

Improvement in the risk assessment of oral leukoplakia through morphology-related copy number analysis

Xiaotian Li, Lu Liu, Jianyun Zhang, Ming Ma, Lisha Sun, Xuefen Li, Heyu Zhang, Jianbin Wang, Yanyi Huang and Tiejun Li

Citation: [SCIENCE CHINA Life Sciences](#); doi: 10.1007/s11427-021-1965-x

View online: <https://engine.scichina.com/doi/10.1007/s11427-021-1965-x>

Published by the [Science China Press](#)

Articles you may be interested in

[Assessment of risk factors associated with oral polio vaccine refusal in Rahim Yar Khan District, Pakistan \(2017\)](#)

Journal of Biosafety and Biosecurity **2**, 27 (2020);

[Thirty Days Mortality Following Lung Resection: Risk Factors, Preoperative Assessment and Improvement Program](#)

Journal of Advances in Health **01**, 205 (2019);

[Improvement of pentathiophene/fullerene planar heterojunction photovoltaic cells by improving the organic films morphology through the anode buffer bilayer](#)

European Physical Journal AP(Applied Physics) **74**, 24603 (2016);

[A RECOMBINANT PLASMID WITH INCREASED COPY NUMBER IN *Bacillus subtilis*](#)

Chinese Science Bulletin **30**, 400 (1985);

[Risk assessment at workspace in industrial radiography. Previous dosimetry analysis](#)

Radioprotection **43**, 89 (2008);

Improvement in the risk assessment of oral leukoplakia through morphology-related copy number analysis

Xiaotian Li^{1,5†}, Lu Liu^{2†}, Jianyun Zhang^{1,5}, Ming Ma^{1,5}, Lisha Sun^{4,5}, Xuefen Li^{4,5}, Heyu Zhang^{4,5*}, Jianbin Wang^{3*}, Yanyi Huang^{2,6,7*} & Tiejun Li^{1,5*}

¹Department of Oral Pathology, Peking University School and Hospital of Stomatology & National Center of Stomatology & National Clinical Research Center for Oral Diseases & National Engineering Laboratory for Digital and Material Technology of Stomatology & Beijing Key Laboratory of Digital Stomatology & Research Center of Engineering and Technology for Computerized Dentistry Ministry of Health & NMPA Key Laboratory for Dental Materials, Beijing 100081, China;

²Beijing Advanced Innovation Center for Genomics (ICG), Biomedical Pioneer Innovation Center (BIOPIC), School of Life Sciences, Peking University, Beijing 100871, China;

³School of Life Sciences and Beijing Advanced Innovation Center for Structural Biology, Tsinghua University, Beijing 100084, China;

⁴Central Laboratory, Peking University School and Hospital of Stomatology, Beijing 100081, China;

⁵Research Unit of Precision Pathologic Diagnosis in Tumors of the Oral and Maxillofacial Regions, Chinese Academy of Medical Sciences (2019RU034), Beijing 100081, China;

⁶College of Chemistry and Molecular Engineering and Beijing National Laboratory for Molecular Sciences, Peking University, Beijing 100871, China;

⁷Institute for Cell Analysis, Shenzhen Bay Laboratory, Shenzhen 528107, China

Received April 25, 2021; accepted June 8, 2021; published online August 2, 2021

Oral leukoplakia is the most common type of oral potentially malignant disorders and considered a precursor lesion to oral squamous cell carcinoma. However, a predictor of oral leukoplakia prognosis has not yet been identified. We investigated whether copy number alteration patterns may effectively predict the prognostic outcomes of oral leukoplakia using routinely processed paraffin sections. Comparison of copy number alteration patterns between oral leukoplakia with hyperplasia (HOL, $n=22$) and dysplasia (DOL, $n=21$) showed that oral leukoplakia with dysplasia had a higher copy number alteration rate (86%) than oral leukoplakia with hyperplasia (46%). Oral leukoplakia with dysplasia exhibited a wider range of genomic variations across all chromosomes compared with oral leukoplakia with hyperplasia. We also examined a retrospective cohort of 477 patients with oral leukoplakia with hyperplasia with detailed follow-up information. The malignant transformation (MT, $n=19$) and leukoplakia recurrence (LR, $n=253$) groups had higher frequencies of aneuploidy events and copy number loss rate than the free of disease (FD, $n=205$) group. Together, our results revealed the association between the degree of copy number alterations and the histological grade of oral leukoplakia and demonstrated that copy number alteration may be effective for prognosis prediction in oral leukoplakia patients with hyperplasia.

oral leukoplakia, copy number alteration, prognosis prediction

Citation: Li, X., Liu, L., Zhang, J., Ma, M., Sun, L., Li, X., Zhang, H., Wang, J., Huang, Y., and Li, T. (2021). Improvement in the risk assessment of oral leukoplakia through morphology-related copy number analysis. *Sci China Life Sci* 64, <https://doi.org/10.1007/s11427-021-1965-x>

†Contributed equally to this work

*Corresponding authors (Heyu Zhang, email: zhangheyu1983@sina.cn; Jianbin Wang, email: jianbinwang@tsinghua.edu.cn; Yanyi Huang, email: yanyi@pku.edu.cn; Tiejun Li, email: litiiejun22@vip.sina.com)

INTRODUCTION

Oral squamous cell carcinoma (OSCC), a major pathological type of head and neck cancer, accounts for more than 90% of all oral malignancies (Ng et al., 2017; Jemal et al., 2010). While the diagnosis and treatment of many cancers have significantly improved over the last decades, the early detection and effective management of OSCC remain challenging (Jemal et al., 2004; Haddad and Shin, 2008; Goodson and Thomson, 2011; Almangush et al., 2021; Chai et al., 2020; Lindemann et al., 2018). Oral potentially malignant disorders are thought to be the precancerous lesions of OSCC, and research has been focused on identifying reliable indicators to identify patients with these lesions. The development of new strategies for disease prediction based on molecular pathology is considered to have promising potential for the improvement of patient survival.

Oral leukoplakia (OL) is the most common oral potentially malignant disorder and is defined by the World Health Organization as a white plaque of questionable risk after exclusion of other known diseases or disorders that carry no increased risk for cancer (Warnakulasuriya and Ariyawardana, 2016; Lee et al., 2000; El-Nagger et al., 2017; Yanik et al., 2015). Currently, the standard diagnostic and prognostic assessments of OL are clinical examinations as well as biopsy, if necessary, to exclude other oral mucosal disorders and determine the presence or absence of epithelial dysplasia. Dysplasia-free OL is often designated as simple hyperplasia or hyperkeratosis. OL lesions with varying degrees of epithelial dysplasia are determined by histopathological examination based on architectural and cytological characteristics (Staines and Rogers, 2017). Importantly, OL with epithelial dysplasia shows a higher risk of malignant transformation (up to 13%) compared with dysplasia-free OL (2.1%) (Lee et al., 2000; Gao et al., 2012). Based on the presence and severity of epithelial dysplasia (i.e., mild, moderate, or severe) determined by pathological examinations, clinicians decide whether to monitor the lesion or to surgically intervene (Gomes et al., 2015; Monteiro et al., 2017). Lesions with a higher dysplasia grade show a higher risk of malignant transformation (Mehanna et al., 2009; Warnakulasuriya et al., 2011). Therefore, clinical therapy for these lesions tends to be surgical removal or even more radical approaches. However, histological grading of epithelial dysplasia is often subjective and displays low levels of intra- and inter-observer agreements (Holmstrup et al., 2007). Moreover, recent longitudinal studies have yielded contradictory results, suggesting that morphological assessment alone may not be a reliable predictor of OL prognosis (Saintigny et al., 2011; Bosman, 2001). Furthermore, no study has confirmed that surgical excision sufficiently prevents OL recurrence and/or malignant transformation (Lodi et al., 2016), indicating that both treated and untreated pa-

tients require close surveillance. Therefore, to better stratify patients and clarify individualized plans for clinical treatment and follow-up, the identification of prognostic indicators of OL is of great clinical significance.

In recent decades, substantial effort has been made to identify key molecular markers for the malignant transformation and prognosis of OL. Several genomic alterations have been identified in OL, including single nucleotide variations and mutation hotspots in genes related to the cell cycle and cancer regulatory pathways, such as *TP53*, *Ki-67*, and *NOTCH1* (Zhang et al., 2017; Ding et al., 2018; Yagyuu et al., 2017; Gissi et al., 2015). Loss of heterozygosity and changes in DNA methylation are also significantly associated with clinical outcomes in OL patients (Tsao et al., 2009; Zhou et al., 2011; Zhang et al., 2012; Türke et al., 2017). However, the frequencies of these mutations and altered biomarkers are low, and only a small number of OL cases harbor these genetic alterations (Mello et al., 2020). Thus, the intervention methods involving these markers have not yet proven to be practical (Dionne et al., 2014). Consequently, an effective marker to determine OL prognostic outcome has not yet been identified.

