Received: 15 January 2022 | Revised: 13 May 2022 | Accepted: 20 May 2022

DOI: 10.1002/gcc.23076

RESEARCH ARTICLE

WILEY

Whole genome duplication in oral squamous cell carcinoma in patients younger than 50 years: implications for prognosis and adverse clinicopathological factors

Laveniya Satgunaseelan[1,2] | Dario Strbenac[3] | Cali Willet[4] | Tracy Chew[4] | Rosemarie Sadsad[4] | James Wykes[2,5] | Tsu-Hui (Hubert) Low[2,5] | Wendy A. Cooper[1,2] | C. Soon Lee[1,2,7,8,9] | Carsten E. Palme[2,5] | Jean Y. H. Yang[3,10] | Jonathan R. Clark[2,5,6] | Ruta Gupta[1,2,5]

[1]Department of Tissue Pathology and Diagnostic Oncology, NSW Health Pathology, Royal Prince Alfred Hospital, Sydney, NSW, Australia

[2]Sydney Medical School, Faculty of Medicine and Health Sciences, The University of Sydney, Sydney, NSW, Australia

[3]School of Mathematics and Statistics, The University of Sydney, Sydney, NSW, Australia

[4]The Sydney Informatics Hub, Core Research Facilities, The University of Sydney, Sydney, NSW, Australia

[5]Sydney Head and Neck Cancer Institute, Department of Head and Neck Surgery, Chris O'Brien Lifehouse, Sydney, NSW, Australia

[6]Royal Prince Alfred Institute of Academic Surgery, Sydney Local Health District, Sydney, NSW, Australia

[7]Discipline of Pathology, School of Medicine, Western Sydney University, Penrith, NSW, Australia

[8]Genomics & Molecular Pathology Laboratory, Department of Anatomical Pathology, Liverpool, NSW, Australia

[9]South Western Sydney Clinical School, University of New South Wales, Kensington, NSW, Australia

[10]Charles Perkins Centre, The University of Sydney, Sydney, NSW, Australia

Correspondence
Ruta Gupta, Anatomical Pathologist, Department of Tissue Pathology and Diagnostic Oncology, Royal Prince Alfred Hospital, Building 94, Johns Hopkins Drive via Missenden Road, Camperdown, NSW 2050. Australia.
Email: ruta.gupta@health.nsw.gov.au

Funding information
Neal Wald Family; Cancer Institute New South Wales, Grant/Award Number: CINSW2020/2081

Abstract

Introduction: Oral squamous cell carcinoma (OSCC) in the young (<50 years), without known carcinogenic risk factors, is on the rise globally. Whole genome duplication (WGD) has been shown to occur at higher rates in cancers without an identifiable carcinogenic agent. We aimed to evaluate the prevalence of WGD in a cohort of OSCC patients under the age of 50 years.

Methods: Whole genome sequencing (WGS) was performed on 28 OSCC patients from the Sydney Head and Neck Cancer Institute (SHNCI) biobank. An additional nine cases were obtained from The Cancer Genome Atlas (TCGA).

Results: WGD was seen in 27 of 37 (73%) cases. Non-synonymous, somatic *TP53* mutations occurred in 25 of 27 (93%) cases of WGD and were predicted to precede WGD in 21 (77%). WGD was significantly associated with larger tumor size ($p = 0.01$) and was frequent in patients with recurrences (87%, $p = 0.36$). Overall survival was significantly worse in those with WGD ($p = 0.05$).

Conclusions: Our data, based on one of the largest WGS datasets of young patients with OSCC, demonstrates a high frequency of WGD and its association with adverse

pathologic characteristics and clinical outcomes. *TP53* mutations also preceded WGD, as has been described in other tumors without a clear mutagenic driver.

1 | INTRODUCTION

Alterations in ploidy are common occurrences in the evolution of nearly all cancer types.[1] Whole genome duplication (WGD), or tetraploidisation, occurs when the entire complement of diploid chromosomes is doubled.[2] WGD is thought to be a precursor event to aneuploidy and further chromosomal instability with increased somatic copy number alterations.[3,4] The tetraploidy that develops following WGD is more tolerant of chromosomal loss as compared with diploidy and thus permits development of aneuploidy more easily.[5] The incidence of WGD is highly variable amongst the various cancer types and ranges from 5% in non-Hodgkin lymphomas and gastrointestinal neuroendocrine tumors to 58% in germ cell tumors.[2,6] WGD can occur at an early stage in tumor development and transitions cells from a pre-malignant to malignant state, as has been described in lung, oesophageal, and cervical adenocarcinoma.[4,7,8] In some tumors, it can occur at later stages and may only be seen in some subclonal populations, as documented in renal cell carcinoma,[9] myeloma,[10] and melanoma.[11]

The timing of the gatekeeper Tumor Protein 53 (*TP53*) mutations and onset of WGD during carcinogenesis have also been studied and *TP53* mutations have been found to precede WGD in tumors that lack a clear mutagenic driver such as germ cell tumors.[2] WGD may provide a survival advantage to the cancer cell[12] and is an adverse prognostic factor in several malignancies.[2] Recently, a study of WGD+ and WGD- breast cancer cell lines treated with drugs that induce replication stress demonstrated reduced viability in the WGD+ cell lines as compared to their WGD- counterparts suggesting that WGD status may guide treatment strategies in the future.[13]

In head and neck squamous cell carcinoma (HNSCC), WGD has been reported in approximately 25% of recurrent and/or metastatic cancers.[14] HNSCC encompasses a heterogeneous range of carcinomas with varying aetiology and biological behavior depending on anatomic location including oral cavity squamous cell carcinoma (OSCC), human papillomavirus (HPV)-associated oropharyngeal squamous cell carcinoma, and solar damage associated cutaneous squamous cell carcinoma.[15,16] OSCC has traditionally affected elderly patients with a history of tobacco use,[17–19] with a SEER database study of OSCC citing a median age of 61 years at diagnosis.[20] However, in the past decade, a disturbing rise in the incidence of OSCC has been observed in a much younger population both in the Asia-Pacific region as well as the northern hemisphere.[20–22] These patients lack traditional lifestyle risk factors, including smoking and HPV infection.[23–25]

The prevalence of WGD has not been explored in patients developing OSCC at a younger age, with lower exposure to smoking associated carcinogens.[26–28] The temporal relationship of WGD with *TP53* mutation has also not been evaluated in OSCC. This may be due to the lack of samples from young patients with OSCC in international databases such as The Cancer Genome Atlas (TCGA) which predominantly include conventional smoking or HPV-associated HNSCC[29].

