Human Genetics (2019) 138:1379–1389
https://doi.org/10.1007/s00439-019-02083-9

ORIGINAL INVESTIGATION

Mutational signatures and mutagenic impacts associated with betel quid chewing in oral squamous cell carcinoma

Shih-Chi Su[1,2] · Lun-Ching Chang[3] · Chiao-Wen Lin[4,5] · Mu-Kuan Chen[6] · Chun-Ping Yu[7] · Wen-Hung Chung[1,2,8] · Shun-Fa Yang[9,10]

Received: 30 July 2019 / Accepted: 26 October 2019 / Published online: 2 November 2019
© Springer-Verlag GmbH Germany, part of Springer Nature 2019

Abstract

Betel quid (BQ) chewing is a prevailing risk for oral squamous cell carcinoma (OSCC) in Southeast Asia. Yet, the detailed mechanisms by which BQ chewing damages the genome are still not fully understood. Through exome sequencing of tumor–normal pairs from 196 male patients with OSCC, including 95 habitual BQ chewers and 101 non-BQ users, we conducted a quantitative survey of mutational signatures and genomic aberrations and explored their association with BQ chewing. We found that BQ-associated elevation in mutation rate was seen in cancers of the tongue, but not in overall OSCC. Additionally, we identified a mutational signature that is enriched in tumors from BQ users. Moreover, the numbers of small insertions and deletions (INDELs) and breakpoints derived from structural variations (SV) were increased, whereas the extent of loss of heterozygosity was decreased in BQ-related OSCC genomes. However, neither the number of base substitutions and microsatellite instability events nor the extent of copy-number alterations differed between BQ-related and -unrelated OSCC. In conclusion, consistent with the proposition that BQ chewing increases OSCC risk as a mutagen, our results unveil a BQ-associated mutational signature and indicate mutagenic impacts of BQ chewing on preferentially eliciting INDELs and SV-related breakpoints in OSCC genomes.

Introduction

Oral squamous cell carcinoma (OSCC), accounting for the majority (approximately 90%) of oral cancers (Jemal et al. 2008), is a common malignancy across the world, especially in Central and Southeast Asia (Krishna Rao et al. 2013). Such variation in incidence rate is linked to etiological parameters of this disease associated with the unique cultural practice and lifestyle of different ethnic populations.

Shih-Chi Su and Lun-Ching Chang contributed equally to this work.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00439-019-02083-9) contains supplementary material, which is available to authorized users.

✉ Shun-Fa Yang
ysf@csmu.edu.tw

1 Whole-Genome Research Core Laboratory of Human Diseases, Chang Gung Memorial Hospital, Keelung, Taiwan

2 Department of Dermatology, Drug Hypersensitivity Clinical and Research Center, Chang Gung Memorial Hospital, Linkou, Taiwan

3 Department of Mathematical Sciences, Florida Atlantic University, Boca Raton, FL, USA

4 Institute of Oral Sciences, Chung Shan Medical University, Taichung, Taiwan

5 Department of Dentistry, Chung Shan Medical University Hospital, Taichung, Taiwan

6 Department of Otorhinolaryngology-Head and Neck Surgery, Changhua Christian Hospital, Changhua, Taiwan

7 Biodiversity Research Center, Academia Sinica, Taipei, Taiwan

8 School of Medicine, College of Medicine, Chang Gung University, Taoyuan, Taiwan

9 Institute of Medicine, Chung Shan Medical University, 110 Chien-Kuo N. Road, Section 1, Taichung 402, Taiwan

10 Department of Medical Research, Chung Shan Medical University Hospital, Taichung, Taiwan

Human papillomavirus (HPV) infection (Gillison et al. 2000) and habitual exposure of cancer-causing substances (Awang 1988; Scully and Bagan 2009) have been recognized as major predisposing factors to OSCC. In addition to tobacco and alcohol use, betel quid (BQ) chewing is a prevailing risk for oral tumorigenesis in Asian countries with high OSCC prevalence, including Taiwan where the prevalence of oncogenic HPV infection (~ 16%) was at a lower rate compared with the worldwide rate (~ 24%) (Lee et al. 2013, 2015; Ndiaye et al. 2014).

Betel quid is a mixture of areca nut, slaked lime, catechu, and several condiments (IARC Working Group on the Evaluation of Carcinogenic Risks to Humans 2004). Chewing of BQ can induce pharmacological responses, such as well-being sensations, euphoria, and heightened alertness (Chu 2001), and yet this popular habit has been strongly implicated in the development of OSCC through prolonged exposure of various BQ-derived carcinogens (Nair et al. 2004). It is known that nitrosation of arecoline, the most abundant alkaloid in areca nuts, may produce a variety of nitrosamines, which can interact with DNA, proteins or other molecules to mediate carcinogenic activities (IARC Working Group on the Evaluation of Carcinogenic Risks to Humans 2004). Moreover, reactive oxygen species (ROS) are generated in substantial amounts in the oral cavity during BQ chewing to cause oxidative damage in the DNA of buccal mucosa tissues (Nair et al. 1995). These metabolically activated reactive species of BQ-derived carcinogens collectively can trigger genotoxic effects in humans, such as formation of safrole-like DNA adducts (Chen et al. 1999) or chromosome breaks (Nair et al. 1991), ultimately leading to oral carcinogenesis.

Cancers of the oral cavity carry somatic mutations that were accumulated through multiple mutational processes. Different mutational processes give rise to specific combinations of mutation types, termed mutational signatures (Alexandrov et al. 2013). Recent genomic analyses have comprehensively characterized HPV− and HPV+ head and neck cancers in the Western countries, providing a foundation for advanced molecular diagnoses, identification of potential biomarkers, and therapeutic insights (Cancer Genome Atlas 2015; Gillison et al. 2019). Since OSCC etiology and a significant number of genomic alterations in OSCC genome were not uniformly observed across geographical regions (Beck and Golemis 2016), it is conceivable that the outcome of mutagenic processes operative in Asian OSCC, especially for those caused by BQ chewing, may be unique. Yet, the comprehensive genomic footprints of BQ chewing-related OSCC are still not fully understood. In this study, we examined 196 OSCC genome through exome sequencing (95 BQ users and 101 non-BQ users), with the goal of identifying mutational signatures and types of genomic alterations associated with BQ chewing.

Materials and methods

Subject enrollment and sample collection

The cohort of this study comprising 196 male OSCC patients who had been neither previously treated nor had proven metastatic disease at the time of diagnosis was accrued from 2008 to 2017, with the approval by the institutional review board of Chung Shan Medical University Hospital, Taichung, Taiwan. Tumors and paired whole blood samples of cases with enriched tumor cell populations (to ensure dissected regions > 70% tumor purity) selected by certified pathologists were collected. All tumor samples were immediately frozen in liquid nitrogen and stored at − 80 °C. All cases were staged clinically at the time of diagnosis according to the TNM staging system of the American Joint Committee on Cancer (AJCC) (Edge and Compton 2010). Tumor differentiation was examined by a pathologist and rated according to the AJCC classification. All participants provided informed written consent at enrollment. Information regarding age, gender, alcohol drinking, cigarette smoking, and BQ chewing was recorded for each participant. BQ chewing and alcohol drinking are defined as behavioral use of BQ (chewing one or more BQ-related products daily for at least 1 year) and substantial alcohol intake (consuming more than two alcoholic beverages per day in average), respectively. Cigarette smoking is defined as current smoking of at least one cigarette per day during the latest 3 months.

