Associations between cigarette smoking and counts of mutational signatures in HPV-negative OSCC, using WGS data. Scatter plots show associations in HPV-negative OSCC (n = 45). *X-axis*, pack-years of cigarette smoking; *y*-axis, $\log_{10}$-transformation ($\log_{10}$ (SNVs + 1)) of counts of mutational signature per tumor (*red dots*). Pearson's correlation coefficient and adjusted p-values using multiple testing correction with FDR were calculated for each signature (see also **Supplemental Table S2M**).



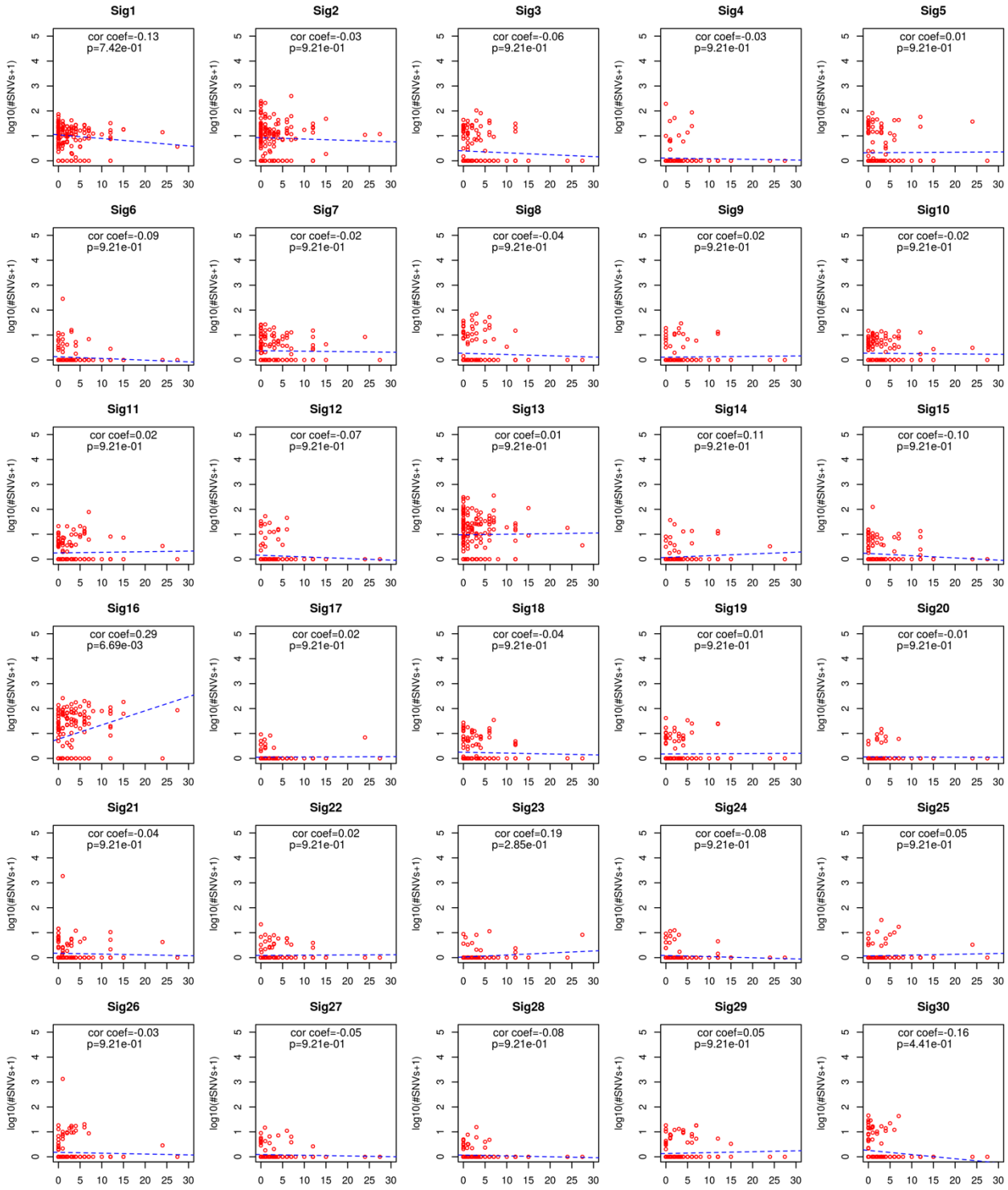
Supplemental Figure S2N7. Associations between cigarette smoking and fractions of mutational signatures in HPV-negative OSCC, using exome data. Scatter plots show associations in HPV-negative OSCC (n = 272). *X-axis*, pack-years of cigarette smoking; *y-axis*, fraction of mutational signatures per tumor (*red dots*). Pearson's correlation coefficient and adjusted p-values using multiple testing correction with FDR were calculated for each signature (see also **Supplemental Table S2N**).



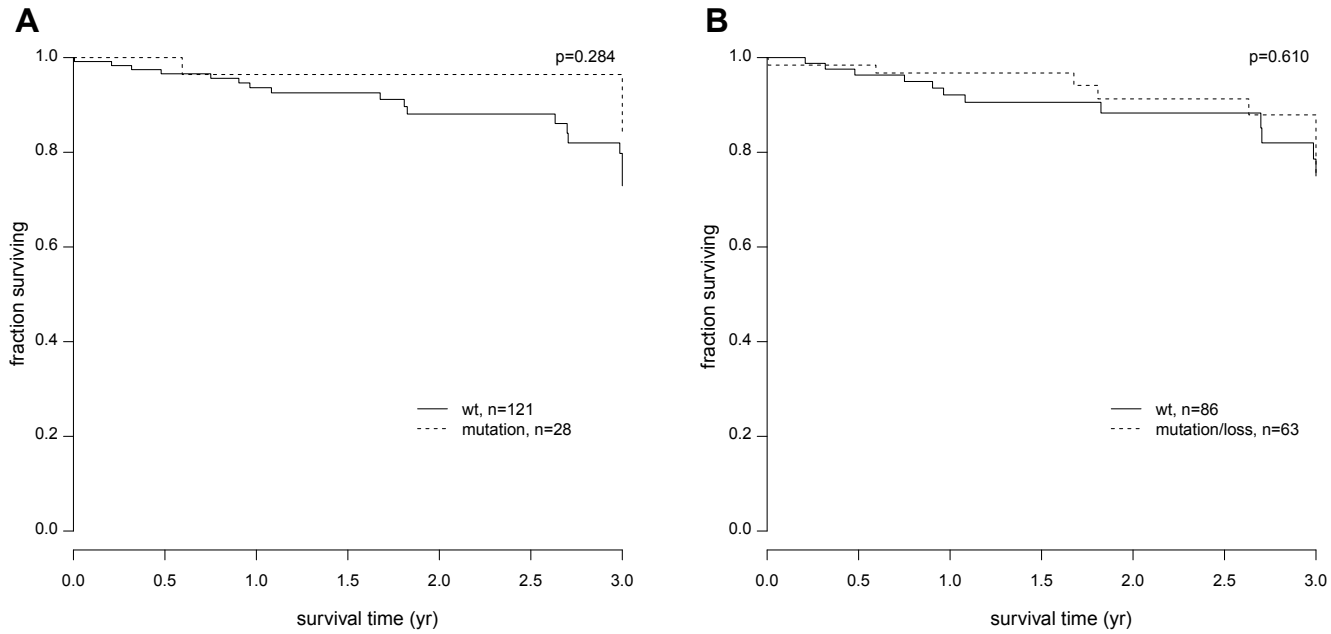
Supplemental Figure S2N8. Associations between cigarette smoking and counts of mutational signatures in HPV-negative OSCC, using exome data. Scatter plots show associations in HPV-negative OSCC (n = 272). *X-axis*, pack-years of cigarette smoking; *y-axis*, $\log_{10}$-transformation of count of mutational signatures, per tumor (*red dots*). Pearson's correlation coefficient and adjusted p-values using multiple testing correction with FDR were calculated for each signature (see also **Supplemental Table S2N**).



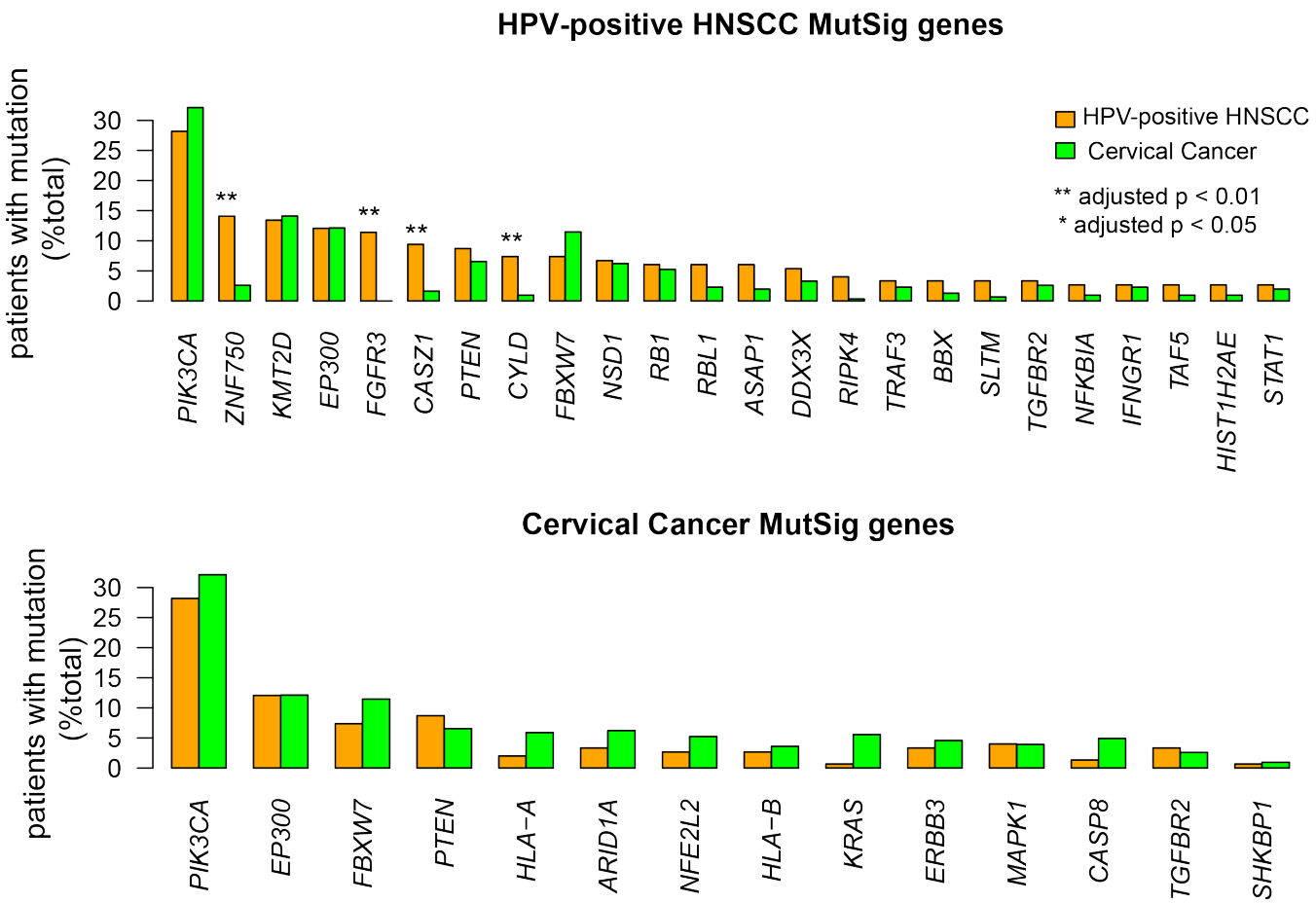
Supplemental Figure S2N9. Correlation between alcohol consumption (amount per day) and signature fraction in exons from 156 HPV-negative samples (exons). X-axis shows amount of alcohol consumption per day. Y-axis shows COSMIC signature fraction. We calculated Pearson's correlation coefficient. P-values are adjusted for multiple testing by FDR method. None of the 30 signatures showed statistically significant correlation with drinking in HPV-positive tumors (data not shown). See also **Supplemental Table S2O**.