The primary aim of this study is to evaluate the prevalence of WGD in patients developing HPV independent OSCC at a younger age using whole genome sequencing (WGS), along with validation through orthogonal testing using clinically validated *in situ* hybridization techniques. The temporal relationship of *TP53* mutations to WGD has also been examined. The secondary aim is to evaluate the association of WGD with clinicopathologic prognostic factors and patient outcomes.

2 | MATERIALS AND METHODS

2.1 | Cohort and tissue selection

The median age at diagnosis of OSCC is 61 years based on SEER data. Schmidt et al. describe a 49.4 pack year smoking history in 67% patients with OSCC.[20,30,31] Thus, for this study, we defined younger age with low exposure to conventional risk factors as patients with diagnosis of OSCC earlier than 50 years of age. Following institutional human research ethics committee approval (Sydney Local Health District Human Research Ethics Committee X19-0282), patients treated with curative intent between 2012 and 2018 were identified from the prospectively collected Sydney Head and Neck Cancer Institute (SHNCI) biobank. Inclusion criteria were: (a) patient age under 50 years at the time of diagnosis of HPV independent OSCC; (b) anatomical location in the oral cavity (buccal mucosa, upper and lower alveolar ridges, retromolar trigone, floor of mouth, hard palate mucosa and anterior two-thirds of the tongue); (c) availability of appropriate samples, including tumor with >30% cellularity on histologic evaluation and matched non- tumor tissue, and computational analyses demonstrating tumor purity ranging from 31–90%; (d) clinicopathological parameters, including clinical follow-up; (e) DNA of sufficient quantity and quality.

Clinical details were obtained, including age, gender, smoking history, and details of adjuvant therapy. Data obtained regarding alcohol consumption was found to be unreliable and was therefore excluded. The histopathology material was reviewed for evaluation of adverse histopathological characteristics, including depth of invasion,[32] perineural invasion,[33] pattern of invasion[32] lymphovascular involvement, nodal involvement (Figure S1) and American Joint Commission on Cancer Staging 8th edition staging categories.[34]

2.2 | Nucleic acid extraction

DNA was extracted from fresh frozen OSCC tissue (tumor) and oral cavity mucosa (matched normal). DNA extraction was performed using the Qiagen AllPrep extraction kit (Qiagen, Germantown, MD, USA), as per the manufacturer's instructions. Library preparations were processed using the TruSeq Nano preparation kit (Illumina, San Diego, CA, USA).

2.3 | Quality control

The quality and quantity of DNA were assessed using Qubit V2.0 (Invitrogen, Carlsbad, CA, USA); spectrophotometric methods (Nanodrop, Thermo Scientific, Waltham, MA, USA; Epoch MicroPlate Spectrophotometer, BioTek Instruments, Winooski, VT, USA), Quant-iT PicoGreen dsDNA assay (ThermoFisher, Waltham, MA, USA), and 0.8% agarose gel electrophoresis. DNA concentration of the SHNCI samples ranged between 35 and 190 ng/μl, with 350 to 1000 ng submitted for whole genome sequencing (WGS).

2.4 | Nucleic acid sequencing

Tumoral and normal DNA were submitted for WGS at a target coverage of 60× and 30×, respectively. DNA sequencing was undertaken on the Illumina NovaSeq 6000 platform (NovaSeq Xp platform). The sequencing was paired-end and each read was 150 base pairs in length.

2.5 | Case selection from The Cancer Genome Atlas (TCGA)

The library of head and neck cancers (referred to as 'HNSC' in TCGA nomenclature) was accessed via the Genomic Data Commons (GDC) Portal (https://portal.gdc.cancer.gov/projects/TCGA-HNSC; dbGaP Study Accession #20551, accessed 7 October 2019). Cases below the age of 50 years with HPV independent OSCC were initially accessed from the library. The cases were filtered according to availability of WGS data of coverage at least 30X and the above inclusion criteria. Thus, nine cases from TCGA cohort were included with computed tumor purity of 38–67%. Clinical features including age, smoking status where available, and mortality status, were also obtained from the GDC portal.

2.6 | Short read alignment and short variant calling

FASTQ files for both the SHNCI and TCGA cohort were processed using the same alignment (BWA-kit version 0.7.17) and variant calling tools (Mutect2) to allow comparisons. BWA-MEM read aligner[35] was used to align reads to the hg38 reference genome and its alternate contigs. The read alignments were refined using Genome Analysis Tool Kit (GATK) 4.1.2.0[36] and its BaseRecalibrator tool. Single nucleotide polymorphisms (SNPs) and insertion–deletion (indel) variants were called by implementing GATK's Best Practices Workflow for "Germline short variant discovery (SNPs + Indels)" and "Somatic short variant discovery (somatic SNVs + Indels)". These workflows use HaplotypeCaller to identify germline variants[37,38] and Mutect2 for somatic variants. Annotation of variants was performed via ENSEMBL Variant Effect Predictor version 99.2 (default settings were used, with the exception of pick where one effect per variant was selected using established criteria[39]). Detailed descriptions of preprocessing and variant discovery are available on the Sydney Informatics Hub Github repository.[40–42]

2.7 | Purity and ploidy

The purity and ploidy of each tumor sample's genome were inferred using the AMBER-COBALT-PURPLE pipeline stipulated in HMF Tools.[43,44] A summary file was created by PURPLE (PURity and Ploidy Estimator, version 2.4.1) for each tumor sample, in which WGD was described as 'true' or 'false'. If the major allele ploidy surpassed 1.5 for at least half of the bases in at least 11 of 22 autosomes, a value of 'true' for WGD was returned. PURPLE determines that if a sufficient number of independent chromosomes is duplicated, the duplication event is likely to have occurred in a single genome-wide event.

Where the copy number estimate was a non-integer, the copy number was rounded to the closest integer. A deletion was defined as a copy number of 0 or 1. High-level amplification was defined as a copy number of 6 or greater.

2.8 | Mutation timing

PURPLE estimates every somatic SNV's purity-adjusted variant and minor allele ploidies. If the purity-adjusted ploidy of a variant (PURPLE_PLOIDY value) is more than 1.5, the variant is considered more likely to occur before WGD. Alternatively, if the purity-adjusted minor allele ploidy (PURPLE_MAP value) surrounding a variant is less than 1.5 and the copy number of the surrounding region is greater than one, the variant most likely happened after WGD. For other values, the timing is indeterminate.