Copy number alterations (CNAs) include deletions, insertions, and duplications of DNA segments (Beroukhim et al., 2010). Unlike the analysis of single hotspot genes, whole genome scale CNA profiling can determine a full spectrum of large-scale genomic variation events, which allows for the identification of common genetic alterations. Recent research has reported that genomic copy number classification can be used in the early diagnosis of esophageal cancer (Killcoyne et al., 2020).

In the present study, we examined the potential of applying CNA analysis through low-depth whole-genome sequencing to predict prognosis and/or malignant potential in OL patients. Our approach integrated pathological findings with CNA profiling of OL tissues. We used laser capture microdissection (LCM) to acquire small-scale tissue samples (200–500 cells) from OL epithelium containing morphological information (foci taken from epithelial hyperplasia or dysplasia) using routinely processed paraffin sections. We applied this approach to elucidate an association between copy number instability and OL prognosis prediction by retrospectively analyzing OL patients with follow-up data. We identified high-risk CNA-harboring hyperplastic tissues, the molecular signature of which significantly improved the effectiveness of malignancy prediction in morphologically normal lesions.

RESULTS

Generation of spatial copy number maps

We combined LCM with a direct whole-genome library

construction method to determine the correlation between pathological information and genomic alterations in OL. To improve the throughput while minimizing the amplification bias of the whole-genome sequencing of mini-bulk samples, we developed a scalable library construction method that used Tn5 transposase to tagment genomic DNA without pre-amplification (Figure 1).

The sections were 10 μm thick and stained with hematoxylin-eosin, which allowed for morphological analysis (Figure 1A). LCM was directly performed on the sections; we obtained cells with clearly defined morphology and high purity at the size of hundreds of micrometers. Each dissected sample contained 200–500 cells and was collected in tubes with lysis buffer to release the DNA from histones (Figure 1B). DNA fragments were amplified using PCR primers containing sample barcodes (Figure 1C). In each sequencing run, we commonly pooled 200–400 sample libraries.

Considering the number of cells and DNA degradation in formalin-fixed paraffin-embedded (FFPE) samples, the quality of constructed libraries was very similar to single-cell libraries. Therefore, an approximate sequencing depth of 0.3 Gb ($0.1\times$) was obtained for each sample, which was similar to depths reported in single-cell CNA studies (Baslan et al., 2012; Mallory et al., 2020). A circular binary segmentation algorithm (Olshen et al., 2004) was used to determine the copy number profiles (resolution 2 Mb, $\alpha=0.0001$, $\text{min.width}=5$, $\text{undo.SD}=2$). We mainly focused on arm- or chromosome-level copy number changes with high confidence to eliminate false identifications of small-sized CNAs. Median absolute pairwise difference (MAPD) and the number of mapping reads were the major parameters used to filter out low-quality samples. CNA profiles for qualified samples could be mapped to original spatial locations to infer the relationship between the molecular signatures of cells with distinct morphologies and phenotypes (Figure 1D).

CNA patterns of OL with hyperplasia or dysplasia

We collected biopsy samples from 529 OL patients. Among these patients, 52 OL patients, including 26 OL with dysplasia (DOL) and 26 OL with hyperplasia (HOL), were diagnosed during 2018–2019 and did not have prognosis information, while 477 HOL patients diagnosed during 2012–2015 had follow-up information. Among the 477 HOL patients with prognosis information, 205 (43.0%) were free of disease, while 253 (53.0%) had recurring symptoms, and 19 (4.0%) exhibited malignant transformation (Table S1 in Supporting Information).

We laser-captured 1,650 samples from the 529 OL patients and obtained valid sequencing data for 998 samples. We used an $\text{MAPD}<0.25$ and a number of mapping reads $>100\text{ k}$ as the major criteria to further filter out unqualified sequencing

data (Figure S1 in Supporting Information). Finally, 580 samples from 260 patients were found to be suitable for subsequent CNA analysis.

We analyzed the CNA ratio according to patient characteristics in the overall OL patient group (Table S2 in Supporting Information). We found that the CNA ratio varied according to patient age, with older patients showing a higher CNA ratio compared with younger patients ($\chi^2=12.043$, $P=0.017$). Sex and site had no significant correlation with CNAs.

We then evaluated whether there were differences in copy number profiles between DOL and HOL patients by analyzing the data from the 26 HOL and 26 DOL patients diagnosed during 2018–2019. Overall, 43 samples from 22 HOL patients and 61 samples from 21 DOL patients passed the quality filtering. HOL is considered a benign proliferative lesion. We found that 18 samples (from 10 patients) of the 43 HOL samples harbored arm-level CNAs, most of which were aneuploidies (Figure 2A). Copy number gains (77.9%) were more frequent than losses (22.1%), with a preference for certain chromosomes (Figure S2A–C in Supporting Information). Chr 1, Chr 8, Chr 20, Chr 2, Chr 6, and Chr 5 tended to have a high frequency of copy number gain events, while copy number losses mainly occurred at a low frequency on Chr 5, Chr 8, Chr 13, and Chr 16.

DOL patients showed even higher rates of CNAs and aneuploidies than HOL patients (Figure 2B). Importantly, in DOL patients, the genomic variation events occurred on all chromosomes with a specific location preference (Figure S2A and B in Supporting Information). More than 25% of the samples harbored CNAs at Chr 20, Chr 8 and Chr 3. Notably, copy number breakpoints were found in Chr 3, Chr 5, Chr 7, Chr 9, and Chr 10 at a high frequency; in contrast, copy number breakpoints were not often detected in HOL samples. These breakpoints were repeated across patients, indicating a higher level of genome reorganization in DOL patients as compared with that in HOL patients. While HOL samples tended to display copy number gains, in DOL samples, the occurrence rate of copy number gains (66.2%) was closer to that of copy number loss (33.8%) (Figure S2C in Supporting Information).

Among the 22 HOL patients possessing qualified data, 10 patients (46%) were found to have CNAs, while 18 out of 21 (86%) DOL patients showed changes in copy number ($\chi^2=7.667$, $P=0.006$) (Figure 2C). The rate of copy number changes was similar between male and female patients in both the HOL and DOL groups (Figure S2D in Supporting Information).

We further used two numerical indices, CNAScore and COUNTScore, to quantitatively depict the landscape of copy number changes in each sample and evaluate the CNA differences between the HOL and DOL groups. CNAScore is an empirical index that combines two major components that

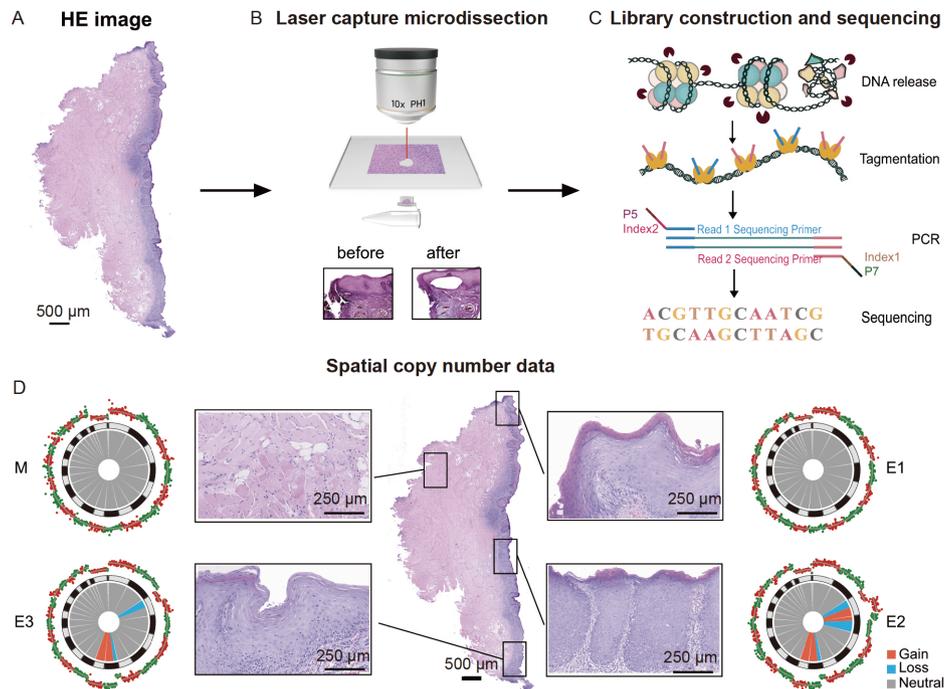


Figure 1 Overview of the copy number profiling method. A, Hematoxylin and eosin (H&E) stained images were used to scan whole tissues and identify cell morphologies. B, Mini-bulk samples were directly dissected by LCM. C, Work-flow of library construction. DNA molecules were released from histones and tagged by the Tn5 enzyme. Each library was barcoded and amplified using PCR. Barcoded libraries were pooled and subjected to next-generation sequencing. D, Copy number profiles were mapped to their original coordinates in tissue sections. A representative example of three epithelial samples (E1, E2 and E3) and one muscle sample (M) is shown.

reflect the degree of copy number changes and the general copy number deviation from a neutral value (see MATERIALS AND METHODS). COUNTscore evaluates the number of CNA segments in each sample and also reflects the frequency of the structural variation in the genome (see MATERIALS AND METHODS). We calculated the frequency distributions of CNAscore and COUNTscore values in the HOL and DOL samples (Figure 2D). The DOL group showed a relatively uniform distribution, with more samples possessing a higher CNAscore and COUNTscore. Conversely, most HOL samples showed low a CNAscore and COUNTscore. Considering the limitations of CNAscore and COUNTscore (see MATERIALS AND METHODS), we believe it is necessary to consider the results of both scores to establish a comprehensive conclusion when estimating the severity of CNAs.

