Human and viral whole genome sequencing identify HPV and APOBEC as oncogenic drivers in sinonasal squamous cell carcinoma

Harrison B. Chong[1,13], Michael E. Bryan[2,3,4], Maoxuan Lin[1,4], William C. Faquin[5,6,7], Lisa Mirabello[8], Sambit K. Mishra[8,9], James S. Lewis Jr[10,11,12], Michael Lawrence[1,4,5*], Daniel Faden[2,4,6,7*¶]

[1] Krantz Family Center for Cancer Research, Massachusetts General Hospital, Charlestown, MA, United States
[2] Department of Otolaryngology-Head and Neck Surgery, Harvard Medical School, Boston, MA, United States
[3] Nuffield Department of Medicine, University of Oxford, Oxford, United Kingdom
[4] Broad Institute of MIT and Harvard, Cambridge, MA, United States
[5] Department of Pathology, Massachusetts General Hospital, Boston, MA, United States
[6] Massachusetts General Hospital Cancer Center, Boston, MA, United States
[7] Massachusetts Eye and Ear, Boston, MA, United States
[8] Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Rockville, MD
[9] Cancer Genomics Research Laboratory, Leidos Biomedical Research, Inc., Frederick, MD, United States
[10] Department of Otolaryngology-Head and Neck Surgery, Vanderbilt University Medical Center, Nashville, TN, United States
[11] Department of Pathology, Microbiology and Immunology, Vanderbilt University Medical Center, Nashville, TN, United States
[12] Department of Laboratory Medicine and Pathology, Mayo Clinic, Scottsdale, AZ, United States
[13] Children's Medical Center Research Institute, University of Texas Southwestern Medical Center, Dallas, TX, USA

Authors' Disclosures: D. L. Faden reports grants from the NIH/NIDCRK23DE029811, NIH/NIDCRR03DE030550, and NIH/NCIR21CA267152 during the conduct of the study as well as grants from Bristol Myers Squibb, Calico, and Haystack(Quest); other support from Predicine, BostonGene, and Neogenomics; and personal fees from Merck, Chrysalis Biomedical Advisors, and Arcadia outside the submitted work; in addition, D. L. Faden has a patent for 169511-00058-MEEI2024-32 pending. No disclosures were reported by the other authors.

Abstract

Sinonasal squamous cell carcinoma (SNSCC) is an aggressive head and neck cancer of the sinonasal cavity which has not benefitted from therapeutic advances over decades[1]. Though historically attributed to inhaled carcinogens such as hardwood dust and tobacco smoking[2], SNSCC is incidentally associated with human papillomavirus (HPV)[3,4]. Importantly, HPV is the primary oncogenic driver of >80% of anatomically adjacent oropharyngeal cancers[5]. While viral status drives clinical staging and treatment guidelines in these malignancies[6,7], the potentially oncogenic consequences and prognostic value of host–virus interactions in SNSCC remain incompletely defined. Here, through paired host and viral whole-genome sequencing (WGS), we map the genomic footprint of HPV in SNSCC. Strikingly, lesser studied strains such as HPV45, 51, and 39 constitute driver infections in this rare but clinically credentialed cancer, where extrachromosomal DNA (ecDNA)-associated viral integration and APOBEC mutagenesis are shown to underpin somatic tumor evolution.

Statement of Significance

Paired host viral and whole-genome sequencing of SNSCC nominates HPV as a primary oncogenic driver of SNSCC. HPV–human ecDNA amplicons harboring noncanonical strains such as HPV45, 51 mediate viral carcinogenesis. Routine clinical diagnostic HPV panels should be expanded to capture the activity of lesser studied strains.

Introduction

Our group and others have demonstrated that HPV is detectable in cancers in the nasopharynx and sinonasal tract[8,9], possibly due to refluxed secretions from the oropharynx. Notably, these cancers are rising both in incidence and within the same geographic locale as HPV+OPC[4]. Previous profiling suggests that HPV^{+} SNSCC shares mutational profiles with associated cervical and head and neck squamous cell carcinoma[10]. However, definitive genomic mapping of potentially oncogenic host-virus interactions through systematic sequencing of paired tumor-normal specimens remains under-explored. Here, we leverage

joint whole genome sequencing (WGS) of both human and HPV genomes to interrogate the mutational signatures and structural events underpinning viral oncogenesis in SNSCC.

Results

Utilizing a cohort of 21 SNSCC paired tumor-normal specimens, we subjected each sample to orthogonal tumor WGS (mean depth, 59x), matched normal WGS (mean depth, 30x), ultra-deep WGS (10,000x) of the viral genome, and a broad HPV DNA PCR panel (see Methods, Supplementary Table 11). Eighteen of 21 (86%) cases were HPV-associated by human WGS and/or DNA PCR, with 94.4% concordance between the two technologies (Figure 1b; 17/18 jointly profiled; 95% CI, 74.2–99.0%; binomial p = <0.001, see Methods). Genotype concordance was 81.0% (17/21; 95% CI, 60.0–92.3%; binomial p = 0.001). Sixteen of 21 (81%) were p16-positive (Figure 1c). Among patients reporting a smoking history, 7/14 (50%) had ≥10 pack years (Figure 1c, Supplementary Table 12). Unlike HPV^{+} OPC, where HPV16 predominates (~90%)[11,12], only 5/18 (27%) HPV+ SNSCC harbored single-genotype HPV16 (Supplementary Table 5). We observed nine distinct HPV genotypes (Figure 2a, Supplementary Table 5), including two multi-genotype infections and two tumors with low-risk HPV11 infection. Both low-risk (e.g. HPV11) infections and a plurality of high-risk genotypes which vary in prevalence across human cancers (e.g. HPV16,18,35,39,45,51,56,59) were present. Moreover, chimeric reads revealed an ecDNA species fusing viral DNA from HPV45, a rare genotype in HPV^{+}OPC, to host DNA from the *TP63* oncogene (Figure 1d, Supplementary Table 9), strongly suggesting an oncogenic role of less-common HPV genotypes in SNSCC. Interestingly, a broad representation of high-risk HPV genotypes other than HPV16 has also been observed in nasopharyngeal cancer, an oropharynx-adjacent site where HPV appears to play a role[13]. One HPV11^{+} tumor lacked any high-risk genotype and showed no evidence of viral integration but harboured a TP53 mutation and was p16-negative, consistent with HPV11 as a bystander in this case[14] (Supplemental Tables 1-2, 11).

Somatic retro-integration of all or part of the HPV genome is a well-described carcinogenic mechanism in HPV^{+} OPC, as well as cervical and anal cancers[15,16]. We identified high-confidence integration breakpoints in most HPV^{+} tumors, with as many as 7 distinct breakpoints per tumor. In total, 28 unique, high evidence viral-host junctions were mapped (Figure 2b, Supplemental Tables 5-9). In cases harboring multi-genotype infections, only a single dominant integrating genotype was observed per tumor, suggesting that one genotype primarily drives oncogenic progression, even amid co-infection by other HPV types. This observation is consistent with findings in cervical and anal cancers[17,18].

Our cohort also exhibited novel patterns of HPV integration (Figure 2c,d). In several cases, integration was accompanied by profound local copy-number amplification of host and viral sequences, including a retro-integration event involving the 3′ region of TP63 in an HPV45^{+} tumor (Figure 2d, top). Focal segmentation of the somatic breakpoints delimiting this fusion revealed a ~50kb segment of chromosome 3q which achieved amplification to ~15–20X genomic median (i.e. “copies”). This segment included the HPV45 E6 URR sequences fused to TP63. The structure and amplitude of this amplicon resemble oncoviral integration-linked extrachromosomal DNA (ecDNA)[19,20] (Figure 2d, top). To our knowledge, this is the first demonstration of an HPV ecDNA-like amplification in SNSCC.