2.9 | Fluorescent in situ hybridization (FISH)

Orthogonal testing with FISH was undertaken to confirm the ploidy status reported by the bioinformatics pipeline, using a tissue microarray (TMA) of the SHNCI OSCC cases. In an accredited diagnostic laboratory setting, interphase FISH for centromeric enumeration probes for chromosomes seven (Vysis *EGFR*/CEP7 Probe Kit, Abbott, Wiesbaden, Germany), 12 (ZytoLight SPEC *MDM2*/CEN12 Dual Colour Probe, ZytoVision, Bremerhaven, Germany), and 17 (PathVysion *HER2*

TABLE 1 Clinicopathological data for (A) SHNCI[a] cohort ($n = 28$), (B) TCGA[e] cohort ($n = 9$)

Variable	n	%
Whole genomic duplication (WGD)		
WGD[b]+	19	68
WGD[b]−	9	32
Age		
≥40, ≤50	14	50
<40	14	50
Sex		
Male	15	54
Female	13	46
Smoking		
Yes	10	36
No	18	64
Anatomical subsite		
Tongue	21	75
Floor of mouth	4	13
Buccal	1	4
Alveolar crest	1	4
Hard palate	1	4
Depth of invasion		
<5 mm	4	14
5–10 mm	11	39
>10 mm	13	47
Perineural invasion		
Yes	16	57
Lymphovascular invasion		
Yes	13	46
Pathological T[c] category		
pT1	4	14
pT2	12	43
pT3	7	25
pT4	5	18
Pathological N[d] category		
pN0	15	46
pN1	2	7
pN2	8	29
pN3	3	11
Treatment[f]		
Surgery alone	12	42
Surgery + radiotherapy	4	14
Surgery + radio/chemotherapy	9	32
Recurrence		
Locoregional recurrence	8	29
Disease related death	7	25
Whole genomic duplication		
WGD[b]+	8	89
WGD[b]−	1	11
Age		
≥40, ≤50	3	33
<40	6	66
Sex		
Male	7	78
Female	2	22
Smoking		
Yes	4	44
History not available	5	56
Death[g]	5	56

[a]Sydney Head and Neck Cancer Institute.
[b]Whole genome duplication.
[c]T = tumor.
[d]N = lymph node involvement.
[e]The Cancer Genome Atlas.
[f]Adjuvant treatment records were unavailable for three patients.
[g]Cause of death not specified in TCGA.

DNA Probe Kit, Abbott, Wiesbaden, Germany), was performed. Briefly, unstained 4 μm formalin-fixed paraffin-embedded tissue sections were deparaffinized, followed by heat-induced epitope retrieval (HIER) pretreatment using a pH 7 buffered solution (SPoT-Light Tissue Pretreatment Solution, Thermo Scientific, Waltham, MA, USA). Proteolytic digestion of tissue sections using Protease 1 (Abbott Molecular, Des Plaines, IL, USA) was then undertaken, after which a saline-sodium citrate (SSC) buffer rinsing step was performed. Application of the aforementioned probes, with probe denaturation of the target chromosomes and gene for 5 min at 95°C and overnight hybridization at 37°C was completed. After hybridization, unbound probe was removed using 2xSSC/0.3% NP-40 solution. The FISH slides were dehydrated, and counterstained with 4,6-diamidino-phenyldinol (SlowFade™ Gold DAPI, Invitrogen, Waltham, MA, USA). Interphase signals were enumerated in at least 50 tumor cell nuclei for the TMA using an epifluorescence microscope (Zeiss, Dublin, CA, USA).

2.10 | Statistical analysis

Categorical data was analyzed using the Fisher's exact test and the Wilcoxon rank-sum test for numerical variables. Survival was calculated using the Kaplan Meier method and comparisons were made using the log-rank test. Overall survival was calculated from the date of surgery to the date of death. Disease-specific survival and disease-free survival were not analyzed due to lack of relevant datapoints (including recurrence and cause of death) in the TCGA dataset.

3 | RESULTS

The study includes 37 patients, including 28 from the SHNCI and nine from TCGA. The median genome coverage for the SHNCI tumor

samples was 64× (range 54–84×) and 35× (range 27–42×) for the matched normal samples. For the nine cases in the TCGA cohort, the median genome coverage was 75× (range 34–85×) for the tumor tissue and 40× for (range 30–49×) for normal tissue.

The clinicopathologic characteristics are summarized in Table 1. The SHNCI cohort (Table 1A) includes 15 males and 13 females with a median age of 41 years (range 19–50 years) at the time of diagnosis. None of the cohort had a prior history of cancer or previous radiotherapy. The median follow up was 16.5 months (3–175 months). Eight patients developed locoregional recurrence and six died from OSCC during this period.

For the cases obtained from TCGA (Table 1B), there were seven men and two women with a median age of 34 years (range 19–49 years). History regarding clinicopathological factors and prior oncologic treatment if any were not available for this cohort. Data regarding recurrence was not available for TCGA cohort. Five of nine patients were deceased at 11.6 months.

3.1 | WGD is common in OSCC patients <50 years

The FASTQ files of both the SHNCI and TCGA cohorts were processed together and the genomic profiles of all young patients, from both the SHNCI and TCGA cohorts, were examined together (Figure S2). WGD was seen in 27 patients (73%) (Figure 1A). The presence of WGD was further validated in the SHNCI cohort, for whom tissue was available for orthogonal FISH testing. FISH demonstrated concordant results with the computational analyses and a mean ploidy for chromosomes 7, 12, and 17 of 3.55, 2.78, and 2.65, respectively, was observed in those with WGD (Figure 1A–D). For patients without WGD, the mean ploidy for chromosomes 7, 12, and 17 was 1.8, 1.7, and 1.4 (Figure 1E, F, G, H).

Gender did not differ significantly between those with WGD (17 males, 10 females) and those without WGD (four males, six females) ($p = 0.27$). Smoking status also was not significantly different between those with WGD (eight smokers of 27 patients), and those without WGD (four smokers of 10 patients) ($p = 0.45$). Mutation signature analysis was performed (Figure S3). No evidence of Catalogue of Somatic Mutations in Cancer (COSMIC) smoking signatures (single base substitution signature [SBS] 4 and SBS 92) was seen in any of the cohort, regardless of WGD status.

