CANCER GENETICS AND EPIGENETICS

Unraveling most abundant mutational signatures in head and neck cancer

Michaela Plath[1] | Johanna Gass[1] | Mario Hlevnjak[2] | Qiaoli Li[1] | Bohai Feng[1] |
Xavier Pastor Hostench[3,4] | Matthias Bieg[3,4] | Lea Schroeder[5] | Dana Holzinger[5] |
Marc Zapatka[2] | Kolja Freier[6] | Wilko Weichert[7] | Jochen Hess[1,8] | Karim Zaoui[1]

[1]Section Experimental and Translational Head and Neck Oncology, Department of Otolaryngology, Head and Neck Surgery, University Hospital Heidelberg, Heidelberg, Germany

[2]Division Molecular Genetics, German Cancer Research Center (DKFZ), Heidelberg, Germany

[3]Division of Theoretical Bioinformatics, German Cancer Research Center (DKFZ), Heidelberg, Germany

[4]Heidelberg Center for Personalized Oncology, DKFZ-HIPO, German Cancer Research Center (DKFZ), Heidelberg, Germany

[5]Division of Molecular Diagnostics of Oncogenic Infections, Infection, Inflammation and Cancer Program, German Cancer Research Center (DKFZ), Heidelberg, Germany

[6]Department of Oral and Cranio-Maxillofacial Surgery, University Hospital Heidelberg, Heidelberg, Germany

[7]Institute of Pathology, Technical University Munich, Munich, Germany

[8]Molecular Mechanisms of Head and Neck Tumors, German Cancer Research Center (DKFZ), Heidelberg, Germany

Correspondence
Karim Zaoui, Section Experimental and Translational Head and Neck Oncology, Department of Otolaryngology, Head and Neck Surgery, University Hospital Heidelberg, Im Neuenheimer Feld 400, 69120 Heidelberg, Germany.
Email: karim.zaoui@med.uni-heidelberg.de

Abstract

Genomic alterations are a driving force in the multistep process of head and neck cancer (HNC) and result from the interaction of exogenous environmental exposures and endogenous cellular processes. Each of these processes leaves a characteristic pattern of mutations on the tumor genome providing the unique opportunity to decipher specific signatures of mutational processes operative during HNC pathogenesis and to address their prognostic value. Computational analysis of whole exome sequencing data of the HIPO-HNC (Heidelberg Center for Personalized Oncology-head and neck cancer) (n = 83) and TCGA-HNSC (The Cancer Genome Atlas-Head and Neck Squamous Cell Carcinoma) (n = 506) cohorts revealed five common mutational signatures (Catalogue of Somatic Mutations in Cancer [COSMIC] Signatures 1, 2, 3, 13 and 16) and demonstrated their significant association with etiological risk factors (tobacco, alcohol and HPV16). Unsupervised hierarchical clustering identified four clusters (A, B, C1 and C2) of which Subcluster C2 was enriched for cases with a higher frequency of signature 16 mutations. Tumors of Subcluster C2 had significantly lower p16[INK4A] expression accompanied by homozygous *CDKN2A* deletion in almost one half of cases. Survival analysis revealed an unfavorable prognosis for patients with tumors characterized by a higher mutation burden attributed to signature 16 as well as cases in Subcluster C2. Finally, a LASSO-Cox regression model was applied to prioritize clinically relevant signatures and to establish a prognostic risk score for head and neck squamous cell carcinoma patients. In conclusion, our study provides a proof of concept that computational analysis of somatic mutational signatures is not only a powerful tool to decipher environmental and intrinsic processes in the pathogenesis of HNC, but could also pave the way to establish reliable prognostic patterns.

1 | INTRODUCTION

Head and neck cancer (HNC) represents the seventh most common cancer with an annual incidence of more than 800 000 cases worldwide in 2018 (http://globocan.iarc.fr/). HNCs are a heterogeneous group of cancers and more than 90% are diagnosed as head and neck squamous cell carcinoma (HNSCC) originating from the epithelial mucosa at the upper aerodigestive tract.[1] Main etiological risk factors are tobacco, excessive alcohol consumption and infection with high-risk human papillomavirus (HPV),[2] in particular HPV16.[3,4] In the last few decades, the incidence of oropharyngeal squamous cell carcinoma (OPSCC) has been increasing in developed countries and the role of HPV is emerging as an important factor in the rise of OPSCCs.[5]

Due to the lack of symptoms in the early stage and effective screening techniques, the majority of HNSCC patients are diagnosed at an advanced stage.[6] Despite our current knowledge on underlying mutational and biological processes as well as implementation of an aggressive and multimodal therapy consisting of surgery, radiotherapy and platinum-based chemotherapy, the prognosis of patients with advanced HNSCC remains dismal with a 5-year survival rate of less than 50%.[7-9]

Hence, new concepts are urgently needed that will advance our understanding on the complex interplay between etiology, mutational and biological processes, which operate during HNSCC pathogenesis. HNSCC develops in a multistep process that involves different molecular alterations including accumulation of multiple genetic and epigenetic changes with tumor progression.[2] Somatic mutations in a cancer genome are the cumulative result of mutational processes as a consequence of the intrinsic infidelity of the DNA replication machinery, exogenous or endogenous mutagen exposures, enzymatic modification of DNA or defective DNA repair.[10] However, our understanding of the processes that cause somatic mutations in HNC is poorly understood.[11] Genomic analyses revealed the presence of several mutational signatures in HNSCC.[12,13] However, it is worth noting that most individual cancer genomes exhibit more than one mutational signature and many different combinations of signatures were observed.[14] In addition, global gene expression and DNA methylome profiling analysis elucidated distinct HNSCC subgroups with characteristic features concerning clinical and pathological traits as well as patient prognosis.[15-17]

Up to date, little progress has been made in utilizing this information for improved diagnostic tools or therapeutic interventions in the clinically highly heterogeneous group of HNSCC patients. Hence, potential implications of mutational signatures for a better understanding of cancer etiology and their translational impact on prevention, prognostic risk assessment and personalized therapeutic concepts remain to be addressed in HNC. The main object of our study was to decipher common signatures of mutational processes

What's new?

Progression of head and neck squamous cell carcinoma (HNSCC) is associated with the accumulation of multiple genetic and epigenetic alterations. Elucidating these mutational processes and their connections to tumor etiology and progression could have significant implications for HNSCC diagnosis and treatment. Here, using various computational analyses, the authors identified five mutational signatures in the COSMIC database associated specifically with HNSCC etiological risk factors and clinical outcome. Tumors with a high frequency of COSMIC signature 16 mutations had reduced p16^{INK4A} expression with homozygous *CDKN2A* deletion, which may represent a critical step in the pathogenesis of a distinct HNSCC subgroup.

operating during HNSCC pathogenesis and to evaluate their prognostic significance.

2 | METHODS

2.1 | Patient population

Patients of the Heidelberg Center for Personalized Oncology-head and neck cancer (HIPO-HNC) cohort (n = 83) were treated between 2012 and 2016 at the University Hospital Heidelberg, Germany, and the cohort consists primarily of advanced HNSCC from the oropharynx (n = 34, 40.9%), oral cavity (n = 21, 25.3%), nasal cavity (n = 11, 13.3%) and laryngeal/hypopharyngeal sites (n = 17, 20.5%). Median age at the time of diagnosis was 61.4 years (range: 39.7-82.5 years), most patients were male (n = 64, 77.1%) and smokers (n = 57, 68.7%). HPV-related tumors were almost exclusively found in the subgroup of OPSCC (n = 21 out of 34, 61.8%), compared to 4.1% (n = 2 out of 49) in non-OPSCC. Demographic and clinical data were collected by chart review and are summarized in Supplemental Table S1.