We further found that 18 out of 61 DOL samples (30%) displayed severe copy number changes (CNAscore>4.3 and COUNTscore>7) compared with only 1 out of 43 (2%) of HOL samples (Figure S3 in Supporting Information). These results are in accordance with previous observations that DOL cases have more structural genomic variations than HOL cases.

Taken together, these results showed that compared with HOL samples, DOL samples exhibited more serious genome rearrangements in terms of a higher CNA occurrence rate, a broader distribution of CNAs, and a higher frequency of

copy number breakpoints. Such histologically associated CNA profiles suggested that genome instability may correlate with lesion severity.

Association between CNA events and disease progression of HOL

As one of the cancer hallmarks, CNA events in the HOL samples may have the potential to serve as molecular signatures to predict the risk of disease progression or even further malignant transformation in certain patients morphologically diagnosed as hyperplasia. Therefore, we next examined the possible correlation between the CNA events in the epithelial samples collected from HOL patients and disease progression. We conducted a retrospective analysis in HOL patients who had been diagnosed in our hospital from 2012 through 2015 and possessed a detailed follow-up. Of the 1,449 LCM samples from the 477 HOL patients, 433 LCM samples passed the quality filtering after sequencing, and 219 patients were thus further analyzed (Table S1 in Supporting Information). We divided these patients into three categories based on the phenotype and severity of their 5-year outcomes: free of disease (FD), leukoplakia recurrence (LR), and malignant transformation (MT). FD indicated a group of patients with no white plaque during the follow-up period after removal of the primary lesions; LR represented patients with recurrence of the white plaque in

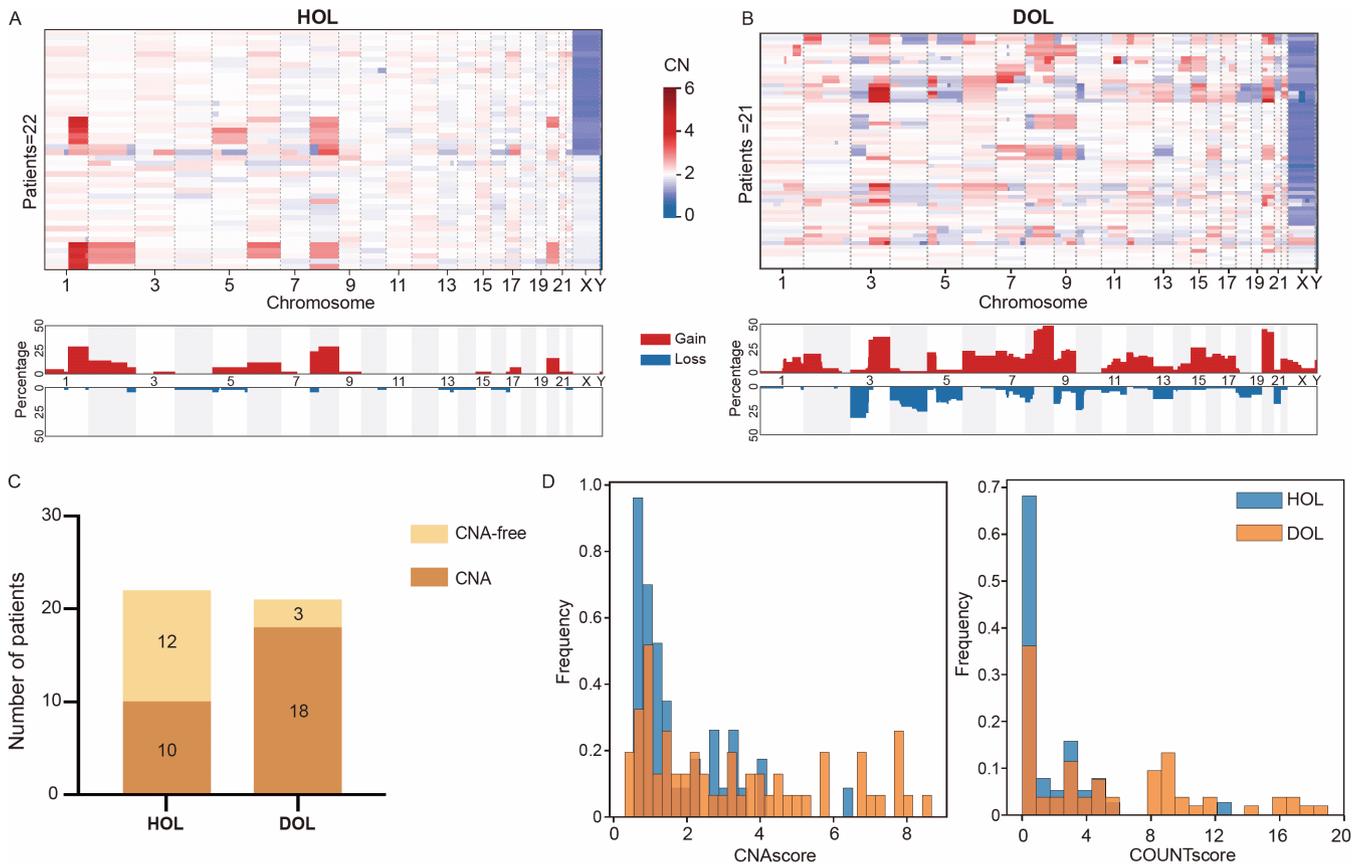


Figure 2 Comparison of copy number alteration between HOL and DOL. A and B, CNA profiles of 43 samples from 22 HOL patients (A) and 61 samples from 21 DOL patients (B) enrolled during 2018–2019. Heatmap (upper panel) and aggregation of CNAs (lower panel) showing the key CNA events in specific chromosomes. In the lower panel, the y-axis indicates the percentage of samples harboring CNAs. C, Proportions of HOL and DOL patients harboring CNAs. D, Normalized frequency distributions of CNAscore (left panel) and COUNTscore (right panel) for HOL and DOL samples.

the same site of the primary lesion; and MT included patients with lesions that showed malignant transformation to OSCC after removal of the primary leukoplakia.

One of the main genomic characteristics of HOL tissues was that most CNA events occurred at the whole-chromosome level. We found that such aneuploidies showed significantly high occurrence rates in Chr 1, Chr 2, Chr 7, Chr 8 and Chr 20 (Figure 3A and B). Notably, the hotspot aneuploidies did not occur at similar rates among the patient groups. Patients in the FD group had a significantly lower incidence than those in the groups with disease progression (LR and MT) (Figure S4A in Supporting Information).

We further observed four major differences in the CNA profiles among the FD, LR, and MT groups (Figure 3A and B). First, although aneuploidy events of Chr 1, Chr 2, Chr 7, Chr 8, and Chr 20 were observed in all groups, they were most frequent in the MT group followed by the LR group, and both were significantly more frequent than those in the FD group (Figure S4A and B in Supporting Information). For example, 50% of the Chr 8 gain events were identified in MT samples, followed by 36% in LR samples, and 11% in FD samples. Second, copy number losses were rare in the FD

group (10.1%), but occurred at a relatively higher rate in the MT (18.2%) and LR (16.5%) groups (Figure S4C in Supporting Information). Third, in addition to the shared aneuploidy hotspots, the MT and LR groups contained more CNA events on many other chromosomes, while the FD group rarely showed CNA events outside of the hot spots (Figure S4B in Supporting Information). Fourth, compared with the LCM samples in the FD group, a number of LCM samples in the LR and MT groups contained breakpoints, similar to the observation in DOL samples. Moreover, we found that 5 out of 76 (7%) LR patients and 2 out of 10 (20%) MT patients harbored the signature Chr 3 breakpoints identical to those seen in DOL cases. Both LR and MT patients harbored Chr 5 breakpoints: 4% in the LR group and 30% in the MT group. Furthermore, three (4%) LR patients showed breakpoints in both Chr 4 and Chr 6, while six (60%) MT patients showed breakpoints in Chr 1 and one (10%) MT patient showed breakpoints in Chr 7, Chr 11, and Chr 12.