3.2 | The majority of WGD patients had pathogenic non-synonymous mutation in TP53 in its DNA binding domain preceding the WGD

Somatic single nucleotide variants (SNVs) in *TP53* were seen in 32 patients (84%) overall, including 17 missense, three frameshift, three splicesite, three nonsense variants, and six cases with multiple

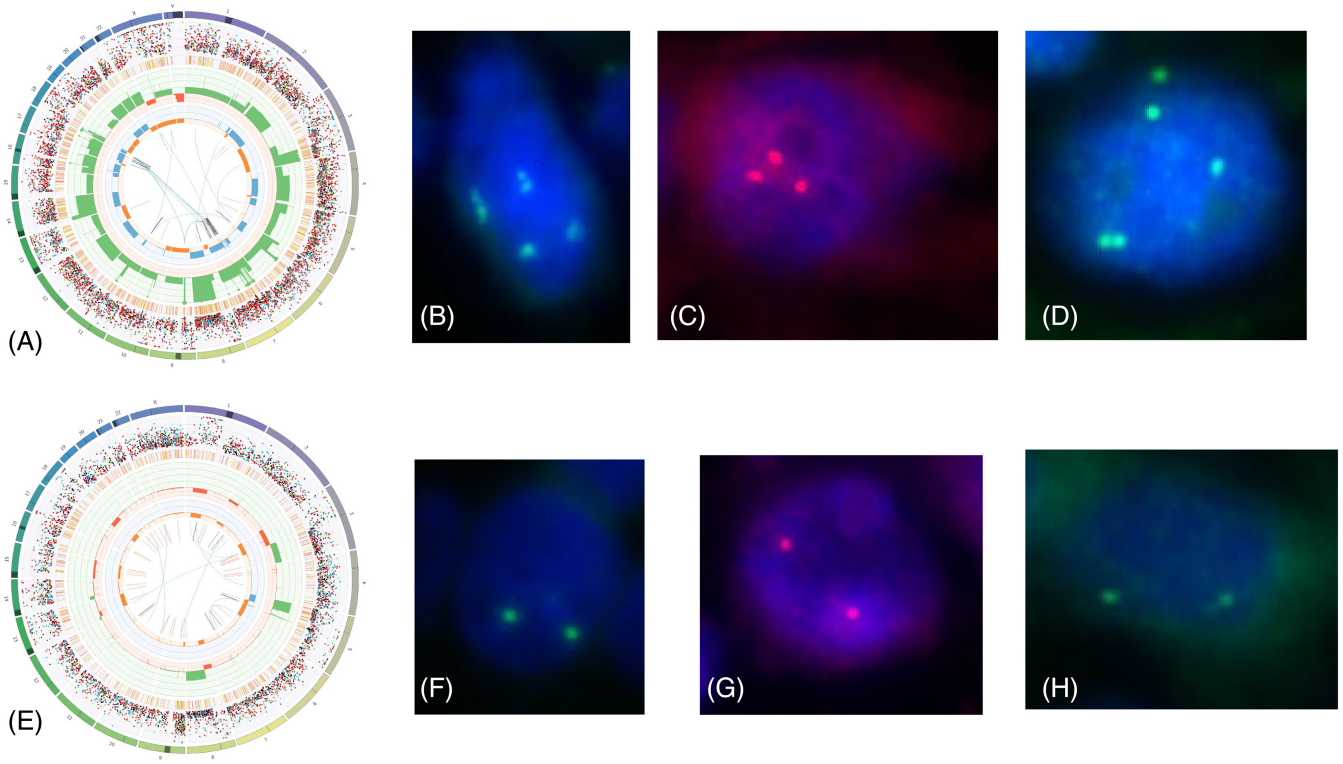

FIGURE 1 CIRCOS plots (PURity and Ploidy Estimator, version 2.4.1) demonstrating (A) presence of whole genome duplication. The inner red ring of the CIRCOS plot shows chromosomal loss, while the adjacent green ring shows chromosomal gain. (B) to (D) FISH CEP of chromosomes 7, 12, and 17 showing increased ploidy; (E) absence of whole genome duplication, with the majority of autosomes showing no notable ploidy changes, confirmed by (F) to (H) FISH CEP of chromosomes 7, 12, and 17

TP53 SNVs (Figures 2 and 3A).[45] There were 27 patients with loss of heterozygosity (LOH) encompassing *TP53*, 19 of which showed biallelic somatic mutation of *TP53*. The remaining eight samples either showed copy number neutral LOH (CN-LOH) with monoallelic *TP53* variant, CN-LOH with no somatic SNVs, or deletion overlapping the *TP53* region (chromosome 17p whole arm or partial deletion). Of the 32 *TP53*-mutant samples, 25 were located in the DNA binding domain of the p53 protein (Figure 2).[45]

Of the 27 patients who had WGD, 25 also had a non-synonymous mutation in *TP53* (93%). Amongst the 25 patients with WGD and non-synonymous *TP53* SNV, 21 (84%) were inferred to have preceded WGD. One patient also had a second non-synonymous mutation, predicted to have stop-gain effect, and inferred to have occurred after WGD.

Amongst the ten patients without WGD, *TP53* somatic SNVs were seen in seven patients (Figures 2 and 3A). Of these, three *TP53* somatic variants occurred in the DNA binding domain, one of which was a frameshift variant, and the other two of which were missense variants also found in patients with WGD (Figure 2). While not reaching statistical significance, largely due to the small number of patients without WGD, *TP53* somatic variants in the DNA binding domain were more frequently observed in patients with WGD (22 [82%]) vs. three (30%) non-WGD patients, $p = 0.07$) Figure 2).

3.3 | Most frequent genomic alterations observed in OSCC patients with and without WGD

Tumor mutation burden did not differ significantly between those with WGD (3.34 mutations/mb) and those without WGD (3.24 mutations/mb) (Figure S4). None of the patients showed microsatellite instability.

Figure 3A,B summarize the most frequent genomic changes observed in patients with and without WGD. Somatic SNVs in *CSMD2*, a gene that may act as a tumor suppressor in gastrointestinal cancers,[46] were seen in four patients with WGD (15%), but were not identified in those without WGD (Figure 3A). Similarly, *EGFR* amplification was seen exclusively in patients with WGD (8/27, 30%), and not seen in patients without WGD ($p = 0.07$, Figure 3B).