We further observed amplification in a tumor with HPV51 with three distinct HPV51 breakpoints clustered at the 3′ end of MAN1A2. Each breakpoint was associated with a distinct copy-number profile, but all showed high-level amplification (~15–50 copies). This suggests either sequential chromosomal rearrangements or the presence of multiple “nested” ecDNA species containing segments of HPV51 and MAN1A2 (Figure 2d, bottom). Initial integration may have been followed by stepwise extrusion of some segments via iterative breakage-fusion cycles. Yet, not all HPV integrations cause detectable copy-number changes. For instance, one tumor showed history of HPV18 integration into the 3’ end of N4BP3 without associated copy gain or loss (Figure 2d, middle).

We next interrogated the host somatic mutational landscape. The overall mutation burden was comparable to that reported in HPV^{+}OPC[21], and clonal mutations were observed in common driver genes regulating mitogenic signaling and genomic integrity (e.g. *PIK3CA, KDM6A, MLL3, BCOR, CHEK2*) (Figure 1c, g; Supplemental Tables 1,2). Among high-risk HPV+ cases, 4/17 (23%) harbored *PIK3CA* missense mutations (Figure 1e). Three of these were in the helical domain, a hotspot enriched for APOBEC3-mediated mutations[22]. *TP53* and *CDKN2A* mutations were mutually exclusive with high-risk HPV infection, whereas *PIK3CA* mutations were seen only in high-risk HPV+ cases. These observations are consistent with reports of depletion of *TP53/CDKN2A* mutation in HPV+OPC, as well as enrichment of mutations in the PIK3CA helical domain[23]. FGFR3 S249C mutations, recently noted in SNSCC[10], were not detected in this cohort. Genome-wide, SNSCCs exhibited a mutational rate like that of previously profiled HPV-independent head and neck squamous cell carcinoma (HPV- HNSCC) (Extended Data Fig. 2a). Despite reasonable sequencing depth and tumor purity, we did not recover any confidently-called somatic mutations in the tumors of two patients (tumor purity MS09 = 0.64, 21530 = 0.46).

Knowing that APOBEC mutagenesis is a primary driver in HPV-associated carcinomas[24], we interrogated mutational signatures in both host and viral genomes. Factorization of somatic mutations identified predominant APOBEC mutagenesis in a subset of SNSCC (Figure 1e). We observed low burden of the smoking-related signature SBS4 in patients with a history of smoking[25,26], arguing against smoking as a primary etiology of SNSCC[27]. When projected alongside TCGA HNSCC, SNSCC showed a greater diversity of APOBEC3 activity and a similar overall mutational burden (Extended Data Fig. 2c, 2d). HPV^{+} SNSCC appears to have less APOBEC3 mutagenesis on average than its less histologically specialized counterpart (HPV^{+} HNSC), but 2/21 (9%) of HPV^{+} SNSCC tumors ranked among the most enriched APOBEC-mutators in the TCGA-ICGC combined cohort (Extended Data Fig. 2b, left; Extended Data Fig. 2c, bottom).

Ultra-deep sequencing of intratumoral HPV genomes showed that APOBEC activity extends to the virus (Figure 2E, Supplementary Tables 6-7). Highly clonal virus mutations likely represent mutations present in the infecting virion; however, a small peak of lower allele frequency subclonal mutations in some instances represents new APOBEC-associated mutations that occurred in the current host (Figure 2e, f). Specifically, we detected the contribution of clonal APOBEC mutagenesis near the L1 domain in a subset of HPV genomes, corroborating ongoing APOBEC3 activity throughout the viral genome in tumors (Figure 2e, f). Concordance of host–virus APOBEC mutagenesis has to our knowledge not previously been demonstrated as a principle in SNSCC[28,29].

Knowing that structural variants (SVs) are well-represented in HNSCC[30], we next asked whether this principle extends to SNSCC (Extended Data Fig. 2a, Supplementary Tables 3, 10). We identified 221 somatic SVs, with an average of ~10 SVs per tumor. Consistent with patterns reported in other cancers[30], simple deletions were the most common class of SV. We observed focal deletions affecting known tumor suppressor loci – for example, loss of PTEN was recurrent in multiple tumors (Extended Data Fig. 2b), suggesting that PTEN inactivation may play a role in some SNSCCs. Moreover, PTEN loss is also seen in a subset of HPV^{+} OPC and could contribute to PI3K/AKT activation[31].

SVs targeting genes were common, where 41% of SVs localized to introns, exons, or other gene-regulatory elements (i.e. promoters), sometimes creating chimeric, in-frame fusion transcripts with epigenetic regulators such as NIPBL or BRWD1 (Extended Data Fig. 2c,e). We also identified chromosomal catastrophe in specific tumors. In one HPV^{+} SNSCC, copy-number oscillations and clustered breakpoints along a single chromosome arm characteristic of chromothripsis were detected, notably producing an in-frame ERG-BRWD1 fusion product[32]. Analysis of whole chromosomes also identified kataegis in a separate case[33] (Extended Data Fig. 2d, e). This suggests that as double-strand breaks were salvaged,

APOBEC might have been concurrently active, peppering exposed single-stranded DNA with C→T mutations. Additionally, a few tumors with otherwise low mutation burden demonstrated chromoplexy-like rearrangements, i.e. showers of interrelated intrachromosomal SVs (Extended Data Fig. 2e).

Discussion

In summary, HPV drives tumorigenesis in SNSCC, and in a subset of cases, the mutagenic activities of the APOBEC3 family of anti-viral DNA cytosine deaminases further accelerates tumor evolution. Moreover, viral WGS directly evidences APOBEC targeting to the viral genome. Viral integration events, particularly the extrachromosomal amplification of known oncogenes such as *TP63*, as well as genomic breakpoints and translocations, contribute to extensive structural variation. These structural variations collectively demonstrate the cooperative roles of both viral and non-viral axes of instability in tumor evolution. The diversity of HPV genotypes other than HPV16 provides a striking contrast to HPV+ OPC, suggesting less evolutionary advantage for HPV16 outside of the oropharynx. Importantly, nontraditional, high-risk HPV genotypes play an underrecognized role in SNSCC, pinpointing a need for diagnostic inclusion in routine clinical assays.