Exome sequencing

Genomic DNA was extracted by using the DNeasy Blood & Tissue Kit (Qiagen) according to the manufacturer's instructions and loaded on a 1% agarose gel for quality control. In-solution enrichment of coding exons and flanking intronic sequences for 196 pairs of qualified DNA samples was performed by the SureSelect Human All Exon V5 or V6 kit (Agilent Technologies). Library construction and sequencing on an Illumina HiSeq 2500 instrument with paired reads of 90–100 bp were performed by Novogene.

Detection of single-nucleotide variation (SNV) and insertion/deletion (INDEL)

Raw sequence data were merged, quality trimmed, aligned and mapped to the reference genome (hg19) by Burrows–Wheeler aligner (BWA) using default options (Li and Durbin 2009). We applied MuTect2 through the Genome Analysis Toolkit (GATK) 4.0 (McKenna et al. 2010) to identify somatic single-nucleotide variations (SNVs) and

small insertions/deletions (INDELs) in target regions plus 5′ and 3′ flanking fragment of 20 bp. All somatic calls detected were filtered against a panel of 300 germline DNA samples (Chang Gung Human Database, an unpublished whole-genome database of normal controls). The remaining somatic mutations constituted the mutational catalog for each matched tumor–normal pair.

Analysis of mutational signatures

We deciphered the mutational signatures by using signeR (Rosales et al. 2017). This R package used an empirical Bayesian treatment of the non-negative matrix factorization (NMF) model to express each signature as a matrix that is derived from all mutation data according to 96 possible mutation types (6 types of substitution × 4 types of 5′ base × 4 types of 3′ base). The optimal number of mutational signatures was identified based on the Bayesian information criterion (BIC) box plots obtained by increasing the number of signatures from one to ten. The identified signatures were compared to previously extracted consensus signatures (Alexandrov et al. 2013) by cosine similarity. Alternatively, we decomposed each tumor exome into eight signatures (Sig.1, 2, 4, 5, 6, 7, 13, and 29) previously found in head and neck cancers by decompTumor2Sig (Kruger and Piro 2019), which estimates the contribution of the corresponding mutational processes to the somatic mutations identified in the tumor.

Analysis of copy-number alteration (CNA) and detection of chromosomal breakpoint

To detect somatic copy-number alterations (CNAs), we applied Control-FREEC (Boeva et al. 2012) to the read count profiles of sequence data. The read count ratios of tumors to paired normal samples were calculated, normalized for GC content and mappability, and used as the proxy of the copy-number ratios. For determining chromosomal breakpoints of somatic structural variations (SVs), BreakDancer (Chen et al. 2009) was used to detect the breakpoints of large insertion, deletion, inversion, and translocation. The number of somatic breakpoints was determined by removing calls with a confidence score ≤ 90 and supported by ≤ 10 reads from the matched normal sample.

Measurement of microsatellite instability (MSI) and loss of heterozygosity (LOH)

Detection of MSI events was performed by MSIsensor (Niu et al. 2014) using default parameters. This program interrogates the aligned sequencing data for available microsatellite areas with sufficient coverage in tumor–normal pairs where it detects the variation in length distributions of microsatellites per site. For measurement of LOH, We applied LOHAS (Yang et al. 2011) to estimate the proportion of homozygous variant calls within target regions plus 5′ and 3′ flanking fragment of 20 bp. Homozygous intensities for the tumors and matched non-cancer samples were estimated by a sliding window containing 100 nearest-neighbor variants to each anchor variant using local polynomial weighted regression to fit the model. In each sliding window, we calculated the 90th percentile of estimated homozygous intensities from a panel of normals, and then the region of LOH was defined by the genomic region where a tumor's estimated homozygous intensities were greater than the 90th percentile of estimated homozygous intensities from both the panel of normals and their matched non-cancer specimen.

Statistical analysis

Comparisons of demographic and clinical parameters between BQ-related and -unrelated samples were performed by using Fisher's exact test or Mann–Whitney *U* test. Differences in the signature exposures, number of base substitutions, INDELs, breakpoints, and MSI events, and fraction of CNV and LOH between BQ-related and -unrelated OSCC were estimated by Student's *t* test. Data were analyzed by using the SAS statistical software (Version 9.1, 2005; SAS Institute Inc., Cary, NC). All reported *p* values were two tailed, and a *p* value of < 0.05 was considered significant.

Results

Characteristics and genomic analyses of OSCC subjects

To determine the mutational signatures and conduct a quantitative survey of genomic alterations associated with BQ chewing in OSCC, we performed exome sequencing of tumor–normal (fresh-frozen tumors and matched whole blood) pairs from 196 male OSCC patients, with the majority bearing cancers of the buccal mucosa (80/196) and tongue (60/196) (Table 1). Of the samples we studied, 95 pairs were reported to be from BQ users and 101 from non-BQ users. There was no significant difference in the distribution of age, tumor location, clinical stage, tumor size, metastasis, and status of tumor cell differentiation detected between the two study groups. A higher portion of patients who had the habit of BQ chewing reported a history of tobacco and alcohol use, compared with that of non-BQ chewers.

We sequenced 196 OSCC exomes with 89.1% and 90.6% of targeted regions covered at > 20-fold in tumors and matched normal samples, respectively (Figure S1), and detected 26, 917 somatic variants, of which 24,718 (91.83%)

Table 1 Demographic characteristics and clinical features of male patients with OSCC

Variables	OSCC without BQ chewing ($n=101$)	OSCC with BQ chewing ($n=95$)	p value
Age (years)			
	57.20 ± 11.32	54.70 ± 10.16	0.106
< 40	4 (4.0%)	7 (7.4%)	0.380
40–49	23 (22.8%)	23 (24.2%)	
50–59	34 (33.7%)	35 (36.8%)	
60–69	24 (23.8%)	23 (24.2%)	
≥ 70	16 (15.8%)	7 (7.4%)	
Smoking status			
No	41 (40.6%)	9 (9.5%)	0.001
Yes	60 (59.4%)	86 (90.5%)	
Drinking status			
No	82 (81.2%)	41 (43.2%)	0.001
Yes	19 (18.8%)	54 (56.8%)	
Cancer location			
Buccal mucosa	43 (42.6%)	37 (38.9%)	0.767
Tongue	34 (33.7%)	26 (27.4%)	
Lip	8 (7.9%)	11 (11.6%)	
Gingiva	8 (7.9%)	8 (8.4%)	
Others	8 (7.9%)	13 (13.7%)	
Stage			
I	31 (30.7%)	25 (26.3%)	0.231
II	22 (21.8%)	27 (28.4%)	
III	14 (13.9%)	6 (6.3%)	
IV	34 (33.7%)	37 (38.9%)	
Tumor T status			
T1	38 (37.6%)	30 (31.6%)	0.560
T2	34 (33.7%)	33 (34.7%)	
T3	2 (2.0%)	5 (5.3%)	
T4	27 (26.7%)	27 (28.4%)	
Lymph node status			
N0	72 (71.3%)	68 (71.6%)	0.085
N1	16 (15.8%)	7 (7.4%)	
N2	13 (12.9%)	20 (21.1%)	
Metastasis			
M0	100 (99.0%)	95 (100%)	0.331
M1	1 (1.0%)	0 (0%)	
Cell differentiation			
Well	13 (12.9%)	20 (21.1%)	0.126
Moderately + poorly	88 (87.1%)	75 (78.9%)	

Mann–Whitney U test was used to compare demographic and clinical parameters between BQ-related and -unrelated cohorts

were single-nucleotide variations (SNVs) and the remaining were small insertions and deletions (INDELs), corresponding to a mean somatic mutation rate of 2.71 per megabase in target regions. This is compatible with the mutation rate observed in previous studies (Agrawal et al. 2011; India Project Team of the International Cancer Genome 2013; Stransky et al. 2011). To possibly restrain the impact of other environmental and intrinsic factors associated with the tumorigenic hypermutator phenotype (Roberts and Gordenin 2014), eight hypermutated samples (> 1000 mutations, a heuristic for limiting the number of discarded samples in such cancer type determined as samples with a mutation count exceeding Tukey's outlier condition (Bailey et al. 2018)) were excluded in further analyses.