2.2 | Isolation of analytes

As described by Schmitt et al,[18] fresh-frozen tumor samples of the HIPO-HNC cohort were obtained from surgical resection and were evaluated by a pathologist (W.W.) to confirm the diagnosis and to estimate neoplastic cell content by microscopic inspection upon

histological staining. Blood samples were collected prior to surgery. DNA and RNA from tumor specimens and DNA from blood samples were isolated at the central DKFZ-HIPO Sample Processing Laboratory using the AllPrep DNA/RNA/Protein Mini Kit (Qiagen) and the QIAamp DNA Blood Mini QIAcube Kit (Qiagen) according to manufacturer's protocols. Quality control and quantification were conducted using a Qubit 2.0 Fluorometer (Thermo Fisher Scientific), the Agilent 2100 Bioanalyzer (Agilent Technologies) and the NanoDrop spectrophotometer (NanoDrop Technologies).[18]

2.3 | Whole-exome sequencing and data analysis

Exome capturing was performed using SureSelect Human All Exon in-solution capture reagents version 4 (n = 58) and version 6 (n = 25, Agilent Technologies) including UTRs, and sequencing was carried out with a HiSeq 2500 instrument (Illumina). We did neither observe a statistically significant difference in total mutation counts nor in the prioritization of the top five most abundant signatures among both subgroups (version 4 vs version 6), indicating no major batch effect. The workflow for mapping of paired-end short reads, single nucleotide variation (SNV) calling, annotation of the variants and assessment of functional relevance was described previously.[18] Small insertions and deletions were obtained from Platypus (version 0.7.4) and further annotated and confidence assessed similar as in single nucleotide variation calling. Average read coverage for tumor samples was 110-fold with a range from 70 to 190.

2.4 | Analysis of mutational signatures

Supervised mutational signature analysis of high-confidence somatic SNVs in individual samples was performed based on non-negative matrix factorization formalism as described previously.[10,19] More specifically, using quadratic programming, the mutational profile (ie, catalog of somatic SNVs in a 96-trinucleotide context) of each tumor was decomposed into individual contributions (ie, exposures) of the reference set of 30 canonical mutational signatures available in the Catalogue of Somatic Mutations in Cancer (COSMIC database; http://cancer.sanger.ac.uk/cosmic/signatures). Furthermore, for exome samples, canonical mutational signatures were renormalized using the ratio of observed trinucleotide frequency in the human exome (as defined by the target region of the used enrichment kit) to the one of the human genome. Samples with cosine distance >0.3 between the original mutational catalog and the reconstructed catalog (ie, the one obtained by multiplying the matrix of exposures with the matrix of canonical signatures) were excluded from downstream analysis.

2.5 | HPV16 status

HPV16 status for the HIPO-HNC cohort was determined by BSGP5 +/6+-PCR/MPG and E6*I mRNA detection as described previously.[20-22] HPV16 DNA- and RNA-positive cases were considered as HPV16-driven while all other cases (DNA-negative and DNA-positive but RNA-negative) were considered as non-HPV16-driven. In the Cancer Genome Atlas (TCGA) cohort, the HPV16 status was determined using the bioinformatic system "ViruScan".[23]

2.6 | TCGA data analysis

TCGA-HNSC aggregated data were downloaded on 17 August 2017. The three lip tumors were excluded along with four samples, which were excluded as described under the section "Analysis of mutational signatures" (n = 506). Exclusion criteria for overall survival (OS) analysis were cM1, pM1 or a history of synchronous or prior malignancy leaving n = 485 patients for further analysis.

2.7 | Global gene expression analysis

The RNA-Seq data of TCGA-HNSC were transformed using the voom function of the limma package, and differentially expressed genes (DEGs) were identified with the limma package in R (http://bioconductor.org/packages/release/bioc/html/limma.html). DEGs with a q-value < 0.05 adjusted for the False Discovery Rate (FDR) and $-1 > log2FC > 1$ were selected for further analysis. Private DEGs of distinct subclusters were selected and visualized using the online tool VENNY2.1 (http://bioinfogp.cnb.csic.es/tools/venny/index.html).

2.8 | Statistical analysis and hierarchical clustering

All patient data were collected and documented using the program Microsoft Excel for Mac (Version 16.23, 190309). It was also used to easily visualize the distributions of patient mutation counts using bar charts. All other statistical analyses were performed using the SPSS Statistics 25 statistics program (IBM Corp. Released 2017. IBM SPSS Statistics for Macintosh, Version 25.0. Armonk, NY: IBM Corp.). The demographic and clinical-pathological characteristics of the patients were examined using descriptive statistics. For the subsequent analyses, the median value was used as the separator for our measurements of the HIPO-HNC and the 75% quartile for the TCGA-HNC cohort. P values of <.05 were considered statistically significant. The association between mutation signatures and etiological risk factors as well as histopathological and clinical characteristics of the patients was analyzed by cross tables using the Pearson's chi-squared test. The following factors were investigated: age, sex, HPV status, tobacco and alcohol consumption of the patients; tumor localization; tumor size; and lymph node status; as well as extracapsular scattering in positive lymph nodes, therapy, and resection margin. Where data were depicted using Tukey's box-and-whiskers plot, the box spans from the 25th to the 75th percentile, with the line dissecting the box denoting the median. The whiskers denote the 25th percentile minus 1.5 times the interquartile range, and the 75th percentile plus 1.5 times the interquartile range. Any values outside this range are depicted as individual points.

OS, disease-specific survival (DSS) and PFS rates were calculated using the Kaplan-Meier method. OS was defined as time from the date of cancer diagnosis to the date of death, DSS as the time from the date of cancer diagnosis to the date of death from HNC. Progression-free survival (PFS) was defined as time from the date of cancer diagnosis to the date of recurrence. The log-rank test was used to test the significant influence of these groups on survival. Univariate and multivariate Cox proportional hazard models were used to explore associations of patient characteristics and biomarker results with PFS, DSS or OS. Hazard ratios (HRs) with 95% confidence intervals (CI) were calculated. The Cox proportional hazard model was used for multivariate analysis and was performed with all parameters possessing a P value <.13 in the univariate analysis and total mutation counts.

Unsupervised hierarchical clustering and visualization by heatmaps were performed with https://biit.cs.ut.ee/clustvis/.[24] Rows are centered and unit variance scaling is applied to rows. Both rows and columns are clustered using correlation distance and average linkage.

2.9 | Mutational landscape of subclusters

In order to compare the distribution of somatic mutations between patient subgroups, candidate genes were selected from the TCGA-HNSC data sets using MutSig 2.0 available on cBioPortal (http://www.cbioportal.org/). Genes with a q value of <0.05 were considered to be significantly mutated, and thus included in the analysis. The mutation frequency of each significantly mutated gene in each of the patient subgroups was determined. The differences in mutation frequencies between the patient subgroups were determined in a principal component analysis and depicted in a heat map as described above using ClustVis (https://biit.cs.ut.ee/clustvis/). Mutational oncomaps were performed by the R package "ComplexHeatmap".[25]

The TCGA_HNSC copy number variation (CNV) data (Level_3_segmented_scna_minus_germline_cnv_hg19_seg) were downloaded from Firebrowse (http://www.firebrowse.org/). We defined the value of Segment_Mean bigger than 0.2 as gain and less than −0.2 as loss. The CNV summary plots are conducted by IGV_2.4.19 (Integrative Genomics Viewer_2.4.19).[26]

2.10 | LASSO Cox regression model

The LASSO Cox regression algorithm was applied to prioritize most relevant prognostic candidates of mutational signatures for the TCGA-HNSC cohort. The risk score was computed by "glmnet" (Lambda.min_value = 0.04384824 and type = "response"), and the analytical formula for risk assessment was derived on the basis of seven mutational signatures (coefficient of candidates: Signature 1 = −0.00106054437 852817, Signature 3 = 0.00482984300797927, Signature 16 = 0.003 67514371412884, Signature 22 = −0.00898280402004339, Signature 27 = 0.0521616095070655, Signature 28 = −0.0549190877110363 and Signature 29 = −0.0161229340270006).