METHODS DETAILS

Patient Cohort and Sample Collection

Primary tumor and matched normal blood samples were collected from 21 patients diagnosed with sinonasal squamous cell carcinoma (SNSCC)[52]. Of these, 16 were from Mass General Brigham and 5 were from Vanderbilt University Medical Center. All patients provided written informed consent under protocols approved by both institutions. Clinicopathological features, including age, sex, smoking history, anatomical subsite, histologic subtype (keratinizing vs. non-keratinizing), and inverted papilloma association, were extracted from medical records. In addition to standard clinical assays (p16 immunohistochemistry and HPV PCR), all patients

underwent gold-standard HPV genotyping performed at the U.S. National Institutes of Health (NIH). HPV detection and typing were performed using the TypeSeq v1 and NCI Carcinogenic HPV All Next Generation Sequencing (NCI CHANGeS) assays. TypeSeq v1 detects 51 HPV genotypes, while NCI CHANGeS provides high-resolution genotyping of 12 high-risk types (HPV16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, and 59).[52]

WGS

DNA was extracted from the tumor and blood using the [Qiagen DNeasy Blood & Tissue Kit] according to the manufacturer's protocol. DNA quality and quantity were assessed with [Qubit fluorometry / TapeStation]. Libraries for whole-genome sequencing (WGS) were prepared using the Illumina TruSeq DNA PCR-Free Library Prep Kit. Paired-end sequencing (151 bp) was performed on an Illumina platform to an average depth of ~59x for tumor DNA and ~30x for matched normal DNA. Raw reads were trimmed to remove adaptors and low-quality bases with TrimGalore (v0.6.6, a wrapper of the Cutadapt program[34]. Reads were aligned to a hybrid human-HPV reference composed of human genome (GRCh37/hg19) and all HPV genomes downloaded from PaVE[35] using BWA-MEM[36] (v0.7.15) with default settings. Sorted, indexed BAM files were generated with SAMtools[37] (v1.4.1). PCR duplicates were marked with Sambamba[38] (v0.4.7).

HPV DNA was captured using a custom Ion AmpliSeq panel with 545 primer pairs, covering overlapping amplicons across the 12 high-risk HPV types considered in this study. The barcoded libraries were prepared using Ion AmpliSeq Library Kit 2.0 (Thermo Fisher Scientific, Carlsbad, CA). Primers were divided into two separate PCR reactions to avoid cross-reaction and primer dimerization. Amplicon lengths ranged from 69 bp to 324 bp, with a median length of 199 bp. Whole-genome single-end sequencing was performed for all 12 HR-HPV types concurrently on the Ion PGM platform. Raw reads were trimmed to remove adaptors and low-quality bases with TrimGalore (v0.6.10; a wrapper of the Cutadapt program[34]. Reads were aligned to the hybrid human-HPV reference using BWA-MEM[35] (v0.7.18) with default settings.

Sorted, indexed BAM files were generated with SAMtools[36] (v1.21). PCR duplicates were removed with Picard (v3.3.0; "Picard Toolkit." 2019. Broad Institute, GitHub Repository. https://broadinstitute.github.io/picard/; Broad Institute). HPV mutations were called using Mutect1[39] (v1.1.7) in unpaired mode using the hybrid human-HPV reference described previously. Low quality mutations with number of alternate alleles <2 and total number of reads < 20 were removed. The remaining mutations were manually reviewed with the Integrative Genomics Viewer[40] (IGV, v2.18.4) to exclude false positives. HPV APOBEC mutations were annotated into three categories: 1) high-confidence APOBEC mutations Tp(C->G/A)pN, 2) medium-confidence APOBEC mutations Tp(C->T)pH, and 3) low-confidence APOBEC mutations Tp(C->T)pG.

Mutation calling

Somatic single-nucleotide variants (SNVs) and small insertions and deletions (indels) were called using MuTect1[39] (v1.1.7) and Strelka2[41] (v2.9.10) using matched tumor/normal pairs. Low quality variants were first removed with progressively stricter filtering thresholds combining Panel of Normal (PoN) and minimum counts of alternate alleles in tumors (t_alt_count) for variants in different ranges of tumor allele fractions (AF): 1) for mutations with 1%<=AF < 2%, PoN <0.05% and t_alt_count >=5, 2) for mutations with 2% <=AF<3%, PoN < 0.1% and t_alt_count >=4, 3) for mutations with 3% <=AF<5%, PoN < 0.15% and t_alt_count >=3, and 4) for mutations with AF>=5%, PoN < 0.2% and t_alt_count >=2, in addition to a total read count of five in tumor, were applied. To remove FFPE artifacts in the remaining mutations, a combination of three tools: FiNGs[42] with default filters of the International Cancer Genome Consortium (ICGC), SOBDetector[43], and a perl tool that detects false positives with biased read diversity[44]. High-confidence variants were functionally annotated with ANNOVAR[45]. Coding variants with allele fraction > 10% in known driver genes and all coding indels were manually reviewed with the Integrative Genomics Viewer[40] (IGV, v2.18.4) to exclude false positives.

Mutational signature attribution

Non-negative matrix factorization (NMF) was applied to pooled samples from our cohort and The Cancer Gene Atlas (TCGA) to decompose mutations into respective single base substitution signatures. Loadings of dominant signatures with acceptable cosine similarity >= 0.85 were projected onto a three-dimensional manifold via the umap.UMAP function in the umap.umap_ Python package (n_neighbors=15, random_state=42, n_components=3).

Identification of structural variants and copy number alterations

Somatic SVs were identified using dRanger and BreakPointer as described previously[46]. Copy number segments were obtained via FACETS and subsequently fed into GISTIC analysis via the Broad pipeline. All HPV genomes were passed through SvABA for human-viral breakpoint detection[51], followed subsequently by manual review for high evidence (read lattering, sufficient read depth) of chimeric reads suggestive of human-HPV breakpoints.

ecDNA detection

Extrachromosomal DNA (ecDNA) amplifications were detected using Amplicon-pipeline[47] v1.3.9. Briefly, focal DNA copy-number variants (CNV) were first identified using CNVKit[48] v0.9.12 . Regions with copy number >=4.5 and size >50kbp were selected as seed regions. These seed regions were then passed to AmpliconArchitect v1.5.r2 to construct cyclic paths, and AmpliconClassifier v1.3.3 was used to determine whether these paths were likely to be ecDNA[47,49]. ecDNA candidates were manually reviewed with the Integrative Genomics Viewer[40] (IGV, v2.18.4) to identify real ecDNA amplifications.

References

1. Consonni, D. *et al.* Survival of Patients with Sinonasal Cancers in a Population-Based Registry, Lombardy, Italy, 2008-2023. *Cancers* **16**, 896 (2024).

2. Fukuda, K. & Shibata, A. Exposure-response relationships between woodworking, smoking or passive smoking, and squamous cell neoplasms of the maxillary sinus. *Cancer Causes Control CCC* **1**, 165–168 (1990).
3. Oliver, J. R. *et al.* Human papillomavirus and survival of patients with sinonasal squamous cell carcinoma. *Cancer* **126**, 1413–1423 (2020).
4. Chang Sing Pang, K. J. W., Mur, T., Collins, L., Rao, S. R. & Faden, D. L. Human Papillomavirus in Sinonasal Squamous Cell Carcinoma: A Systematic Review and Meta-Analysis. *Cancers* **13**, 45 (2020).
5. D'Souza, G. *et al.* Differences in the Prevalence of Human Papillomavirus (HPV) in Head and Neck Squamous Cell Cancers by Sex, Race, Anatomic Tumor Site, and HPV Detection Method. *JAMA Oncol.* **3**, 169–177 (2017).
6. Ang, K. K. *et al.* Human papillomavirus and survival of patients with oropharyngeal cancer. *N. Engl. J. Med.* **363**, 24–35 (2010).
7. Fakhry, C. *et al.* Improved survival of patients with human papillomavirus-positive head and neck squamous cell carcinoma in a prospective clinical trial. *J. Natl. Cancer Inst.* **100**, 261–269 (2008).
8. Hung, S.-H., Yang, T.-H., Cheng, Y.-F., Chen, C.-S. & Lin, H.-C. Association of Nasopharynx Cancer with Human Papillomavirus Infections. *Cancers* **15**, 4082 (2023).
9. Bryan, M. E. *et al.* Direct Comparison of Alternative Blood-Based Approaches for Early Detection and Diagnosis of HPV-Associated Head and Neck Cancers. *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.* OF1–OF11 (2025) doi:10.1158/1078-0432.CCR-24-2525.
10. Zamuner, F. T. *et al.* Molecular patterns and mechanisms of tumorigenesis in HPV-associated and HPV-independent sinonasal squamous cell carcinoma. *Nat. Commun.* **16**, 5285 (2025).
11. Gillison, M. L., Chaturvedi, A. K., Anderson, W. F. & Fakhry, C. Epidemiology of Human Papillomavirus–Positive Head and Neck Squamous Cell Carcinoma. *J. Clin. Oncol.* **33**, 3235–3242 (2015).

