**Description of Supplementary data files:**

**Supplementary Figure S1. Workflow showing the processing and usage of HNSCC samples.** Samples with more that 60% tumor content were included for HPV consensus testing and next generation sequencing and other analyses**.**

**Supplementary Figure S2. Selection of mutated genes sorted by HPV association (q-values).** Histogram shows percentage of cases with mutations occurring in HPV(-) and HPV(+) samples.

**Supplementary Figure S3. Overall survival (OS) and progression free survival (PFS) in different subgroups. (A)** Subgroup analysis by HPV status and oropharyngeal primary site, **(B)** subgroup analysis by PIK3CA mutational status and oropharyngeal primary site, and **(C)** Subgroup analysis by *TP53* mutational status and oropharyngeal primary site.

**Supplementary Figure S4. Disruption of various key cellular processes by genomic alterations.** Genomic alterations that affect: **(A)** apoptosis, **(B)** Differentiation, **(C)** Oxidative stress and (D) Nucleic acid processing/modification.

**Supplementary Table S1:** **List of genes and significant copy number changes.** **(A)** List of genes included in the sequencing analysis/capture, **(B)** Validation rate by variant class, **(C)** Validation rate by gene **(D)** Significant copy number gain in HPV negative, **(E)** Significant copy number loss in HPV negative, **(F)** Significant copy number gain in HPV positive, **(G)** Significant copy number loss in HPV positive **(H)** Validation of significant copy number changes.

**Supplementary Table S2: Comparison of mutation events across different subgroups. (A)** Summary of mutation events across different subgroups.Mutation event difference between: **(B)** Smokers vs non-smokers, **(C)** Drinkers vs non-drinkers, **(D)** Large primary vs small primary, **(E)** Node positive vs node negative, (F) Failure vs non-failure.

**Supplementary Data 1**: **Somatic mutations in 120 HNC patients in MAF format.** List of somatic mutations with the mutated gene name, genomic location, dbSNP id and predicted change in protein.

**Supplementary Methods:**

**Sample preparation and DNA purification**

Sample originated from surgical biopsies or resections of HNSCC prior to definitive therapy with chemo- and radiotherapy. Selection was based on availability of sufficient tumor frozen tissue; tissues were derived from patients treated on six FHX-based chemoradiotherapy protocols. Information including survival (PFS/OS) was extracted from clinical databases.

Matched normal DNA was obtained from blood (n=24) or other “normal” tissue (e.g. uninvolved lymph nodes, skin biopsies, etc)(n=96). An overview of the tissue-processing is provided in *Figure S1*

A section was cut from frozen OCT or FFPE blocks and stained with hematoxylin and eosin (H&E). HE slides were reviewed by a head and neck pathologist (ML) to confirm histology and tumor content. Samples/areas of tumor that contained ≥60% cancer cells were processed using macro-dissection of tumor. Guided by the H&E stained slides, the region with normal tissue or the highest tumor content was cut from the tissue specimen or scraped from the FFPE slide.

 OCT blocks were cut, pulverized using CryoPrep (Covaris, Woburn, MA) and homogenized in lysis buffer from an AllPrepRNA/DNA/Protein Mini kit (Qiagen, Valencia, CA) or from an RNA/DNA/Protein Purification Kit (Norgen Biotek, Thorold, Canada) using an Ultrasonicator (Covaris, Woburn, MA). DNA, RNA and protein were isolated from each sample using the respective kit and following the manufacturer’s protocol. Depending on the size of the encircled area up to five FFPE slides with a thickness of 10µm each were used for DNA extraction. The tissue was scraped from the slides; material was put in 100% ethanol, spun down for two minutes at room temperature (RT) and maximum speed. The supernatant was removed and the pellet was dried over night at RT. Subsequent purification of nucleic acids was performed using an AllPrep DNA/RNA FFPE kit (Qiagen, Valencia, CA) following the manufacturer’s protocol for extraction with xylene with the modification of using Citrisolv (Fisher) instead of xylene. DNA isolated from blood samples (as obtained by pathology) was isopropanol precipitated in a final concentration of 0.3M sodium acetate to concentrate and clean up the samples. DNA concentrations and quality were determined through Quant-iT PicoGreen (Invitrogen, Carlsbad, CA) and spectrofluorometric measurements (Nanodrop). Whole genome amplification (WGA) was performed when nucleic acid yield after purification was not sufficient for library preparation. We used a REPLI-g Mini kit (Qiagen, Valencia, CA) to perform WGA on DNA from frozen samples and a GenomePlex Complete Whole Genome Amplification (WGA) kit (Sigma-Aldrich) for DNA purified from FFPE following the instructions of the manufacturers.