3 | RESULTS

3.1 | Somatic mutation frequency in the HIPO-HNC cohort

Total somatic mutation counts per cancer genome were determined based on whole-exome sequencing data, which were available for n = 83 cases of the HIPO-HNC cohort. Mutation counts were slightly higher in smokers as compared to nonsmokers (P = .054), which did
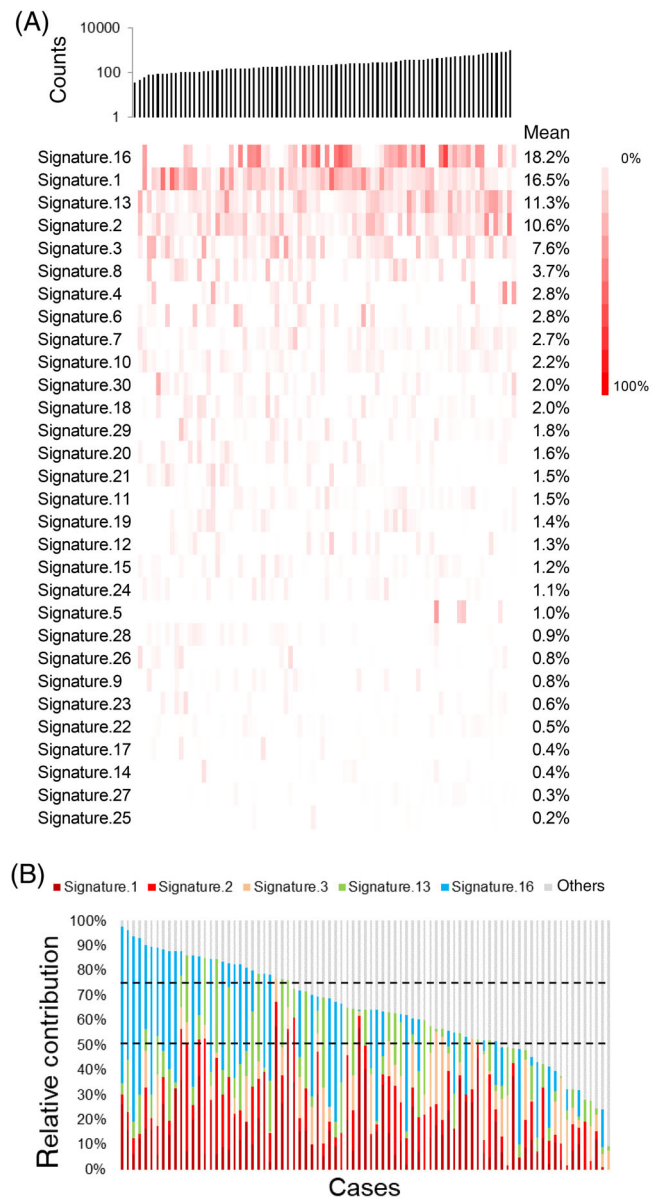

FIGURE 1 Common mutational signatures for the HIPO-HNC cohort. A, Relative contribution of the most frequent mutational signatures (defined by COSMIC—Catalogue of Somatic Mutations in Cancer) sorted by the absolute mutation counts. B, Graph depicts the relative proportion (%) of the top five signatures as compared to all other signatures for individual cases of the HIPO-HNC cohort (n = 83). HIPO-HNC, Heidelberg Center for Personalized Oncology-head and neck cancer [Color figure can be viewed at wileyonlinelibrary.com]

not reach statistical significance (Supplemental Figure S1A). No major difference was found considering HPV16 status, alcohol consumption, clinical or histopathological features (age, gender, tumor size, lymph node metastasis, pathological grading, resection margin, PFS or DFS), except for primary tumor site with a higher total mutation count in laryngeal SCC (Supplemental Figure S1A). Similar data were obtained for mutation counts of the TCGA-HNSC cohort, except for age, which revealed a weak but significant positive correlation with mutation counts (Supplemental Figure S1B).

3.2 | Most abundant mutation signatures in the HIPO-HNC cohort

We computed the relative contribution of distinct mutational signatures for individual cancer genomes (Figure 1A, Supplemental Table S2). This analysis identified signatures 1, 2, 3, 13 and 16 (nomenclature according to COSMIC)[27] as the most abundant mutational signatures in the HIPO-HNC cohort (Figure 1A, Supplemental Table S3). In 64 out of 83 cases (77.1%), the relative mutation burden attributed
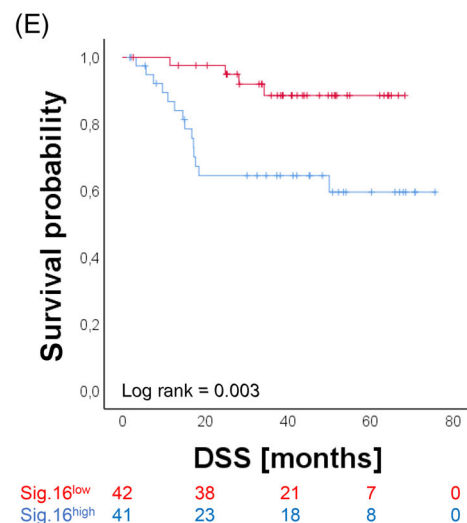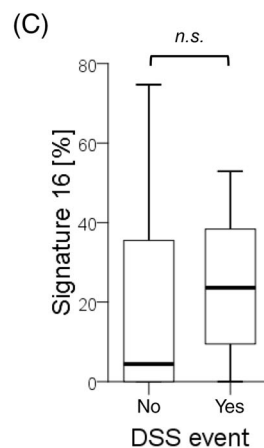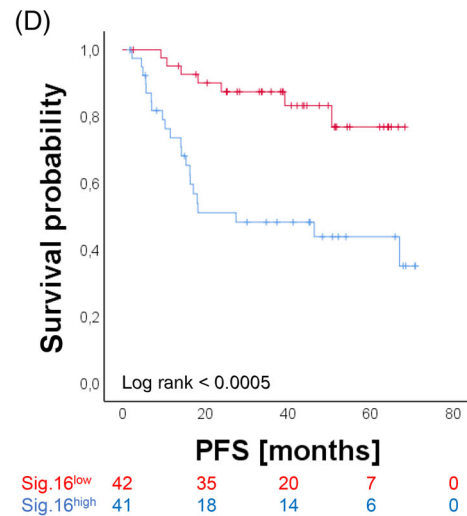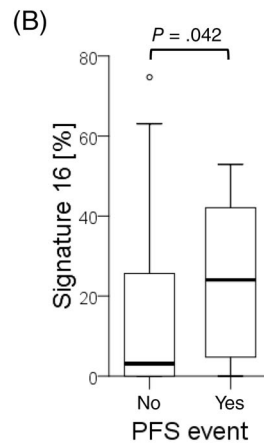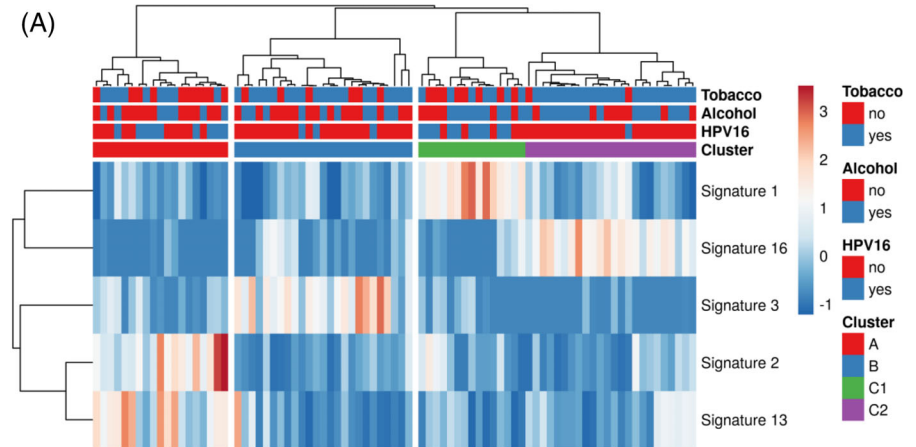

FIGURE 2 Prognostic value of Signature 16 for the HIPO-HNC cohort. A, The heatmap represents distinct patient subgroups based on unsupervised hierarchical clustering of the relative contribution for the five most common mutational signatures and provides information on most relevant risk factors (tobacco, alcohol, HPV16). B,C, Box plots present the association between Signature 16 and either PFS (no [n = 55], yes [n = 28]) or DSS (no [n = 65], yes [n = 18]) for the HIPO-HNC cohort. Kaplan-Meier graphs illustrate the survival probability of subgroups with a either high (≥median) or low (<median) relative contribution of somatic mutations attributed to Signature 16 for PFS (D) or DSS (E) for the HIPO-HNC cohort. Numbers below the graph represent patients at risk at the indicated time points. DSS, disease-free survival; HIPO-HNC, Heidelberg Center for Personalized Oncology-head and neck cancer; PFS, progression-free survival [Color figure can be viewed at wileyonlinelibrary.com]

to the top five signatures was larger than 50%. In 29 cases (34.9%), it was even larger than 75% (Figure 1B). The frequency of somatic mutation signatures was also analyzed in subgroups categorized by well-known risk factors (tobacco, alcohol or HPV16). Though the ranking of individual signatures was variable, all subgroups shared signatures 1, 2, 3, 13 and 16 as top five candidates (Supplemental Table S3). Moreover, unsupervised hierarchical clustering revealed three distinct patient subgroups (clusters A-C) based on the relative distribution of the top five mutation signatures with high relative contribution of the apolipoprotein B mRNA editing enzyme catalytic polypeptide (APOBEC)-related Signatures 2 and 13 in Cluster A, Signature 3 in Cluster B, Signatures 1 and 16 in Cluster C (Figure 2A; Supplemental Figure S2A). Cluster C was further divided in Subcluster C1

(enriched for Signature 1) and Subcluster C2 (enriched for Signature 16). A similar enrichment was also observed considering the quantity of somatic mutations attributed to individual signatures (Supplemental Figure 2B).