Despite the small number of patients in the MT group, all 10 (100%) MT patients had CNAs, while 31 out of 130 (23.8%) FD patients and 43 out of 76 (56.6%) LR patients harbored CNAs ($\chi^2=41.157$, $P<0.0001$). We also used

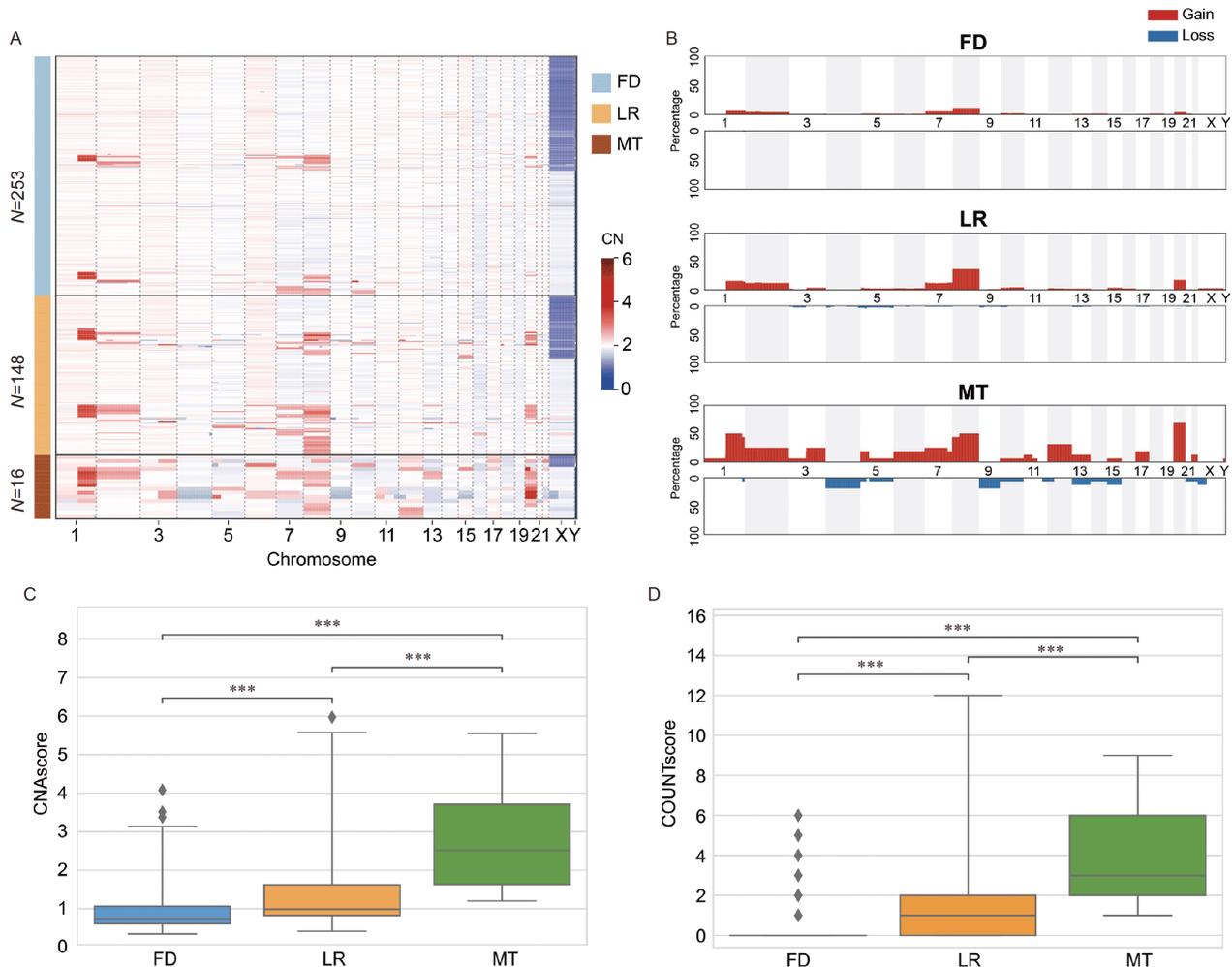


Figure 3 CNAs are associated with disease progression of HOL. A, Heatmap showing the epithelial CNA profiles of HOL patients categorized to three groups (FD, LR and MT) based on 5-year outcomes. B, Accumulation of copy number changes in samples from the three categories. y-axis indicates the percentage of samples harboring CNAs. C and D, Boxplots showing the distributions of CNAScore (C) and COUNTScore (D), indicating an association between the degree of CNAs and disease progression. (*t*-test, ***, $P < 0.001$).

CNAScore and COUNTScore to quantitatively depict the degree of copy number changes of the three groups. Both scores could separate the FD, LR, and MT groups, with MT samples showing the highest CNAScore and COUNTScore followed by LR samples (Figure 3C and D). Notably, both scores of HOL tissue samples in the MT group were comparable with those obtained in the DOL samples (Figure S4D in Supporting Information), which displayed a strong tendency to develop into OSCC. This observation indicates an association between the severity of copy number changes and the development of leukoplakia.

CNA evolution during disease progression

We further examined the CNA profiles of the hyperplastic tissues obtained from the same individuals prior to and following the recurrence of leukoplakia (Figure 4A and B; Figure S5A in Supporting Information). There were four

such patients, and 21 LCM samples passed quality filtering. For each individual, both the CNAScore and COUNTScore were similar for the primary and recurrent leukoplakia tissues (Figure S5B in Supporting Information); however, neither score was sensitive enough for small-sized variations. Such indices could not reveal the finer scale difference in CNA events. Newly occurring small variation features in the recurrent samples, such as Chr 8 gains in patient LR-P509 and Chr 6 breakpoints in patient LR-P258, did not cause a significant change in either score (Figure 4B).

Four of the enrolled HOL patients were later diagnosed with OSCC. We evaluated the CNA profiles of these MT group patients using 26 LCM samples (Figure 4C and D; Figure S5C in Supporting Information). As an intra-individual control, CNA profiling was performed on muscle cells captured from the same section that exhibited no evidence of aneuploidy (Figure S5C in Supporting Information). The hyperplastic epithelial samples collected from

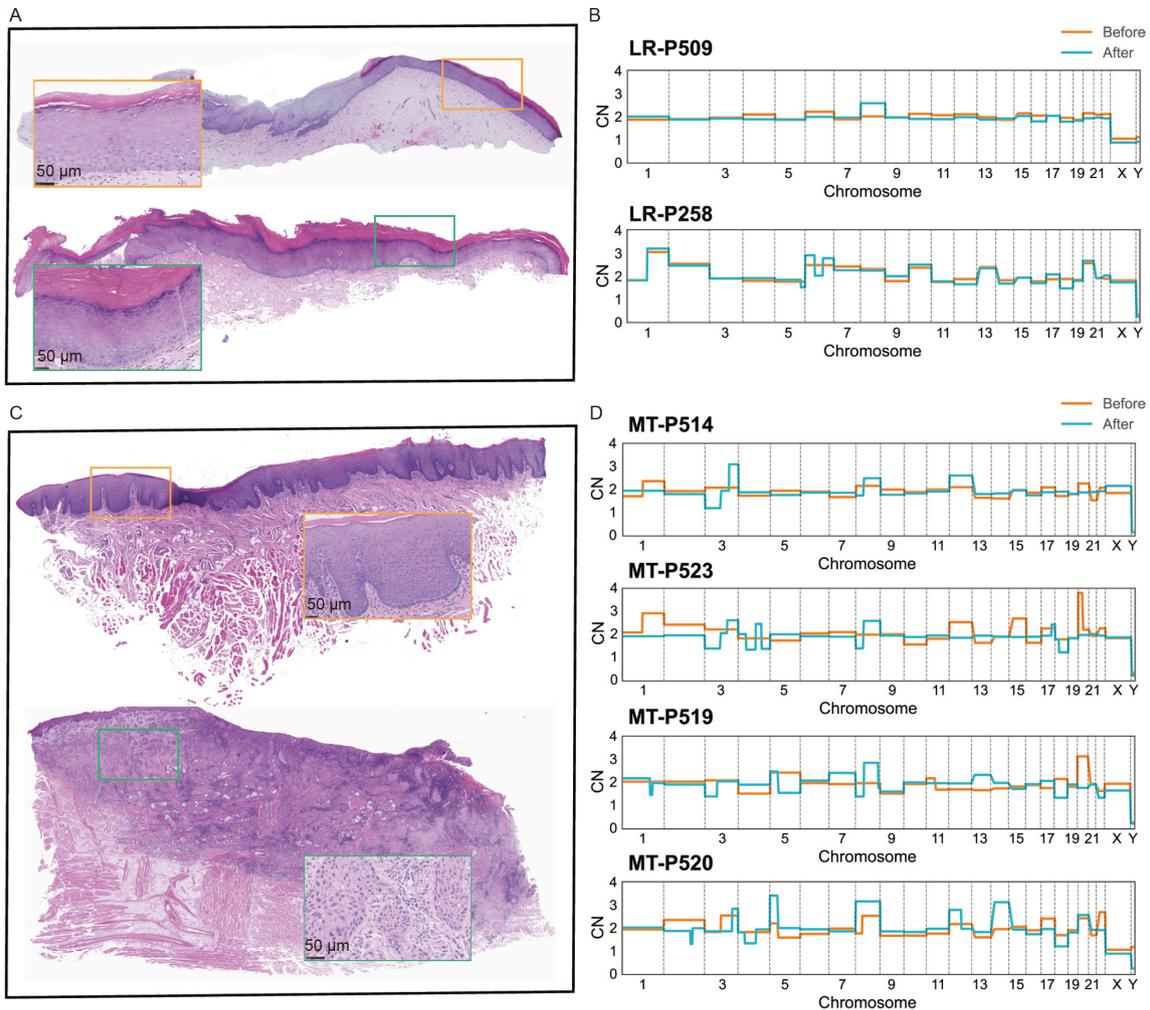


Figure 4 Copy number profiling during disease progression. A, H&E images showing primary epithelial tissues (upper panel) and recurrent epithelial tissues (lower panel) from patient LR-P509. Enlarged images indicate regions isolated by LCM. B, Copy number profiles of primary and recurrent tissues in patients LR-P509 and LR-P258. Copy number in each bin was calculated as the median value of all samples of the same cell type from each patient. C, H&E images showing primary epithelial tissues (upper panel) and tumor tissues following malignant transformation (lower panel) from patient MT-P520. Enlarged images indicate regions isolated by LCM. D, Copy number profiles of primary epithelial tissues and malignant tumor tissues in patients MT-P514, MT-P523, MT-P519 and MT-P520. Copy number in each bin was calculated as the median value of all the samples of the same cell type from each patient.

different locations of the tissue section showed a pattern of abnormal copy numbers (Figure 4D, Figure S5C in Supporting Information), indicating that genome structural variations had developed, and probably clonally expanded, for a period of time without noticeable morphological changes.