Supplemental Figure S2N10. Correlation between alcohol consumption (amount per day) and the number of signature SNVs in exons from 156 HPV-negative samples (exons). X-axis shows amount of alcohol consumption per day. Y-axis shows log transformed signature SNV counts. We calculated Pearson's correlation coefficient. P-values are adjusted for multiple testing by FDR method. None of the 30 COSMIC signatures show statistically significant correlation with drinking in HPV-positive tumors (data not shown). See also **Supplemental Table S2O.**

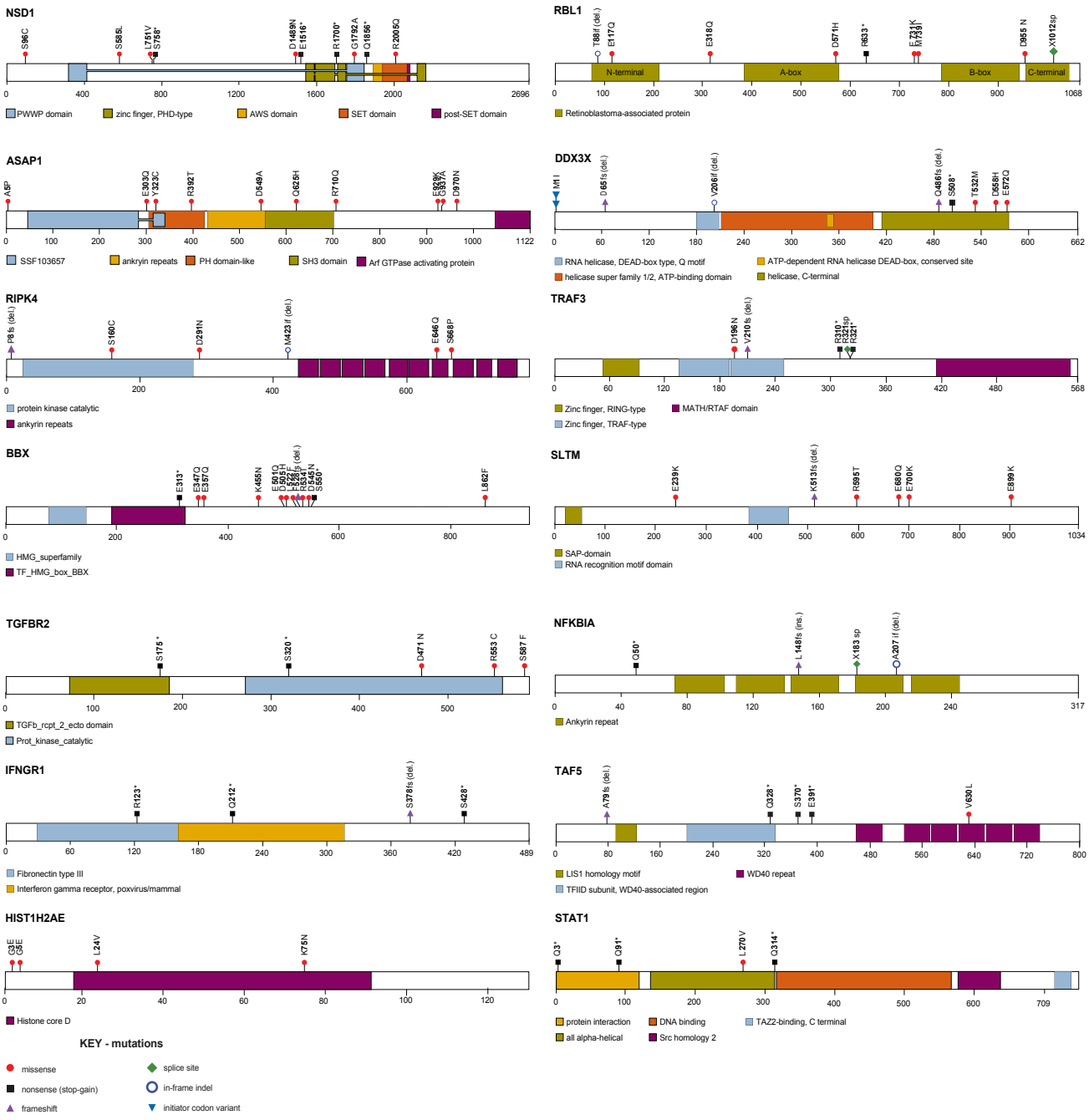


Supplemental Figure S2O. Lack of association between particular NF-κB pathway mutations and patient survival outcomes. Shown here are Kaplan-Meier survival plots for patients with HPV-positive OSCC having evaluable survival data (n = 149) harboring (A) mutations in vs. wildtype (wt) *CYLD*, *TRAF3*, *DDX3X*, *NFKBIA*; or (B) mutations and/or local copy number losses in vs. wildtype (wt) *CYLD*, *TRAF3*, *DDX3X*, *NFKBIA*. Statistical differences between survival curves (p-values) were tested with the G-rho family of tests using the surv-diff function of the survival package in R.



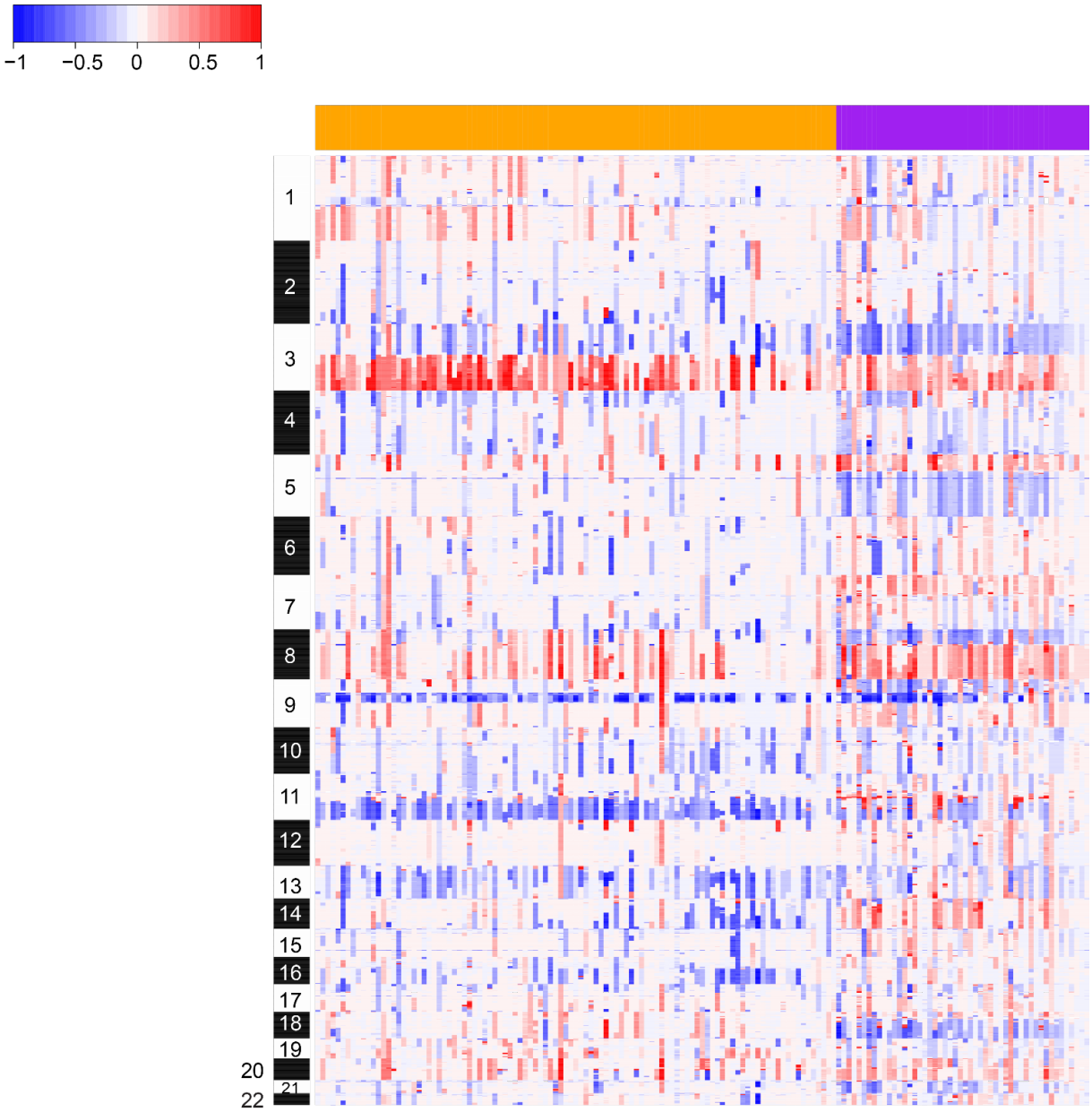
Supplemental Figure S3. Comparison of mutation frequency in MutSig identified significantly mutated genes (HPV-positive OSCC vs. cervical cancer)

Bar graphs display fractions of all (*gold*) HPV-positive OSCC and (*green*) cervical cancers bearing somatic variants (SNVs and small indels) in (*top*) 24 MutSig genes most frequently mutated in HPV-positive OSCC and in (*bottom*) 14 genes most frequently mutated in cervical cancer. *Asterisk*, adjusted p value < 0.05; *double asterisks*, adjusted p value < 0.01, Fisher's Exact test.

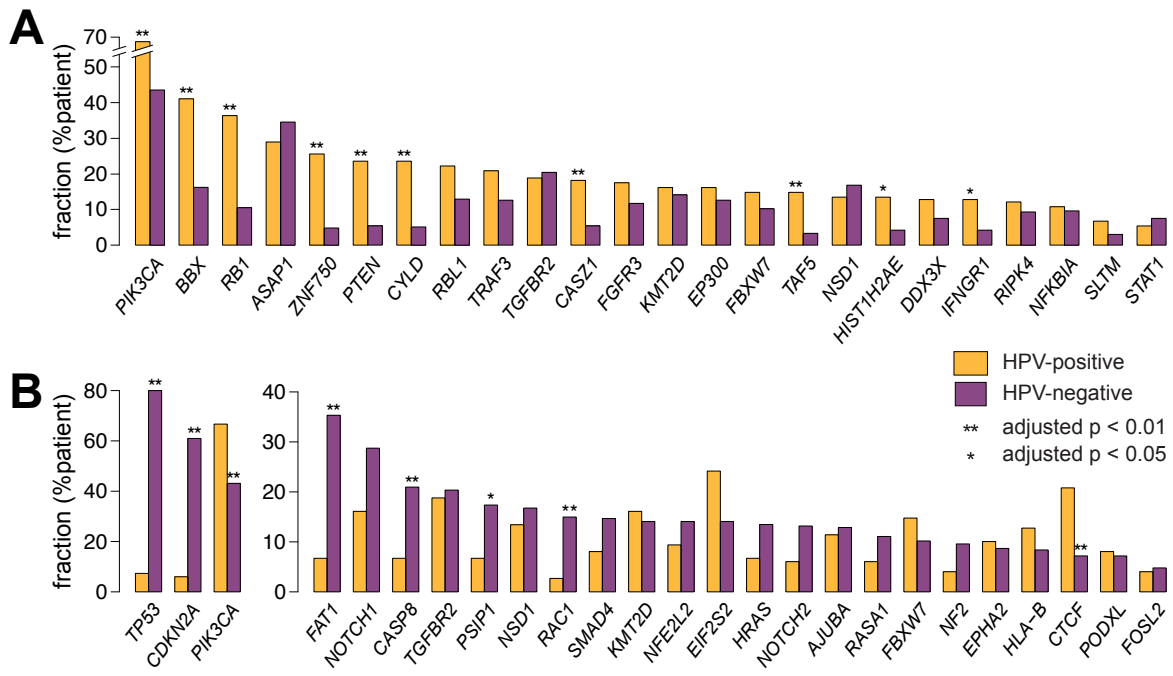


Supplemental Figure S4. Somatic mutations alter protein structures.

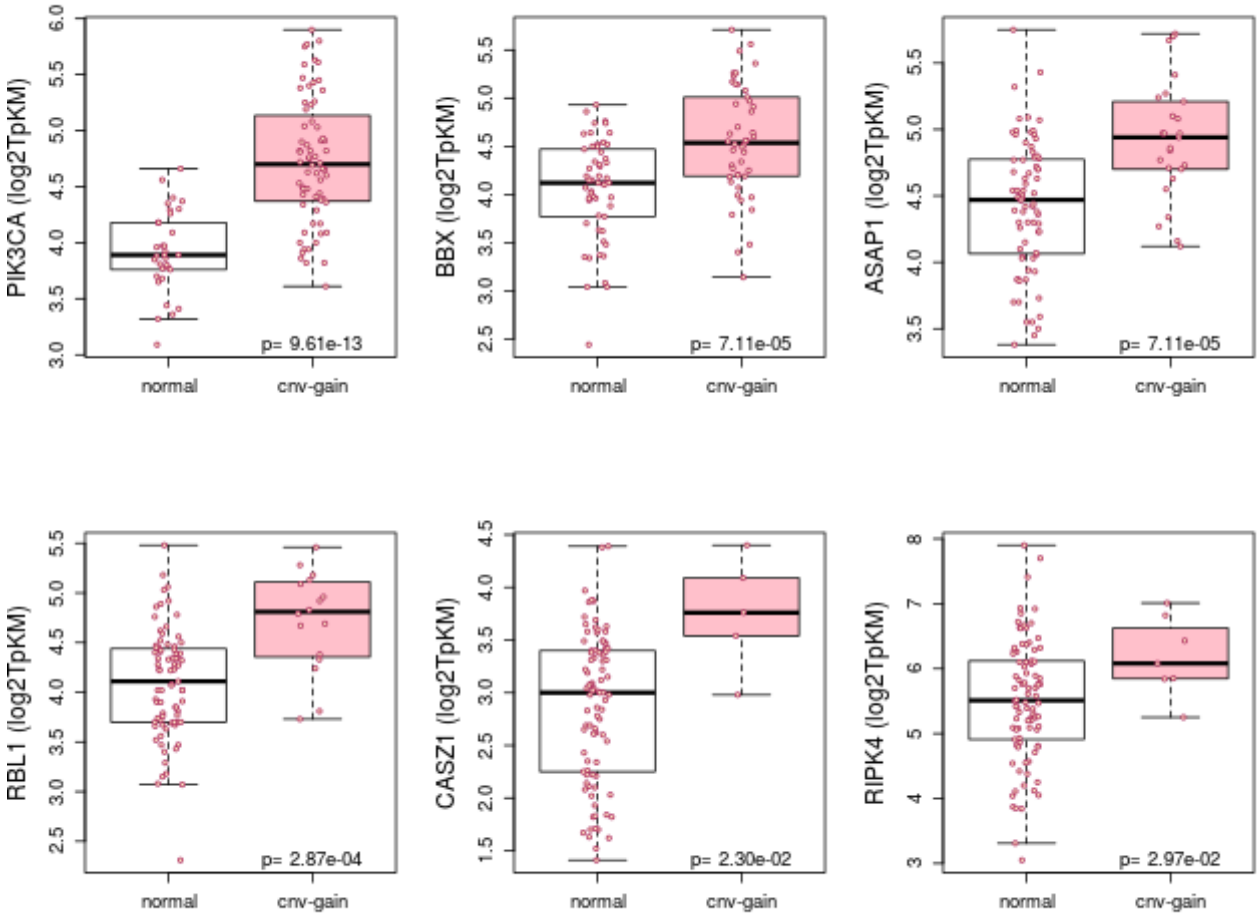
Schematics of protein domains, displaying sites of mutations identified in 14 frequently mutated genes identified in HPV-positive OSCC (also cf. **Figure 4**). *Insets*: color-coded, annotated protein domains. *Key, bottom right*: types of mutations identified in each gene product: *red circle*, missense mutation; *black square*, stop-gain (nonsense); *purple triangle*, frameshift; *green diamond*, splice site; *open blue circle*, in-frame indel (see also **Figure 4; Supplemental Table S4**).