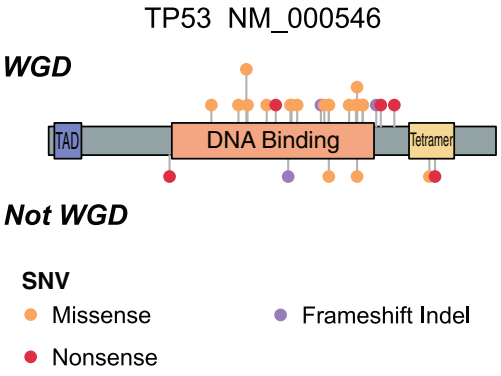

FIGURE 2 Lollipop plots demonstrating location of *TP53* mutations in relation to p53 protein in those with WGD and without WGD

3.4 | WGD is significantly associated with presence of adverse clinicopathological features

Information regarding histopathological prognostic features was available for evaluation only for the 28 SHNCI patients. WGD was significantly more frequent in tumors with a larger tumor size (median size of 30 mm in WGD patients [range 12–95 mm]), versus median size of 15 mm in non-WGD patients (range 5–40 mm), ($p = 0.01$, Figure 4A). There was a trend toward more advanced pT category ($p = 0.08$) and a higher rate (68%) of perineural invasion in patients with WGD ($p = 0.11$).

There was no association between WGD and lymphovascular invasion ($p = 0.43$), depth of invasion ($p = 0.24$, Figure S5), infiltrative pattern of invasion ($p = 1.0$), the presence of nodal metastases ($p = 1.0$), and the presence of extranodal extension ($p = 1.0$).

3.5 | Survival is significantly poorer in patients with WGD

For those whom treatment and outcome data were available, the rate of post-operative radiotherapy ($p = 0.41$) and chemotherapy ($p = 1.0$) was similar amongst those with and without WGD.

Eight patients developed locoregional recurrence. Of these, seven harbored WGD (87%). Overall survival was significantly worse in those with WGD ($p = 0.05$) (Figure 4B). The 24-month overall survival of those with WGD was 67% compared to 100% for those without WGD.

4 | DISCUSSION

Our study presents WGS data from one of the largest cohorts of OSCC in patients younger than 50 years with low cumulative burden of traditional risk factors, such as smoking. A high prevalence (73%) of WGD in patients younger than 50 years with OSCC is demonstrated. The computational analyses have been confirmed through orthogonal testing using clinically validated *in situ* hybridization techniques. In the vast majority (84%) of these patients, mutations in the gatekeeper *TP53* DNA binding domain were found to precede the WGD event. Our data demonstrate that patients with WGD were more likely to adverse pathological features such as larger tumor size. Most notably, patients with WGD were more likely to recur and had poor overall survival.

Morris et al. previously reported WGD in 25% of recurrent/metastatic head and neck cancers using a cohort of HPV-positive and -negative mucosal HNSCC (all sites), salivary gland cancers, and cutaneous carcinomas using targeted sequencing.[14] The median age in their study was 59 years with a male preponderance (2.8:1). Of the

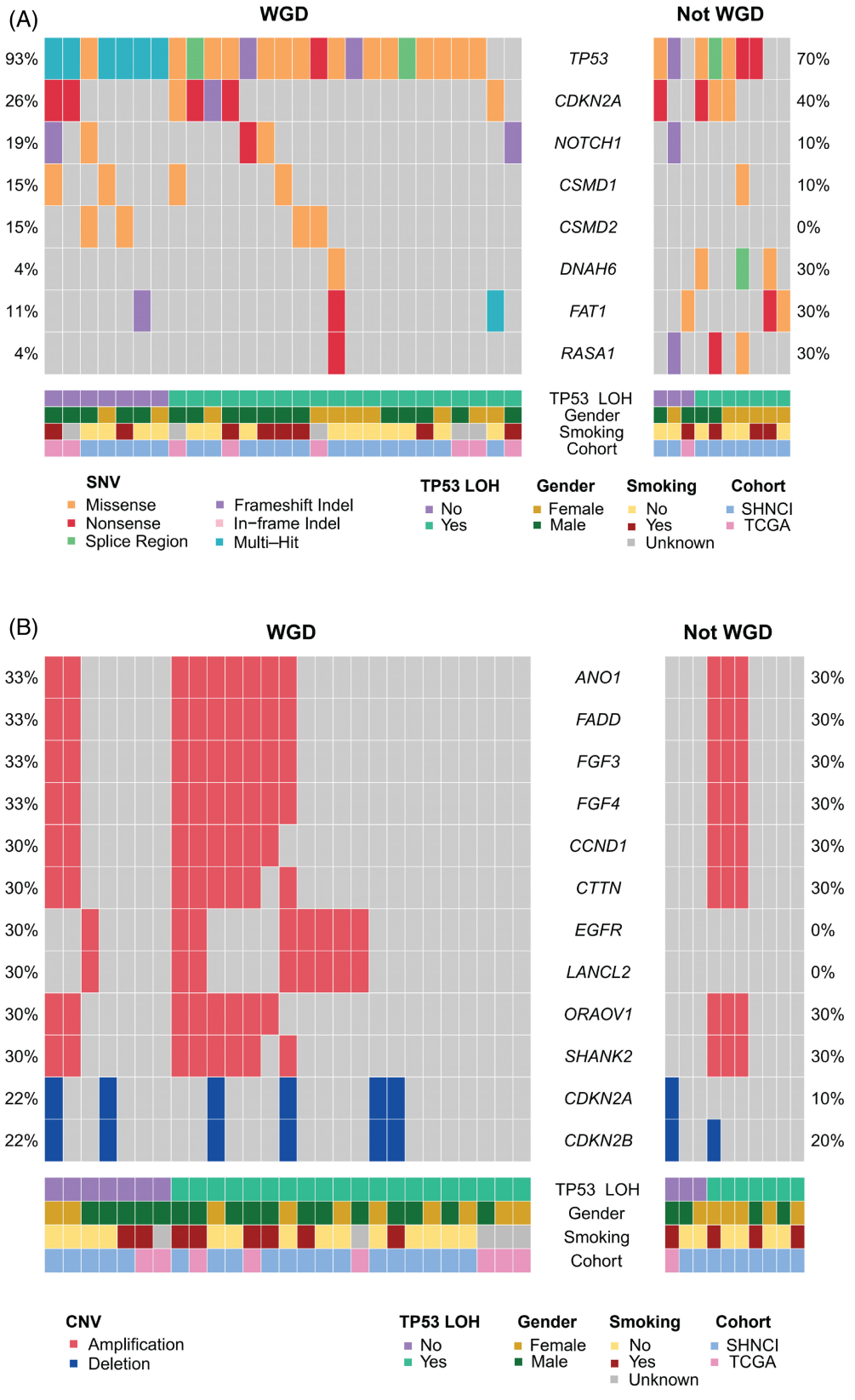

FIGURE 3 Oncoplot of patients with WGD and without WGD, demonstrating most common (A) somatic SNVs, and (B) CNVs. Figure 3A shows cases which lack *TP53* variants but show *TP53* LOH, due to deletion events overlapping the *TP53* region