Comparison of these clusters according to clinical or histopathological features revealed significant differences in mutation burden attributed to distinct signatures in relation to etiological risk factors, but not to any other clinical or histopathological feature tested (Figure 2A, Supplemental Table S4). In particular, a high relative contribution of either Signature 2 in Cluster A or Signature 1 in Cluster C1 was a characteristic feature of HPV16-related cancers, while Signature 16 (enriched in cluster C2) was almost absent in these cases (Supplemental Figure S2C). In contrast, a high relative contribution or
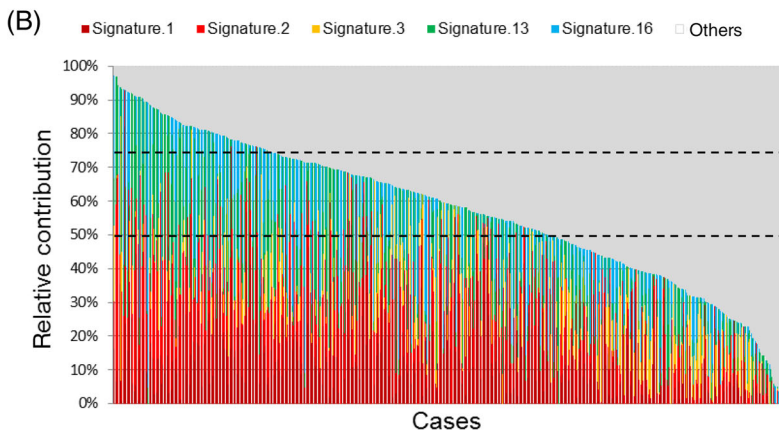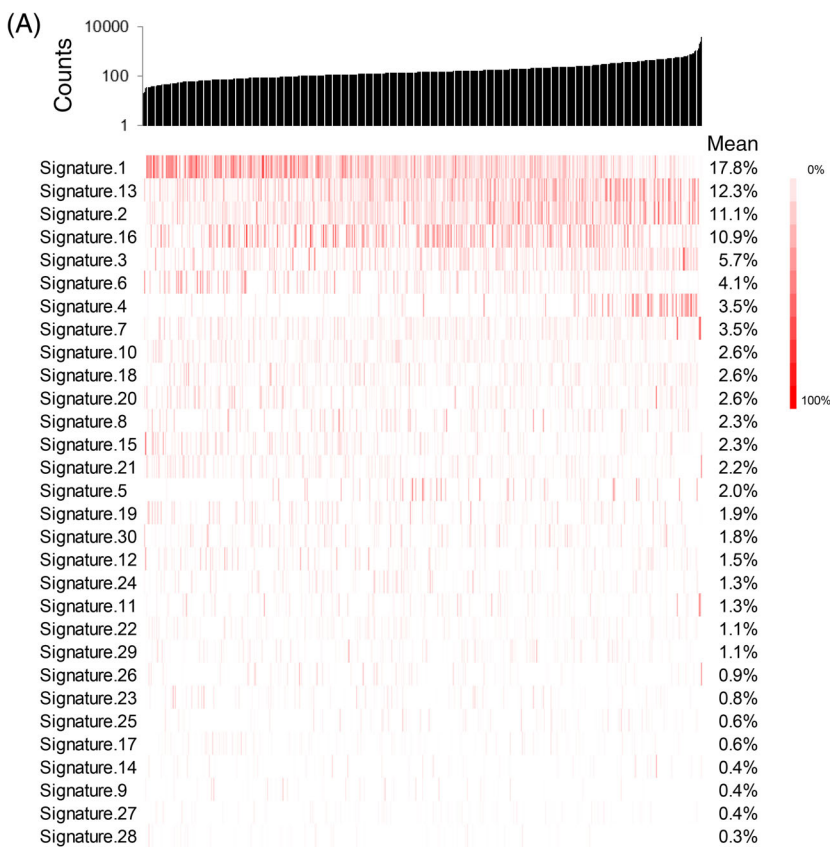

FIGURE 3 Common mutational signatures for the TCGA-HNSC cohort. A, Relative contribution of the most frequent mutational signatures (defined by COSMIC—Catalogue of Somatic Mutations in Cancer) sorted by the absolute mutation counts. B, Graph depicts the relative proportion (%) of the top five signatures as compared to all other signatures for individual cases of the TCGA-HNSC cohort (n = 506). TCGA-HNSC, The Cancer Genome Atlas-Head and Neck Squamous Cell Carcinoma [Color figure can be viewed at wileyonlinelibrary.com]

quantity of somatic mutations attributed to Signature 16 was observed in cancers related with tobacco consumption, though it was not significantly correlated with the quantity of pack-years (Supplemental Figure S2C,D), which might be due to the limited amount of cases (n = 36) for which this information was available.

A higher relative contribution of Signature 16 to the mutational burden was found for cases with tumor progression and disease-specific death (Figure 2B,C), and Kaplan-Meier analysis confirmed a significantly shorter PFS and DSS for patients with a high mutational burden attributed to Signature 16 (Figure 2D,E). Unfavorable
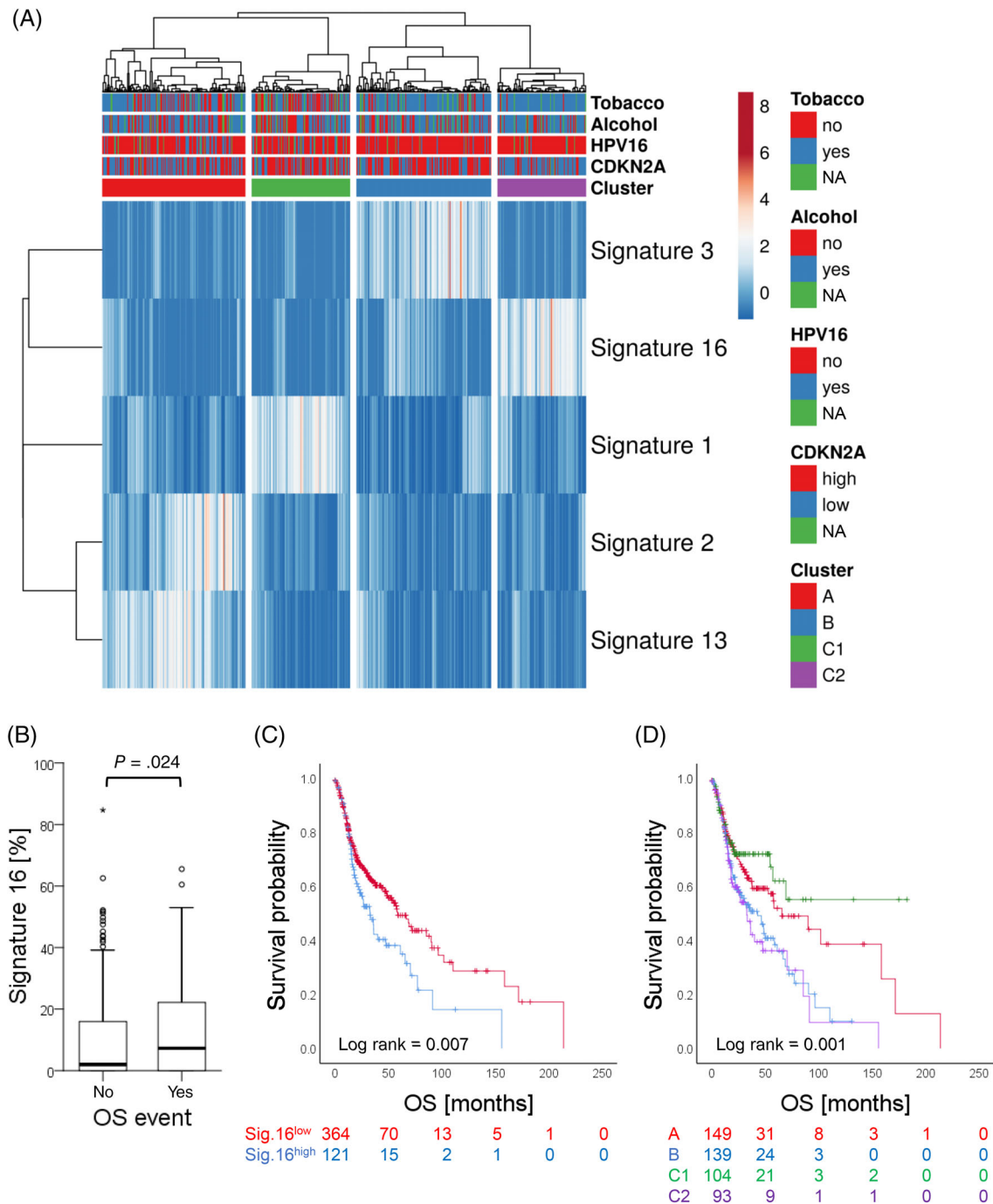