Supplemental Figure S5A. Copy number gains and losses in OSCC detected in WGS data. Heat map shows chromosomal segment gains and losses in (*left*) 103 HPV-positive and (*right*) 50 HPV-negative OSCC, based on WGS data. Somatic genomic copy number alterations were detected by CNAnorm in 2 kb windows and were visualized by R heatmap2 package in 500 kb windows. *X-axis*, individual OSCC studied by WGS (*top*, *orange*, 103 HPV-positive OSCC; *purple*, 50 HPV-negative OSCC); *y-axis*, genomic coordinates in Chromosomes 1 – 22 (*left*). *Colors*, *top left*, heat map scale of $\log_2$ transformation of estimated ploidy, $\log_2$(ploidy estimate / 2); *red*, amplifications; *blue*, deletions.

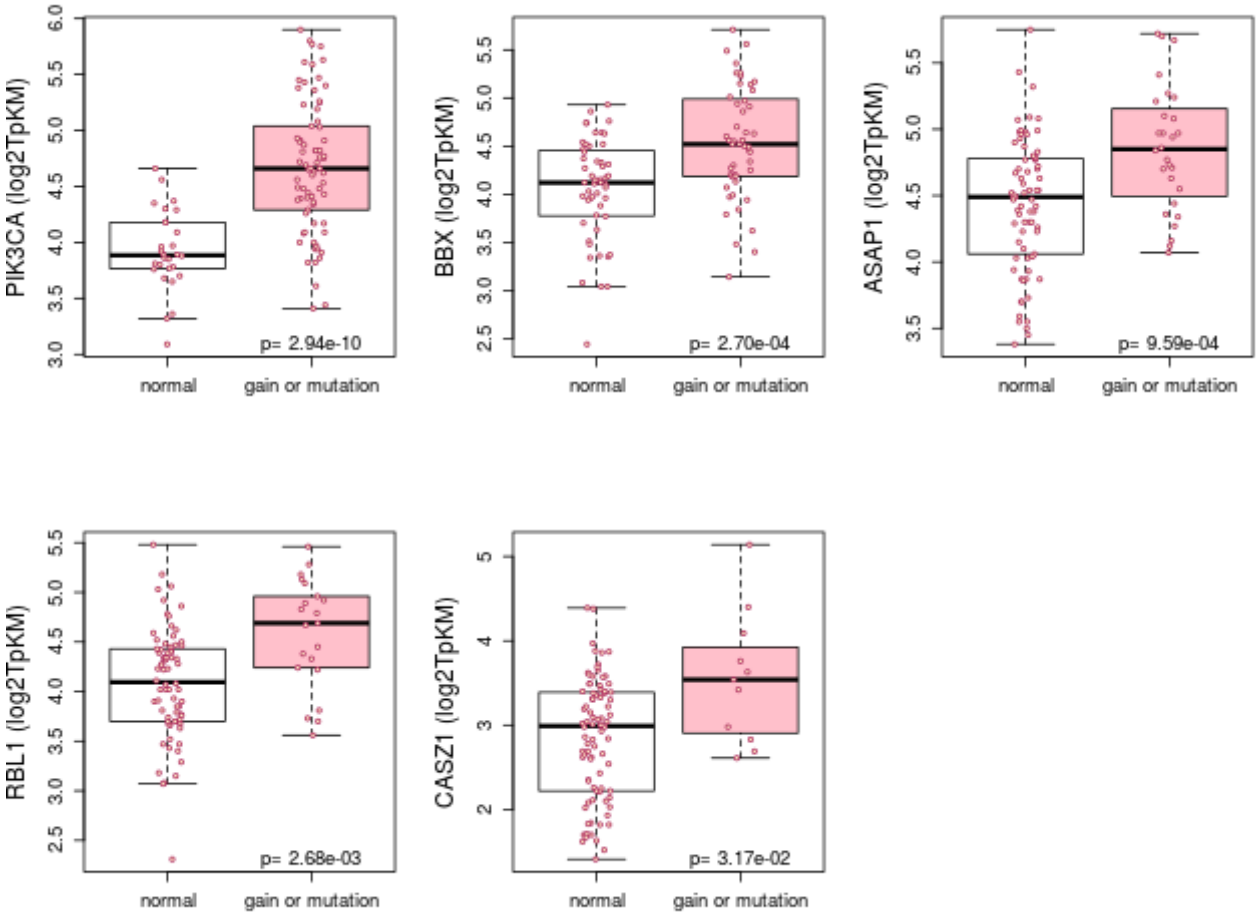


Supplemental Figure S5B. Frequencies of somatic variants and CNVs in HPV-positive vs. HPV-negative OSCC. Bar graphs display fractions of all (*gold*) HPV-positive and (*purple*) HPV-negative OSCC bearing somatic variants (SNVs and small indels) and CNVs in (A) 24 MutSig genes detected in 149 HPV-positive OSCC and in (B) 25 genes in 335 HPV-negative OSCC. *Key: asterisks*, genes with significantly different mutation frequencies in HPV-positive vs. HPV-negative OSCC (**, $p < 0.01$; *, $p < 0.05$ by Fisher's exact test); *gold*, frequency of mutations in HPV-positive cancers, *purple*, frequency of mutations in HPV-negative cancers (see also **Supplemental Tables S5A-E**).



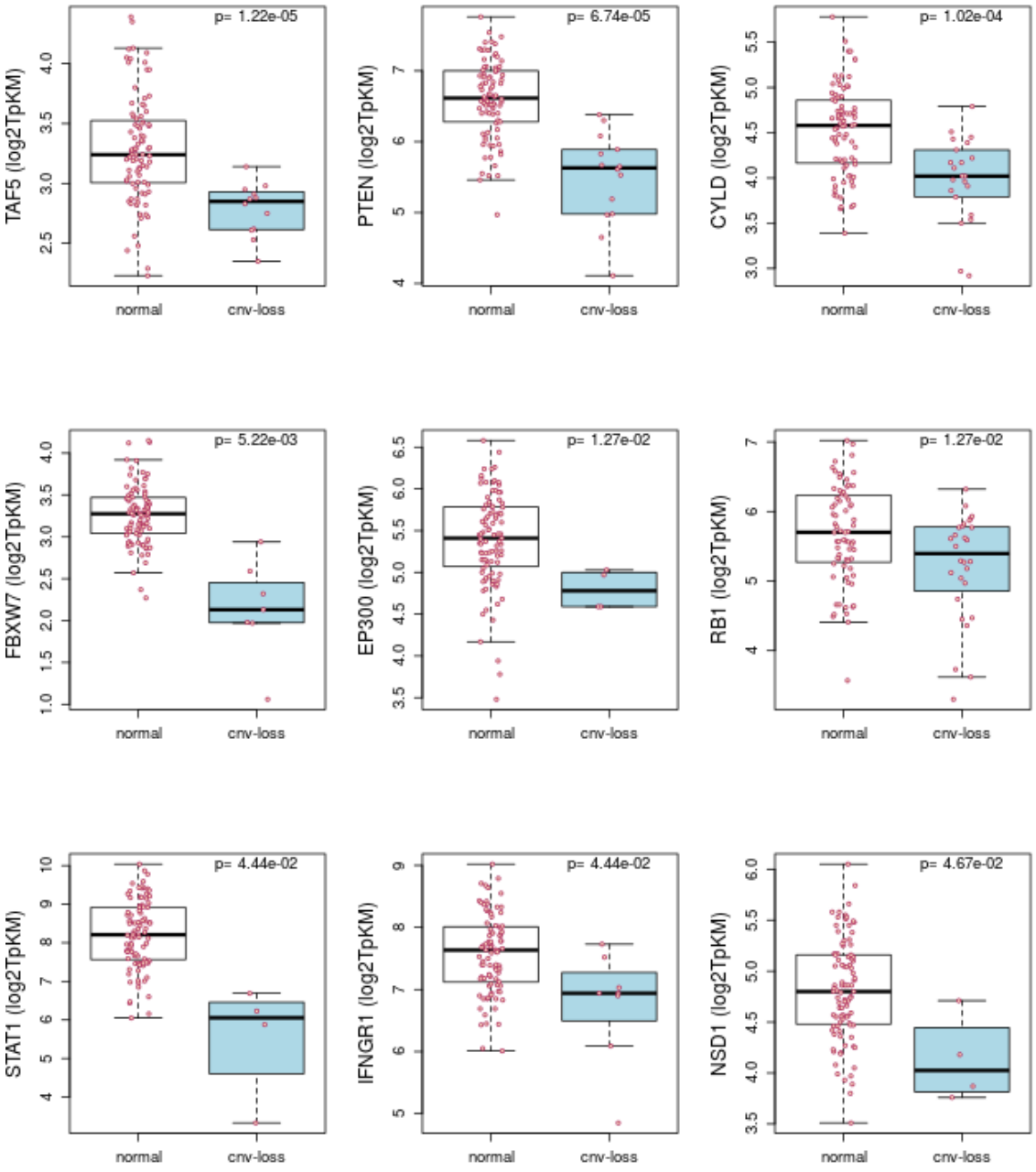
Supplemental Figure S5C. Association between CNV gains and gene expression changes in HPV-positive OSCC.

Shown here in box and whisker plots of gene expression levels (*y-axis*) in HPV-positive OSCC samples (*red dots*), presented as the $\log_2$-transformation of TPM values, for 6 of 24 highly mutated MutSig genes, comparing (*x-axis*) normal ploidy vs. local CNV gains (ploidy > 2.5 N). Out of the 13 genes with three or more samples having copy number gains using t-test (one-tailed), the 6 genes presented here showed significant differences in expression (adj. $p < 0.05$; t-test with false discovery rate multiple testing correction).