Single-nucleotide variation (SNV) and insertion/deletion (INDEL)

We first compared the mutation rate between BQ and non-BQ users and found no difference in the total number of somatic mutations, including SNVs and INDELs, for all OSCC types together (Fig. 1a). Notably, cancers of the buccal mucosa from BQ users exhibited more INDELs than those from non-BQ users, while the number of both SNVs and INDELs was higher in BQ users compared with non-BQ users for tongue carcinoma. Further analysis of INDEL size revealed that BQ chewing tends to cause shorter INDELs (< 3 bp) found at mono/polynucleotide repeats, but not longer ones (≥ 3 bp) with overlapping microhomology in cancers of the tongue, buccal mucosa, and overall oral cavity (Fig. 1b), yet the difference was marginal considering multiple testing.

Mutational signature

Mutational signatures deciphered from tumor genomes can provide valuable insights into cancer etiology. We then extracted mutational signatures, estimated the contributions of each signature to each tumor, and compared the numbers of mutations attributable to each signature in BQ users and non-BQ users. A panel of three mutational signatures was identified in our cohort (Fig. 2a), with each corresponding to previously reported signature 1, 7, and 13, respectively, in the Catalogue Of Somatic Mutations In Cancer (COSMIC) (Figure S2 and Table S1) (Alexandrov et al. 2013). Among them, exposures to signature 1, defined as the contribution of this mutational signature to the total number of mutations found in each OSCC genome, were higher in tumors from BQ users for overall OSCC and tongue carcinoma (Fig. 2b), yet the difference was marginal considering multiple comparisons. Signature 1, featured with predominant C > T mutations at CpG sites, is initiated by spontaneous deamination of 5-methylcytosine and correlated with the age of cancer diagnosis (Alexandrov et al. 2013). Notably, a certain level of T > C and C > A mutations that could be derived from other consensus signature(s) not deciphered in our dataset seemed to be blended in the signature 1 identified here (Fig. 2a). We observed an increase in T > C, but

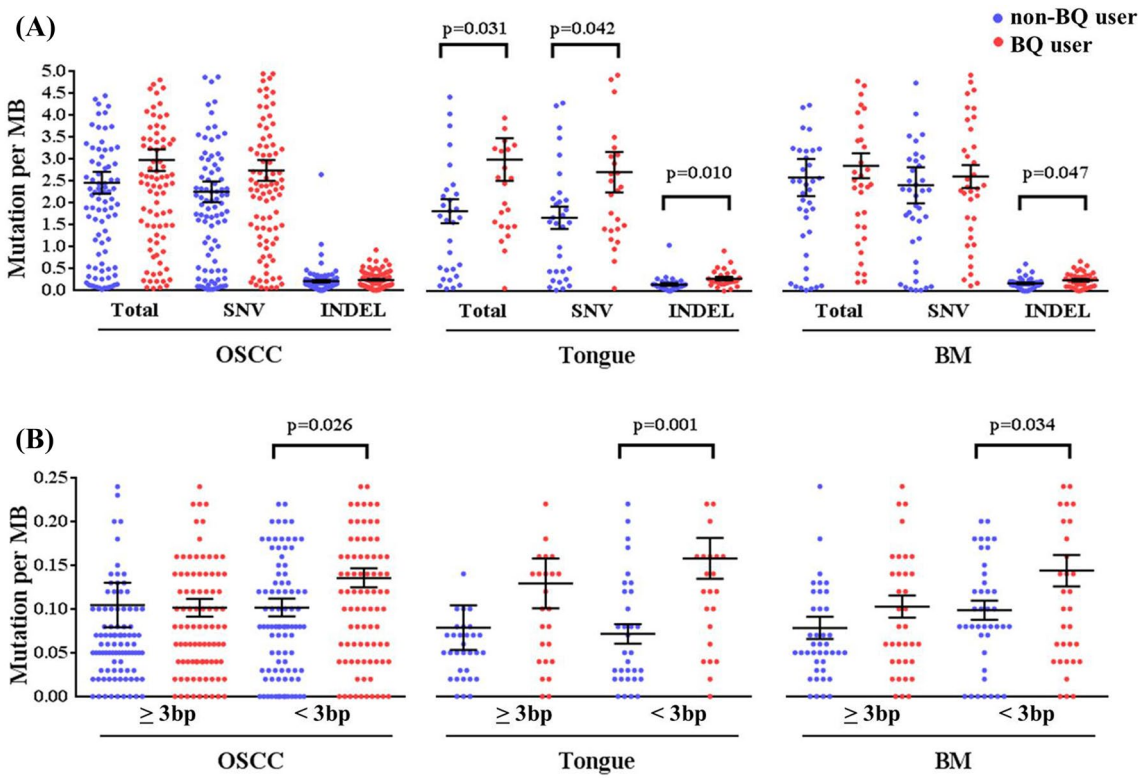

Fig. 1 Comparison of somatic mutation rate between habitual BQ users and non-BQ users. **a** Dots are used to display the numbers of somatic single-nucleotide variations (SNV), insertions and deletions (INDEL), and their combination (Total) per megabase (MB) detected in cancers of the tongue (Tongue, $n=26$ and 32 for BQ-related and -unrelated tumors, respectively), buccal mucosa (BM, $n=37$ and 41 for BQ-related and -unrelated tumors, respectively), and overall oral cavity (OSCC, $n=94$ and 94 for BQ-related and -unrelated tumors, respectively). **b** INDELs were further divided into longer INDELs (≥ 3 bp) with overlapping microhomology and shorter ones (<3 bp) found at mono/polynucleotide repeats for comparisons. Lines represent mean ± SD

not C > A mutations in overall OSCC (Fig. 2c), as tongue carcinoma from BQ users carried a marginally higher frequency of C > T mutations than those from non-BQ users (Fig. 2d). To improve the resolution of mutational signature identification, we decomposed our mutational catalog into eight signatures previously observed in head and neck cancers (Figure S3A) (Agrawal et al. 2011; India Project Team of the International Cancer Genome 2013; Su et al. 2017) to further address this issue (Figure S3B). After signature refitting, we found that in addition to signature 1, the contribution of signature 5 to the mutation load was also higher in BQ-related OSCC (Figure S3C). These data indicate the identification of mutational signatures linked to BQ chewing in OSCC.