30 HPV-negative OSCC cases older than 50 years, five cases (17%) had WGD.[14] This rate of WGD reported in the study by Morris et al. for all metastatic and recurrent head and neck cancers irrespective of site and histologic type is significantly lower than that observed in the WGS data of a homogenous cohort of HPV independent OSCC patients younger than 50 years of age in our data. This study also found that WGD occurred more frequently in HPV-negative versus primary HPV-positive disease (odds ratio 4.8), than in HPV-negative

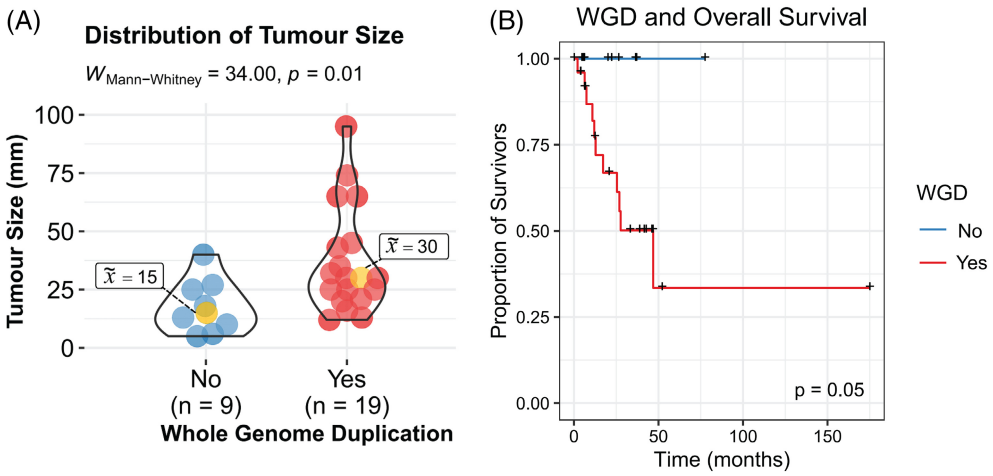

FIGURE 4 (A) Violin plot demonstrating association of WGD with tumor size; (B) Kaplan-Meier curve of overall survival in those with and without WGD

versus recurrent/metastatic HPV-positive disease (odds ratio 0.71). (A relationship between *TP53*-mutant status and WGD was not examined in this study.) The differences in the sequencing methods (targeted vs WGS) may account for some of this difference. However, other factors such as differences in age and cumulative risk factor exposure, such as smoking, HPV, and ultraviolet damage are more likely to account for the significant disparity. Indeed, Morris et al. report that WGD occurred more frequently in HPV-negative versus primary HPV-positive disease (odds ratio 4.8) in their cohort.

OSCC in those under the age of 50 is a rare cancer. However, an increasing incidence is being documented globally.[19,47,48] Despite the rising incidence, the pathways to carcinogenesis have not been explored, and a consistent mutagenic risk factor for OSCC in young patients has not been identified.[47,49] While our cohort includes 14 (37%) smokers, they are younger than 50 years of age at diagnosis and unlikely to have the 49.4 pack year smoking history observed by Schmidt et al. in older patients with OSCC[31].This is further strengthened by the absence of COSMIC mutational signature 4 associated with smoking in this cohort. While it is difficult to precisely delineate the relationship between carcinogenic risk factors and WGD,[50] WGD has been described as an early event, preceding subclonal proliferation in tumors that lack a mutagenic risk factor such as germ cell tumors and lung adenocarcinoma in non-smokers.[51,52] In contrast, WGD has been described as a late event in malignancies with known mutagenic risk factors or driver point mutations.[10,11,50] For example, WGD was seen to occur after the accumulation of a number of ultraviolet light exposure-related mutations including *BRAF* and *KRAS* alterations, and prior to the occurrence of copy number alterations in melanoma, a UV damage driven malignancy. Similarly, WGD is also a late event in myeloma, where point mutations often occur early. Interestingly, nearly all of the patients ith germ cell tumors and lung adenocarcinoma in non-smokers tend to be younger at the time of diagnosis[53,54]; a demographic feature shared with patients in this cohort. Thus, WGD may play an important role in OSCC carcinogenesis in young patients with limited cumulative exposure to risk factors or in those where a causative agent has not been identified.

The majority of patients with WGD in our study demonstrated somatic non-synonymous *TP53* alterations. This is in keeping with previous pan-cancer analyses of WGD, where *TP53*-mutant tumors are twice as likely to undergo WGD as compared with *TP53*-wildtype tumors.[2,6] Indeed, inactivation of the p53 protein has been shown to increase tolerance of WGD.[55] Most *TP53* mutations in our cohort occurred in the DNA binding domain of the p53 protein, a region essential for recognizing transcription errors and maintaining genome integrity during cell cycle.[56] Most somatic *TP53* mutations were inferred to have occurred before the WGD event. *TP53* mutations have been shown to precede WGD, which would be consistent with the cell cycle regulator function of the intact p53 protein, preventing a tetraploid cancer cell from re-entering the cell cycle.[2,6] Interestingly, not all samples with WGD had TP53 LOH in our cohort. Thus, there is likely to be another cause of WGD yet to be discovered. Zack et al. describe amplification of *CCNE1*, or cyclin E1, a gene that encodes for a cyclin protein involved in cell cycle transition, in all patients with wild-type p53. Interestingly, *CCNE1* amplification was not observed in this cohort which shows a predominance of early *TP53* mutations.

WGD has been shown to be associated with poor prognosis in a diverse range of cancers.[2,3,50] In a pan-cancer study of advanced cancer patients,[2] WGD was an independent predictor of adverse outcome after adjusting for well-established adverse clinicopathologic features and genetic alterations. For example, WGD was found to be an independent predictor of death in *KRAS*-mutant colorectal cancer and estrogen receptor positive breast cancer.[2] Correlation of WGD with survival has not been previously performed in OSCC. Similar to Bielski and colleagues[2] in other cancer types, our data demonstrate for the first time in OSCC, that overall survival is significantly reduced in OSCC patients with WGD.