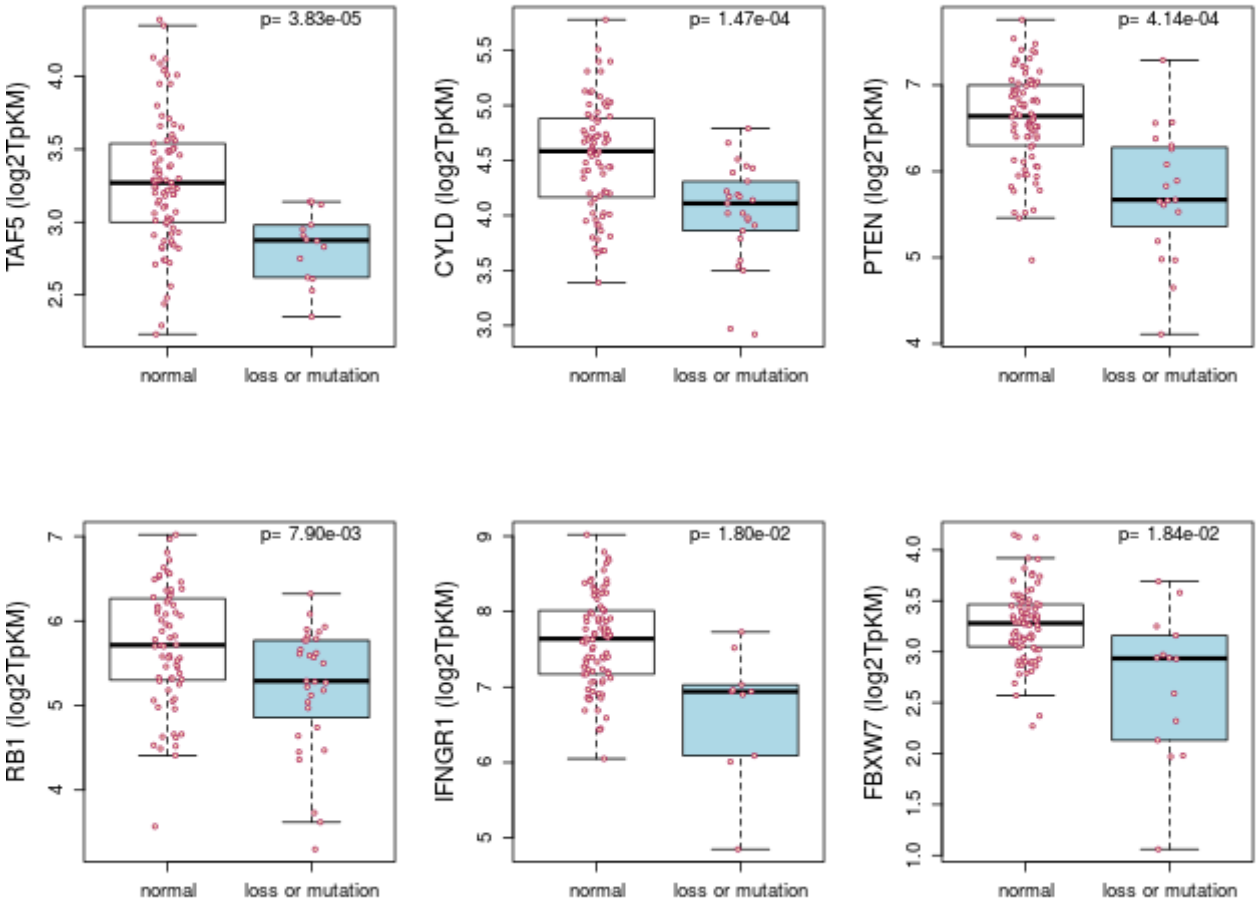
Supplemental Figure S5D. Association between CNV gain and/or mutation and gene expression changes in HPV-positive OSCC.

Shown here in box and whisker plots are gene expression levels (*y-axis*) in 101 HPV-positive OSCC studied by WGS and RNA-seq (*red dots*), presented as the $\log_2$-transformation of TPM values, for 5 of 24 highly mutated MutSig genes, comparing (*x-axis*) normal ploidy or lack of mutations vs. local CNV gains (ploidy > 2.5 N) or presence of mutations. Out of the 24 genes with three or more samples having copy number gains and/or mutations using t-test (one-tailed), the 5 genes presented here showed significant differences in expression (adj. $p < 0.05$; t-test with false discovery rate multiple testing correction).

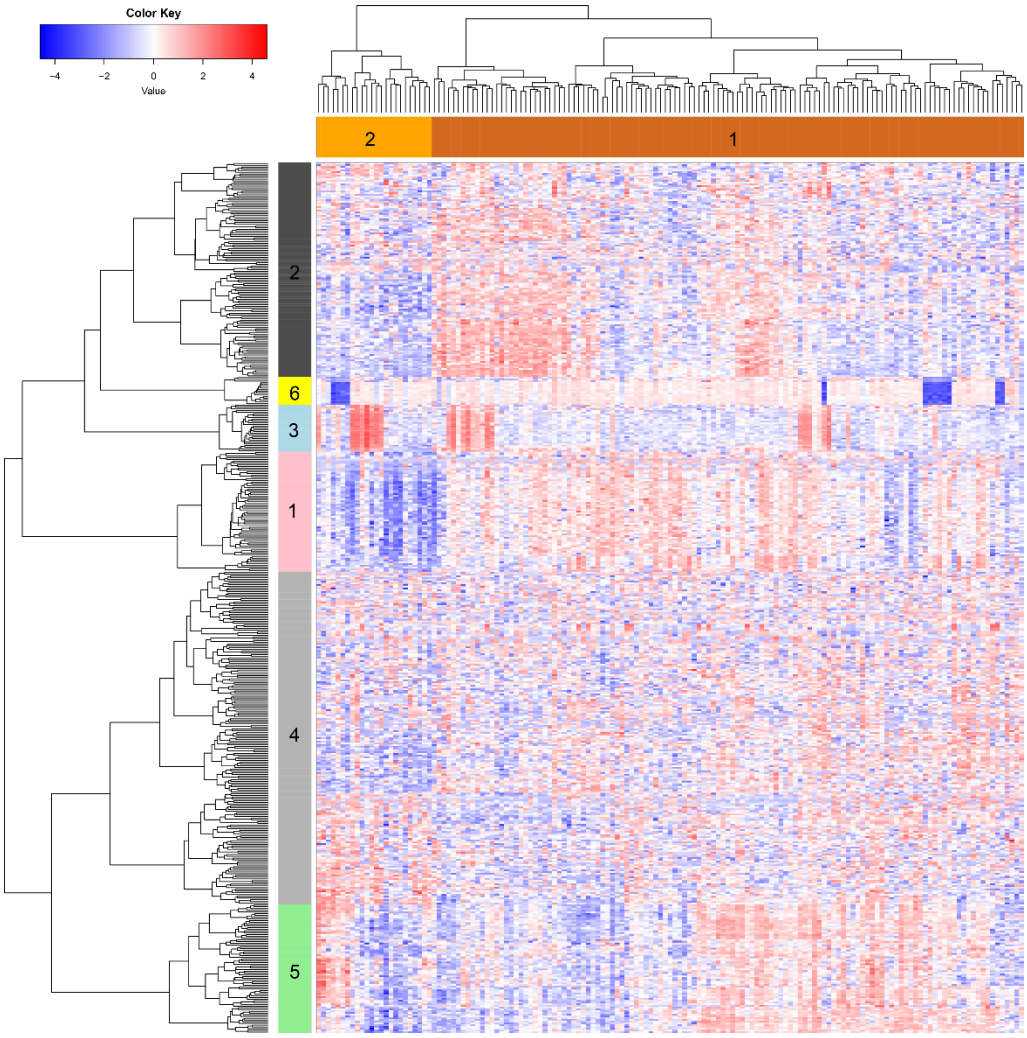


Supplemental Figure S5E. Association between CNV loss and gene expression changes in HPV-positive OSCC.

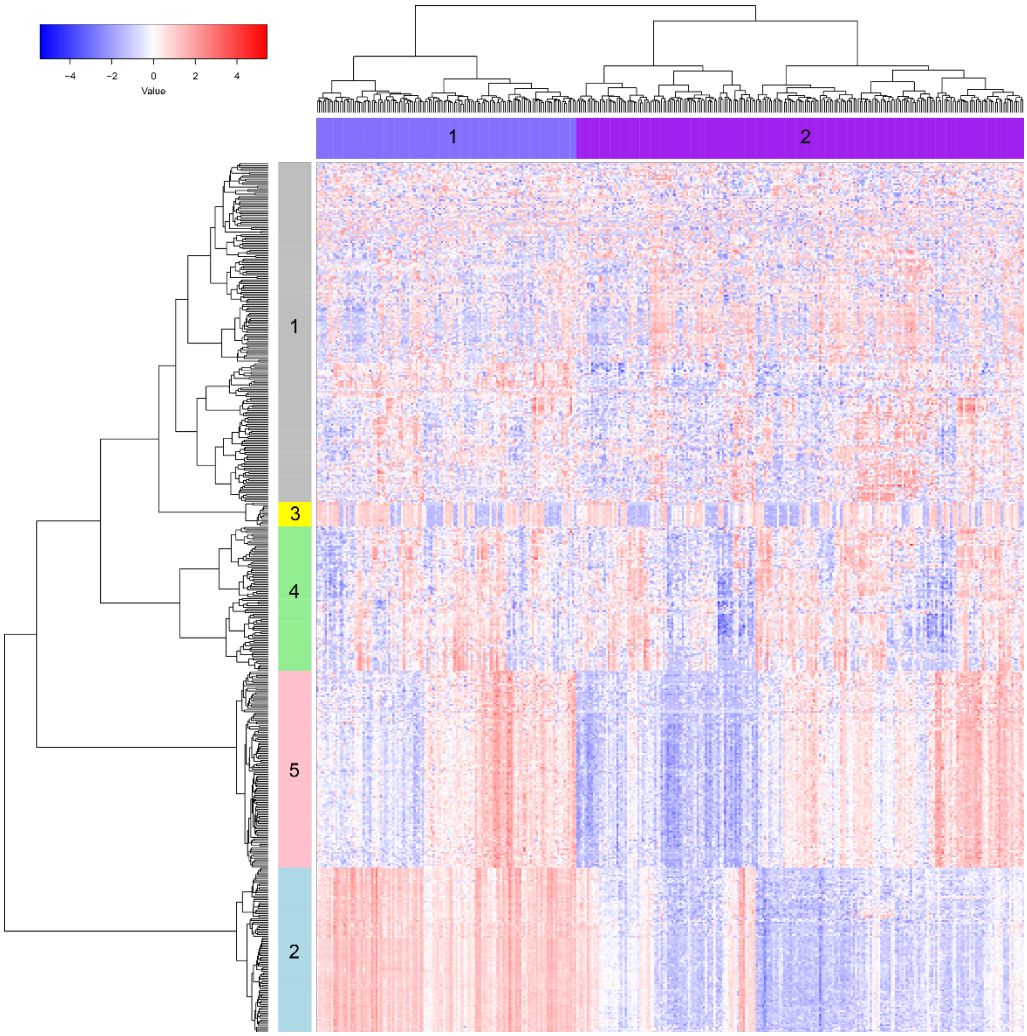
Shown here in box and whisker plots of gene expression levels (*y-axis*) in 101 HPV-positive OSCC studied by WGS and RNA-seq (*red dots*), presented as the log2-transformation of TPM values, for 9 of 24 highly mutated MutSig genes, comparing (*x-axis*) normal ploidy vs. local CNV losses (ploidy < 1.5 N). Out of the 20 MutSig genes with three or more samples having copy number losses using t-test (one-tailed), the 9 genes presented here showed significant differences in expression (adj. p < 0.05; t-test with false discovery rate multiple testing correction).



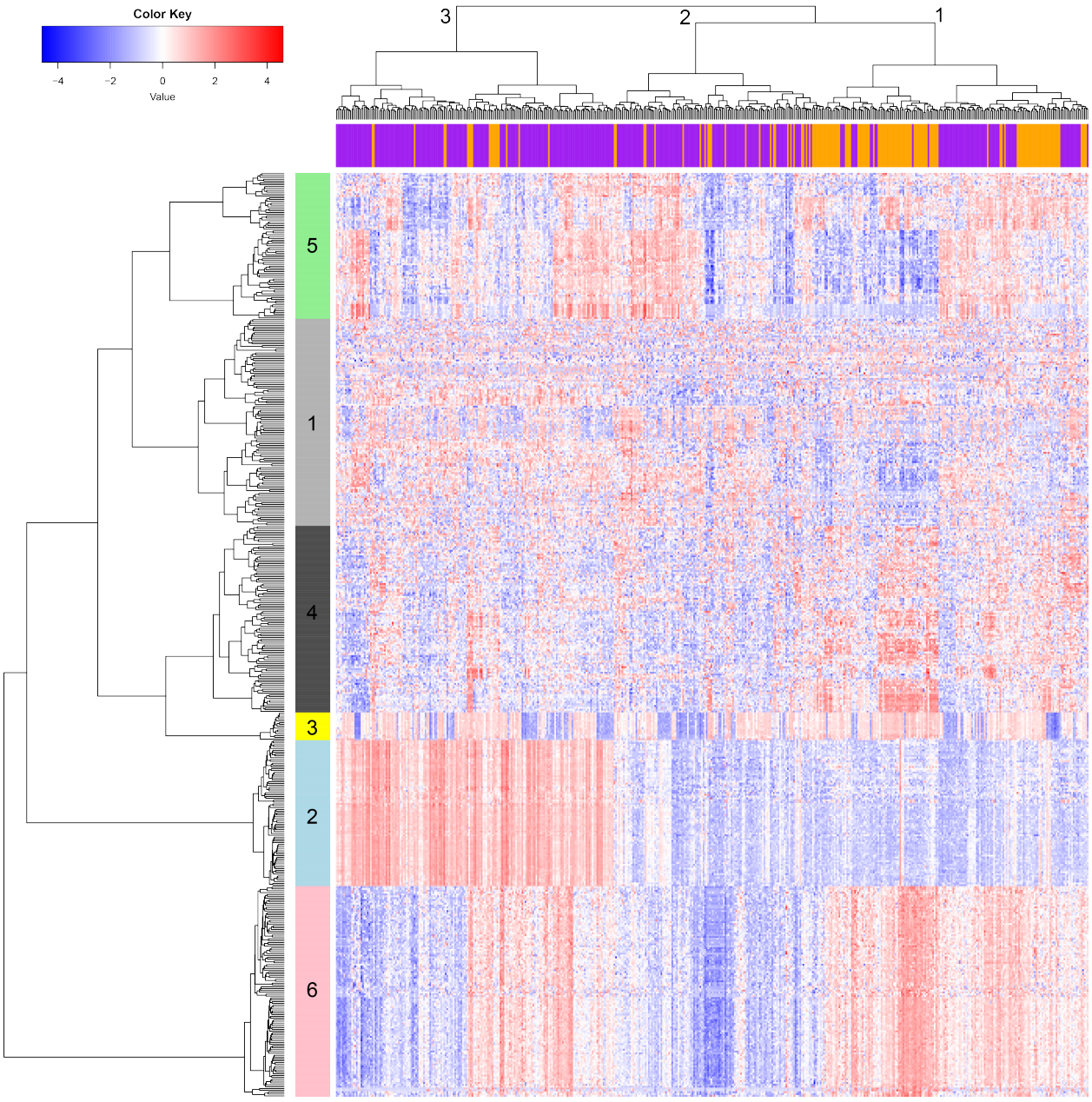
Supplemental Figure S5F. Association between CNV loss and/or mutations and gene expression changes in HPV-positive OSCC.
Shown here in box and whisker plots of gene expression levels (*y-axis*) in 101 HPV-positive OSCC studied by WGS and RNA-seq (*red dots*), presented as the log2-transformation of TPM values, for 6 of 24 highly mutated MutSig genes, comparing (*x-axis*) normal ploidy and lack of mutations vs. local CNV losses (ploidy < 1.5 N) and/or mutations. The 6 genes presented here showed significant differences in expression (adj. p < 0.05; t-test with false discovery rate multiple testing correction).