Copy-number variation (CNV) and chromosomal breakpoint

In addition to SNV and INDEL, determination of somatic copy-number variation (CNV) and structural variation (SV) at the nucleotide level is important to reveal the genetic causes for head and neck cancers (Beck and Golemis 2016; Chang et al. 2017; Pickering et al. 2013). Next, we compared the fraction of the genomic regions that is copy-number aberrant and the numbers of SV-related breakpoints in OSCC genomes between BQ users and non-BQ users. Strikingly, chromosomal breaks were more frequently detected in cancers of the tongue, buccal mucosa, and overall oral cavity of BQ users (Fig. 3a). In contrast, increased extent of copy-number-altered areas was only seen in cancers of the buccal mucosa, but not in overall OSCC or tongue carcinoma regarding the status of BQ chewing (Fig. 3b). Our data suggest that BQ-related tumors from distinct anatomic sites of the oral cavity may undergo diverse mutational processes and exhibit differential levels of vulnerability to mutagenic impacts of BQ chewing.

Microsatellite instability (MSI) and loss of heterozygosity (LOH)

Microsatellite instability (MSI) and loss of heterozygosity (LOH) are of value in detection of oral premalignant lesions (Kang et al. 2015), even though the majority of OSCC is not MSI-high tumors (Cortes-Ciriano et al.

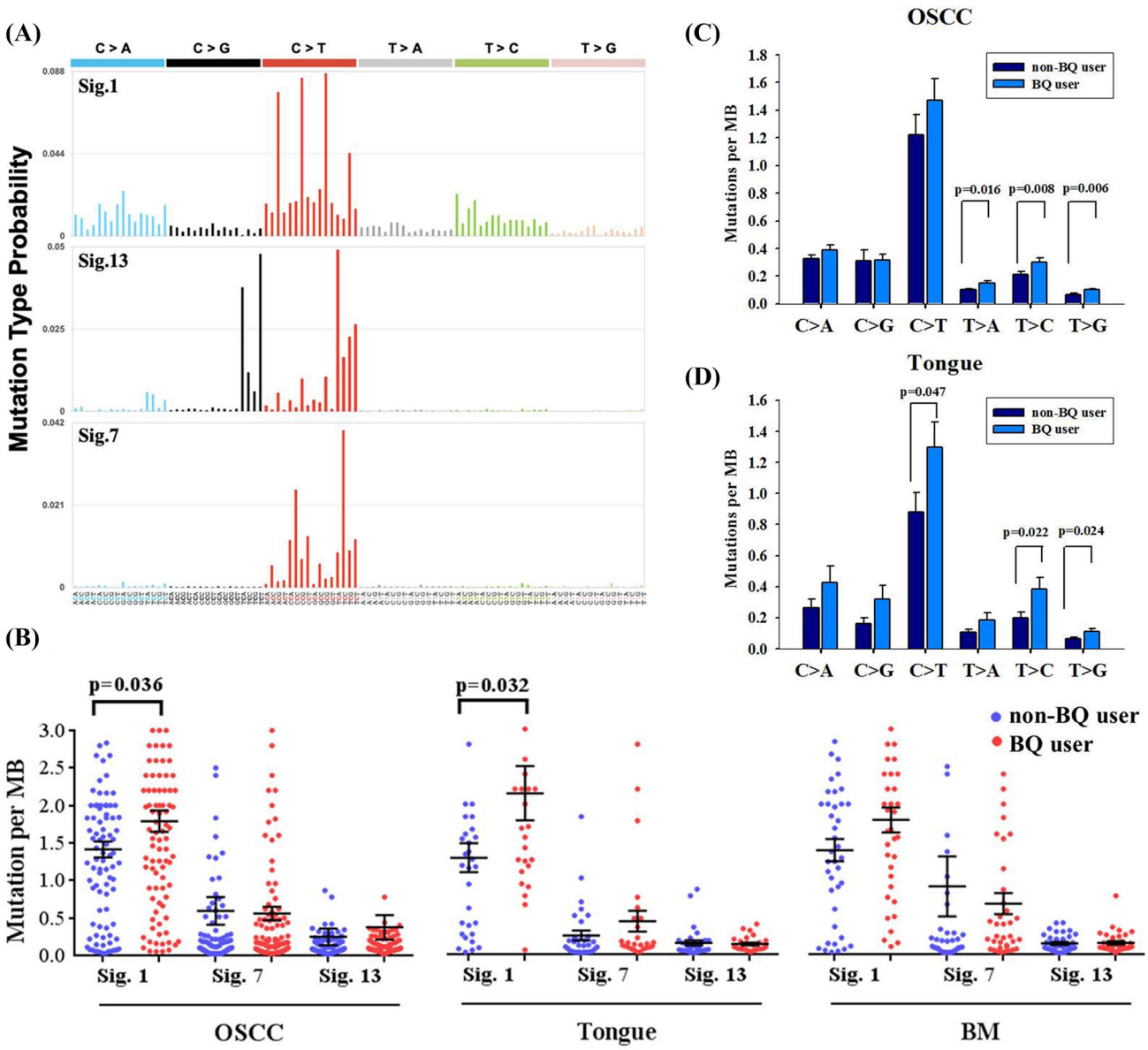

Fig. 2 Mutational signature associated with BQ chewing. **a** Three mutational signatures were deciphered in the OSCC cohort and corresponding to the updated consensus signatures (Alexandrov et al. 2013). Signatures are depicted using a 96-substitution classification defined by the substitution type and sequence context immediately 5′ and 3′ to the mutated base. Substitution types are shown in different colors on the horizontal axis. The vertical axis illustrates the percentage of mutations attributed to a specific mutation type. **b** Comparison of the signature exposures between habitual BQ users and non-BQ users. Dots represent numbers of mutations per MB attributed to mutational signatures found in cancers of the tongue (Tongue), buccal mucosa (BM), and overall oral cavity (OSCC). Lines represent mean ± SD. **c**, **d** Comparison of somatic base substitutions between BQ-related and -unrelated OSCC (**c**) and tongue carcinoma (**d**). Bars represent the average numbers of substitution types per MB detected in each group

2017). To determine whether BQ chewing affects these two molecular tumor phenotypes, the numbers of MSI events and the proportion of target regions bearing LOH in OSCC genomes were calculated. Although MSI events were detectable in slightly more OSCC samples from BQ users (35.1%, 33/94 and 29.8%, 28/94 for BQ-related and -unrelated OSCC, respectively; 40.5%, 15/37 and 24.4%, 10/41 for BQ-related and -unrelated cancers of buccal mucosa, respectively), there is no difference in the average number of MSI events between the tumors of BQ users and of non-BQ users (Fig. 4). In addition, we detected greater LOH regions in tongue cancers and overall OSCC from non-BQ chewers as compared with those from BQ chewers.