FIGURE 4 Prognostic value of Signature 16 for the TCGA-HNSC cohort. A, The heatmap represents distinct patient subgroups based on unsupervised hierarchical clustering of the relative contribution for the five most common mutational signatures and provides information on most relevant risk factors (tobacco, alcohol, HPV16). B, Box plot presents the association between Signature 16 and overall survival events (OS, no [n = 276], yes [n = 209]) for the TCGA-HNSC cohort. Kaplan-Meier graphs illustrate the survival probability of subgroups with a either high (≥75% quartile) or low (<75% quartile) relative contribution of somatic mutations related to Signature 16 (C) or cases in distinct clusters for OS (D). Numbers below the graphs represent patients at risk at the indicated time points. TCGA-HNSC, The Cancer Genome Atlas-Head and Neck Squamous Cell Carcinoma [Color figure can be viewed at wileyonlinelibrary.com]

prognosis of HNSCC with a high frequency of Signature 16 mutations was also evident for subgroups with a smoking history, HPV16 negative or positive tumors and distinct primary tumor sites (Supplemental Figure S3A-F). Univariate analysis revealed a trend for smoking history, negative HPV16 status and lymph node metastasis with unfavorable PFS and for larger tumor size, lymph node metastasis and positive resection margins for unfavorable DSS (Supplemental Tables S5). Multivariate Cox models adjusted for these variables and total mutation counts indicated that a high mutation burden attributed to Signature 16 serves as an independent risk factor for an

unfavorable PFS (Supplemental Table S5). In contrast, no significant difference in survival was detected for any other more abundant signature (Supplemental Figure S3G,H).

3.3 | Most abundant mutation signatures in the TCGA-HNC cohort

To confirm our findings in an independent and larger patient cohort, we computed the relative contribution of distinct mutational
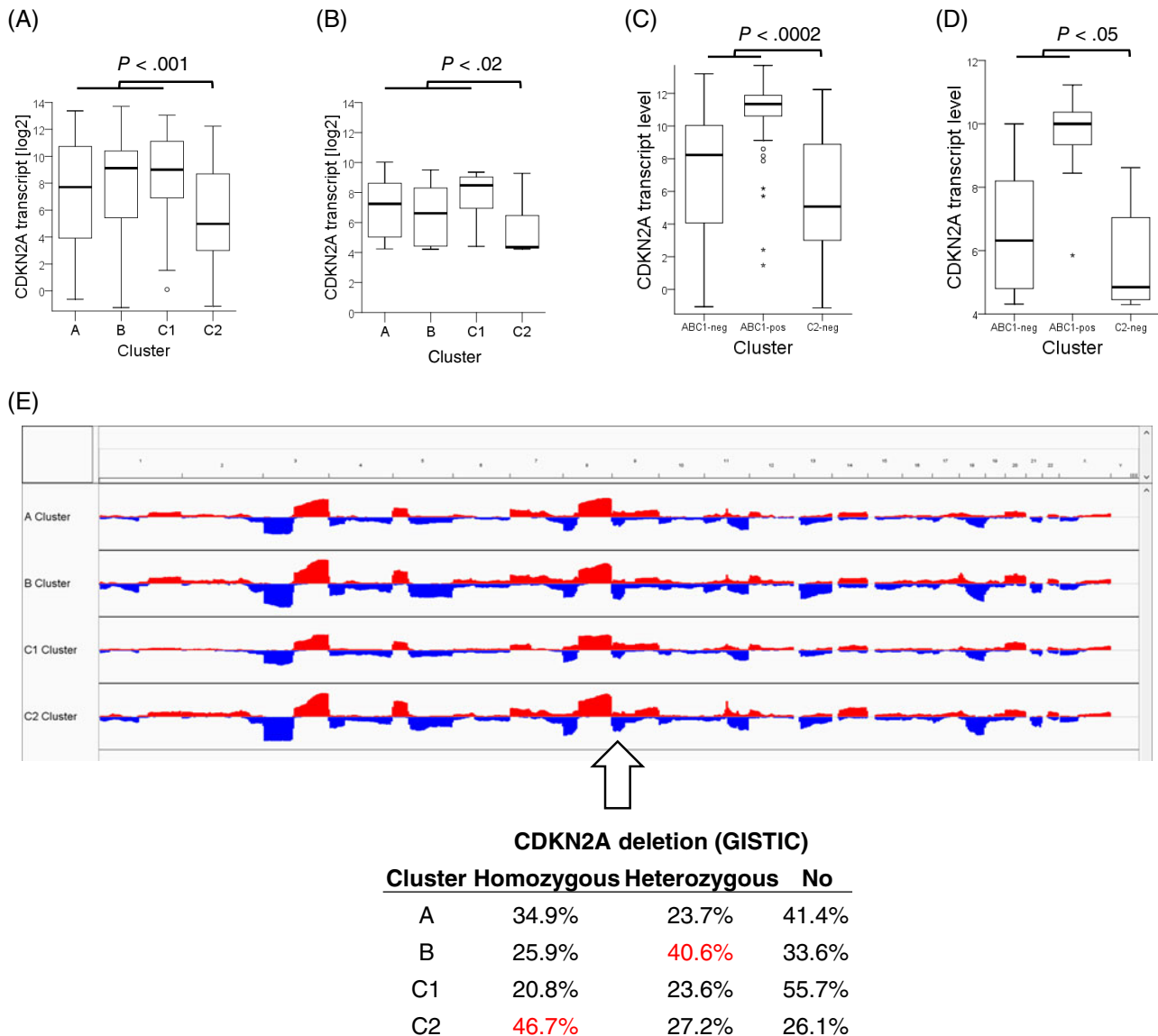

FIGURE 5 Association between Signature 16 and CDKN2A expression or copy number variations for the TCGA-HNSC cohort. Box plots present mean values and 25% or 75% quartiles of CDKN2A transcript levels for either distinct clusters of the TCGA-HNSC (A, Cluster A [n = 155], B [n = 146], C1 [n = 107], C2 [n = 96]), or HIPO-HNC cohorts (B, cluster A [n = 19], B [n = 25], C1 [n = 15], C2 [n = 24]), or Clusters A, B and C1 with (ABC1-pos) or without (ABC1-neg) HPV16-related tumors and Cluster C2 without HPV16-related tumors (C2-neg) of the TCGA-HNSC (C, ABC1-neg [n = 317], ABC1-pos [n = 66], C2-neg [n = 87]) or HIPO-HNC cohorts (D, ABC1-neg [n = 37], ABC1-pos [n = 22], C2-neg [n = 23]). E, Oncomap illustrates the chromosomal position and relative frequency of copy number variations (gains in red and deletion in blue) for distinct clusters of the TCGA-HNSC cohort. The table below the oncomap summarizes the relative distribution of homozygous or heterozygous CDKN2A deletions at chromosome 9 in distinct clusters according to GISTIC 2.0. HIPO, Heidelberg Center for Personalized Oncology; HNC, head and neck cancer; TCGA-HNSC, the Cancer Genome Atlas-head and neck squamous cell [Color figure can be viewed at wileyonlinelibrary.com]

signatures in individual cancer genomes for the TCGA-HNSC cohort (n = 506) (Figure 3A, Supplemental Table S6). Patient cases with total mutation counts ≤30 were excluded. Again, Signatures 1, 2, 3, 13 and 16 were the most abundant mutational signatures with 330 out of 506 cases (65.2%) exhibiting a relative contribution larger than 50%. In 117 cases (23.1%), it was larger than 75% (Figure 3B). Most subgroups of the TCGA-HNSC cohort based on risk factor stratification shared Signatures 1, 2, 3, 13 and 16 as top five candidates