Certain location-specific CNA events, such as breakpoints in Chr 3, Chr 4, Chr 5, and Chr 8, were shared among the four OSCC patients. Interestingly, comparison of HOL and OSCC samples from the same individual showed that the CNA profiles were not identical (Figure 4D; Figure S5C in Supporting Information). For example, aneuploidy events of Chr 14 and Chr 18 were identified in tumor samples but absent in HOL samples, and several aneuploidy events in HOL samples were not found in tumor samples. This longitudinal inconsistency in CNA patterns suggests that there may be certain key events of genome reorganization that are tightly associated with tumorigenesis and reflects the pos-

sible selection of specific clones during this process.

Taken together, the association of the CNA profile and the risk of leukoplakia recurrence or malignant transformation provides an additional layer of quantitative information, via the CNAscore and COUNTscore, to the conventional pathological assessment based on morphological features. Such integration of genomic variation data with pathological identification may offer a unique tool to predict the potential outcome in HOL patients.

Delineating spatially clonal genotypes

Using LCM, we mapped the copy number profile of each sample to its original spatial coordinate and illustrated the high-resolution heterogeneity between histologically similar lesions. Mini-bulk sequencing, aided with pathological identification, offered high cellular purity of each sample and

hence would allow for the identification of CNA that only occurred as small clones.

Spatially adjacent sampling of the DOL lesion can also generate different CNA patterns, and we found two such cases among all the DOL patients (Figures 1D and 5A). For example, in patient P36, the normal epithelial tissues showed no evidence of copy number changes (Figure 5A and B); however, four CNA clones were identified in the four adjacent DOL samples (D1, D2, D3, and D4). The four clones shared similar aneuploidy of Chr 8, Chr 9, Chr 10, Chr 16, Chr 19, and Chr 20 and exhibited different features of Chr 3, Chr 6, Chr 7, Chr 13, and Chr 22 (Figure 5B).

A similar phenomenon was also found in HOL patients. For example, two CNA clones were also identified in patient P22 (Figure 5C); one of the clones was formed by morphologically normal cells and the other contained three hyperplastic cell clusters. Two clones shared the same CNAs in Chr 1, Chr 2, and Chr 20, and yet the copy numbers in Chr 6 and Chr 8 were different (Figure 5D). Although we had dissected multiple hyperplastic cell clusters that were scattered throughout the HOL tissue, we observed no intra-tissue heterogeneity from CNA profiling.

These two examples unveil the possible clone expansion of leukoplakia cells. Furthermore, the existence of aneuploidy in morphologically normal cells also agrees with the recent reports on the various types of cancers. Together, our observations suggested that genomic variations may provide an additional facet to help predict the developmental fate of leukoplakia cells.

DISCUSSION

Recent research has demonstrated that genomic changes in many cancers may occur prior to other alterations that can be detected by conventional pathological techniques (Gerstung et al., 2020). Given that chromosome structural variations are considered a hallmark of cancer and were recently found in morphologically noncancerous tissues (Zhou et al., 2020), we believe that such CNAs in morphologically normal cells may provide new insights into the prognosis of OL, the most common premalignant lesion of OSCC.

Previous studies have reported that a variety of CNAs, such as aberrations in Chr 5q, 7p, 7q, and 8p, are involved in the development and malignant transformation of OL (Wood et al., 2017; Bhattacharya et al., 2011; Bhosale et al., 2011; de Boer et al., 2019). However, these studies have two major limitations. First, most of the studies focused on specific genomic chromosomes, and second, these studies used large bulk samples as input for sequencing, which resulted in unavoidable mixing of various cell types.

In the present study, the combination of LCM and mini-bulk sequencing allowed us to link histopathological in-

formation with genomic alterations in OL samples. The copy number profiling of 529 OL patients yielded comprehensive and detailed findings showing that HOL is chromosomally distinct from DOL. Moreover, we identified different CNA clones within relatively close epithelial regions, which are typically obscured by bulk tissue sequencing approaches. Although there is some evidence of possible correlations between DOL and OSCC through common genomic variations (Ho et al., 2013), it remains unclear which variations could serve as markers for diagnosis or prognosis monitoring. Furthermore, because HOL has been conventionally considered a benign lesion with negligible malignant potential, the majority of previously published OL studies have focused on DOL. Therefore, the prediction of outcomes for HOL is even more challenging.

From the perspective of pathology, hyperplasia has long been thought to be a harmless and mild lesion that mainly exhibits epithelial hyperkeratosis and disappearance of nuclei. However, hyperplasia may lead to completely different fates of disease progression, and its malignant potential remains one of the most challenging topics for clinical pathologists. Our study provides a new strategy to quantitatively evaluate the probability of transformation from HOL to OSCC through whole genome CNA landscape assessment using paired HOL and OSCC samples.

Examination of the frequencies of CNA events in hyperplasia and dysplasia samples showed a clear association between genomic changes and pathological grades. In our retrospective study of the correlation between CNA profiles and prognostic outcomes of HOL, we observed that patients with poor prognostic outcomes showed a higher degree of genomic variations. Intriguingly, some morphologically hyperplastic samples with abnormal copy number profiles transformed into OSCC. Importantly, we found that morphologically different HOL and DOL shared many common genomic variations, suggesting that these two types of lesions may share the same causes.

Currently, the prognostic assessment of OL still mainly depends on clinical features and histopathological diagnosis based on atypical epithelial performance (Warnakulasuriya et al., 2008; Karatayli-Ozgursoy et al., 2015). However, the grading systems based only on cytological changes are limited in providing reliable results to predict malignant potential and prognostic outcomes. Because morphological grading is relatively subjective and relies largely on the pathologists' experience, an objective and reliable prognostic indicator of OL is urgently needed (Pitiyage et al., 2009; Rivera et al., 2017). The correlation between CNA occurrence and malignant transformation that we observed through whole-genome shallow sequencing in the present study shows that CNA may be an effective auxiliary indicator for the disease outcome of precancerous lesions. We recommend a novel follow-up program for OL patients via a