12. Nichols, A. C. *et al.* Does HPV type affect outcome in oropharyngeal cancer? *J. Otolaryngol. - Head Neck Surg. J. Oto-Rhino-Laryngol. Chir. Cervico-Faciale* **42**, 9 (2013).

13. Zhao, B. Y., Hirayama, S., Goss, D., Zhao, Y. & Faden, D. L. Human papillomavirus-associated nasopharyngeal carcinoma: A systematic review and meta-analysis. *Oral Oncol.* **159**, 107057 (2024).

14. Muñoz, N. *et al.* Epidemiologic classification of human papillomavirus types associated with cervical cancer. *N. Engl. J. Med.* **348**, 518–527 (2003).

15. Pinatti, L. M., Walline, H. M. & Carey, T. E. Human Papillomavirus Genome Integration and Head and Neck Cancer. *J. Dent. Res.* **97**, 691–700 (2018).

16. Zhang, Y., Qiu, K., Ren, J., Zhao, Y. & Cheng, P. Roles of human papillomavirus in cancers: oncogenic mechanisms and clinical use. *Signal Transduct. Target. Ther.* **10**, 44 (2025).

17. Schmitt, M. *et al.* Abundance of multiple high-risk human papillomavirus (HPV) infections found in cervical cells analyzed by use of an ultrasensitive HPV genotyping assay. *J. Clin. Microbiol.* **48**, 143–149 (2010).

18. Wong, A. K. *et al.* Human papillomavirus genotypes in anal intraepithelial neoplasia and anal carcinoma as detected in tissue biopsies. *Mod. Pathol. Off. J. U. S. Can. Acad. Pathol. Inc* **23**, 144–150 (2010).

19. Bafna, V. & Mischel, P. S. Extrachromosomal DNA in Cancer. *Annu. Rev. Genomics Hum. Genet.* **23**, 29–52 (2022).

20. McBride, A. A. & White, E. A. HPV Integration Can Drive the Formation of Virus-Host Extrachromosomal DNA in Tumors. *Cancer Discov.* **13**, 814–816 (2023).

21. Lim, Y. X., Mierzwa, M. L., Sartor, M. A. & D'Silva, N. J. Clinical, morphologic and molecular heterogeneity of HPV-associated oropharyngeal cancer. *Oncogene* **42**, 2939–2955 (2023).

22. Henderson, S., Chakravarthy, A., Su, X., Boshoff, C. & Fenton, T. R. APOBEC-mediated cytosine deamination links PIK3CA helical domain mutations to human papillomavirus-driven tumor development. *Cell Rep.* **7**, 1833–1841 (2014).

23. Cancer Genome Atlas Network. Comprehensive genomic characterization of head and neck squamous cell carcinomas. *Nature* **517**, 576–582 (2015).

24. Argyris, P. P. *et al.* Endogenous APOBEC3B overexpression characterizes HPV-positive and HPV-negative oral epithelial dysplasias and head and neck cancers. *Mod. Pathol. Off. J. U. S. Can. Acad. Pathol. Inc* **34**, 280–290 (2021).

25. Alexandrov, L. B. *et al.* The repertoire of mutational signatures in human cancer. *Nature* **578**, 94–101 (2020).

26. Wellcome Sanger Institute. *Mutational Signatures (v3.4 - October 2023) SBS4 GRCh37 COSMIC V102*. https://cancer.sanger.ac.uk/signatures/sbs/sbs4/ (2023).

27. Elgart, K. & Faden, D. L. Sinonasal Squamous Cell Carcinoma: Etiology, Pathogenesis, and the Role of Human Papilloma Virus. *Curr. Otorhinolaryngol. Rep.* **8**, 111–119 (2020).

28. Faden, D. L. *et al.* APOBEC Mutagenesis Is Concordant between Tumor and Viral Genomes in HPV-Positive Head and Neck Squamous Cell Carcinoma. *Viruses* **13**, 1666 (2021).

29. Vartanian, J.-P., Guétard, D., Henry, M. & Wain-Hobson, S. Evidence for editing of human papillomavirus DNA by APOBEC3 in benign and precancerous lesions. *Science* **320**, 230–233 (2008).

30. Sasa, N. *et al.* Intratumor heterogeneity of HPV integration in HPV-associated head and neck cancer. *Nat. Commun.* **16**, 1052 (2025).

31. Chiosea, S. I. *et al.* PIK3CA, HRAS and PTEN in human papillomavirus positive oropharyngeal squamous cell carcinoma. *BMC Cancer* **13**, 602 (2013).

32. Stephens, P. J. *et al.* Massive genomic rearrangement acquired in a single catastrophic event during cancer development. *Cell* **144**, 27–40 (2011).

33. Nik-Zainal, S. *et al.* Mutational processes molding the genomes of 21 breast cancers. *Cell* **149**, 979–993 (2012).

34. Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. journal*, *17*(1), 10-12.

35. Van Doorslaer, K., Li, Z., Xirasagar, S., Maes, P., Kaminsky, D., Liou, D., ... & McBride, A. A. (2017). The Papillomavirus Episteme: a major update to the papillomavirus sequence database. *Nucleic acids research*, *45*(D1), D499-D506.

36. Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv preprint arXiv:1303.3997*.

37. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., ... & 1000 Genome Project Data Processing Subgroup. (2009). The sequence alignment/map format and SAMtools. *bioinformatics*, *25*(16), 2078-2079.

38. Tarasov, A., Vilella, A. J., Cuppen, E., Nijman, I. J., & Prins, P. (2015). Sambamba: fast processing of NGS alignment formats. *Bioinformatics*, *31*(12), 2032-2034.

39. Cibulskis, K., Lawrence, M. S., Carter, S. L., Sivachenko, A., Jaffe, D., Sougnez, C., ... & Getz, G. (2013). Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nature biotechnology*, *31*(3), 213-219.

40. Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., & Mesirov, J. P. (2011). Integrative genomics viewer. *Nature biotechnology*, *29*(1), 24-26.

41. Kim, S., Scheffler, K., Halpern, A. L., Bekritsky, M. A., Noh, E., Källberg, M., ... & Saunders, C. T. (2018). Strelka2: fast and accurate calling of germline and somatic variants. *Nature methods*, *15*(8), 591-594.

42. Wardell, C. P., Ashby, C., & Bauer, M. A. (2021). FiNGS: high quality somatic mutations using filters for next generation sequencing. *BMC bioinformatics*, *22*(1), 77.

43. Diossy, M., Sztupinszki, Z., Krzystanek, M., Borcsok, J., Eklund, A. C., Csabai, I., ... & Szallasi, Z. (2021). Strand Orientation Bias Detector to determine the probability of FFPE sequencing artifacts. *Briefings in Bioinformatics*, *22*(6).