(Supplemental Table S7). The only exception was the subgroup without smoking history, in which Signatures 3 and 16 were replaced by Signatures 6 and 7. However, the latter two signatures had no impact on clinical outcome in both cohorts, including subgroups with or without a smoking history (Supplemental Figure S4A, data not shown), only a minor impact on the stratification of clusters and subclusters (Supplemental Figure S4B,C), and were not related to strong enrichments of somatic mutations in MutSig genes (Supplemental Figure S4D,E).
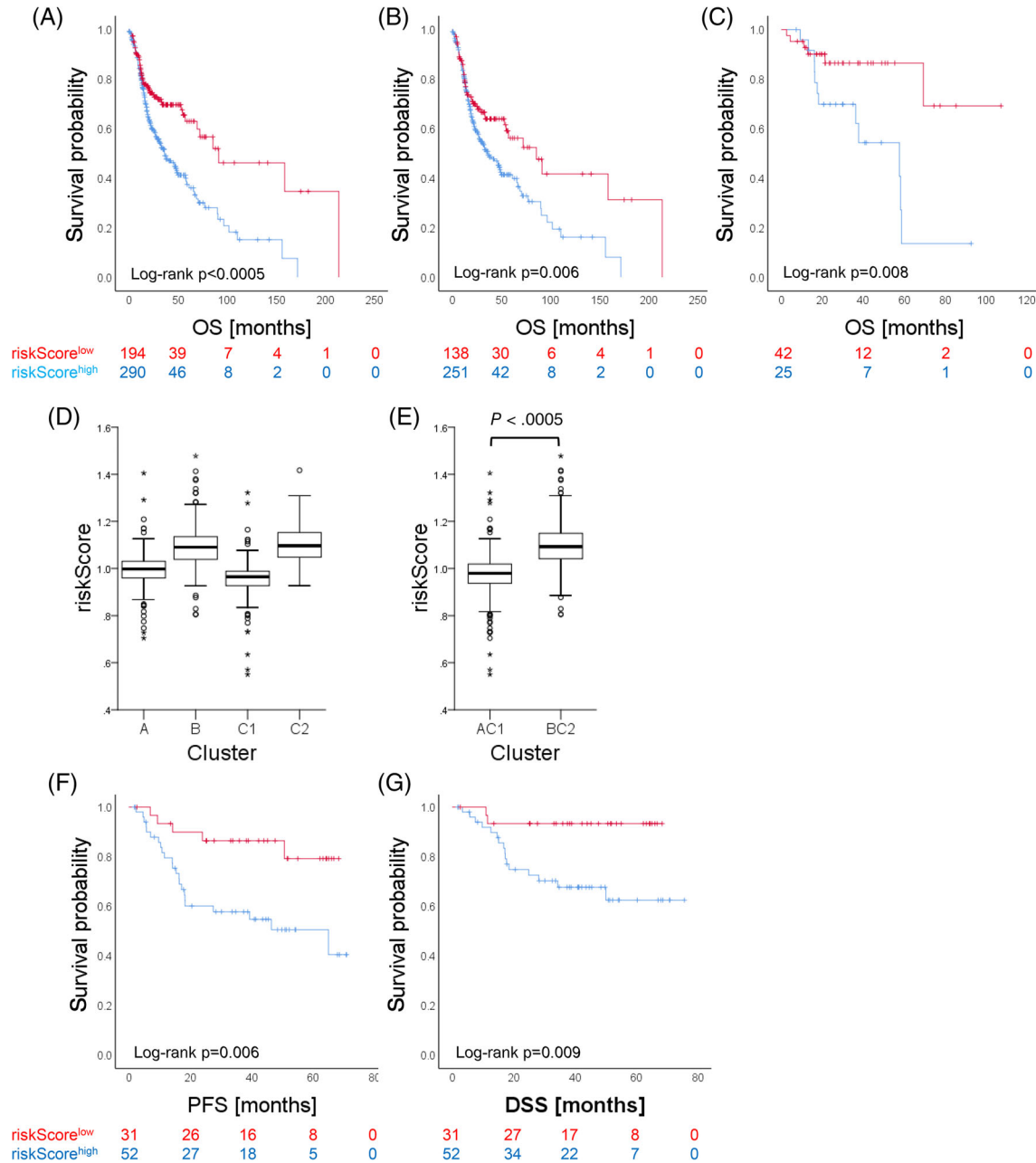

FIGURE 6   Prognostic value of the risk model. Kaplan-Meier plots illustrate the overall survival probability of subgroups with an either high-risk or low-risk score for the TCGA-HNSC cohort (A) or subgroups with HPV16-negative (B) or -positive HNSCC (C). Box plots present mean values and 25% or 75% quartiles of the risk score for distinct clusters (D, cluster A [n = 149], B [n = 139], C1 [n = 104], C2 [n = 94]) or combined clusters AC1 and BC2 (E, cluster AC1 [n = 253], BC2 [n = 233]) of the TCGA-HNSC cohort. Kaplan-Meier plots show differences in progression-free survival (PFS, F) and disease-specific survival (DSS, G) for subgroups with an either high-risk or low-risk score of the HIPO-HNC cohort. Numbers below the graphs represent patients at risk at the indicated time points. HIPO, Heidelberg Center for Personalized Oncology; HNC, head and neck cancer; TCGA-HNSC, the Cancer Genome Atlas-head and neck squamous cell [Color figure can be viewed at wileyonlinelibrary.com]

Unsupervised hierarchical clustering upon unit variant scaling revealed two main clusters with two subclusters, respectively, with a similar enrichment in mutational signatures, as observed for the HIPO-HNC cohort (Figure 4A; Supplemental Figure S5A,B). Comparison with etiological risk factors confirmed the enrichment of HPV16-positive HNSCC in Cluster A attributed to Signatures 2 and 13 as well as Cluster C1 attributed to Signature 1, and a high relative contribution as well as higher quantity of somatic mutations attributed to Signature 16 in cancers related with tobacco consumption (Supplemental Table S8; Supplemental Figure S5C). In the group of smokers, pack-years were slightly but significantly associated with the relative fraction of somatic mutations attributed to signature 16 (Supplemental Figure S5D). Patients with a high mutational burden attributed to Signature 16 had a significantly shorter OS (Figure 4B,C), which was most prominent for oral SCC and OPSCC (Supplemental Figure S5E-G). Accordingly, patients in cluster C2 with a high relative contribution of Signature 16, but also patients in cluster B, which are enriched for Signature 3, showed an unfavorable prognosis as compared to patients in Cluster A (enriched for Signatures 2 and 13) or cluster C1 (enriched for Signature 1) (Figure 4D).

Univariate analysis demonstrated a significant association between negative HPV16 status, female gender, positive resection margin, higher tumor mutation counts and the combined Clusters B and C2 (BC2) with unfavorable OS (Supplemental Table S9). The unfavorable OS of patients in the BC2 as compared to AC1 subgroup was independent of risk factors (tobacco, HPV) or the primary tumor sites (Supplemental Figure S6), and together with a positive resection margin, high tumor mutation counts served as an independent risk factor for OS in a multivariate Cox regression model adjusted for tobacco, HPV16, gender, age, tumor size and lymph node metastasis (Supplemental Table S9).

3.4 | Association between CDKN2A expression and Signature 16

Global gene expression analysis was conducted to unravel DEGs among individual clusters of the TCGA-HNSC cohort (Supplemental Tables S10). We selected private DEGs of Subcluster C2 for further analysis (Supplemental Figure S7A) due to the unfavorable prognosis of patients in both cohorts. A similar trend of gene expression was confirmed for 8 out of 17 candidate genes with induced and 3 out of 6 candidate genes with reduced transcript levels in cases of Subcluster C2 as compared to all others in the HIPO-HNC cohort (Supplemental Figure S7B,C). Strikingly, a highly significant decrease in CDKN2A transcript level, encoding p16[INK4A] was found for Subcluster C2 in both cohorts independent of the HPV16 status (Figures 4A and 5A-D), indicating that low p16[INK4A] expression is either a consequence or contributes to an enrichment of somatic mutations attributed to Signature 16. Indeed, both cohorts shared an inverse correlation between CDKN2A transcript levels and relative frequency of signature 16 mutations (Supplemental Figure S8A). Differences in transcript levels among clusters were not related to the frequency of somatic mutations in

CDKN2A (range 17%-25%), which was the lowest for Subcluster C2 (Supplemental Figure S8B). However, analysis of copy number alterations revealed a higher frequency of CDKN2A deletions in Cluster B and Subcluster C2 (Figure 5E). Although Cluster B exhibits more heterozygous deletions, almost half of all cases in Subcluster C2 had a homozygous CDKN2A deletion. Moreover, HNSCC with homozygous CDKN2A deletion shared a significantly higher frequency of mutations attributed to Signature 16 in TCGA-HNSC, including HPV16-positive or HPV16-negative subgroups (Supplemental Figure S9).