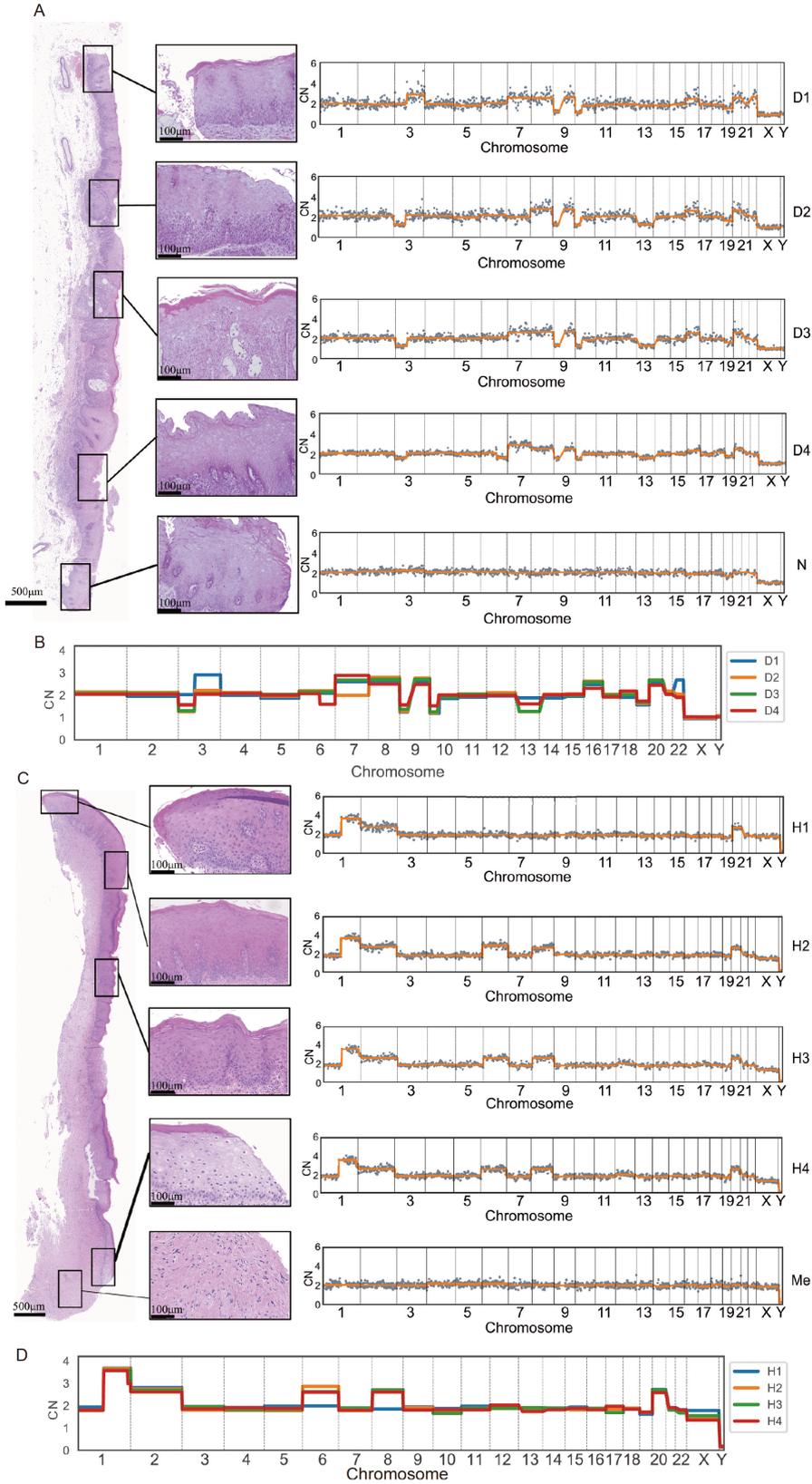


Figure 5 Identification of spatially clonal genotypes. A, CNA profiles (right panel) of four epithelial dysplasia samples (D1, D2, D3, and D4) and one normal epithelial sample (N) isolated from the same tissue slice from DOL patient P36. H&E images (left panel) showing regions isolated by LCM. B, Four CNA subclones were identified in D1, D2, D3, and D4. C, CNA profiles (right panel) of four epithelial samples (H1, H2, H3 and H4) and one mesenchymal sample (Me) isolated from the same tissue slice from HOL patient P22. H&E images (left panel) showing regions isolated by LCM. D, Two CNA subclones were identified in H1, H2, H3, and H4.

combination of morphological analysis and genomic copy number profiling. The morphological HOL patients lacking arm-level CNA tended to have a relatively better prognosis, and therefore regular re-examination is sufficient in these patients. However, HOL patients harboring arm-level CNAs are better treated as DOL patients, with intensive follow-up to monitor possible malignant transformation.

Our findings also have important implications for clinical surgery. One of the key factors of *in situ* recurrence after surgical resection is the incomplete removal of lesion tissues. During traditional surgery, safety margins are largely determined by the subjective judgment of pathologists. However, CNA examination may be a more sensitive method to identify safe margins and evaluate genome alterations as an additional assessment of future surgical risks. Our results indicated that arm-level CAN-containing cells can appear as morphologically normal, suggesting that submicroscopic-level molecular pathology investigation should be performed for many clinical cases. Complete removal of dysplastic regions as well as morphologically normal epithelia with genomic alterations is needed, using quantitative scores as guidelines for the arrangement of marginal resection.

Although all the samples in the present study were collected from a single hospital, we believe that CNA profiling of morphologically normal HOL tissues has general advantages over conventional approaches for risk evaluation. A multi-center study will further improve the reliability of risk assessment and optimize the quantitation of the degree of change in copy number. We conducted our investigation mainly using FFPE samples, which have limitations compared with fresh or frozen samples because of degraded DNA in FFPE samples. Approximately half of the samples did not pass the quality filter for inclusion data analysis. Further improvement should be planned to optimize sequencing library construction using low-quality FFPE samples.

MATERIALS AND METHODS

Sample collection, preparation, and staining

This research was approved by the Institutional Review Board of Peking University Hospital of Stomatology, and written informed consent was obtained from all patients (approval NO. PKUSSIRB-201949116). A total of 1,650 samples were collected from 529 OL cases; all samples were FFPE samples (Table S1 in Supporting Information) retrieved from the Department of Pathology, Peking University Hospital of Stomatology. Among the 529 OL patients, 52 had enrolled in the past two years and therefore did not have 5-year prognosis information. The remaining OL patients had been diagnosed as hyperplasia OL from 2012 to 2015 and possessed detailed follow-up information.

FFPE samples were sectioned into 10- μ m tissue slices

using a tissue microtome (REM710, Yamato, Japan), and placed on PEN-membrane glass slides (Leica, USA). Staining procedures were performed as previously described with slight modifications (Martelotto et al., 2017). Briefly, FFPE slides were soaked in xylene three times for 10 min each and subsequently rehydrated in a series of ethanol immersions for 1 min each (twice in 100% ethanol, followed by 95%, 70%, and 50% ethanol). After washing with DEPC-treated water, slides were subjected to hematoxylin and eosin (H&E) staining by standard protocol. Finally, sequential ethanol solutions were used to dehydrate the samples. Slides were scanned using a digital slice scanning device (NanoZoomer 2.0T).

LCM and cell lysis

Stained slides were micro-dissected using an LCM system (LMD7, Leica). The areas of interest, containing morphologically normal epithelia, hyperplasia, and dysplasia, were captured and reviewed by three independent experienced pathologists. Muscle or mesenchyme tissues distant from the lesion regions were also obtained as germline control. LCM was performed using a 10 \times objective, and the number of cells in each captured tissue sample was maintained at 200–500. Typically, FFPE samples less than two years old displayed better DNA preservation than those over 5 years old, indicating that a greater number of cells may be required for the preparation of sequencing libraries. Captured tissues were lysed in 8 μ L lysis buffer (6.4 μ L nuclease-free water, 0.24 μ L 1 mol L⁻¹ Tris-HCl, 0.16 μ L 500 mmol L⁻¹ NaCl, 0.08 μ L 500 mmol L⁻¹ EDTA, 0.32 μ L 5% Triton and 0.8 μ L Proteinase K) at 50°C for 12 h.

Whole-genome amplification and sequencing

DNA in lysis buffer was tagged using Tn5-transposase (Vazyme, Nanjing, China), amplified by 20 cycles of PCR, and purified using VAHTS DNA clean beads (Vazyme). Approximately 2 μ g final amplified product was obtained for each sample. The concentration of the samples was measured using the Qubit system (Invitrogen, USA) and the fragment size (ranging from 300 to 700 bp) was determined by the Fragment Analyzer (Agilent, USA). Samples that failed in library construction were excluded before sequencing. Libraries with low concentration or incorrect size distribution as determined by the Agilent Fragment Analyzer were identified as unqualified and excluded before sequencing. Libraries were then sequenced on an Illumina HiSeq 4000 sequencer with PE150.

Bioinformatics analysis

We filtered out unqualified samples in wet-lab quality con-

trol. Sequenced samples without sufficient sequencing data (<0.1 M paired-end reads) were excluded before performing bioinformatic analysis. In total, 652 unqualified samples were removed leaving 998 samples for bioinformatic analysis. Adapter trimming was first performed on 2×150 paired-end reads using Cutadapt (version 2.10) (Martin, 2011) under default setting, which were subsequently aligned to the human reference genome (hg19) by Bowtie2 aligner (version 2.2.9) (Langmead and Salzberg, 2012) using default setting. The hg19 genome was downloaded from the UCSC Genome Browser: <http://hgdownload.cse.ucsc.edu/goldenPath/hg19/bigZips/>. Approximately 1 M mapped reads were obtained for each sample. Reads were tabulated into non-overlapping dynamic bins (2 Mb resolution) across the genome. Lowess regression normalization was performed to reduce the GC bias of bin counts. Copy number was called by R package DNACopy (version 1.44.0) (Seshan and Olshen, 2020) using circular binary segmentation algorithm (alpha=0.0001, min.width=5, undo.SD=2).

Statistical analysis

As widely used in previous CNA studies (Kader et al., 2016; Ning et al., 2015), we calculated the MAPD (Affymetrix, 2008) of sampling bins to judge the dispersion of data points which reflects the quality of CNA profiles and filter out low-quality samples (MAPD≥0.25). If x_i is the copy number value of the i th bin, then

$$\text{MAPD} = \text{median}(|x_{i+1} - x_i|),$$

where i is ordered by genomic position. The determination of the MAPD threshold was based on the distribution of MAPD shown in Figure S1B in Supporting Information. Passing-filter samples were also examined manually. Unqualified examples either with low coverage (mapping reads<100 k) or noisy bins (MAPD>0.25) are shown in Figure S1C in Supporting Information.