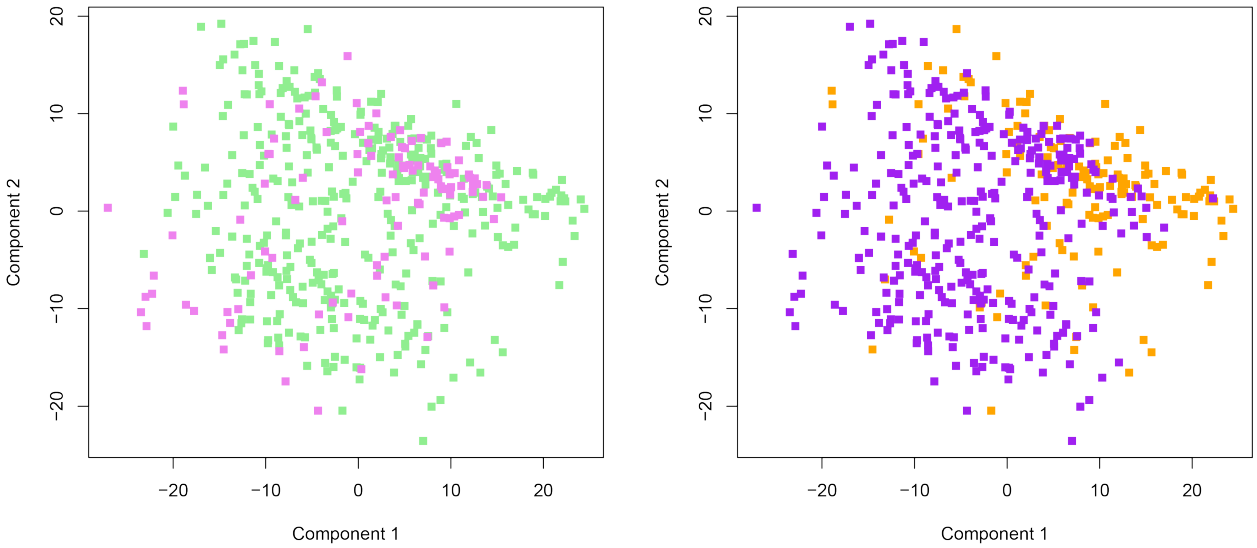
Supplemental Figure S5G. Variation in gene expression in HPV-positive OSCC.
Heat map shows the 500 most differentially expressed genes (*y-axis*) in HPV-positive OSCC with available RNA-seq data (*x-axis*; n = 147; including 84 Ohio cohort and 63 TCGA cancers). Numbers of RNA-seq reads aligning to each gene model were quantified, normalized and batch-corrected, and expression values were calculated by the $\log_2$-transformation of TpKM. We calculated variation in expression for GenCode v. 17 gene models. Unsupervised hierarchical clustering of the most highly variable was performed using the Ward D2 method, resulting in 6 groups of genes (*left*): *pink*, n = 69, group 1, complement activation, immunoglobulin production; *black*, n = 123, group 2, skeletal system morphogenesis; *light blue*, n = 27, group 3, muscle filament sliding, skeletal muscle contraction; *gray*, n = 191, group 4, epithelial development, cell differentiation; *green*, n = 74, group 5, cornification, keratinization; *yellow*, n = 16, group 6, no significantly enriched ontology terms. The HPV-positive OSCC were grouped into 2 groups (*top*): *dark orange*, n = 123, group 1, higher expression of complement activation genes; *orange*, n = 24, group 2 (see also **Supplemental Table S5U**).



Supplemental Figure S5H. Differential gene expression in HPV-negative OSCC.
Heat map shows the 500 most differentially expressed genes (*y-axis*) in HPV-negative OSCC with available RNA-seq data (*x-axis*; n = 335; including 26 Ohio cohort and 309 TCGA cancers). Numbers of RNA-seq reads aligning to each gene model were quantified, normalized and batch-corrected, and expression values were calculated by the $\log_2$-transformation of TpKM. Unsupervised hierarchical clustering was performed using the Ward D2 method, resulting in 6 groups of genes analyzed using Panther overrepresentation test to determine enrichment in biological process gene ontology terms (*left*): *gray*, n = 195, group 1, fat soluble vitamin metabolic process, cellular hormone metabolic process; *light blue*, n = 95, group 2, muscle filament sliding, skeletal muscle contraction; *yellow*, n = 14, group 3, no significantly enriched ontology terms; *green*, n = 83, group 4, cornification, keratinization; *pink,* n = 113, group 5, complement activation, immunoglobulin production. HPV-negative OSCC were clustered into 2 groups (*top*): *blue*, n = 123, group 1, higher expression of skeletal muscle contraction genes; *purple*, n = 212, group 2 (see also **Supplemental Table S5V**).

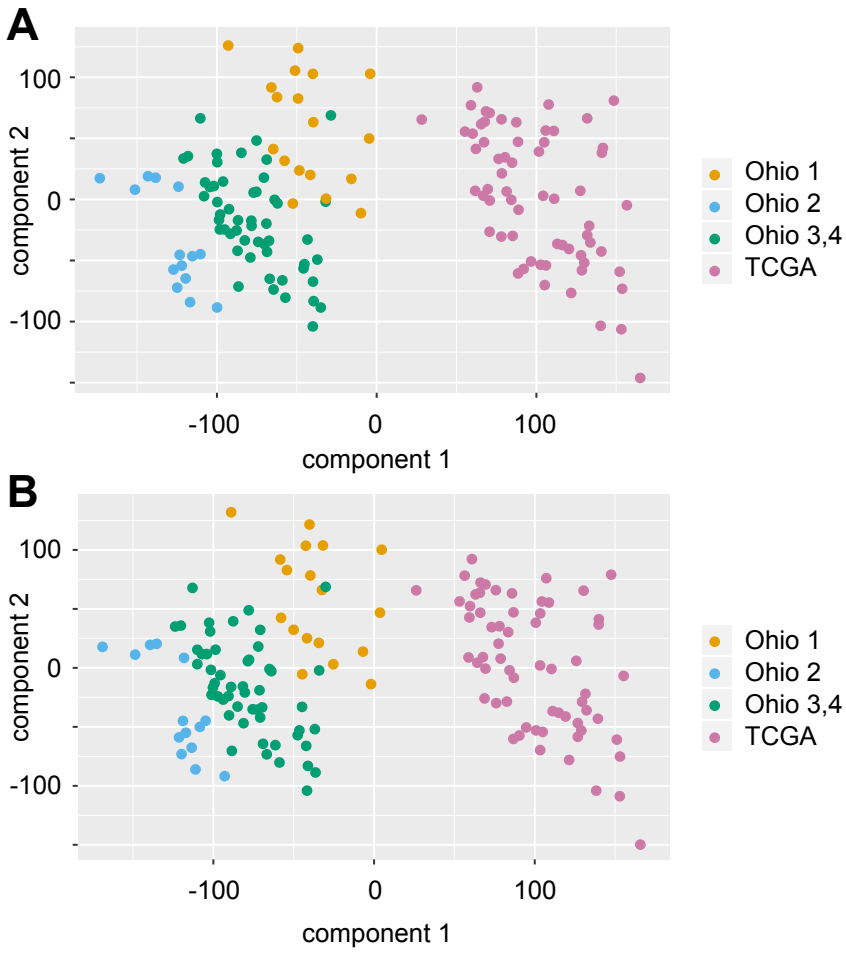


Supplemental Figure S5l. Differential gene expression in OSCC. Heat map shows the 500 most differentially expressed genes (*y-axis*) in both (*yellow*) HPV-positive and (*purple*) HPV-negative OSCC with available RNA-seq data (*x-axis*; n = 482; including 110 Ohio cohort and 372 TCGA cancers). Numbers of RNA-seq reads aligning to each gene model were quantified, normalized and batch-corrected, and expression values were calculated by the log2-transformation of TpKM. Unsupervised hierarchical clustering was performed using the Ward D2 method, resulting in 6 groups of genes analyzed using Panther overrepresentation test to determine enrichment in biological process gene ontology terms (*left*): *gray*, n = 112, group 1, no significantly enriched ontology terms; *light blue*, n = 79, group 2, muscle filament sliding, skeletal muscle contraction; *yellow*, n = 15, group 3, no significantly enriched ontology terms; *black*, n = 101, group 4, retinoic acid metabolic process, neuron fate specification; *green,* n = 79, group 5, cornification, keratinization; *pink*, n = 114, group 6, complement activation, immunoglobulin production. The OSCC were clustered into 3 groups (*top*): cluster 1, n = 169; cluster 2, n = 135; cluster 3, n = 178. Representation of HPV-positive OSCC in these clusters (*top*) was: cluster 1, 58% (96 out of 169); cluster 2, 24% (32 out of 135); cluster 3, 11% (19 out of 178). See also **Supplemental Table S5W.**



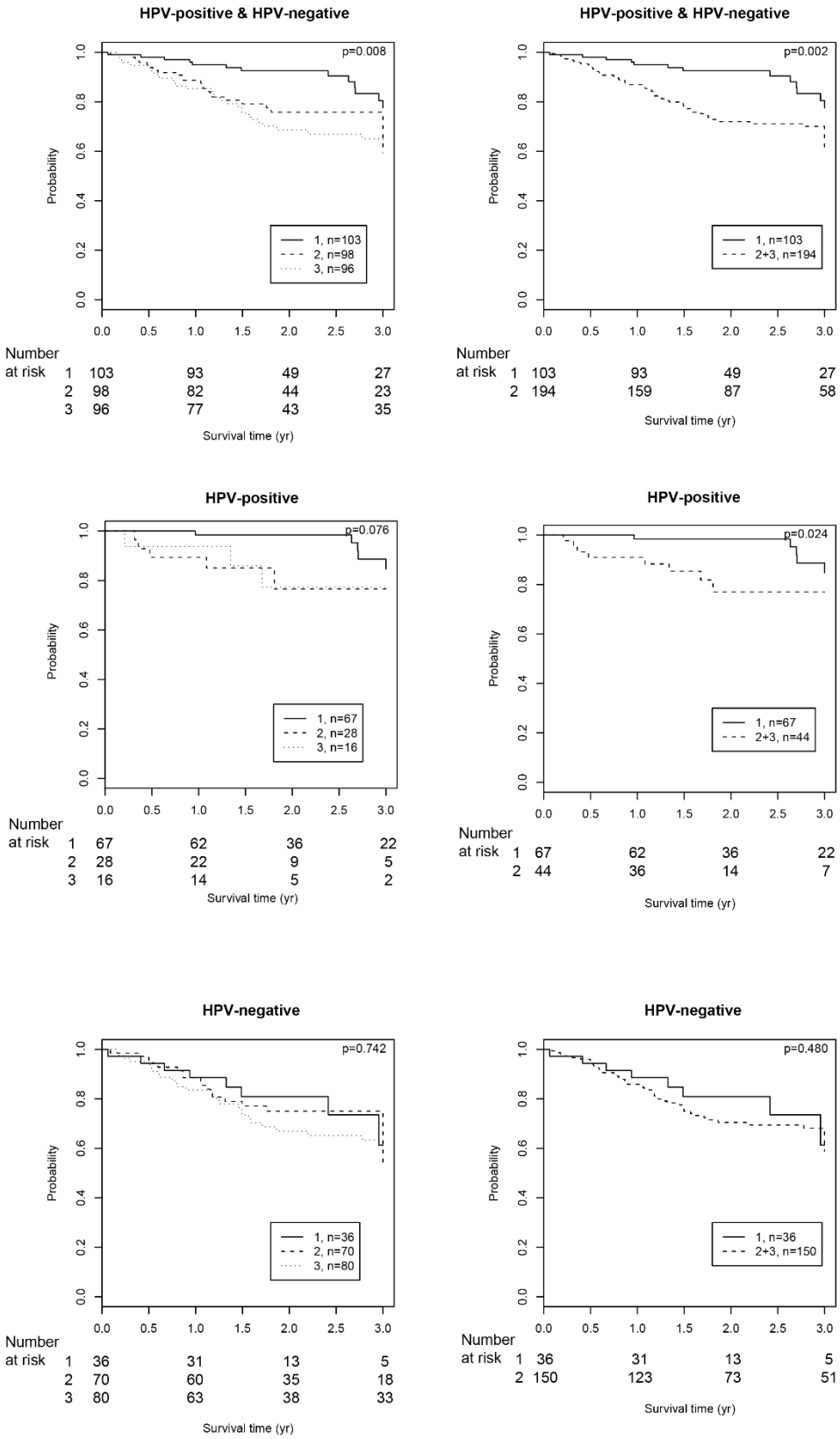
Supplemental Figure S5J. Principal components analysis of the most variably expressed genes in OSCC categorized by sample source or HPV status.