3.5 | Risk model based on combination of clinically relevant signatures

Finally, we conducted a LASSO Cox regression model based on OS of the TCGA-HNSC cohort to prioritize most relevant signatures and to address the question of whether less abundant signatures in combination with Signatures 3 and 16 enable the establishment of a risk model for unfavorable survival in HNSCC. The model confirmed the prognostic value of Signatures 3 and 16 and also identified an impact of Signatures 1, 22, 27, 28 and 29. A risk score computed on the basis of those signatures served as a highly significant and independent prognosticator for the TCGA-HNSC cohort, independent of the HPV16 status (Figure 6A-C), and any other demographic or clinical variables tested (Supplemental Table S11). As expected, the risk score was higher in Cluster B and C2 and significantly different between Clusters BC2 and AC1 (Figure 6D,E). Furthermore, a higher risk score was significantly associated with unfavorable PFS and DFS of the HIPO-HNC cohort (Figure 6F,G).

4 | DISCUSSION

The main objective of our study was the profiling of somatic mutations to enable a comprehensive and system-based interrogation of mutational landscapes and to evaluate their prognostic impact on survival of HNSCC patients. We conducted a whole-exome sequencing analysis of samples from the HIPO-HNC cohort, for which complete information on clinical features was available, and confirmed our findings in the larger TCGA-HNC cohort.

In the HIPO-HNC cohort, we identified five prevalent mutational signatures: the age-related Signature 1,[10] APOBEC-associated Signatures 2 and 13,[28] the BRCA1/2-associated Signature 3[10] and Signature 16, which has been detected so far in smokers with liver cancer[29] or esophageal squamous cell carcinomas.[13,30] These mutational signatures were also highly abundant in tumors of the TCGA-HNC cohort, representing the most comprehensive integrative genomic analysis for HNSCC.[27] Based on these results, we were able to define patient subgroups with different prognostic impacts. In the TCGA-HNC cohort, OS analysis confirmed an unfavorable prognosis for Subgroups B and C2, representing either a high relative mutational burden attributed to Signature 3 or Signature 16, and a better outcome for Subgroups C1

and A, with a high contribution related to either the APOBEC-associated Signatures 2 and 13 or Signature 1. In line with previous studies,[31,32] we observed an association between Signatures 2 and 13 and a positive HPV status. It is now well accepted that HPV-positive HNCs have improved prognosis and survival in comparison with HPV-negative cancer,[33] explaining the better prognostic outcome for the cluster A and C1. However, we could not show a statistically better survival for patients with a high contribution related to signatures 1, 2 and 13 as individual stratification criteria.

A major finding of our study is the unfavorable survival of patients in Subgroup C2, which is enriched for cases with high prevalence of Signature 16, smoking history and presents predominantly HPV16-negative tumors. It is well established that HPV16 status and tobacco exposure are the main determinants of survival in HNSCC.[5] Signature 16 is predominantly characterized by T>C mutations at ApTpA, ApTpG and ApTpT trinucleotides, shows strong transcriptional strand bias, and possibly reflects the involvement of transcription-coupled nucleotide excision repair acting on bulky DNA adducts due to exogenous carcinogens.[10] More recently, Gillison et al reported a high prevalence of Signature 16 in HPV-negative as compared to HPV-positive HNSCC from two cohorts (TCGA-HNSC and Ohio cohort).[13] In addition to an association with smoking and HPV-negative cancers, Signature 16 is also associated with alcohol intake.[30,34-36]

In a recent study, Signature 4 was found mainly in cancers derived from epithelia directly exposed to tobacco smoke and was most abundant in lung and laryngeal cancers.[29] Signature 4 mutations were also found in oral cavity and pharynx cancers, albeit in much smaller numbers most likely due to less exposure to tobacco smoke or more efficient clearance. A higher abundance of Signature 4 mutations based on our analysis was confirmed for the TCGA-HNSC and HIPO-HNC cohorts (data not shown), but as laryngeal cancers represent a minor subgroup within both cohorts, Signature 4 was not ranked as one of the top five signatures in our study.

Global gene expression profiling demonstrated that tumors with a highly relative mutational burden attributed to Signature 16 exhibit significantly lower p16^{INK4A} expression, accompanied by homozygous deletion of *CDKN2A*. *CDKN2A* deletions have been associated with HPV-negative HNSCC[13,37-40] and with a poor prognosis.[41,42] CDKN2A is a well-established tumor suppressor gene in squamous cell carcinoma and other types of human cancer.[43,44] The loss of p16^{INK4A} function could explain the unfavorable outcome of *CDKN2A*-deleted HNSCC with high frequency of Signature 16. Further studies should investigate whether this implies an absence of the primary inhibitory brake on CDK4/6 kinases mediating transition from G0/G1 phase to S phase of the cell cycle.[45] Up to date, it is still unclear whether low p16^{INK4A} expression, caused by *CDKN2A* deletion or other modes of action, is either a consequence or contributes to an enrichment of somatic mutations attributed to Signature 16. However, these survival differences retained prognostic significance in multivariate analyses, suggesting that a high mutational burden of *Signature 16*, connected with *CDKN2A* copy number loss, may have clinical practice as an independent prognostic factor for advanced HNSCC. Further studies are

necessarily offering a therapeutic approach for patients with a high mutational burden of Signature 16, for example, the use of checkpoint kinase inhibitors.[46,47] Targeting of CDK4 and 6 is subject of many clinical early phase III trials.[45] Although our sample size was limited, our integration of exomic sequencing data, combined with the validation of our findings in a larger cohort of clinical samples, provides a complementary breadth and depth of molecular information. In order to further unravel the prognostic impact of mutational signatures in HNSCC, especially for Signatures 3 and 16, prospective studies are warranted. Another attractive avenue in clinical translation of data presented in our study is the combination of mutation frequencies attributed to clinically relevant and highly abundant signatures with less abundant signatures to establish a reliable prognostic risk score for treatment intensification or de-escalation of HNSCC patient of distinct category (HPV-driven or not).

CONFLICT OF INTEREST

J.H. reports receiving a commercial research support from CureVac AG and is consultant/advisory board member for Bristol-Myers Squibb and MSD Sharp & Dohme. W.W. reports receiving a commercial research support and is consultant/advisory board member from Roche, MSD, BMS, AstraZeneca, Pfizer, Merck, Lilly, Boehringer, Novartis, Takeda, Amgen, Astellas and receives research fundings from Roche, MSD, BMS, AstraZeneca and Bruker. No potential conflicts of interest were disclosed by other authors.

DATA AVAILABILITY STATEMENT

The data discussed in this publication have been deposited in NCBI's Gene Expression Omnibus and are accessible through GEO Series accession number GSE117973 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE117973).

ETHICS STATEMENT

Patient samples were obtained under protocol S-206/2011, approved by the Ethics Committee of Heidelberg University, with written informed consent from all participants. Our study was conducted in accordance with the Declaration of Helsinki.

ORCID

Karim Zaoui https://orcid.org/0000-0002-6676-2273

REFERENCES

1. Thompson L. World Health Organization classification of tumours: pathology and genetics of head and neck tumours. *Ear Nose Throat J.* 2006;85:74.

2. Leemans CR, Snijders PJF, Brakenhoff RH. The molecular landscape of head and neck cancer. *Nat Rev Cancer*. 2018;18:269-282.

3. Gillison ML, Chaturvedi AK, Anderson WF, Fakhry C. Epidemiology of human papillomavirus-positive head and neck squamous cell carcinoma. *J Clin Oncol*. 2015;33:3235-3242.

4. Lefevre M, Rousseau A, Rayon T, et al. Epithelial to mesenchymal transition and HPV infection in squamous cell oropharyngeal carcinomas: the papillophar study. *Br J Cancer*. 2017;116:362-369.

5. Ang KK, Harris J, Wheeler R, et al. Human papillomavirus and survival of patients with oropharyngeal cancer. *N Engl J Med*. 2010;363: 24-35.