Association of WGD and adverse histopathological features, representing *in vivo* characteristics of the tumor have not been previously examined in OSCC, or in other cancer types. The patients in our cohort with WGD showed a significantly larger tumor size, with a trend toward higher pT category at clinical presentation, and the presence of perineural invasion. Wangsa et al. describe similar findings in their study of the effect of WGD and polyploidy in single-cell derived

tetraploid clones, and demonstrate increased invasive and migratory ability of cancer cells *ex vivo*[57]. Tumor recurrence was also frequent in those OSCC patients with WGD. Thus, our data indicate that WGD is an adverse prognostic indicator in OSCC. If verified, this information could be harnessed to guide prognostic stratification of OSCC in light of the increasing availability, accuracy, and decreasing cost of next-generation sequencing.

One of the main limitations of the study is its small cohort size. HPV-independent OSCC is a rare cancer in young people and the clinicopathologic data of the SHNCI cohort has been carefully curated and the DNA selected for sequencing has undergone stringent quality controls. Of note, search of publicly available databases such as TCGA yielded only 11 cases of HPV-independent OSCC in patients younger than 50 years. Of these, nine had sufficient read depth and tumor purity that could be utilized for further analyses with very limited clinicopathologic and follow up data. In addition, another limitation of using TCGA data was the lack of detailed carcinogenic exposure data, particularly with regards to smoking history (Table 1B). The need for well-annotated clinical data in cancer genomic datasets, as obtained for the SHNCI cohort, is highlighted by our study. The paucity of WGS data for HPV-independent OSCC in young patients in one of the largest genomic consortia highlights the unique nature of the current study.

The current study is the first to demonstrate the high prevalence of WGD among young patients who develop HPV-independent OSCC, in the absence of a conventionally recognized mutagen. The majority of cases with WGD demonstrated *TP53* mutations and these predominantly occurred prior to WGD. Of clinical relevance, the presence of WGD may represent a potential prognostic biomarker of adverse outcome in young HPV-independent OSCC patients. The frequency of adverse prognostic histopathological features and poor outcome in OSCC patients with WGD highlight the need for targeted therapies. Novel agents targeting the mitotic spindle-assembly checkpoints in WGD-driven cancers are under development, and may represent a new therapeutic horizon in the treatment of OSCC[13].

ACKNOWLEDGMENTS

The authors acknowledge the technical assistance provided by the Sydney Informatics Hub, a Core Research Facility of the University of Sydney, and the Australian BioCommons which is enabled by NCRIS via Bioplatforms Australia. The authors acknowledge the use of the National Computational Infrastructure (NCI) supported by the Australian Government and the Sydney Informatics Hub HPC Allocation Scheme, supported by the Deputy Vice-Chancellor (Research), University of Sydney and the ARC LIEF, 2019: Smith, Muller, Thornber et al., Sustaining and strengthening merit-based access to National Computational Infrastructure (LE190100021). The authors wish to thank Cancer Institute New South Wales (CINSW 2020/2081) and the Neal Wald family for their funding support for the bioinformatic analysis for this project. The authors acknowledge the help of Ms Maggie Y Lee, Dr Grace H Wei and Associate Professor Michael E Buckland of the Department of Neuropathology, Royal Prince Alfred Hospital, for their assistance in nucleic acid extractions; and Mr Isuru Hewage and Ms Mikaela Holmes of the Department of Tissue Pathology and Diagnostic Oncology, Royal Prince Alfred Hospital, for their assistance in the construction of the TMAs and FISH studies respectively. The results published here are in whole or part based upon data generated by The Cancer Genome Atlas managed by NCI and NHGRI. Information about TCGA can be found at http://cancergenome.nih.gov. Open access publishing facilitated by The University of Sydney, as part of the Wiley - The University of Sydney agreement via the Council of Australian University Librarians.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

ORCID

Laveniya Satgunaseelan https://orcid.org/0000-0002-7435-0834
Rosemarie Sadsad https://orcid.org/0000-0003-2488-953X
Ruta Gupta https://orcid.org/0000-0003-2940-313X

REFERENCES

1. Taylor AM, Shih J, Ha G, et al. Genomic and functional approaches to understanding cancer aneuploidy. *Cancer Cell*. 2018;33(4):676-689.e673.
2. Bielski CM, Zehir A, Penson AV, et al. Genome doubling shapes the evolution and prognosis of advanced cancers. *Nat Genet*. 2018;50(8):1189-1195.
3. Dewhurst SM, McGranahan N, Burrell RA, et al. Tolerance of whole-genome doubling propagates chromosomal instability and accelerates cancer genome evolution. *Cancer Discov*. 2014;4(2):175-185.
4. Zack TI, Schumacher SE, Carter SL, et al. Pan-cancer patterns of somatic copy number alteration. *Nat Genet*. 2013;45(10):1134-1140.
5. Ben-David U, Amon A. Context is everything: aneuploidy in cancer. *Nat Rev Genet*. 2020;21(1):44-62.
6. McGranahan N, Favero F, de Bruin EC, Birkbak NJ, Szallasi Z, Swanton C. Clonal status of actionable driver events and the timing of mutational processes in cancer evolution. *Sci Transl Med*. 2015;7(283):283ra254.
7. Olaharski AJ, Sotelo R, Solorza-Luna G, et al. Tetraploidy and chromosomal instability are early events during cervical carcinogenesis. *Carcinogenesis*. 2006;27(2):337-343.
8. Maley CC, Galipeau PC, Finley JC, et al. Genetic clonal diversity predicts progression to esophageal adenocarcinoma. *Nat Genet*. 2006;38(4):468-473.
9. Gerlinger M, Rowan AJ, Horswell S, et al. Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N Engl J Med*. 2012;366(10):883-892.
10. Maura F, Bolli N, Angelopoulos N, et al. Genomic landscape and chronological reconstruction of driver events in multiple myeloma. *Nat Commun*. 2019;10(1):3835.
11. Birkeland E, Zhang S, Poduval D, et al. Patterns of genomic evolution in advanced melanoma. *Nat Commun*. 2018;9(1):2665.
12. Davoli T, Uno H, Wooten EC, Elledge SJ. Tumor aneuploidy correlates with markers of immune evasion and with reduced response to immunotherapy. *Science*. 2017;355(6322):eaaf8399.
13. Quinton RJ, DiDomizio A, Vittoria MA, et al. Whole-genome doubling confers unique genetic vulnerabilities on tumour cells. *Nature*. 2021;590(7846):492-497.