44. Yost, S. E., Smith, E. N., Schwab, R. B., Bao, L., Jung, H., Wang, X., ... & Frazer, K. A. (2012). Identification of high-confidence somatic mutations in whole genome sequence of formalin-fixed breast cancer specimens. *Nucleic acids research*, *40*(14), e107-e107.

45. Wang, K., Li, M., & Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic acids research*, *38*(16), e164-e164.

46. Isozaki, H., Sakhtemani, R., Abbasi, A., Nikpour, N., Stanzione, M., Oh, S., ... & Hata, A. N. (2023). Therapy-induced APOBEC3A drives evolution of persistent cancer cells. *Nature*, *620*(7973), 393-401.

47. Luebeck, J., Huang, E., Kim, F., Liefeld, T., Dameracharla, B., Ahuja, R., ... & Bafna, V. (2024). AmpliconSuite: an end-to-end workflow for analyzing focal amplifications in cancer genomes. *bioRxiv*, 2024-05.

48. Talevich, E., Shain, A. H., Botton, T., & Bastian, B. C. (2016). CNVkit: genome-wide copy number detection and visualization from targeted DNA sequencing. *PLoS computational biology*, *12*(4), e1004873.

49. Deshpande, V., Luebeck, J., Nguyen, N. P. D., Bakhtiari, M., Turner, K. M., Schwab, R., ... & Bafna, V. (2019). Exploring the landscape of focal amplifications in cancer using AmpliconArchitect. *Nature communications*, *10*(1), 392.

50. Ashkenazy, H., Abadi, S., Martz, E., Chay, O., Mayrose, I., Pupko, T., & Ben-Tal, N. (2016). ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic acids research*, *44*(W1), W344-W350.

51. Wala, J. A., Bandopadhayay, P., Greenwald, N. F., O'Rourke, R., Sharpe, T., Stewart, C., ... & Beroukhim, R. (2018). SvABA: genome-wide detection of structural variants and indels by local assembly. *Genome research*, *28*(4), 581-591.

52. Wagner, S., Roberson, D., Boland, J., Kreimer, A. R., Yeager, M., Cullen, M., ... & Wentzensen, N. (2019). Evaluation of TypeSeq, a novel high-throughput, low-cost, next-generation sequencing-based assay for detection of 51 human papillomavirus genotypes. *The Journal of infectious diseases*, *220*(10), 1609-1619.

Figure Legends

Figure 1. **A portrait of genomic aberrations in SNSCC** (a) Previous studies utilizing different modalities (PCR, p16 stain, ISH, etc.) report varying rates of HPV positivity in SNSCC[4]. (b) HPV positivity rate as detected in the current study (21 SNSCC), by WGS. (c) Focal segmentation of HPV-encoding extrachromosomal DNA hijacking the *TP63* oncogene, resulting in local copy number gain as estimated by Amplicon Architect[47]. A semi-circular arc connecting two loci denotes chimeric reads which map to two sites. (d) Clinical phenotypes and histopathological traits present in a cohort of 21 patients with SNSCC. (e) Decomposition of somatic mutations in individual patients into corresponding mutational signatures (left, see Methods). Oncoplot summarizing the genomic landscape of 21 SNSCCs, including: tumor ploidy, overall mutational burden, HPV genotype by WGS, and the relative contribution of different single base substitution signatures (right). Select examples of patients which are enriched for different SBS signatures (e.g. Patient 12 APOBEC, Patient 15 UV Sig7b) are shown. (f) Assessment of point mutations in known cancer driver genes, highlighting examples of *PIK3CA* mutations which are recurrent in HPV+ cancers as well as TCGA HNSCC, overlaying protein-relative residue-level conservation in PIK3CA.[50] Cartoon visualizes the localization of individual *PIK3CA* mutations among four SNSCC profiled herein (Patients 12, 47180, 52087, 4139). (g) Evaluation of candidate somatic driver genes in SNSCC by assessment of mutational clonality (left), highlighting evidence of clonal mutations in known driver genes (*CHEK2, KDM6A*) in two different patients (right).

Figure 2. **Genomic record of HPV activity in SNSCC** (a) Sequencing coverage of individual HPV genomes which were detected across 21 SNSCC. (b) Linearized representation of the approximate location of viral genes in the HPV genome (top). Distribution of HPV positivity (total number of genotypes per patient) and detectable hg19-HPV breakpoints among HPV^{+} patients (bottom). (c) Circos plot mapping the somatic integration of HPV into the human genome, sub-stratified by HPV genotype and highlighting the identification of a chimeric *TP63-HPV45* amplicon. Each arc represents chimeric reads mapping across two sites (d) High-

resolution characterization of select HPV integration sites in 3 SNSCC patients. Co-option of the *TP63* oncogene into an HPV45-containing amplicon (top). Evidence of subclonal integration of HPV18 (3' UTR of *N4BP3*) without corresponding copy number instability in the human genome (middle). Focal copy number gain (50X genomic median) of *MAN1A2*-HPV51 amplicon (bottom). (e) Clonality of mutations across the HPV genome as detected by viral WGS, highlighting the contribution of APOBEC mutagenesis. (f) Distribution of mutational clonality within the HPV genome (left), and distribution of mutations across the HPV genome, highlighting targeted domains (right).

Figure S1. **Pan-cancer context of mutational signature activity in SNSCC** (a) Scatter-plot depicting the genomic mutational rates of different TCGA-ICGC cancer lineages, highlighting SNSCC. (b) Cartography of mutational signatures across human cancers, highlighting the distribution of SNSCC relative to TCGA HNSCC. UMAP loadings of seven prominent SBS signatures (BRCA, APOBEC, Aging, UV Sig7b, A→G Sig 12, Smoking) are used to partition individual TCGA-ICGC tumors along different mutational "trajectories." Each dot in the cartography represents a single tumor and is colored by the dominant mutational signature. Select cancer subtypes (e.g. HPV^+ HNSCC) are highlighted with symbols. The mutational signatures of two select patients which are profoundly enriched for different signatures at the single base level (top left Patient 47180, APOBEC-enriched; bottom right 64208, UV-enriched) are deconvolved into corresponding histograms and lego plots to highlight the mutational content of each signature. (c) Top, scatter plot directly compares the total mutational burden in HPV^+ SNSCC to HPV^- HNSCC, HPV^+ HNSCC, and HPV^+ CESC. Bottom, scatter plot directly compares APOBEC signature loadings among different HNSCC subtypes, highlighting select SNSCC which are both HPV^+ and APOBEC-enriched. p-values were computed using the MATLAB ttest2() function. (d) Explicit plotting of SNSCC and TCGA-ICGC genomes into SBS signatures, sorted by APOBEC loading.

Figure S2. **Discovery of structural variants in SNSCC** (a) Abundance of SVs of various classes in 21 SNSCC (left). Size distributions of deletions (right). (b) GISTIC (Genomic Identification of Significant Targets in Cancer, Broad pipeline) output showing common focal copy number alterations present among 21 SNSCC, highlighting regions of focal PTEN copy number loss across three patients. (c) Whole genome circos plot defining the constellation of different classes of inter- and intra-chromosomal SVs which are harbored across 21 SNSCC. Detection of an in-frame fusion between exon 24 of NIPBL and exon 10 of DNAJC21 in a single patient, showcasing raw densities of both chimeric and wildtype reads at the mapped genomic breakpoints in both tumor and normal samples (top left). (d) Detection of kataegic storms of APOBEC-related mutations on chromosome 12 of patient 12. Radial scatter plot depicts the inter-mutational distances between mutations. (e) Tracing of structural variations to allelic and copy number imbalance in patient 8 across whole chromosomes (17, 18, 19, 20, 21). Clustered targeting of intra-chromosomal rearrangements resulting in loss of chromosome 21; focal gain on chromosome 17; oscillating copy number states on chromosome 19; as well as a chromothriptic shower of intra-chromosome 21 rearrangements, leading to an in-frame fusion between ERG and BRWD1, are highlighted.