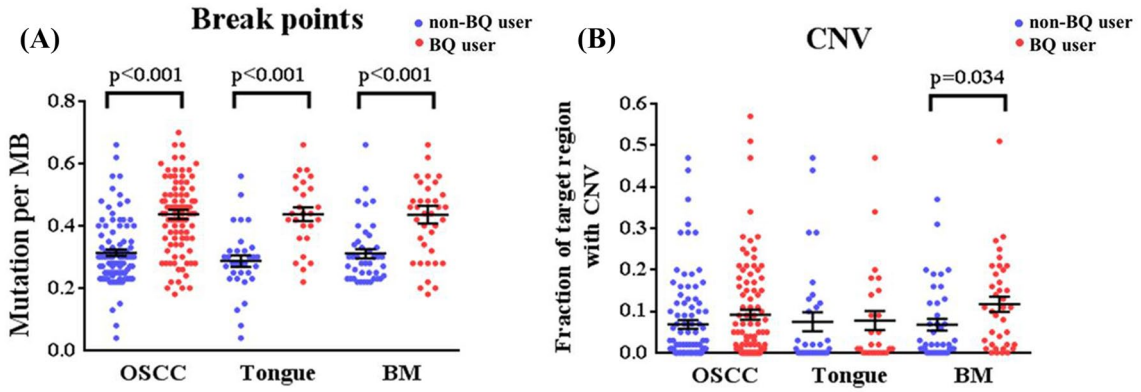

Fig. 3 Comparison of the frequency of breakpoints and the extent of copy-number variations between habitual BQ users and non-BQ users. Dots are the numbers of breakpoints per MB (**a**) or fraction of the target region that shows copy-number changes (**b**) detected in cancers of the tongue (Tongue, $n=26$ and 32 for BQ-related and -unrelated tumors, respectively), buccal mucosa (BM, $n=37$ and 41 for BQ-related and -unrelated tumors, respectively), and overall oral cavity (OSCC, $n=94$ and 94 for BQ-related and -unrelated tumors, respectively). Lines represent mean ± SD

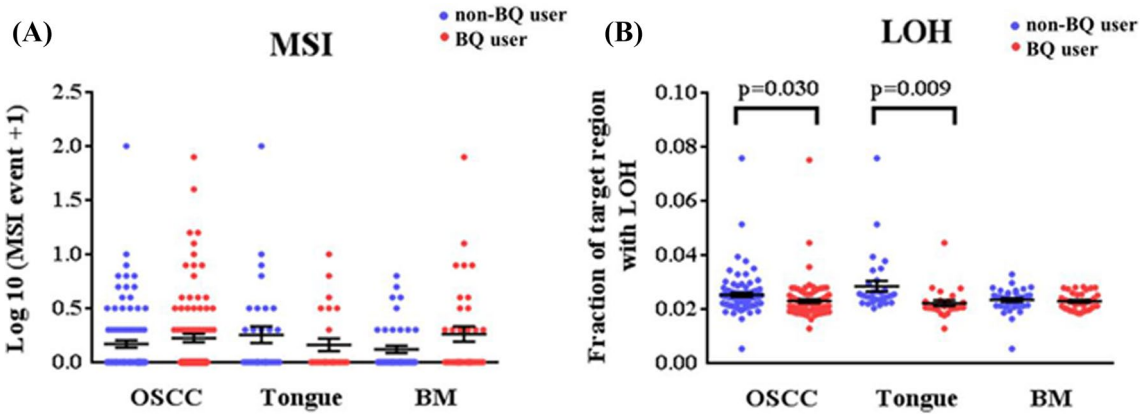

Fig. 4 Comparison of the frequency of microsatellite instability (MSI) events and the extent of loss of heterozygosity (LOH) between habitual BQ users and non-BQ users. Dots represent the numbers of MSI events (**a**) or fraction of the target region bearing LOH (**b**) found in cancers of the tongue (Tongue, $n=26$ and 32 for BQ-related and -unrelated tumors, respectively), buccal mucosa (BM, $n=37$ and 41 for BQ-related and -unrelated tumors, respectively), and overall oral cavity (OSCC, $n=94$ and 94 for BQ-related and -unrelated tumors, respectively). Lines represent mean ± SD

Discussion

The mutational spectrum of head and neck cancer appears to vary depending on the exposure of environmental carcinogens (Beck and Golemis 2016), as habitual consumption of tobacco, alcohol, and BQ is recognized as the major etiologic factors for this type of malignancies. Distinctive mutational features in tumors resulting from carcinogenesis associated with cigarette smoking (Alexandrov et al. 2016) and alcohol drinking (Chang et al. 2017) have been demonstrated. However, the molecular footprints for which BQ chewing damages the genome and creates somatic alterations that ultimately cause OSCC are still not fully understood. In the present study, through conducting a quantitative survey of mutational patterns and different types of genetic aberrations on the BQ-related OSCC exomes, we identified a specific mutational signature that is linked to the use of BQ. In addition, particular types of somatic alterations, such as INDEL and SV-related breakpoint, were found to be enriched in oral cancers from BQ chewers as compared with those from OSCC patients without a history of BQ use. For the first time, the mutagenic impact of BQ chewing has been comprehensively characterized at the genome-wide scale.

BQ-derived carcinogens may promote DNA lesions in oral epithelium, facilitating transformation and genetic instability. Unlike tobacco smoking that has been shown to

increase mutation burdens in multiple cancer types (Alexandrov et al. 2016), we did not observe elevated overall mutation rates (SNV and INDEL) in OSCC with respect to patients' history of BQ chewing. Notably, among oral cancers of different anatomic sites, association of BQ chewing with an increase in the number of somatic mutations was seen merely in cancers of the tongue (Fig. 1), although a recent study with limited sample size failed to detect the difference in the mutation load between BQ-related and -unrelated tongue carcinoma (Zhang et al. 2019). Our results suggest that tumors from distinct anatomic sites of the oral cavity may exhibit differential levels of vulnerability to mutagenic impacts of BQ chewing. Furthermore, since a higher number of somatic mutations were observed previously in large-sized tumors and in elder patients with OSCC (Su et al. 2017), we have enrolled the study cohorts (cancers of the tongue, buccal mucosa, and overall oral cavity) with matched age and TNM staging (Tables S2 and S3) for ruling out potential confounding factors.

On the other hand, similar to cigarette smoke, BQ products are a mixture of carcinogenic chemicals and may generate a variety of nitrosamines. It has been shown that cigarette smoking caused a higher number of INDELs and dinucleotide substitutions and was associated with increased fraction of copy-number aberrant regions in organs directly (lung adenocarcinoma) and indirectly (liver cancer) exposed to cigarette smoke (Alexandrov et al. 2016). Here, we detected more INDELs, especially those shorter ones found at mono/polynucleotide repeats (< 3 bp), but no significant increase in the number of dinucleotide substitutions (Data not shown) and extent of CNV in OSCC from BQ chewers as compared with those from non-BQ users. Strikingly, as the overall survival rate in patients with OSCC is one of the lowest among common malignant neoplasms in BQ-prevalent areas (Chen et al. 2008; Nair et al. 2004), in BQ-related OSCC, we observed more frequent SV-related breakpoints, whose presence is associated with cancer patient survival rates (Ewing and Semple 2018; Inaki and Liu 2012). It has been shown that arecoline, the most abundant alkaloid in areca nuts, induced structural chromosomal aberration, sister chromatid exchange and micronuclei formation in different cell types (Jeng et al. 2001; Shirname et al. 1983). Our results, for the first time, indicate a mutagenic impact of BQ chewing on preferentially eliciting small INDELs and SV-related breakpoints in OSCC genomes. Moreover, MSI, a molecular tumor phenotype diagnosed by the gain or loss of nucleotides in repetitive DNA sequences (microsatellites) resulting from genomic hypermutability, can potentially identify cancer driver genes and reflect the outcomes of cancer immunotherapies and patients' survival (Cortes-Ciriano et al. 2017; Hause et al. 2016). By assessing ten microsatellite markers, the presence of MSI in the surgical margin of head and neck squamous cell carcinoma was shown to be associated with local recurrence in an endemic BQ chewing area (Lin et al. 2016). Yet, although MSI events were detectable in slightly more cancers of buccal mucosa from BQ users, no difference in the average number of MSI events between the tumors from BQ users and from non-BQ users was observed in our genome-wide survey (Fig. 4).