Principal components analysis of gene expression profiles was performed for the 500 most variably expressed genes in 482 OSCC (including 147 HPV-positive and 335 HPV-negative OSCC). The scatterplots display the 2 most significant two components: *x-axis*, component 1; *y-axis,* component 2. *Dots*, RNA-seq data from individual cancers. *(Left*) *pink*, Ohio cohort (n = 110); *green*, TCGA (n=372). (*Right*) *yellow*, HPV-positive (n=147), *purple*, HPV-negative (n=335) OSCC.



Supplemental Figure S5K. Principal components analysis of OSCC transcriptome data before and after reanalysis with harmonized RNA-seq aligner and reference genome.

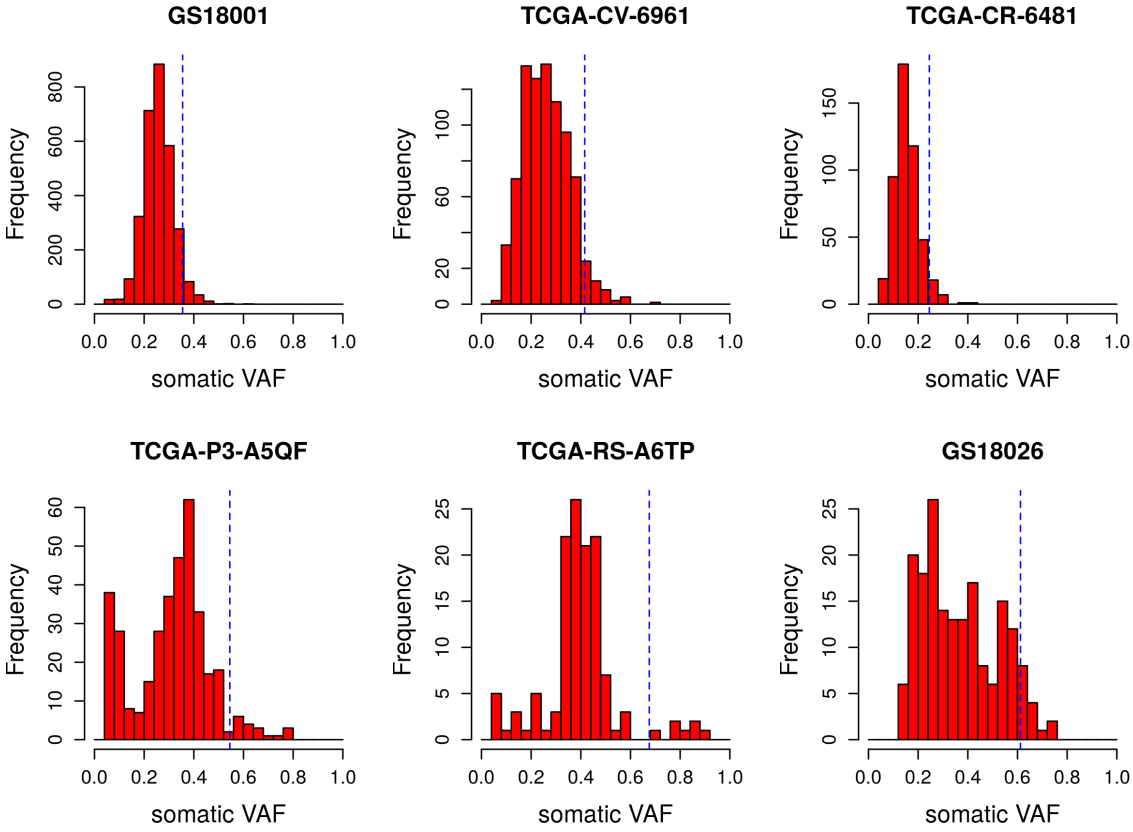
Principal components analysis of OSCC transcriptomes from RNA-seq datasets was performed to assess potential changes in RNA-seq sample data evaluated (A) before and (B) after reanalysis with harmonized pipeline STAR 2.4.2a and single reference genome assembly GRCh37.p13. The results revealed minimal impacts of reanalysis with the single aligner and reference genome assembly version differences. *Key, symbols,* individual OSCC sample transcriptomes; *orange*, batch 1 from Ohio cohort, in (A) aligned to GRCh37 human reference genome using STAR aligner 2.3.1z for 18 HPV-positive and 2 HPV-negative OSCC; *blue*, batch 2, in (A) aligned to GRCh37 using STAR 2.4.0c for 14 HPV-positive OSCC; *green*, batches 3 and 4, in (A) aligned to GRCh37.p13 reference assembly which includes non-canonical chromosomes, using STAR 2.4.2a for 52 HPV-positive and 24 HPV-negative OSCC; *lavender*, TCGA samples, in (A) aligned to GRCh37.p13 reference assembly which includes non-canonical chromosomes, using STAR 2.4.2a for 17 HPV-positive and 24 HPV-negative OSCC.



Supplemental Figure S5L. Survival analysis of OSCC patients grouped by gene expression clusters.

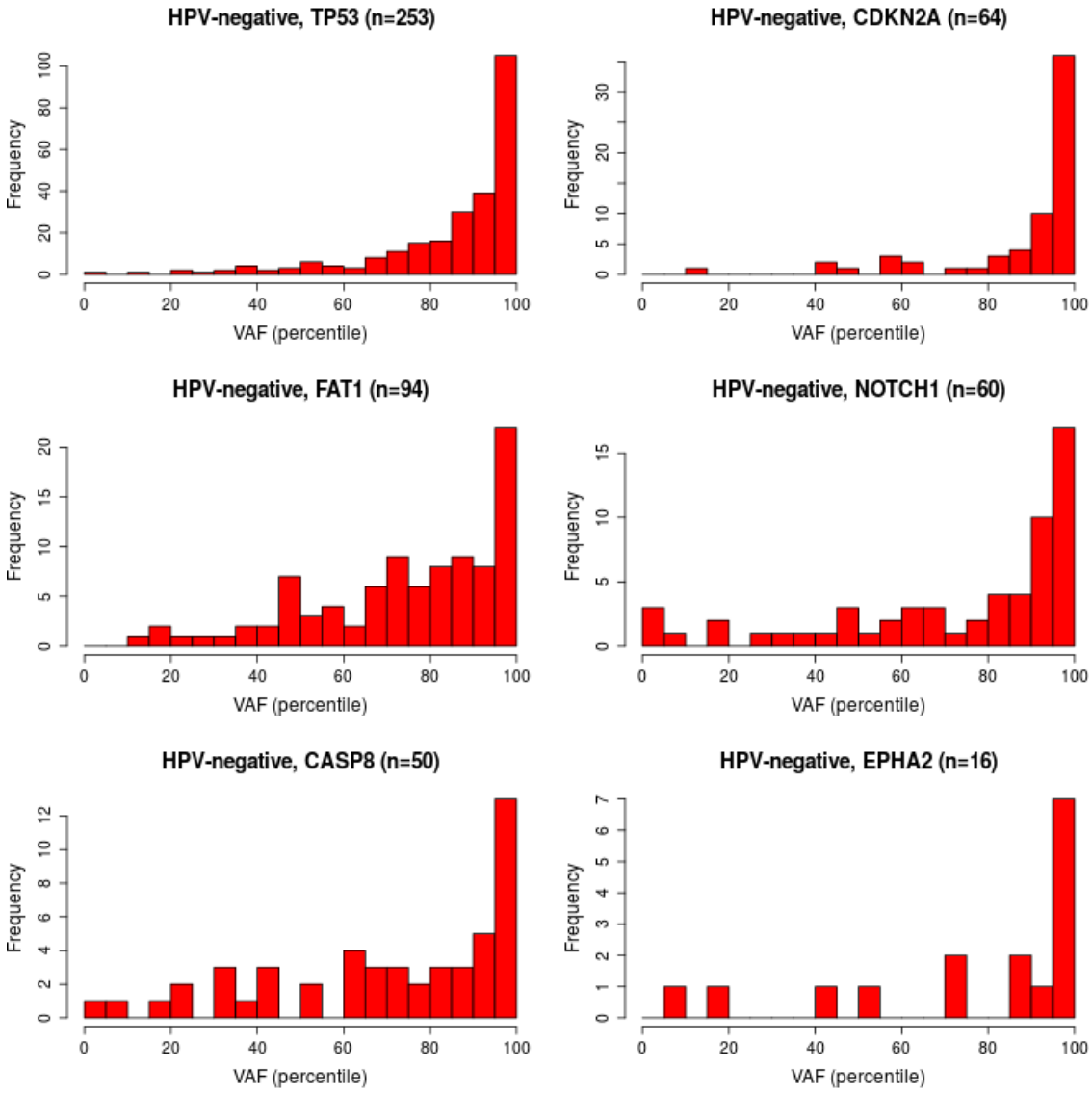
The 500 most variably expressed genes identified by RNA-seq grouped 482 OSCC (including 147

HPV-positive and 335 HPV-negative OSCC) into three sample clusters (Figure S5H). Of these, 297 patients had survival data available. *X-axis*, survival time from diagnosis (years); *y-axis*, probability of survival. Shown here are survival curves documenting differences in patient survival: (*top left*) 3 sample clusters in all OSCC samples; (*top right*) 2 sample clusters where clusters 2 and 3 were combined, resulting in comparison of 1 vs. 2+3, in all OSCC samples; (*middle left*) 3 sample clusters in HPV-positive samples; (*middle right*) 2 sample clusters (1 vs. 2+3) for HPV-positive samples; (*bottom left*) 3 sample clusters for HPV-negative samples; and *(bottom right*) 2 sample clusters (1 vs. 2+3) for HPV-positive samples. Statistical differences between survival curves (p-values) were tested with the G-rho family of tests using the surv-diff function of the survival package in R. *Bottom of each panel*, numbers of surviving patients at risk at each indicated time point. Sample clusters were defined as described in **Supplemental Figure S5I**.

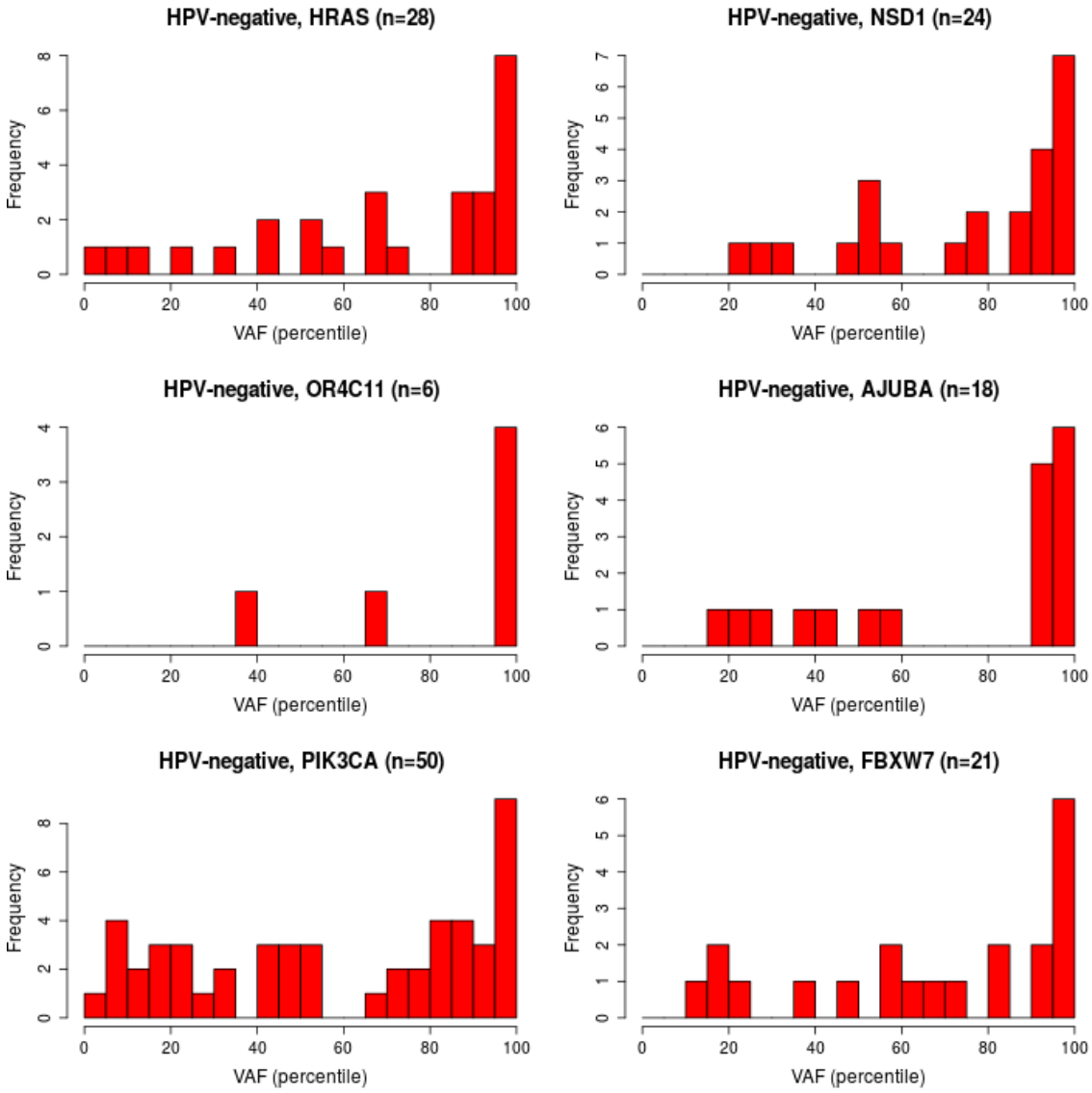


Supplemental Figure S6A. Distributions of somatic variant allelic fractions (VAF) in individual OSCC.