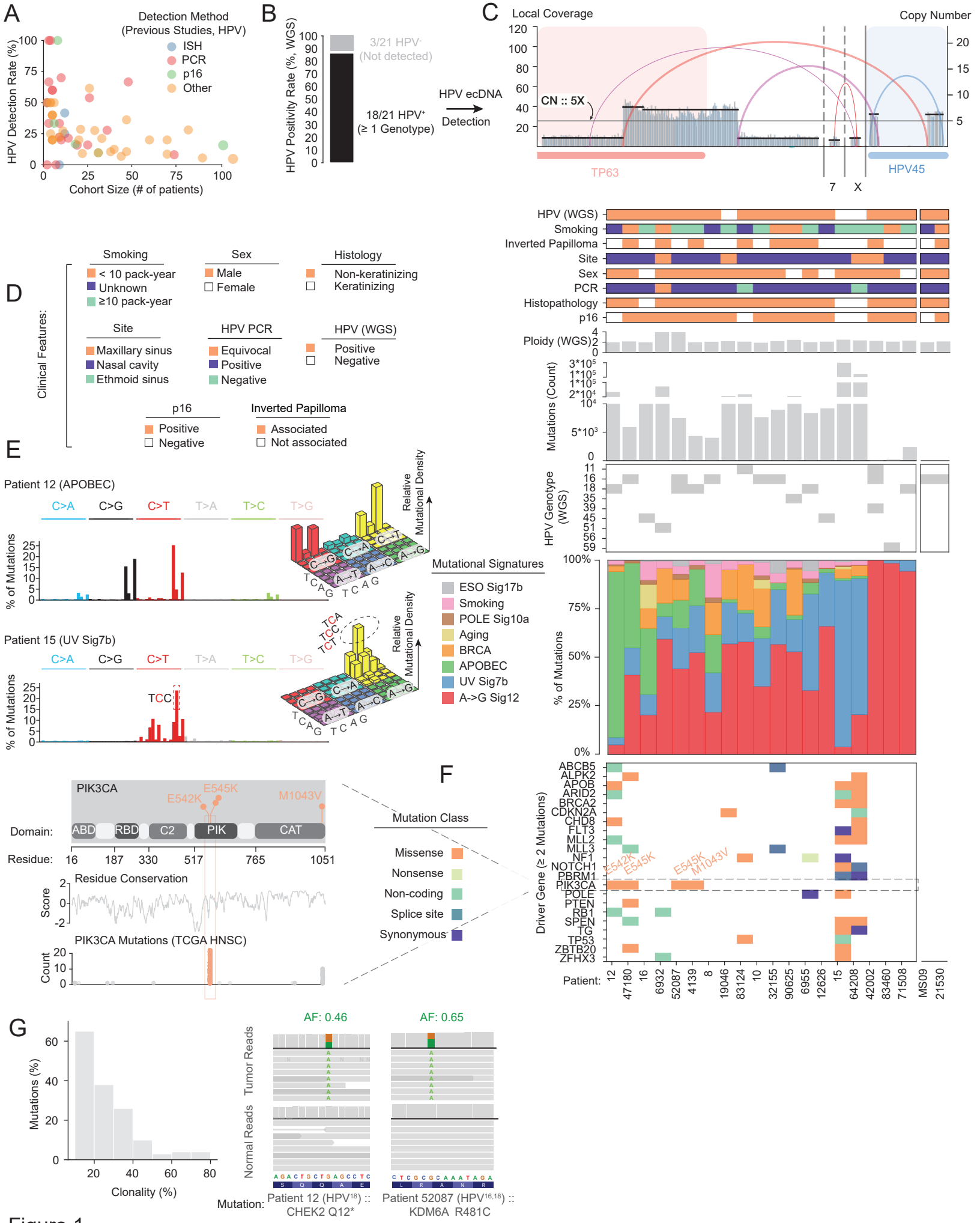
A
Detection Method
(Previous Studies, HPV)
ISH
PCR
p16
Other
HPV Detection Rate (%)
Cohort Size (# of patients)
B
HPV Positivity Rate (%, WGS)
3/21 HPV⁻
(Not detected)
18/21 HPV⁺
(≥ 1 Genotype)
HPV ecDNA
Detection
C
Local Coverage
Copy Number
CN :: 5X
TP63
HPV45
7
X
D
Clinical Features:
Smoking
< 10 pack-year
Unknown
≥10 pack-year
Sex
Male
Female
Histology
Non-keratinizing
Keratinizing
Site
Maxillary sinus
Nasal cavity
Ethmoid sinus
HPV PCR
Equivocal
Positive
Negative
HPV (WGS)
Positive
Negative
p16
Positive
Negative
Inverted Papilloma
Associated
Not associated
HPV (WGS)
Smoking
Inverted Papilloma
Site
Sex
PCR
Histopathology
p16
Ploidy (WGS)
Mutations (Count)
HPV Genotype (WGS)
% of Mutations
E
Patient 12 (APOBEC)
C>A
C>G
C>T
T>A
T>C
T>G
% of Mutations
Relative Mutational Density
Patient 15 (UV Sig7b)
TCC
Mutational Signatures
ESO Sig17b
Smoking
POLE Sig10a
Aging
BRCA
APOBEC
UV Sig7b
A->G Sig12
F
PIK3CA
E542K
E545K
M1043V
Domain:
ABD
RBD
C2
PIK
CAT
Residue:
16
187
330
517
765
1051
Residue Conservation
Score
PIK3CA Mutations (TCGA HNSC)
Count
Mutation Class
Missense
Nonsense
Non-coding
Splice site
Synonymous
Driver Gene (≥ 2 Mutations)
ABCB5
ALPK2
APOB
ARID2
BRCA2
CDKN2A
CHD8
FLT3
MLL2
MLL3
NF1
NOTCH1
PBRM1
PIK3CA
POLE
PTEN
RB1
SPEN
TG
TP53
ZBTB20
ZFHX3
Patient:
12
47180
16
6932
52087
4139
8
19046
83124
10
32155
90625
6955
12626
15
64208
42002
83460
71508
MS09
21530
G
Mutations (%)
Clonality (%)
AF: 0.46
AF: 0.65
Tumor Reads
Normal Reads
Mutation:
Patient 12 (HPV18) ::
CHEK2 Q12*
Patient 52087 (HPV16,18) ::
KDM6A R481C