In addition, by deconvoluting BQ-related OSCC genomes, we identified a mutational signature, whose contribution to the total number of mutations was higher in tumors from BQ users for overall OSCC and tongue carcinoma. This mutational pattern, corresponding to previously reported signature 1 in COSMIC, is described predominately by C > T substitutions at NpCpG sites and believed to be a result of an endogenous mutation process carried out by activation-induced deaminase (AID) or related enzymes (Alexandrov et al. 2013). Here, we detected a distinctive higher frequency of C > T mutations in BQ-related tongue carcinoma as compared with those from BQ-unrelated tumors. Consistently, frequent mutations have been previously found in the CpG islands of BQ-related tongue cancer genomes and such NpCpG > NpTpG pattern was the major contributor of the CpG island mutations (Zhang et al. 2019). A similar observation that such AID-associated mutational signature overlaps with CpG methylation sites was reported in follicular lymphoma and other cancers (Rogozin et al. 2016). Besides, in a recent study where genome-wide DNA methylation profiles of OSCC among patients in India, another BQ-prevalent country, were compared with that of mostly BQ-unrelated samples from The Cancer Genome Atlas program, a set of hypomethylated regions enriched in the promoters of immune response genes that may contribute to OSCC progression was identified (Basu et al. 2017). Our results, for the first time, link the consensus signature 1 with exogeneous exposure of BQ-derived carcinogens in OSCC and implicate a role of this association in changing DNA methylation status through an AID-dependent loss of CpG islands. Furthermore, since a certain level of T > C and C > A mutations that could be derived from other signature(s) not deciphered in our dataset seemed to be blended in the BQ-associated signature identified here, we found an additional signature, signature 5, significantly enriched in BQ-related tumors after signature refitting. Signature 5, but not signature 1, has been associated with smoking in oral cancer, although not a direct consequence of tobacco carcinogens due to its ubiquity in cancer types unrelated to tobacco smoking (Alexandrov et al. 2016). Inevitably, a higher portion of patients who had the habit of BQ chewing reported a history of smoking in our cohort, compared with that of non-BQ chewers. Our results, together with others' suggest that mutations attributed to consensus signature 1 in our dataset may be largely associated with BQ chewing, as those signature 5 mutations are likely due to the confounding risk factor, smoking, in our cohort.

To explore the mutational signatures and mutagenic impacts associated with BQ chewing in OSCC, additional work is needed to address several limitations of the present study. One issue is that the composition of betel quid differs geographically. There are three major types of betel quid used in Taiwan, all of which are free of tobacco and associated with the particular molecular pathology of oral cancer (Chen et al. 2008). In this regard, the observations about the genomic alterations associated with BQ-related OSCC in this study might be confined to particular ethnic or geographic groups. Another drawback is that the impact from covariate behaviors of BQ chewers cannot be excluded. Smoking and alcohol uptake are two prominent covariate behaviors of BQ chewers that have been linked to specific mutational signatures in head and neck cancers. Signature 4 (characterized mainly by C > A mutations) and signature 5 (with a predominance of T > C and C > T mutations) were increased in smokers versus nonsmokers (Alexandrov et al. 2016), as signature 16 (featured by T > C mutations) was associated with alcohol drinking (Chang et al. 2017). These signatures were not extracted in this study presumably due to a limited size of our mutational catalog. However, no difference in overall C > A mutation counts was observed in our cohort (Fig. 2c). In spite of an overall increase of T > C mutations observed in our BQ-related samples, we did not detect the difference in T > C mutation calls at ApTpN sites in our cohort (Figure S4) as signature 5 and 16 are both predominantly represented by T > C mutations at ApTpN sites. These suggest that a possible contribution from excessive alcohol consumption and smoking, to some extent, may be involved, yet BQ chewing itself is most plausibly the major cause of these observed differences based on our current study design. Also not excluded is the role for the differences in the biology of tumors from distinct locations of the oral cavity arising in BQ users compared with non-BQ users since OSCC relates to a myriad of anatomical sites that comprise this disease. Additionally, the actual impact of BQ chewing on OSCC may be underestimated, as higher mutation rates were observed in introns than in exons due to differential mismatch repair (Frigola et al. 2017). A large-scale analysis of oral cancer using whole-genome sequencing will further extend our knowledge of BQ-associated effects on OSCC genomes. Also unavailable was a completed collection of HPV typing data, since BQ chewing was recognized as a prevailing risk of OSCC among males in Taiwan. Moreover, the genomic landscape of BQ-related OSCC and whether there is a transcriptional strand bias within BQ-related mutational signatures remain a puzzle in the OSCC etiology and need future investigations.

Collectively, we identified a specific mutational signature and particular types of somatic aberrations associated with BQ chewing, which permits reassessment of our understanding of how BQ chewing affects oral carcinogenesis at the genome-wide scale.

Acknowledgements This study was supported by research grants from Chung Shan Medical University Hospital, Taiwan (CSH-2019-E-001-Y3) and the Health and Welfare Surcharge on Tobacco Products (grants MOHW107-TDU-B-212-114025; MOHW108-TDU-B-212-124025) from the Ministry of Health and Welfare (MOHW), Executive Yuan, Taiwan to SFY, and from Chang Gung Memorial Hospital (CMRPG2F0421) to SCS. We thank Tissue Bank at Chang Gung Memorial Hospital, Keelung, for sample preparation.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval All procedures performed in studies involving human participants were in accordance with the institutional review board of Chung Shan Medical University Hospital, Taichung, Taiwan (CSMUH NO: CS17132), and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

Informed consent Informed consent was obtained from all individual participants included in the study.

References

Agrawal N, Frederick MJ, Pickering CR, Bettegowda C, Chang K, Li RJ, Fakhry C, Xie TX, Zhang J, Wang J, Zhang N, El-Naggar AK, Jasser SA, Weinstein JN, Trevino L, Drummond JA, Muzny DM, Wu Y, Wood LD, Hruban RH, Westra WH, Koch WM, Califano JA, Gibbs RA, Sidransky D, Vogelstein B, Velculescu VE, Papadopoulos N, Wheeler DA, Kinzler KW, Myers JN (2011) Exome sequencing of head and neck squamous cell carcinoma reveals inactivating mutations in NOTCH1. Science 333:1154–1157. https://doi.org/10.1126/science.1206923

Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SA, Behjati S, Biankin AV, Bignell GR, Bolli N, Borg A, Borresen-Dale AL, Boyault S, Burkhardt B, Butler AP, Caldas C, Davies HR, Desmedt C, Eils R, Eyfjord JE, Foekens JA, Greaves M, Hosoda F, Hutter B, Ilicic T, Imbeaud S, Imielinski M, Jager N, Jones DT, Jones D, Knappskog S, Kool M, Lakhani SR, Lopez-Otin C, Martin S, Munshi NC, Nakamura H, Northcott PA, Pajic M, Papaemmanuil E, Paradiso A, Pearson JV, Puente XS, Raine K, Ramakrishna M, Richardson AL, Richter J, Rosenstiel P, Schlesner M, Schumacher TN, Span PN, Teague JW, Totoki Y, Tutt AN, Valdes-Mas R, van Buuren MM, van't Veer L, Vincent-Salomon A, Waddell N, Yates LR, Zucman-Rossi J, Futreal PA, McDermott U, Lichter P, Meyerson M, Grimmond SM, Siebert R, Campo E, Shibata T, Pfister SM, Campbell PJ, Stratton MR, Australian Pancreatic Cancer Genome I, Consortium IBC, Consortium IM-S, PedBrain I (2013) Signatures of mutational processes in human cancer. Nature 500:415–421. https://doi.org/10.1038/nature12477