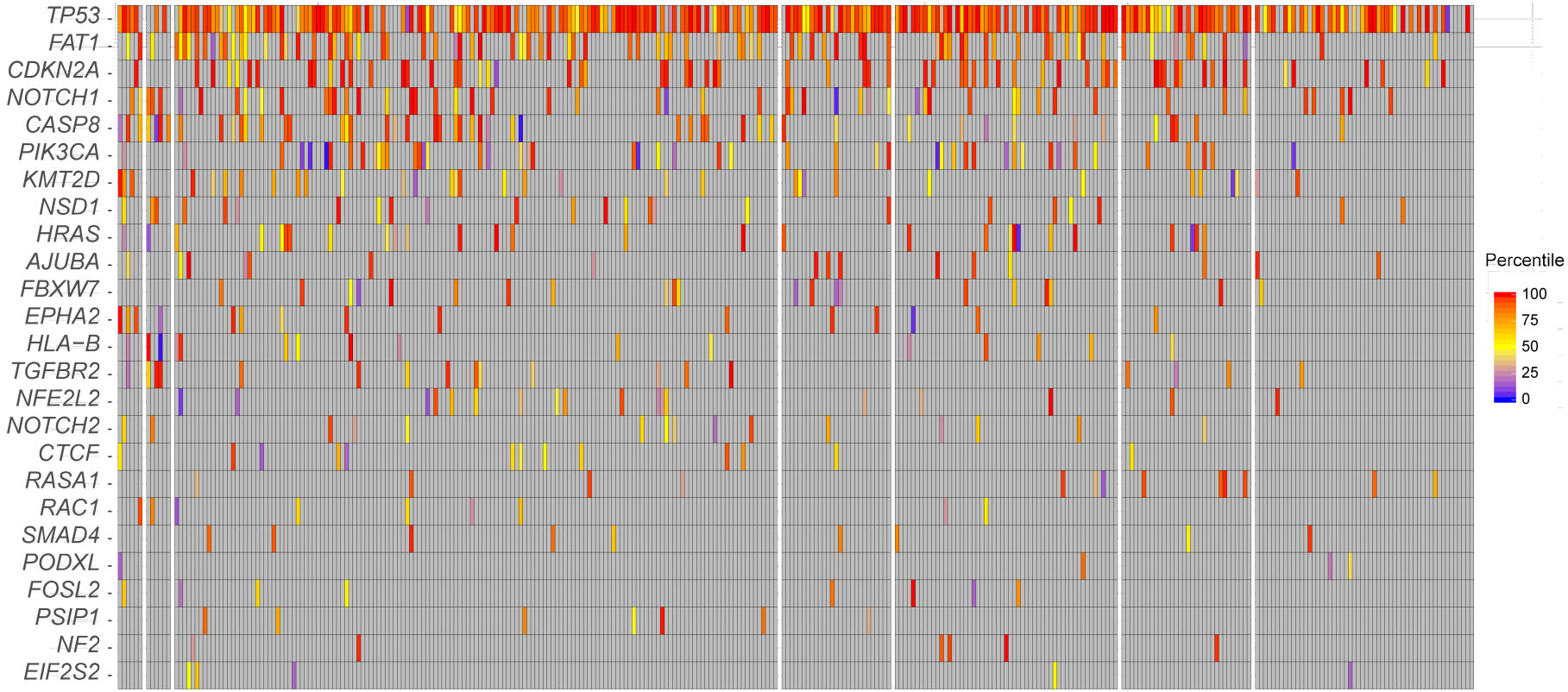
Histograms display widely variable distributions of somatic variant allelic fractions (VAFs) in six arbitrarily chosen examples of OSCC (*top*). *X-axis*, somatic VAFs; *y-axis*, counts of variants with indicated VAFs. Criteria to identify variants included their location in exons having 20x or greater depth of coverage, and at least 5% of reads supported the alternative allele (minimum VAF >5%). *Blue dotted lines*, 95th percentile of ranked VAFs in each sample.



Supplemental Figure S6B (continued on next page).

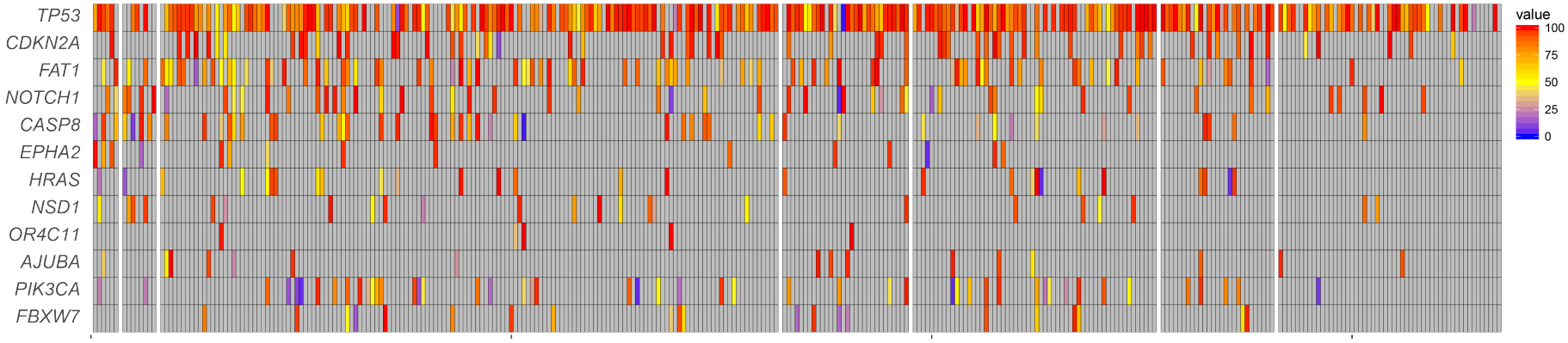


Supplemental Figure S6B. Genes disrupted by mutations with high-ranking VAFs in HPV-negative OSCC.
Histograms display significant bias in the distributions of HPV-negative OSCC sample counts (*y-axis*, frequency) harboring mutations with the indicated rankings of somatic variant allelic fractions (*x-axis*, ranked VAF, percentile) for 12 genes as indicated. Statistical enrichment of coding-change variants with VAFs recurrently ranking in the top 5% was calculated using binomial test, and multiple testing correction was performed using the R qvalue package. Of the 12 most mutated genes, 7 (*TP53, CDKN2A, FAT1, NOTCH1, CASP8, EPHA2, HRAS*) had q-value < 0.01; 3 (*NSD1, OR4C11, AJUBA*) had q-value < 0.05; and 2 (*PIK3CA, FBXW7*) had q-value < 0.1.



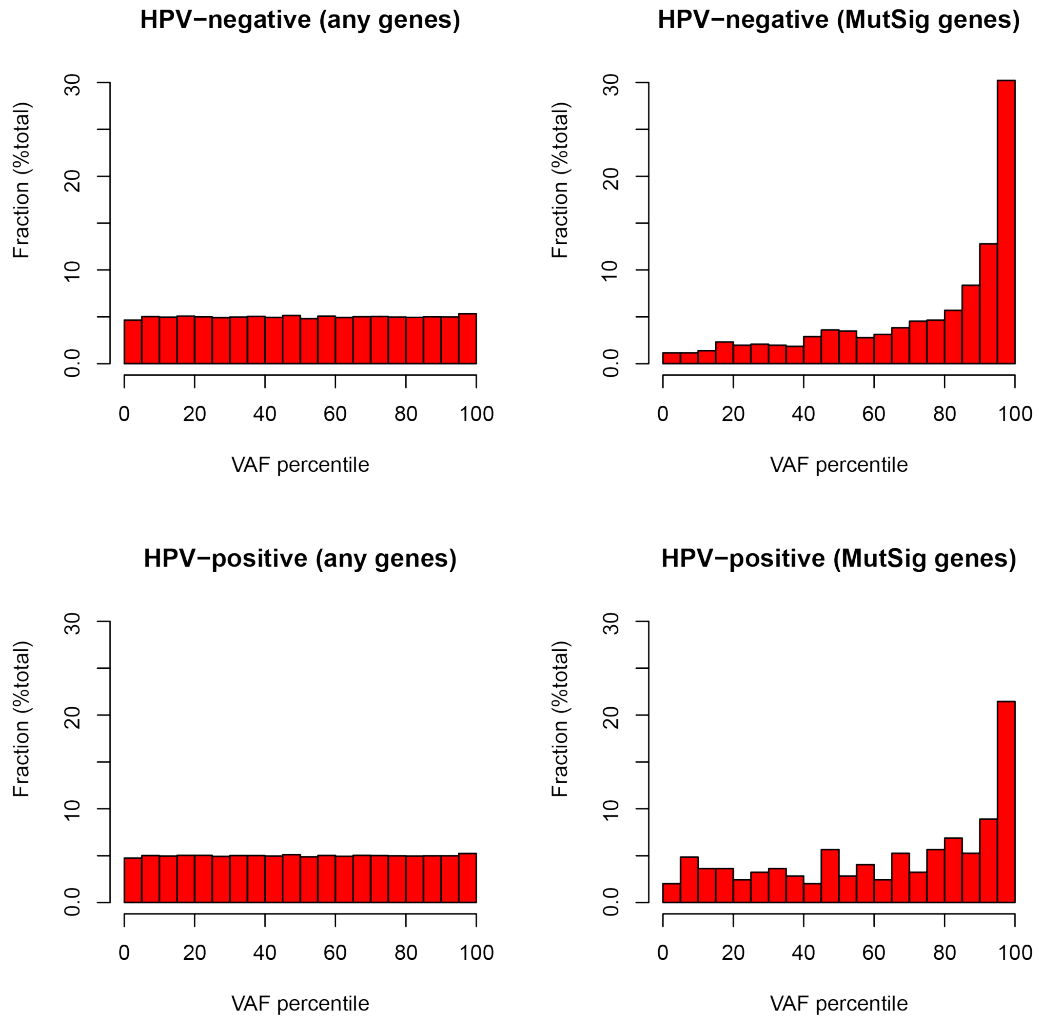
Supplemental Figure S6C. VAFs of somatic variants in significantly mutated MutSig genes in HPV-negative OSCC.

Heatmap displays percentiles (*key*, *right*) of ranked VAFs (thin vertical tiles) for SNVs in exons of 329 HPV-negative OSCC samples (*x-axis*, sorted by mutation rate) for the most highly mutated MutSig genes (*y-axis*, q < 0.2, sorted by mutation rate). *Gray tiles*, no gene variant in sample; *white vertical lines*, 6 HPV-negative samples (out of 335 total) not available from TCGA.