**HPV PCR and HPV consensus testing**

E6/E7 DNA based multiplex, nested PCR for five high risk HPV types (1) using AmpliTaq Gold (Applied Biosystems), as well as an E6/E7-specific HPV16/18 RNA based qRT-PCR were used to determine HPV status. SCC2 (HPV16) and HeLa (HPV18) cell lines were used as positive controls whereas SCC151 and water were used as negative control and no template control respectively. PCR products were analyzed on a 2% agarose gel.

Results were corroborated by p16 IHC for oropharyngeal tumors (where available), *CDKN2A* mRNA expression measurements, *TP53* mutational status, as well as an HPV gene expression signature in order to identify HPV related tumors with very high accuracy and exclude HPV testing errors that may occur with single HPV testing methods e.g. false positive p16 testing or non-etiologically related HPV-bystanders/colonization as previously reported (2).

**Library preparation and target capture**

 DNA sequencing libraries were prepared using 100-1000ng (300-500ng for most samples) of starting material as determined by Quant-iT PicoGreen (Invitrogen, Carlsbad, CA) and following a previously published protocol(3) with some minor modifications: Fragmentation of DNA to 300bp prior to library preparation was performed with an Ultrasonicator (Covaris, Woburn, MA); adapter tagged sequencing libaries were subjected to an e-gel (Invitrogen) for size selection with a subsequent cleanup using a QIAquick PCR Purification kit (Qiagen, Valencia, CA); indexing PCRs were performed in four parallel reactions (18 cycles) with a subsequent QC step using an Agilent 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA).

For target capture on a microarray indexed libraries were pooled 12plex and hybridized on a customized capture slide (Agilent CGH array) following published methodology (4) and applying the modifications suggested by Meyer et al (3). Further modifications: the eluted material was concentrated through ethanol precipitation; for amplification after capture cycle number was reduced to 13 or 14, primers IS5 and IS6 and protocol for indexing PCR according to Meyer et al were used (3). For solution phase capture then manufacturer protocol was followed (Nimblegen EZ Choice, Madison, WI)

**CN aberration/Nanostring**

Copy number aberration was determined using the nCounter Cancer CN Kit as well as a custom designed assay (NanoString Technologies, Seattle, WA) following the instructions of the manufacturer with one minor modification: Instead of the *Alu*I digestion, 300ng of genomic DNA (measured with Quant-iT Picogreen (Invitrogen)) was sheared to ~300bp fragments using a Covaris ultrasonicator S-series (duty cycle 10%, Intensity 4, cycles per burst 200, time 120 seconds) followed by centrifugation in a vacufuge to reduce the sample volume. If the initial volume taken for shearing exceeded the volume needed for the CN assay, the fragmented DNA was cleaned up using a QIAquick PCR Purification Kit (Qiagen, Germany) prior to the aforementioned centrifugation step. Normal DNA isolated from the peripheral blood of four individuals was used as a reference. If required, DNAs were treated with RNaseI (Fermentas Inc., Maryland) to remove RNA and a subsequent column cleanup was performed to remove RNase I.

Information, such as sex, age, tumor stage, smoking/alcohol record and including survival (PFS/OS) was extracted from clinical research databases for this cohort. OS was measured from the date of diagnosis until the date of death databases.

**Sequencing data analysis**

Mutation and indel calling was done using the bioinformatic pipeline as described previously (www.broadinstitute.org/cancer/cga) (5-7).

We used an established machine learning based approach – Cancer-Specific High-Throughput Annotation of Somatic Mutations (CHASM) – to predict and prioritize missense mutations leading to functional changes and thus likely driving tumorigenesis (8). This approach was previously validated to show high specificity, and to a lesser degree sensitivity to identify “drivers” of oncogenesis (8).