Because copy number change events covering larger genome regions may influence more genes and different copy number gains or losses may have different dosage effects in the regulation of cell activities, we assume both factors should be considered when estimating CNA severity. We first developed CNAScore to calculate the level of sample genome rearrangement in samples. CNAScore of the j th chromosome was first calculated by

$$\text{CNAScore}_j = \text{mean}(|s_i - s_{i-1}|) + 0.5 * |\text{mean}(s_i) - \text{norm}|,$$

where i is ordered by the genomic position within a specific chromosome, s_i is the segment value at position i , and norm is the neutral copy number of each segment (norm=0, 1, or 2).

The CNAScore of each sample was calculated as the sum of the values of all 24 chromosomes:

$$\text{CNAScore} = \sum_{j=1}^{24} \text{CNAScore}_j.$$

The CNAScore comprises two parts. The first penalizes the variation between adjacent segments, which reflects the level of fluctuation among CNA segments. The second part penalizes deviation of the estimated average ploidy from the neutral state. As a result, longer CNA or CNA with higher copy number would have a bigger influence on ploidy estimation, thus increasing the value of CNAScore. The drawback of CNAScore is the fluctuation of segment values and that ploidy inferring may inevitably be influenced by library quality; that is, among samples sharing the same CNA profiles, those with noisier bins may have higher CNAScores.

COUNTscore calculates the sum of CNA segments in each sample, and it has been used in previous studies to quantify the severity of CNA changes (Davoli et al., 2017; Taylor et al., 2018). For diploid chromosomes, copy number gain is defined as a copy number>2.3, and copy number loss is defined as a copy number<1.6 based on the copy number distribution of all the bins (Figure S6A in Supporting Information). For haploid chromosomes, copy number gain is defined as copy number>1.5, and copy number loss is defined as copy number<0.5. In comparison with CNAScore, COUNTscore is less likely to be affected by sample quality; however, COUNTscore is limited by the copy number cut-offs, and therefore it may underestimate the copy number state of each sample, especially in bulk cases. Thus, we combined both CNAScore and COUNTscore to fully evaluate the genome complexity of each sample. The cut-offs of defining CNA events are shown in Figure S6 in Supporting Information and found the major conclusions were robust to the cut-off selection. A COUNTscore>0 was used to infer samples with CNAs to reduce false positive calling.

Statistical differences between HOL and DOL were determined using Chi-squared (χ^2) test. Other analyses were performed using independent t -test. $P<0.05$ was considered statically significant.

Compliance and ethics The author(s) declare that they have no conflict of interest. This research was approved by Institutional Review Board of Peking University Hospital of Stomatology, and written informed consent was obtained from all patients (approval NO. PKUSSIRB-201949116).

Acknowledgements This work was supported by the National Natural Science Foundation of China (81671006, 81300894, 22050002, 22050004) and Chinese Academy of Medical Sciences Innovation Fund for Medical Sciences (2019-I2M-5-038).

References

- Affymetrix. (2008). Median of the absolute values of all pairwise differences and quality control on affymetrix genome-wide human SNP array 6.0. Available from: URL: http://tools.thermofisher.com/content/sfs/brochures/mapd_snp6_whitepaper.pdf.
- Almangush, A., Leivo, I., and Mäkitie, A.A. (2021). Biomarkers for immunotherapy of oral squamous cell carcinoma: current status and

- challenges. *Front Oncol* 11, 616629.
- Baslan, T., Kendall, J., Rodgers, L., Cox, H., Riggs, M., Stepansky, A., Troge, J., Ravi, K., Esposito, D., Lakshmi, B., et al. (2012). Genome-wide copy number analysis of single cells. *Nat Protoc* 7, 1024–1041.
- Beroukhi, R., Mermel, C.H., Porter, D., Wei, G., Raychaudhuri, S., Donovan, J., Barretina, J., Boehm, J.S., Dobson, J., Urashima, M., et al. (2010). The landscape of somatic copy-number alteration across human cancers. *Nature* 463, 899–905.
- Bosman, F.T. (2001). Dysplasia classification: pathology in disgrace? *J Pathol* 194, 143–144.
- Bhattacharya, A., Roy, R., Snijders, A.M., Hamilton, G., Paquette, J., Tokuyasu, T., Bengtsson, H., Jordan, R.C.K., Olshen, A.B., Pinkel, D., et al. (2011). Two distinct routes to oral cancer differing in genome instability and risk for cervical node metastasis. *Clin Cancer Res* 17, 7024–7034.
- Bhosale, P.G., Cristea, S., Ambatipudi, S., Desai, R.S., Kumar, R., Patil, A., Kane, S., Borges, A.M., Schäffer, A.A., Beerenwinkel, N., et al. (2011). Chromosomal alterations and gene expression changes associated with the progression of leukoplakia to advanced gingivobuccal cancer. *Transl Oncol* 10, 396–409.
- Chai, A.W.Y., Lim, K.P., and Cheong, S.C. (2020). Translational genomics and recent advances in oral squamous cell carcinoma. *Semin Cancer Biol* 61, 71–83.
- Davoli, T., Uno, H., Wooten, E.C., and Elledge, S.J. (2017). Tumor aneuploidy correlates with markers of immune evasion and with reduced response to immunotherapy. *Science* 355, eaaf8399.
- de Boer, D.V., Brink, A., Buijze, M., Stigter-van Walsum, M., Hunter, K. D., Ylstra, B., Bloemena, E., Leemans, C.R., and Brakenhoff, R.H. (2019). Establishment and genetic landscape of precancer cell model systems from the head and neck mucosal lining. *Mol Cancer Res* 17, 120–130.
- Ding, X., Zheng, Y., Wang, Z., Zhang, W., Dong, Y., Chen, W., Li, J., Chu, W., Zhang, W., Zhong, Y., et al. (2018). Expression and oncogenic properties of membranous Notch1 in oral leukoplakia and oral squamous cell carcinoma. *Oncol Rep* 39, 2584–2594.
- Dionne, K.R., Warnakulasuriya, S., Binti Zain, R., and Cheong, S.C. (2014). Potentially malignant disorders of the oral cavity: Current practice and future directions in the clinic and laboratory. *Int J Cancer* 136, 503–515.
- EI-Nagger, A.K., Chan, J.K.C., Grandis, J.R., Takata, T., and Sliotweg, P.J. (2017). WHO Classification of Head and Neck Tumours. 4th ed. Geneva: World Health Organization.
- Gao, Y., Guo, Z.L., Luo, H.Y., and Wang, J. (2012). Clinicopathological characteristics of malignant transformation in 85 cases of oral leukoplakia (in Chinese). *Chin J Stomatol* 47, 410–413.
- Gerstung, M., Jolly, C., Leshchiner, I., Dentre, S.C., Gonzalez, S., Rosebrock, D., Mitchell, T.J., Rubanova, Y., Anur, P., Yu, K., et al. (2020). The evolutionary history of 2,658 cancers. *Nature* 578, 122–128.
- Gissi, D.B., Gabusi, A., Servidio, D., Cervellati, F., and Montebugni, L. (2015). Predictive role of p53 protein as a single marker or associated with ki67 antigen in oral leukoplakia: a retrospective longitudinal study. *Open Dent J* 9, 41–45.
- Gomes, C.C., Fonseca-Silva, T., Galvão, C.F., Friedman, E., De Marco, L., and Gomez, R.S. (2015). Inter- and intra-lesional molecular heterogeneity of oral leukoplakia. *Oral Oncol* 51, 178–181.
- Goodson, M.L., and Thomson, P.J. (2011). Management of oral carcinoma: Benefits of early precancerous intervention. *Br J Oral Max Surg* 49, 88–91.
- Haddad, R.I., and Shin, D.M. (2008). Recent advances in head and neck cancer. *N Engl J Med* 359, 1143–1154.
- Holmstrup, P., Vedtofte, P., Reibel, J., and Stoltze, K. (2007). Oral premalignant lesions: is a biopsy reliable? *J Oral Pathol Med* 36, 262–266.
- Ho, M.W., Field, E.A., Field, J.K., Risk, J.M., Rajlawat, B.P., Rogers, S.N., Steele, J.C., Triantafyllou, A., Woolgar, J.A., Lowe, D., et al. (2013). Outcomes of oral squamous cell carcinoma arising from oral epithelial dysplasia: rationale for monitoring premalignant oral lesions in a multidisciplinary clinic. *Br J Oral Maxillofac Surg* 51, 594–599.
- Jemal, A., Clegg, L.X., Ward, E., Ries, L.A.G., Wu, X., Jamison, P.M., Wingo, P.A., Howe, H.L., Anderson, R.N., and Edwards, B.K. (2004). Annual report to the nation on the status of cancer, 1975–2001, with a special feature regarding survival. *Cancer* 101, 3–27.
- Jemal, A., Siegel, R., Xu, J., and Ward, E. (2010). Cancer statistics, 2010. *CA Cancer J Clin* 60, 277–300.
- Kader, T., Goode, D.L., Wong, S.Q., Connaughton, J., Rowley, S.M., Devereux, L., Byrne, D., Fox, S.B., Mir Arnau, G., Tothill, R.W., et al. (2016). Copy number analysis by low coverage whole genome sequencing using ultra low-input DNA from formalin-fixed paraffin embedded tumor tissue. *Genome Med* 8, 121.
- Karatayli-Ozgunsoy, S., Pacheco-Lopez, P., Hillel, A.T., Best, S.R., Bishop, J.A., and Akst, L.M. (2015). Laryngeal dysplasia, demographics, and treatment: a single-institution, 20-year review. *JAMA Otolaryngol Head Neck Surg* 141, 313.
- Killcoyne, S., Gregson, E., Wedge, D.C., Woodcock, D.J., Eldridge, M.D., de la Rue, R., Miremadi, A., Abbas, S., Blasko, A., Kosmidou, C., et al. (2020). Genomic copy number predicts esophageal cancer years before transformation. *Nat Med* 26, 1726–1732.
- Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9, 357–359.
- Lee, J.J., Hong, W.K., Hittelman, W.N., Mao, L., Lotan, R., Shin, D.M., Benner, S.E., Xu, X.C., Lee, J.S., Papadimitrakopoulou, V.M., et al. (2000). Predicting cancer development in oral leukoplakia: ten years of translational research. *Clin Cancer Res* 6, 1702–1710.
- Lindemann, A., Takahashi, H., Patel, A.A., Osman, A.A., and Myers, J.N. (2018). Targeting the DNA damage response in OSCC with TP53 mutations. *J Dent Res* 97, 635–644.
- Lodi, G., Franchini, R., Warnakulasuriya, S., Varoni, E.M., Sardella, A., Kerr, A.R., Carrassi, A., MacDonald, L.C.I., and Worthington, H.V. (2016). Interventions for treating oral leukoplakia to prevent oral cancer. *Cochrane Database Syst Rev* 7, 1–72.
- Mallory, X.F., Edrisi, M., Navin, N., and Nakhleh, L. (2020). Methods for copy number aberration detection from single-cell DNA-sequencing data. *Genome Biol* 21, 208.
- Martelotto, L.G., Baslan, T., Kendall, J., Geyer, F.C., Burke, K.A., Spraggon, L., Piscuoglio, S., Chadalavada, K., Nanjangud, G., Ng, C.K. Y., et al. (2017). Whole-genome single-cell copy number profiling from formalin-fixed paraffin-embedded samples. *Nat Med* 23, 376–385.
- Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet journal* 17, 10.
- Mehanna, H.M., Rattay, T., Smith, J., and McConkey, C.C. (2009). Treatment and follow-up of oral dysplasia—A systematic review and meta-analysis. *Head Neck* 31, 1600–1609.
- Mello, F.W., Melo, G., Guerra, E.N.S., Warnakulasuriya, S., Garnis, C., and Rivero, E.R.C. (2020). Oral potentially malignant disorders: A scoping review of prognostic biomarkers. *Crit Rev Oncol/Hematol* 153, 102986.
- Monteiro, L., Barbieri, C., Warnakulasuriya, S., Martins, M., Salazar, F., Pacheco, J.J., Vescovi, P., and Meleti, M. (2017). Type of surgical treatment and recurrence of oral leukoplakia: a retrospective clinical study. *Med Oral* 22, 217–223.
- Ng, J.H., Iyer, N.G., Tan, M.H., and Edgren, G. (2017). Changing epidemiology of oral squamous cell carcinoma of the tongue: a global study. *Head Neck* 39, 297–304.
- Ning, L., Li, Z., Wang, G., Hu, W., Hou, Q., Tong, Y., Zhang, M., Chen, Y., Qin, L., Chen, X., et al. (2015). Quantitative assessment of single-cell whole genome amplification methods for detecting copy number variation using hippocampal neurons. *Sci Rep* 5, 11415.
- Olshen, A.B., Venkatraman, E.S., Lucito, R., and Wigler, M. (2004). Circular binary segmentation for the analysis of array-based DNA copy number data. *Biostatistics* 5, 557–572.
- Pitiyage, G., Tilakaratne, W.M., Tavassoli, M., and Warnakulasuriya, S. (2009). Molecular markers in oral epithelial dysplasia: review. *J Oral Pathol Med* 38, 737–752.
- Rivera, C., Oliveira, A.K., Costa, R.A.P., De Rossi, T., and Paes Leme, A.F.