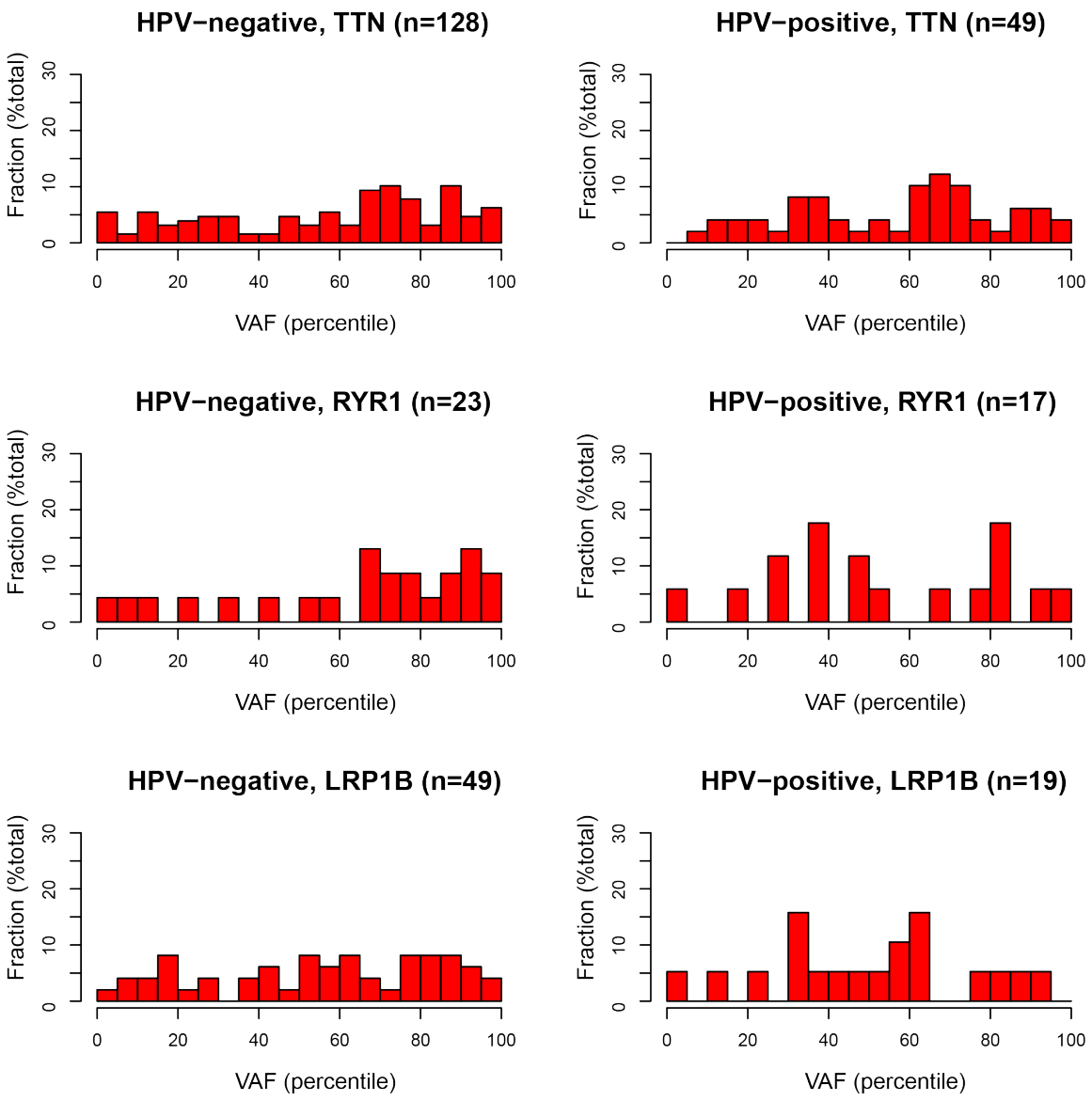


Supplemental Figure S6D. VAFs of somatic coding-change variants in most significantly mutated MutSig genes in HPV-negative OSCC

Heatmap displays percentiles (*key*, *right*) of ranked VAFs (thin vertical tiles) for SNVs in exons of 329 HPV-negative OSCC samples (*x-axis*, sorted by mutation rate) for the most highly mutated MutSig genes (*y-axis*, q < 0.2, sorted by mutation rate). *Gray tiles*, no gene variant in sample; *white vertical lines*, 6 HPV-negative samples (out of 335 total) not available from TCGA. The 12 genes shown here harbored significant enrichment of coding-change variants in the top 5% fraction of ranked VAFs (cf. **Supplemental Figure S6B**).

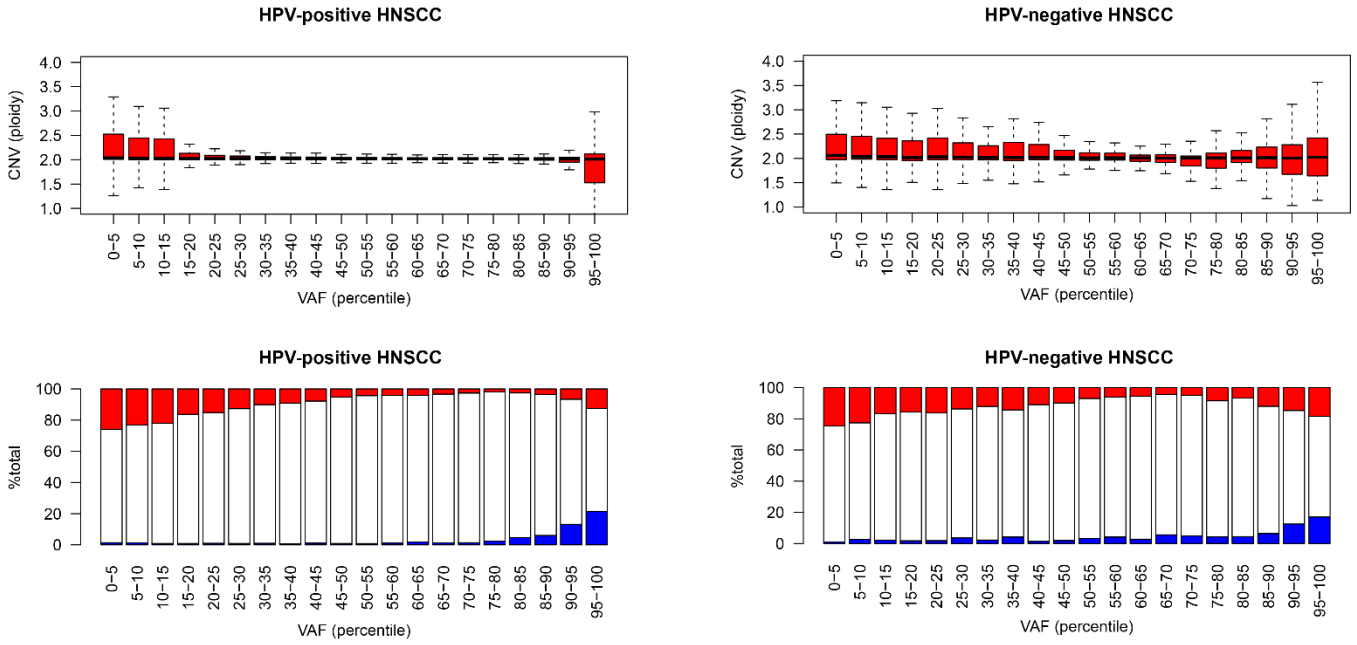


Supplemental Figure S6E. Distributions of ranked VAFs in OSCC.
Histograms display distributions of (*top*) HPV-negative and (*bottom*) HPV-positive OSCC samples (*y-axis*, fractions of samples) harboring mutations with the indicated rankings of somatic variant allelic fractions (*x-axis*, ranked VAF, percentile) for (*left*) all RefSeq and (*right*) significantly mutated MutSig genes. (*Upper left*) Distribution of the ranked VAFs for 46,610 somatic variants in all RefSeq genes having >20x depth of sequencing coverage and >5% VAF, identified in 329 HPV-negative tumors. (*Upper right*) Of the 860 ranked somatic variants in the 25 most significantly mutated MutSig genes, 262 were in the top 5th percentile ($p < 1 \times 10^{-100}$). *(Lower left*) Distribution of the ranked VAFs for 26,646 somatic variants in all RefSeq genes in 145 HPV-positive tumors. (*Lower right*) Of the 247 ranked somatic variants in the top 24 significantly mutated MutSig genes, 54 (21.9%) were in the top 5% percentile ($p=4.5 \times 10^{-20}$). Statistical enrichment of coding-change variants with VAFs recurrently ranking in the top 5% was calculated using binomial test, and multiple testing correction was performed using the Bioconductor qvalue package.



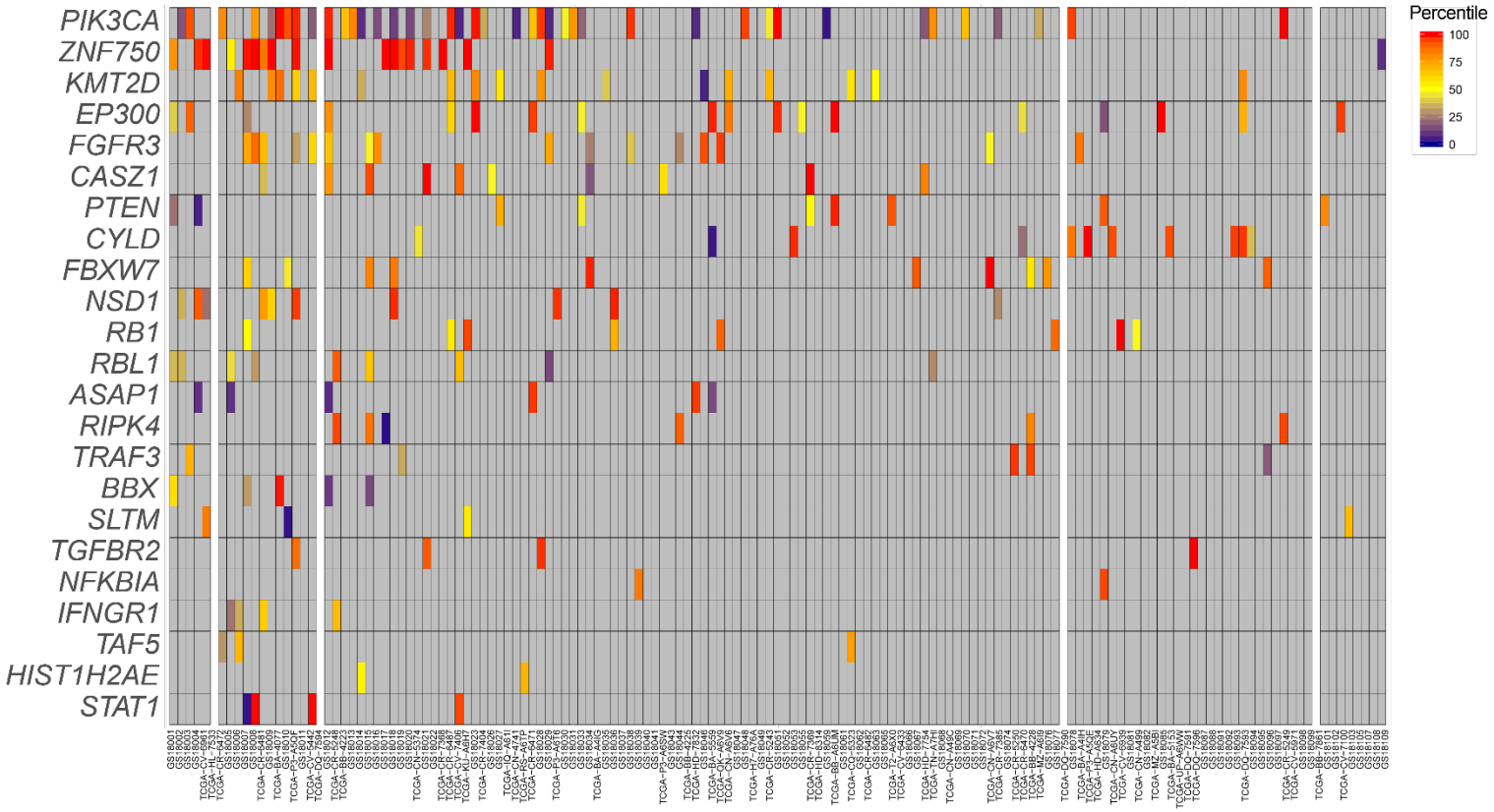
Supplemental Figure S6F. Unbiased VAFs of gene mutations in OSCC.
Histograms display distributions of gene mutation VAFs in HPV-positive or HPV-negative OSCC. These 6 genes were frequently mutated, but as per MutSig analysis the mutations were not statistically significant. Fractions of all affected samples (*y-axis*, percent of total) harboring mutations with the indicated rankings of somatic variant allelic fractions (*x-axis*, ranked VAF, percentile) are shown. Statistical enrichment of coding-change variants with VAFs recurrently ranking in the top 5th percentile was calculated using binomial test, and multiple testing correction was performed using the R qvalue package; no significant enrichment was found for these genes.

.



Supplemental Figure S6G. Associations between ranked VAFs and local ploidy changes in OSCC.

Boxplots and bargraphs show associations between (*x-axis*) ranked VAF percentiles and (*top, y-axis*) local CNV or ploidy or (*bottom, y-axis*) fraction of all samples affected. *Left*, HPV-positive; *right*, HPV-negative OSCC. *Red*, CNV ploidy gain, > 2.5N; *blue,* CNV loss, ploidy < 1.5N; white, normal copy number (1.5N < ploidy <= 2.5N).



Supplemental Figure S6H. VAFs of somatic variants in significantly mutated MutSig genes in HPV-positive OSCC.

Heatmap displays percentiles (*key*, *right*) of ranked VAFs (thin vertical tiles) for SNVs in exons of 145 HPV-positive OSCC samples (*x-axis*, sorted by mutation rate) for the most highly mutated MutSig genes (*y-axis*, q < 0.2, sorted by mutation rate). *Gray tiles*, no gene variant in sample; *white vertical lines*, 4 HPV-positive samples (out of 149 total) not available from TCGA.