6. Argiris A, Karamouzis MV, Raben D, Ferris RL. Head and neck cancer. *Lancet*. 2008;371:1695-1709.

7. Laramore GE, Scott CB, al-Sarraf M, et al. Adjuvant chemotherapy for resectable squamous cell carcinomas of the head and neck: report on intergroup study 0034. *Int J Radiat Oncol Biol Phys*. 1992;23:705-713.

8. LQM C. Head and neck cancer. *N Engl J Med*. 2020;382:60-72.

9. Cooper JS, Pajak TF, Forastiere AA, et al. Postoperative concurrent radiotherapy and chemotherapy for high-risk squamous-cell carcinoma of the head and neck. *N Engl J Med*. 2004;350:1937-1944.

10. Alexandrov LB, Nik-Zainal S, Wedge DC, et al. Signatures of mutational processes in human cancer. *Nature*. 2013;500:415-421.

11. Faden DL, Thomas S, Cantalupo PG, Agrawal N, Myers J, DeRisi J. Multimodality analysis supports APOBEC as a major source of mutations in head and neck squamous cell carcinoma. *Oral Oncol*. 2017;74:8-14.

12. Vossen DM, Verhagen CVM, Verheij M, Wessels LFA, Vens C, van den Brekel MWM. Comparative genomic analysis of oral versus laryngeal and pharyngeal cancer. *Oral Oncol*. 2018;81:35-44.

13. Gillison ML, Akagi K, Xiao W, et al. Human papillomavirus and the landscape of secondary genetic alterations in oral cancers. *Genome Res*. 2019;29:1-17.

14. Alexandrov LB, Stratton MR. Mutational signatures: the patterns of somatic mutations hidden in cancer genomes. *Curr Opin Genet Dev*. 2014;24:52-60.

15. Kostareli E, Hielscher T, Zucknick M, et al. Gene promoter methylation signature predicts survival of head and neck squamous cell carcinoma patients. *Epigenetics*. 2016;11:61-73.

16. Kostareli E, Holzinger D, Bogatyrova O, et al. HPV-related methylation signature predicts survival in oropharyngeal squamous cell carcinomas. *J Clin Invest*. 2013;123:2488-2501.

17. Lechner M, Fenton T, West J, et al. Identification and functional validation of HPV-mediated hypermethylation in head and neck squamous cell carcinoma. *Genome Med*. 2013;5:15.

18. Schmitt K, Molfenter B, Koerich Laureano N, et al. Somatic mutations and promotor methylation of the ryanodine receptor 2 is a common event in the pathogenesis of head and neck cancer. *Int J Cancer*. 2019;145:3299-3310.

19. Grobner SN, Worst BC, Weischenfeldt J, et al. The landscape of genomic alterations across childhood cancers. *Nature*. 2018;555: 321-327.

20. Schmitt M, Bravo IG, Snijders PJ, Gissmann L, Pawlita M, Waterboer T. Bead-based multiplex genotyping of human papillomaviruses. *J Clin Microbiol*. 2006;44:504-512.

21. Schmitt M, Dondog B, Waterboer T, Pawlita M. Homogeneous amplification of genital human alpha papillomaviruses by PCR using novel broad-spectrum GP5+ and GP6+ primers. *J Clin Microbiol*. 2008;46: 1050-1059.

22. Halec G, Schmitt M, Dondog B, et al. Biological activity of probable/possible high-risk human papillomavirus types in cervical cancer. *Int J Cancer*. 2013;132:63-71.

23. Cao S, Wendl MC, Wyczalkowski MA, et al. Divergent viral presentation among human tumors and adjacent normal tissues. *Sci Rep*. 2016; 6:28294.

24. Metsalu T, Vilo J. ClustVis: a web tool for visualizing clustering of multivariate data using principal component analysis and heatmap. *Nucleic Acids Res*. 2015;43:W566-W570.

25. Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics*. 2016;32: 2847-2849.

26. Robinson JT, Thorvaldsdottir H, Wenger AM, Zehir A, Mesirov JP. Variant review with the integrative genomics viewer. *Cancer Res*. 2017;77:e31-e34.

27. Cancer Genome Atlas N. Comprehensive genomic characterization of head and neck squamous cell carcinomas. *Nature*. 2015;517:576-582.

28. Nik-Zainal S, Wedge DC, Alexandrov LB, et al. Association of a germline copy number polymorphism of APOBEC3A and APOBEC3B with burden of putative APOBEC-dependent mutations in breast cancer. *Nat Genet*. 2014;46:487-491.

29. Alexandrov LB, Ju YS, Haase K, et al. Mutational signatures associated with tobacco smoking in human cancer. *Science (New York, NY)*. 2016;354:618-622.

30. Lin DC, Dinh HQ, Xie JJ, et al. Identification of distinct mutational patterns and new driver genes in oesophageal squamous cell carcinomas and adenocarcinomas. *Gut*. 2018;67:1769-1779.

31. Qin T, Zhang Y, Zarins KR, et al. Expressed HNSCC variants by HPV-status in a well-characterized Michigan cohort. *Sci Rep*. 2018;8:11458.

32. Cannataro VL, Gaffney SG, Sasaki T, et al. APOBEC-induced mutations and their cancer effect size in head and neck squamous cell carcinoma. *Oncogene*. 2019;38:3475-3487.

33. O'Rorke MA, Ellison MV, Murray LJ, Moran M, James J, Anderson LA. Human papillomavirus related head and neck cancer survival: a systematic review and meta-analysis. *Oral Oncol*. 2012;48:1191-1201.

34. Letouzé E, Shinde J, Renault V, et al. Mutational signatures reveal the dynamic interplay of risk factors and cellular processes during liver tumorigenesis. *Nat Commun*. 2017;8:1315.

35. Chang J, Tan W, Ling Z, et al. Genomic analysis of oesophageal squamous-cell carcinoma identifies alcohol drinking-related mutation signature and genomic alterations. *Nat Commun*. 2017;8:15290.

36. Li XC, Wang MY, Yang M, et al. A mutational signature associated with alcohol consumption and prognostically significantly mutated driver genes in esophageal squamous cell carcinoma. *Ann Oncol*. 2018;29: 938-944.

37. Lim SM, Cho SH, Hwang IG, et al. Investigating the feasibility of targeted next-generation sequencing to guide the treatment of head and neck squamous cell carcinoma. *Cancer Res Treatm*. 2019;51:300-312.

38. Ludwig ML, Kulkarni A, Birkeland AC, et al. The genomic landscape of UM-SCC oral cavity squamous cell carcinoma cell lines. *Oral Oncol*. 2018;87:144-151.

39. Seiwert TY, Zuo Z, Keck MK, et al. Integrative and comparative genomic analysis of HPV-positive and HPV-negative head and neck squamous cell carcinomas. *Clin Cancer Res*. 2015;21:632-641.

40. Lechner M, Frampton GM, Fenton T, et al. Targeted next-generation sequencing of head and neck squamous cell carcinoma identifies novel genetic alterations in HPV+ and HPV- tumors. *Genome Med*. 2013;5:49.

41. Vossen DM, Verhagen CVM, van der Heijden M, et al. Genetic factors associated with a poor outcome in head and neck cancer patients receiving definitive chemoradiotherapy. *Cancers (Basel)*. 2019;11(4):445.

42. Chen WS, Bindra RS, Mo A, et al. CDKN2A copy number loss is an independent prognostic factor in HPV-negative head and neck squamous cell carcinoma. *Front Oncol*. 2018;8:95.

43. Kamb A, Gruis NA, Weaver-Feldhaus J, et al. A cell cycle regulator potentially involved in genesis of many tumor types. *Science (New York, NY)*. 1994;264:436-440.

44. Nobori T, Miura K, Wu DJ, Lois A, Takabayashi K, Carson DA. Deletions of the cyclin-dependent kinase-4 inhibitor gene in multiple human cancers. *Nature*. 1994;368:753-756.

45. O'Leary B, Finn RS, Turner NC. Treating cancer with selective CDK4/6 inhibitors. *Nat Rev Clin Oncol*. 2016;13:417-430.

46. Gadhikar MA, Zhang J, Shen L, et al. CDKN2A/p16 deletion in head and neck cancer cells is associated with CDK2 activation, replication stress, and vulnerability to CHK1 inhibition. *Cancer Res*. 2018;78: 781-797.

47. Jiang YZ, Ma D, Suo C, et al. Genomic and transcriptomic landscape of triple-negative breast cancers: subtypes and treatment strategies. *Cancer Cell*. 2019;35:428.e5-440.e5.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.