Figure 1

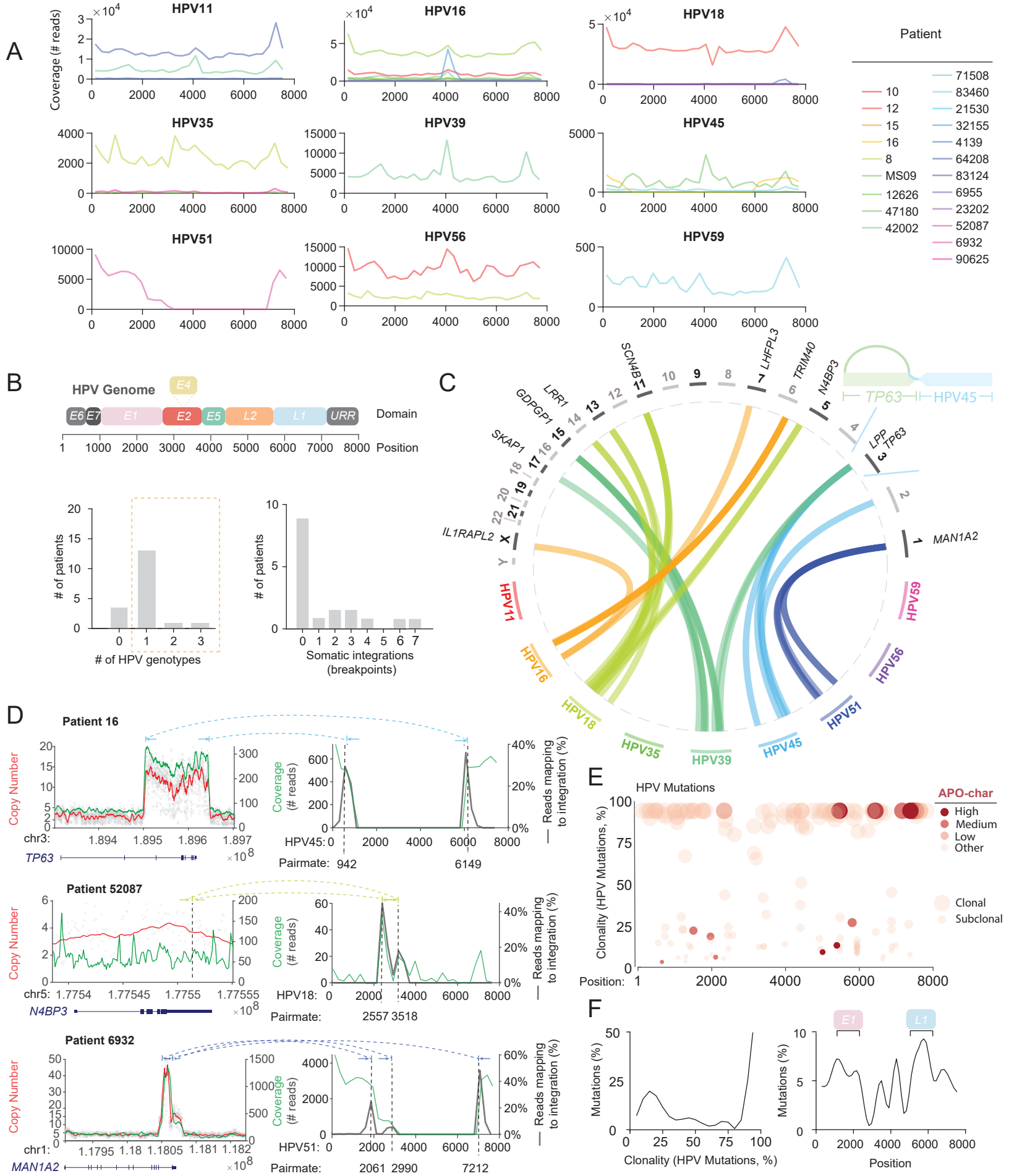
A
HPV11
HPV16
HPV18
HPV35
HPV39
HPV45
HPV51
HPV56
HPV59
Coverage (# reads)
Patient
10
12
15
16
8
MS09
12626
47180
42002
71508
83460
21530
32155
4139
64208
83124
6955
23202
52087
6932
90625
B
HPV Genome
E4
E6 E7 E1 E2 E5 L2 L1 URR
Domain
Position
# of patients
# of HPV genotypes
Somatic integrations (breakpoints)
C
SCN4B
LHFPL3
TRIM40
N4BP3
LRR1
GDPGP1
SKAP1
IL1RAPL2
LPP
TP63
MAN1A2
HPV45
HPV11
HPV16
HPV18
HPV35
HPV39
HPV45
HPV51
HPV56
HPV59
D
Patient 16
Copy Number
Coverage (# reads)
Reads mapping to integration (%)
chr3:
TP63
HPV45:
Pairmate: 942 6149
Patient 52087
chr5:
N4BP3
HPV18:
Pairmate: 2557 3518
Patient 6932
chr1:
MAN1A2
HPV51:
Pairmate: 2061 2990 7212
E
HPV Mutations
APO-char
High
Medium
Low
Other
Clonal
Subclonal
Clonality (HPV Mutations, %)
Position:
F
Mutations (%)
Clonality (HPV Mutations, %)
E1
L1
Position

Figure 2

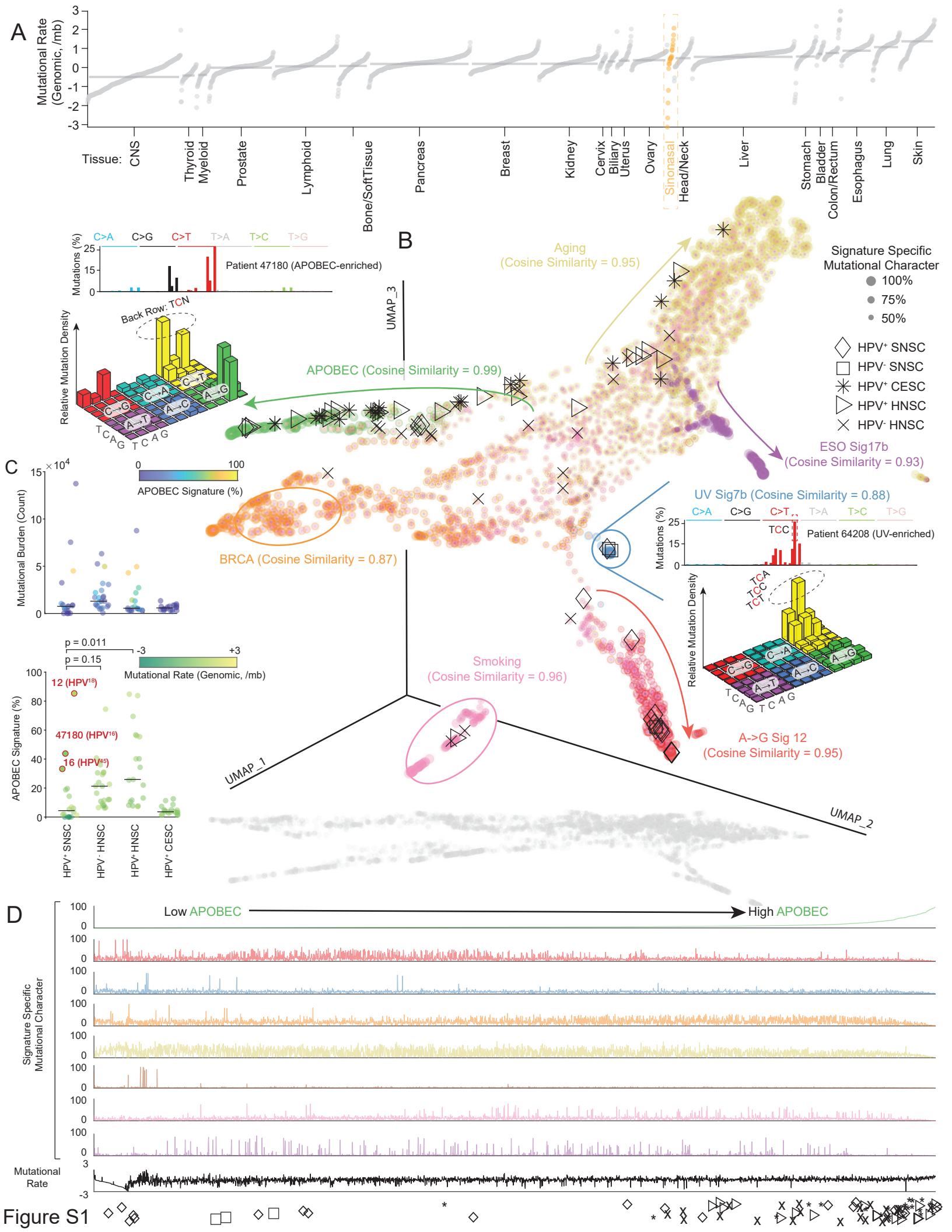
A
Mutational Rate (Genomic, /mb)
Tissue:
CNS
Thyroid
Myeloid
Prostate
Lymphoid
Bone/SoftTissue
Pancreas
Breast
Kidney
Cervix
Biliary
Uterus
Ovary
Sinonasal
Head/Neck
Liver
Stomach
Bladder
Colon/Rectum
Esophagus
Lung
Skin
B
Patient 47180 (APOBEC-enriched)
Back Row: TCN
Relative Mutation Density
Mutations (%)
UMAP_3
Aging
(Cosine Similarity = 0.95)
APOBEC (Cosine Similarity = 0.99)
Signature Specific
Mutational Character
100%
75%
50%
HPV+ SNSC
HPV- SNSC
HPV+ CESC
HPV+ HNSC
HPV- HNSC
ESO Sig17b
(Cosine Similarity = 0.93)
UV Sig7b (Cosine Similarity = 0.88)
Patient 64208 (UV-enriched)
BRCA (Cosine Similarity = 0.87)
Smoking
(Cosine Similarity = 0.96)
A->G Sig 12
(Cosine Similarity = 0.95)
UMAP_1
UMAP_2
C
Mutational Burden (Count)
APOBEC Signature (%)
p = 0.011
p = 0.15
Mutational Rate (Genomic, /mb)
12 (HPV18)
47180 (HPV16)
16 (HPV45)
HPV+ SNSC
HPV- HNSC
HPV+ HNSC
HPV+ CESC
D
Low APOBEC
High APOBEC
Signature Specific
Mutational Character
Mutational
Rate