- (2017). Prognostic biomarkers in oral squamous cell carcinoma: a systematic review. *Oral Oncol* 72, 38–47.
- Saintigny, P., Zhang, L., Fan, Y.H., El-Naggar, A.K., Papadimitrakopoulou, V.A., Feng, L., Lee, J.J., Kim, E.S., Ki Hong, W., and Mao, L. (2011). Gene expression profiling predicts the development of oral cancer. *Cancer Prev Res* 4, 218–229.
- Seshan, V.E., and Olshen, A.B. (2020). DNACopy: DNA copy number data analysis R package Version 1.64.0.
- Staines, K., and Rogers, H. (2017). Oral leukoplakia and proliferative verrucous leukoplakia: a review for dental practitioners. *Br Dent J* 223, 655–661.
- Taylor, A.M., Shih, J., Ha, G., Gao, G.F., Zhang, X., Berger, A.C., Schumacher, S.E., Wang, C., Hu, H., Liu, J., et al. (2018). Genomic and functional approaches to understanding cancer aneuploidy. *Cancer Cell* 33, 676–689.e3.
- Tsao, A.S., Liu, D., Martin, J., Tang, X., Lee, J.J., El-Naggar, A.K., Wistuba, I., Culotta, K.S., Mao, L., Gillenwater, A., et al. (2009). Phase II randomized, placebo-controlled trial of green tea extract in patients with high-risk oral premalignant lesions. *Cancer Prev Res* 2, 931–941.
- Türke, C., Horn, S., Petto, C., Labudde, D., Lauer, G., and Wittenburg, G. (2017). Loss of heterozygosity in FANCG, FANCF and BRIP1 from head and neck squamous cell carcinoma of the oral cavity. *Int J Oncol* 50, 2207–2220.
- Warnakulasuriya, S., and Ariyawardana, A. (2016). Malignant transformation of oral leukoplakia: a systematic review of observational studies. *J Oral Pathol Med* 45, 155–166.
- Warnakulasuriya, S., Reibel, J., Bouquot, J., and Dabelsteen, E. (2008). Oral epithelial dysplasia classification systems: predictive value, utility, weaknesses and scope for improvement. *J Oral Pathol Med* 37, 127–133.
- Warnakulasuriya, S., Kovacevic, T., Madden, P., Coupland, V.H., Sperandio, M., Odell, E., and Møller, H. (2011). Factors predicting malignant transformation in oral potentially malignant disorders among patients accrued over a 10-year period in South East England. *J Oral Pathol Med* 40, 677–683.
- Wood, H.M., Daly, C., Chalkley, R., Senguven, B., Ross, L., Egan, P., Chengot, P., Graham, J., Sethi, N., Ong, T.K., et al. (2017). The genomic road to invasion—examining the similarities and differences in the genomes of associated oral pre-cancer and cancer samples. *Genome Med* 9, 53.
- Yagyuu, T., Hatakeyama, K., Imada, M., Kurihara, M., Matsusue, Y., Yamamoto, K., Obayashi, C., and Kirita, T. (2017). Programmed death ligand 1 (PD-L1) expression and tumor microenvironment: Implications for patients with oral precancerous lesions. *Oral Oncol* 68, 36–43.
- Yanik, E.L., Katki, H.A., Silverberg, M.J., Manos, M.M., Engels, E.A., and Chaturvedi, A.K. (2015). Leukoplakia, oral cavity cancer risk, and cancer survival in the U.S. elderly. *Cancer Prev Res* 8, 857–863.
- Zhang, L., Poh, C.F., Williams, M., Laronde, D.M., Berean, K., Gardner, P. J., Jiang, H., Wu, L., Lee, J.J., and Rosin, M.P. (2012). Loss of heterozygosity (LOH) profiles—Validated risk predictors for progression to oral cancer. *Cancer Prev Res* 5, 1081–1089.
- Zhang, X., Kim, K.Y., Zheng, Z., Bazarsad, S., and Kim, J. (2017). Nomogram for risk prediction of malignant transformation in oral leukoplakia patients using combined biomarkers. *Oral Oncol* 72, 132–139.
- Zhou, J., Cao, J., Lu, Z., Liu, H., and Deng, D. (2011). A 115-bp MethyLight assay for detection of p16 (CDKN2A) methylation as a diagnostic biomarker in human tissues. *BMC Med Genet* 12, 67.
- Zhou, Y., Bian, S., Zhou, X., Cui, Y., Wang, W., Wen, L., Guo, L., Fu, W., and Tang, F. (2020). Single-cell multiomics sequencing reveals prevalent genomic alterations in tumor stromal cells of human colorectal cancer. *Cancer Cell* 38, 818–828.e5.

SUPPORTING INFORMATION

The supporting information is available online at <https://doi.org/10.1007/s11427-021-1965-x>. The supporting materials are published as submitted, without typesetting or editing. The responsibility for scientific accuracy and content remains entirely with the authors.