Alexandrov LB, Ju YS, Haase K, Van Loo P, Martincorena I, Nik-Zainal S, Totoki Y, Fujimoto A, Nakagawa H, Shibata T, Campbell PJ, Vineis P, Phillips DH, Stratton MR (2016) Mutational signatures associated with tobacco smoking in human cancer. Science 354:618–622. https://doi.org/10.1126/science.aag0299

Awang MN (1988) Betel quid and oral carcinogenesis. Singap Med J 29:589–593

Bailey MH, Tokheim C, Porta-Pardo E, Sengupta S, Bertrand D, Weerasinghe A, Colaprico A, Wendl MC, Kim J, Reardon B, Ng PK, Jeong KJ, Cao S, Wang Z, Gao J, Gao Q, Wang F, Liu EM, Mularoni L, Rubio-Perez C, Nagarajan N, Cortes-Ciriano I, Zhou DC, Liang WW, Hess JM, Yellapantula VD, Tamborero D, Gonzalez-Perez A, Suphavilai C, Ko JY, Khurana E, Park PJ, Van Allen EM, Liang H, Godzik A, Lopez-Bigas N, Stuart J, Wheeler D, Getz G, Chen K, Lazar AJ, Mills GB, Karchin R, Ding L, Group MCW, Cancer Genome Atlas Research N, Lawrence MS (2018) Comprehensive characterization of cancer driver genes and mutations. Cell 173(371–385):e18. https://doi.org/10.1016/j.cell.2018.02.060

Basu B, Chakraborty J, Chandra A, Katarkar A, Baldevbhai JRK, Dhar Chowdhury D, Ray JG, Chaudhuri K, Chatterjee R (2017) Genome-wide DNA methylation profile identified a unique set of differentially methylated immune genes in oral squamous cell carcinoma patients in India. Clin Epigenet 9:13. https://doi.org/10.1186/s13148-017-0314-x

Beck TN, Golemis EA (2016) Genomic insights into head and neck cancer. Cancers Head Neck. https://doi.org/10.1186/s41199-016-0003-z

Boeva V, Popova T, Bleakley K, Chiche P, Cappo J, Schleiermacher G, Janoueix-Lerosey I, Delattre O, Barillot E (2012) Control-FREEC: a tool for assessing copy number and allelic content using next-generation sequencing data. Bioinformatics 28:423–425. https://doi.org/10.1093/bioinformatics/btr670

Cancer Genome Atlas N (2015) Comprehensive genomic characterization of head and neck squamous cell carcinomas. Nature 517:576–582. https://doi.org/10.1038/nature14129

Chang J, Tan W, Ling Z, Xi R, Shao M, Chen M, Luo Y, Zhao Y, Liu Y, Huang X, Xia Y, Hu J, Parker JS, Marron D, Cui Q, Peng L, Chu J, Li H, Du Z, Han Y, Tan W, Liu Z, Zhan Q, Li Y, Mao W, Wu C, Lin D (2017) Genomic analysis of oesophageal squamous-cell carcinoma identifies alcohol drinking-related mutation signature and genomic alterations. Nat Commun 8:15290. https://doi.org/10.1038/ncomms15290

Chen CL, Chi CW, Chang KW, Liu TY (1999) Safrole-like DNA adducts in oral tissue from oral cancer patients with a betel quid chewing history. Carcinogenesis 20:2331–2334

Chen YJ, Chang JT, Liao CT, Wang HM, Yen TC, Chiu CC, Lu YC, Li HF, Cheng AJ (2008) Head and neck cancer in the betel quid chewing area: recent advances in molecular carcinogenesis. Cancer Sci 99:1507–1514. https://doi.org/10.1111/j.1349-7006.2008.00863.x

Chen K, Wallis JW, McLellan MD, Larson DE, Kalicki JM, Pohl CS, McGrath SD, Wendl MC, Zhang Q, Locke DP, Shi X, Fulton RS, Ley TJ, Wilson RK, Ding L, Mardis ER (2009) BreakDancer: an algorithm for high-resolution mapping of genomic structural variation. Nat Methods 6:677–681. https://doi.org/10.1038/nmeth.1363

Chu NS (2001) Effects of Betel chewing on the central and autonomic nervous systems. J Biomed Sci 8:229–236. https://doi.org/10.1159/000054038

Cortes-Ciriano I, Lee S, Park WY, Kim TM, Park PJ (2017) A molecular portrait of microsatellite instability across multiple cancers. Nat Commun 8:15180. https://doi.org/10.1038/ncomms15180

Edge SB, Compton CC (2010) The American Joint Committee on Cancer: the 7th edition of the AJCC cancer staging manual and the future of TNM. Ann Surg Oncol 17:1471–1474. https://doi.org/10.1245/s10434-010-0985-4

Ewing A, Semple C (2018) Breaking point: the genesis and impact of structural variation in tumours. Breaking point: the genesis and impact of structural variation in tumours. https://doi.org/10.12688/f1000research.16079.1

Frigola J, Sabarinathan R, Mularoni L, Muinos F, Gonzalez-Perez A, Lopez-Bigas N (2017) Reduced mutation rate in exons due to differential mismatch repair. Nat Genet 49:1684–1692. https://doi.org/10.1038/ng.3991

Gillison ML, Koch WM, Capone RB, Spafford M, Westra WH, Wu L, Zahurak ML, Daniel RW, Viglione M, Symer DE, Shah KV, Sidransky D (2000) Evidence for a causal association between human papillomavirus and a subset of head and neck cancers. J Natl Cancer Inst 92:709–720. https://doi.org/10.1093/jnci/92.9.709

Gillison ML, Akagi K, Xiao W, Jiang B, Pickard RKL, Li J, Swanson BJ, Agrawal AD, Zucker M, Stache-Crain B, Emde AK, Geiger HM, Robine N, Coombes KR, Symer DE (2019) Human papillomavirus and the landscape of secondary genetic alterations in oral cancers. Genome Res 29:1–17. https://doi.org/10.1101/gr.241141.118

Hause RJ, Pritchard CC, Shendure J, Salipante SJ (2016) Classification and characterization of microsatellite instability across 18 cancer types. Nat Med 22:1342–1350. https://doi.org/10.1038/nm.4191

IARC Working Group on the Evaluation of Carcinogenic Risks to Humans (2004) Betel-quid and areca-nut chewing and some areca-nut derived nitrosamines. IARC Monogr Eval Carcinog Risks Hum 85:1–334

Inaki K, Liu ET (2012) Structural mutations in cancer: mechanistic and functional insights. Trends Genet 28:550–559. https://doi.org/10.1016/j.tig.2012.07.002

India Project Team of the International Cancer Genome C (2013) Mutational landscape of gingivo-buccal oral squamous cell carcinoma reveals new recurrently-mutated genes and molecular subgroups. Nat Commun 4:2873. https://doi.org/10.1038/ncomms3873

Jemal A, Siegel R, Ward E, Hao Y, Xu J, Murray T, Thun MJ (2008) Cancer statistics, 2008. CA Cancer J Clin 58:71–96. https://doi.org/10.3322/CA.2007.0010