Figure S1

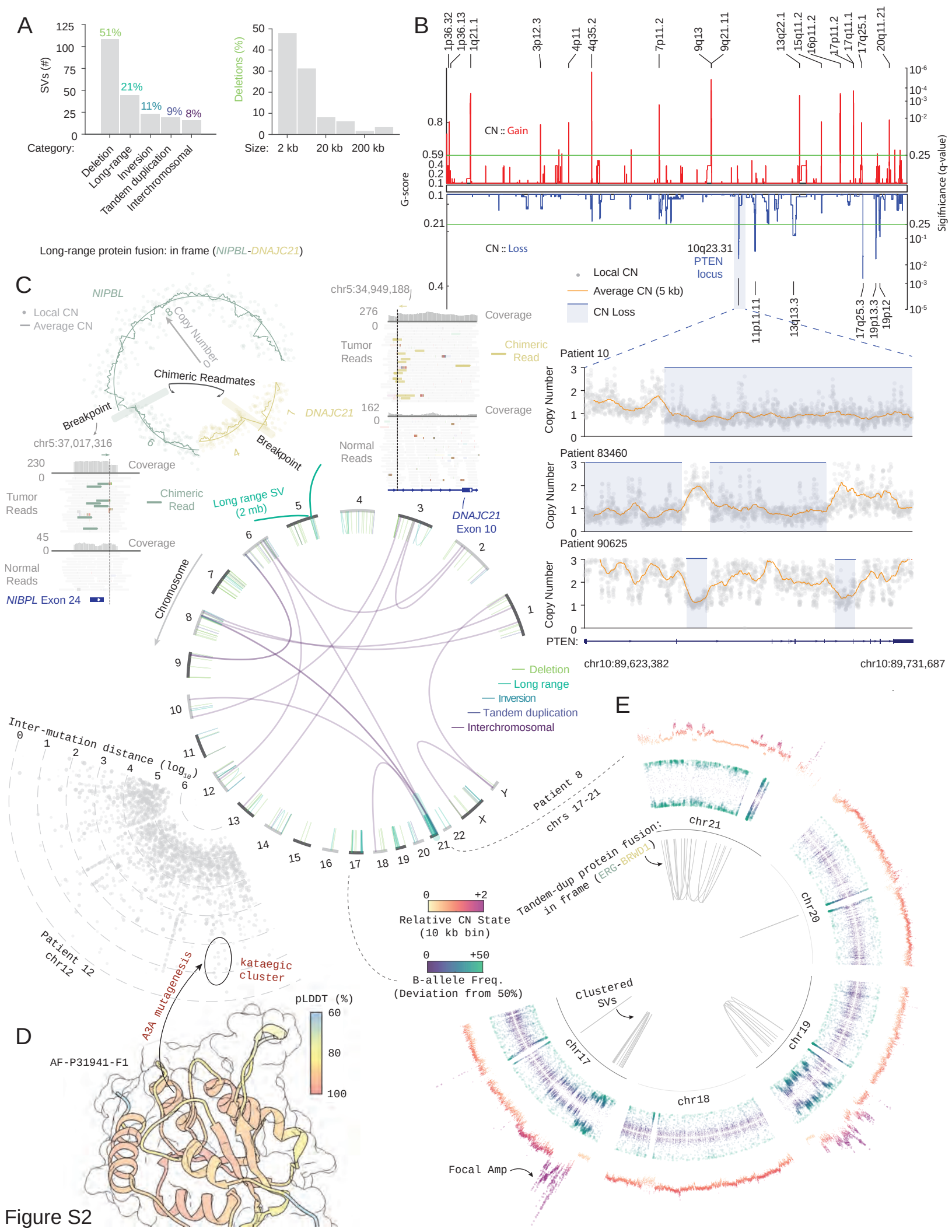
A
SVs (#)
51%
21%
11%
9%
8%
Category:
Deletion
Long-range
Inversion
Tandem duplication
Interchromosomal
Deletions (%)
Size: 2 kb 20 kb 200 kb
Long-range protein fusion: in frame (NIPBL-DNAJC21)
B
CN :: Gain
CN :: Loss
G-score
Significance (q-value)
10q23.31
PTEN
locus
Local CN
Average CN (5 kb)
CN Loss
C
NIPBL
Local CN
Average CN
Copy Number
Chimeric Readmates
Breakpoint
Breakpoint
chr5:37,017,316
Coverage
Tumor Reads
Chimeric Read
Normal Reads
NIBPL Exon 24
Long range SV (2 mb)
chr5:34,949,188
DNAJC21
DNAJC21 Exon 10
Chromosome
Patient 10
Patient 83460
Patient 90625
Copy Number
PTEN:
chr10:89,623,382
chr10:89,731,687
Deletion
Long range
Inversion
Tandem duplication
Interchromosomal
E
Inter-mutation distance (log10)
Patient 8
chrs 17-21
Patient 12
chr12
Relative CN State
(10 kb bin)
B-allele Freq.
(Deviation from 50%)
Tandem-dup protein fusion:
in frame (ERG-BRWD1)
chr21
chr20
chr19
chr18
chr17
Clustered SVs
Focal Amp
D
AF-P31941-F1
A3A mutagenesis
kataegic cluster
pLDDT (%)

Figure S2