Copy number (CN) analysis was performed using sequencing data and the CONTRA algorithm (9). Results were validated on the Nanostring nCounter using the predefined/custom cancer gene panels (Nanostring, Seattle, WA).

VarWalker (10) was used to determine potentially relevant genetic aberrations (mutations and copy number aberrations) using prioritization via a protein-protein interaction network focusing on frequently mutated and cancer gene census genes.

**Statistical analysis**

A GISTIC2.0 like strategy was used to determine the statisitcal significance of CNAs (11). The pair of genes and samples of the CNA data were permuted 1000 times to obtain a null distribution of random G-scores (average amplitudes times aberration frequency), followed by a calucation of the statistical significance of each G-socre in the read dataset (adjusted by multiple test).

Fisher exact test was used to determine the relationship of clinical characteristics to HPV status.

One-sided t-test was used to determine the significance of the difference of mutation number between non-smoker/light smoker and heavy smoker in HPV positive and HPV negative cohort, respectively. The same analysis was performed for the comparison of TP53 mutation and TP53 wild type patients. TP53 mutations were classified into DISRUPTIVE mutations and NON-DISRUPTIVE mutations according to Poetal *et, al* (12)*.*

To assess the relationship of survival to the mutations, the R package (survival) was used to perform statiscal analysis. PFS time was defined as the time from date of initial clinical visit until the date of progression (metastasis or 1st recurrence). The Kaplan-Meier estimator was used to estimate the distribution of the 5-year survival. Log-rank test was used to compare the different survival distributions.

**Supplementary Material References:**

1. Sotlar K, Diemer D, Dethleffs A, Hack Y, Stubner A, Vollmer N, et al. Detection and typing of human papillomavirus by e6 nested multiplex PCR. J. Clin. Microbiol. 2004Jul.;42(7):3176–84.

2. zuo Z, Keck MK, Patel R, Khattri A, Walter K, Lingen MW, et al. Multimodality determination of HPV status in head and neck cancers (HNC) and development of an HPV signature. J Clin Oncol 31, 2013 (suppl; abstr 6008). 2013Jun.3.

3. Meyer M, Kircher M. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. Cold Spring Harb Protoc. 2010Jun.;2010(6):pdb.prot5448.

4. Hodges E, Rooks M, Xuan Z, Bhattacharjee A, Benjamin Gordon D, Brizuela L, et al. Hybrid selection of discrete genomic intervals on custom-designed microarrays for massively parallel sequencing. Nat Protoc. 2009;4(6):960–74.

5. Cibulskis K, McKenna A, Fennell T, Banks E, DePristo M, Getz G. ContEst: estimating cross-contamination of human samples in next-generation sequencing data. Bioinformatics. 2011Sep.15;27(18):2601–2.

6. Hammerman PS, Lawrence MS, Voet D, Jing R, Cibulskis K, Sivachenko A, et al. Comprehensive genomic characterization of squamous cell lung cancers. Nature. 2012Sep.9;489(7417):519–25.

7. Cibulskis K, Lawrence MS, Carter SL, Sivachenko A, Jaffe D, Sougnez C, et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. Nat Biotechnol. 2013Mar.;31(3):213–9.

8. Carter H, Chen S, Isik L, Tyekucheva S, Velculescu VE, Kinzler KW, et al. Cancer-specific high-throughput annotation of somatic mutations: computational prediction of driver missense mutations. Cancer Research. 2009Aug.15;69(16):6660–7.

9. Li J, Lupat R, Amarasinghe KC, Thompson ER, Doyle MA, Ryland GL, et al. CONTRA: copy number analysis for targeted resequencing. Bioinformatics. 2012May15;28(10):1307–13.

10. Jia P, Zhao Z. VarWalker: Personalized Mutation Network Analysis of Putative Cancer Genes from Next-Generation Sequencing Data. PLoS Comput Biol. 2014Feb.1;10(2):e1003460–0.

11. Mermel CH, Schumacher SE, Hill B, Meyerson ML, Beroukhim R, Getz G. GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. Genome Biology. 2011;12(4):R41.

12. Poeta ML, Manola J, Goldwasser MA, Forastiere A, Benoit N, Califano JA, et al. TP53 mutations and survival in squamous-cell carcinoma of the head and neck. N Engl J Med. 2007Dec.20;357(25):2552–61.