Supplementary materials and methods

Cell line, tumor specimen DNA preparation and HPV detection

A total of 18 OSCC cell lines were used in the present study: OSC19, OSC20, and HO-1-N-1 cells were obtained from the Japanese Collection of Research Bioresources (Tokyo, Japan); HSC2, HSC3, HSC4, SAS, and Ho-1-u-1 cells were obtained from the Institute of Development, Aging, and Cancer at Tohoku University (Sendai, Japan); KON, SAT, and Ca9-22 cells were purchased from the Health Science Research Resources Bank (Osaka, Japan); the SCC-25 cell line was purchased from the ATCC (Manassas, VA, USA); OM1, MON2, HOC119, and HOC621 cells were kindly provided by Dr. N. Kamata (Tokyo Medical and Dental University, School of Medicine, Tokyo, Japan); and POT1 cells were kindly provided by Dr. A. Miyazaki (Sapporo Medical University, Sapporo, Japan).

Patients with OSCC were diagnosed and treated in the Oral Surgery Department of Sapporo Medical University from 2007 to 2016. We obtained a total of 137 OSCC samples (95 FFPE and 42 frozen samples). Among FFPE samples, 21 had insufficient DNA and 36 inadequate DNA. NGS sequencing data were available for 80 samples (38 FFPE and 42 frozen samples). Clinical data were reviewed to identify personal habits and tumor stage. Hematoxylin and eosin–stained slides were re-examined to confirm the original diagnosis. Information concerning past or current tobacco smoking and alcohol drinking was recorded. A list of tumor samples and their clinical characteristics is shown in Table 1. Tumor size, nodal metastases and distant metastases (TNM) staging of tumors was performed according to the International Union Against Cancer (Union Internationale Contre le Cancer [UICC]) criteria, and histological grading was performed

according to the World Health Organization classification. DNA was extracted from cell lines and fresh frozen tissue using the QIAamp DNA Mini kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. DNA was also isolated from FFPE sections using the QIAamp DNA FFPE Tissue kit (Qiagen) according to the manufacturer's instructions without modifications. The purified DNA was quantified via real-time PCR of short (87 bp) and long (256 kb) amplicons for the RNase P gene using the TaqMan RNase P Detection Reagent Kit and a custom TaqMan Assay (ThermoFisher Scientific, Waltham, MA, USA), respectively. Detection of human papillomavirus (HPV) 16, 18 and 33 in the extracted tumor DNA was performed using TaKaRa PCR Human Papillomavirus Detection Set and ExTaq polymerase (Takara, Otsu, Japan).

Bioinformatics analyses

Sequencing reads were first automatically processed using Ion Torrent Suite v5.0 on the Ion Torrent Server (ThermoFisher Scientific). The alignment of the trimmed reads with human genome build 19 was performed in Torrent Suite, and somatic mutations (point mutations, insertions, and deletions) were detected using statistical approaches from the Ion Reporter software 5.0 workflow. A sequencing coverage of 20X and a minimum variant frequency of 15% of the total number of distinct tags were used as cutoffs. Mutations were called if they were absent from dbSNP and the 1000 Genomes Project database. For data interpretation and verification, the aligned reads were visualized using the Integrative Genomics Viewer (IGV) software (Broad Institute, http//www.broadinstitute.org/igv) to filter out possible strand-specific errors, such as a mutation that was only detected in the

forward or reverse DNA strand but not in both strands. Copy number variation (CNV) detection was performed with the Ion Reporter software using the control sequence data provided by ThermoFisher Scientific. Genomic regions with CNV were identified using a copy number greater than 3 and a copy number less than 1 for gains and losses, respectively.

Fusion gene detection

Amplicon libraries for individual RNAs from cell lines were prepared using the Ion AmpliSeq RNA Lung Cancer Research Fusion Panel and the Ion AmpliSeq RNA Library Kit (ThermoFisher Scientific) according to the manufacturer's protocol. This panel can detect more than 70 known fusions involving *ALK, ROS1, RET*, and *NTRK1*. Semi-conductor sequencing was performed as described above. The detection of fusion transcripts was performed using the fusion workflow integrated into the Ion Reporter Fusion analysis workflow.

Circulating cell-free DNA (cfDNA) analysis

Cell-free, circulating plasma DNA from patients No.10 and No.27 was extracted from 3 mL of plasma using the QIAamp Circulating Nucleic Acid kit (Qiagen) according to the manufacturer's instructions. Ten nanograms of circulating plasma DNA were used to generate sequencing libraries. Two independent libraries were pooled and sequenced on a PGM 318 chip to achieve an average coverage per target region of approximately 10,000X.

Statistical analysis

Statistical analyses were carried out using GraphPad Prism 6 (GraphPad Software, La Jolla, CA, USA). The significance of correlations between molecular groups and clinico-pathological data were tested using Fisher's exact test and the $\chi 2$ test for trends and the unpaired t-test as indicated. P-values <0.05 were considered to be statistically significant.