Jeng JH, Chang MC, Hahn LJ (2001) Role of areca nut in betel quid-associated chemical carcinogenesis: current awareness and future perspectives. Oral Oncol 37:477–492

Kang H, Kiess A, Chung CH (2015) Emerging biomarkers in head and neck cancer in the era of genomics. Nat Rev Clin Oncol 12:11–26. https://doi.org/10.1038/nrclinonc.2014.192

Krishna Rao SV, Mejia G, Roberts-Thomson K, Logan R (2013) Epidemiology of oral cancer in Asia in the past decade–an update (2000-2012). Asian Pac J Cancer Prev 14:5567–5577

Kruger S, Piro RM (2019) decompTumor2Sig: identification of mutational signatures active in individual tumors. BMC Bioinform 20:152. https://doi.org/10.1186/s12859-019-2688-6

Lee LA, Huang CG, Tsao KC, Liao CT, Kang CJ, Chang KP, Huang SF, Chen IH, Fang TJ, Li HY, Yang SL, Lee LY, Hsueh C, Chen TC, Lin CY, Fan KH, Wang HM, Ng SH, Chang YL, Lai CH, Shih SR, Yen TC (2013) Increasing rates of low-risk human papillomavirus infections in patients with oral cavity squamous cell carcinoma: association with clinical outcomes. J Clin Virol 57:331–337. https://doi.org/10.1016/j.jcv.2013.04.010

Lee LA, Huang CG, Tsao KC, Liao CT, Kang CJ, Chang KP, Huang SF, Chen IH, Fang TJ, Li HY, Yang SL, Lee LY, Hsueh C, Lin CY, Fan KH, Chang TC, Wang HM, Ng SH, Yen TC (2015) Human papillomavirus infections are common and predict mortality in a retrospective cohort study of taiwanese patients with oral cavity cancer. Medicine (Baltimore) 94:e2069. https://doi.org/10.1097/MD.0000000000002069

Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25:1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Lin JC, Wang CC, Jiang RS, Wang WY, Liu SA (2016) Microsatellite alteration in head and neck squamous cell carcinoma patients from a betel quid-prevalent region. Sci Rep 6:22614. https://doi.org/10.1038/srep22614

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA (2010) The genome analysis toolkit: a MapReduce framework

for analyzing next-generation DNA sequencing data. Genome Res 20:1297–1303. https://doi.org/10.1101/gr.107524.110

Nair U, Obe G, Nair J, Maru GB, Bhide SV, Pieper R, Bartsch H (1991) Evaluation of frequency of micronucleated oral mucosa cells as a marker for genotoxic damage in chewers of betel quid with or without tobacco. Mutat Res 261:163–168

Nair UJ, Nair J, Friesen MD, Bartsch H, Ohshima H (1995) Ortho- and meta-tyrosine formation from phenylalanine in human saliva as a marker of hydroxyl radical generation during betel quid chewing. Carcinogenesis 16:1195–1198

Nair U, Bartsch H, Nair J (2004) Alert for an epidemic of oral cancer due to use of the betel quid substitutes gutkha and pan masala: a review of agents and causative mechanisms. Mutagenesis 19:251–262

Ndiaye C, Mena M, Alemany L, Arbyn M, Castellsague X, Laporte L, Bosch FX, de Sanjose S, Trottier H (2014) HPV DNA, E6/E7 mRNA, and p16INK4a detection in head and neck cancers: a systematic review and meta-analysis. Lancet Oncol 15:1319–1331. https://doi.org/10.1016/S1470-2045(14)70471-1

Niu B, Ye K, Zhang Q, Lu C, Xie M, McLellan MD, Wendl MC, Ding L (2014) MSIsensor: microsatellite instability detection using paired tumor–normal sequence data. Bioinformatics 30:1015–1016. https://doi.org/10.1093/bioinformatics/btt755

Pickering CR, Zhang J, Yoo SY, Bengtsson L, Moorthy S, Neskey DM, Zhao M, Ortega Alves MV, Chang K, Drummond J, Cortez E, Xie TX, Zhang D, Chung W, Issa JP, Zweidler-McKay PA, Wu X, El-Naggar AK, Weinstein JN, Wang J, Muzny DM, Gibbs RA, Wheeler DA, Myers JN, Frederick MJ (2013) Integrative genomic characterization of oral squamous cell carcinoma identifies frequent somatic drivers. Cancer Discov 3:770–781. https://doi.org/10.1158/2159-8290.CD-12-0537

Roberts SA, Gordenin DA (2014) Hypermutation in human cancer genomes: footprints and mechanisms. Nat Rev Cancer 14:786–800. https://doi.org/10.1038/nrc3816

Rogozin IB, Lada AG, Goncearenco A, Green MR, De S, Nudelman G, Panchenko AR, Koonin EV, Pavlov YI (2016) Activation induced deaminase mutational signature overlaps with CpG methylation sites in follicular lymphoma and other cancers. Sci Rep 6:38133. https://doi.org/10.1038/srep38133

Rosales RA, Drummond RD, Valieris R, Dias-Neto E, da Silva IT (2017) signeR: an empirical Bayesian approach to mutational signature discovery. Bioinformatics 33:8–16. https://doi.org/10.1093/bioinformatics/btw572

Scully C, Bagan J (2009) Oral squamous cell carcinoma overview. Oral Oncol 45:301–308. https://doi.org/10.1016/j.oraloncology.2009.01.004

Shirname LP, Menon MM, Nair J, Bhide SV (1983) Correlation of mutagenicity and tumorigenicity of betel quid and its ingredients. Nutr Cancer 5:87–91. https://doi.org/10.1080/01635588309513783

Stransky N, Egloff AM, Tward AD, Kostic AD, Cibulskis K, Sivachenko A, Kryukov GV, Lawrence MS, Sougnez C, McKenna A, Shefler E, Ramos AH, Stojanov P, Carter SL, Voet D, Cortes ML, Auclair D, Berger MF, Saksena G, Guiducci C, Onofrio RC, Parkin M, Romkes M, Weissfeld JL, Seethala RR, Wang L, Rangel-Escareno C, Fernandez-Lopez JC, Hidalgo-Miranda A, Melendez-Zajgla J, Winckler W, Ardlie K, Gabriel SB, Meyerson M, Lander ES, Getz G, Golub TR, Garraway LA, Grandis JR (2011) The mutational landscape of head and neck squamous cell carcinoma. Science 333:1157–1160. https://doi.org/10.1126/science.1208130

Su SC, Lin CW, Liu YF, Fan WL, Chen MK, Yu CP, Yang WE, Su CW, Chuang CY, Li WH, Chung WH, Yang SF (2017) Exome sequencing of oral squamous cell carcinoma reveals molecular subgroups and novel therapeutic opportunities. Theranostics 7:1088–1099. https://doi.org/10.7150/thno.18551

Yang HC, Chang LC, Huggins RM, Chen CH, Mullighan CG (2011) LOHAS: loss-of-heterozygosity analysis suite. Genet Epidemiol 35:247–260. https://doi.org/10.1002/gepi.20573

Zhang W, Wang M, Wu Q, Zhu Q, Jiao Y, Zhu Y, Yang B, Ni S, Yu J, Sun H, Zeng YX (2019) Mutational signatures and the genomic landscape of betel quid chewing-associated tongue carcinoma. Cancer Med 8:701–711. https://doi.org/10.1002/cam4.1888

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.