14. Morris LGT, Chandramohan R, West L, et al. The molecular landscape of recurrent and metastatic head and neck cancers: insights from a precision oncology sequencing platform. *JAMA Oncol*. 2017;3(2):244-255.
15. Amin MB, Edge SB, Cancer AJC. *AJCC Cancer Staging Manual*. 8th edition ed. Springer; 2017.
16. WHO Classification of Tumours Editorial Board. Head and neck tumours. *WHO Classification of Tumours*. 2022. https://tumourclassification.iarc.who.int/chapters/52
17. Toporcov TN, Znaor A, Zhang ZF, et al. Risk factors for head and neck cancer in young adults: a pooled analysis in the INHANCE consortium. *Int J Epidemiol*. 2015;44(1):169-185.
18. de Morais EF, Mafra RP, Gonzaga AKG, de Souza DLB, Pinto LP, da Silveira EJD. Prognostic factors of oral squamous cell carcinoma in young patients: a systematic review. *J Oral Maxillof Surg*. 2017;75(7):1555-1566.
19. Warnakulasuriya S. Global epidemiology of oral and oropharyngeal cancer. *Oral Oncol*. 2009;45(4–5):309-316.
20. Mukdad L, Heineman TE, Alonso J, Badran KW, Kuan EC, St John MA. Oral tongue squamous cell carcinoma survival as stratified by age and sex: a surveillance, epidemiology, and end results analysis. *Laryngoscope*. 2019;129(9):2076-2081.
21. Kim Y-J, Kim JH. Increasing incidence and improving survival of oral tongue squamous cell carcinoma. *Sci Rep*. 2020;10(1):7877.
22. Satgunaseelan L, Allanson BM, Asher R, et al. The incidence of squamous cell caricnoma of the oral tongue is rising in young non-smoking women: an international multi-institutional analysis. *Oral Oncol*. 2020;110:104875.
23. Chalmers ZR, Connelly CF, Fabrizio D, et al. Analysis of 100,000 human cancer genomes reveals the landscape of tumor mutational burden. *Genome Med*. 2017;9(1):34.
24. Annertz K, Anderson H, Biorklund A, et al. Incidence and survival of squamous cell carcinoma of the tongue in Scandinavia, with special reference to young adults. *Int J Cancer*. 2002;101(1):95-99.
25. Hussein AA, Helder MN, de Visscher JG, et al. Global incidence of oral and oropharynx cancer in patients younger than 45 years versus older patients: a systematic review. *Eur J Cancer*. 2017;82:115-127.
26. Pickering CR, Zhang J, Neskey DM, et al. Squamous cell carcinoma of the oral tongue in young non-smokers is genomically similar to tumors in older smokers. *Clin Cancer Res*. 2014;20(14):3842-3848.
27. Pickering CR, Zhang J, Yoo SY, et al. Integrative genomic characterization of oral squamous cell carcinoma identifies frequent somatic drivers. *Cancer Discov*. 2013;3(7):770-781.
28. Li R, Faden DL, Fakhry C, et al. Clinical, genomic, and metagenomic characterization of oral tongue squamous cell carcinoma in patients who do not smoke. *Head Neck*. 2015;37(11):1642-1649.
29. The Cancer Genome Atlas (TCGA). *The Cancer Genome Atlas Platform* 2018. 2022. https://www.cancer.gov/about-nci/organization/ccg/research/structural-genomics/tcga
30. Qian X, Nguyen DT, Dong Y, et al. Prognostic score predicts survival in HPV-negative head and neck squamous cell cancer patients. *Int J Biol Sci*. 2019;15(7):1336-1344.
31. Schmidt BL, Dierks EJ, Homer L, Potter B. Tobacco smoking history and presentation of oral squamous cell carcinoma. *J Oral Maxillof Surg*. 2004;62(9):1055-1058.
32. Brandwein-Gensler M, Smith RV, Wang B, et al. Validation of the histologic risk model in a new cohort of patients with head and neck squamous cell carcinoma. *Am J Surg Pathol*. 2010;34(5):676-688.
33. Brandwein-Gensler M, Teixeira MS, Lewis CM, et al. Oral squamous cell carcinoma: histologic risk assessment, but not margin status, is strongly predictive of local disease-free and overall survival. *Am J Surg Pathol*. 2005;29(2):167-178.
34. *AJCC Cancer Staging Manual*. 8th edition ed. Springer International Publishing: American Joint Commission on Cancer; 2017.
35. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv e-prints*. 2013. https://ui.adsabs.harvard.edu/abs/2013arXiv1303.3997L
36. Van der Auwera GA, Carneiro MO, Hartl C, et al. From FastQ data to high-confidence variant calls: the genome analysis toolkit best practices pipeline. *Curr Protoc Bioinformatics*. 2013;43(1):11.10.11-11.10.33.
37. Poplin R, Ruano-Rubio V, DePristo MA, et al. Scaling accurate genetic variant discovery to tens of thousands of samples. *bioRxiv*. 2017:201178.
38. Karczewski KJ, Francioli LC, Tiao G, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*. 2020;581(7809):434-443.
39. McLaren W, Gil L, Hunt SE, et al. The ensembl variant effect predictor. *Genome Biol*. 2016;17(1):122.
40. Willet C, Chew T, Samaha G, et al. Fastq-to-BAM (Version 2.0) [Computer software]. Mapping (Fastq-to-BAM). 2021. 10.48546/workflowhub.workflow.146.1
41. Chew T, Willet C, Samaha G, et al. Germline-ShortV (Version 1.0) [Computer software]. Germline variant calling. 2021. 10.48546/workflowhub.workflow.143.1
42. Chew T, Willet C, Samaha G, et al. Somatic-ShortV (Version 1.0) [Computer software]. Somatic variant calling. 2021. 10.48546/workflowhub.workflow.148.1
43. Priestley P, Baber J, Lolkema MP, et al. Pan-cancer whole-genome analyses of metastatic solid tumours. *Nature*. 2019;575(7781):210-216.
44. Wong M, Mayoh C, Lau LMS, et al. Whole genome, transcriptome and methylome profiling enhances actionable target discovery in high-risk pediatric cancer. *Nat Med*. 2020;26(11):1742-1753.
45. Mayakonda A, Lin DC, Assenov Y, Plass C, Koeffler HP. Maftools: efficient and comprehensive analysis of somatic variants in cancer. *Genome Res*. 2018;28(11):1747-1756.
46. Yang PS, Hsu HH, Hsu TC, et al. Genome-wide scan for copy number alteration association with relapse-free survival in colorectal cancer with liver metastasis patients. *J Clin Med*. 2018;7(11):446.
47. Satgunaseelan L, Allanson B, Asher R, et al. The incidence of squamous cell carcinoma of the oral tongue is rising in young non-smoking women: an international multi-institutional analysis. Submitted. 2020.
48. Michermerhuizen NL, Birkeland AC, Bradford CR, Brenner JC. Genetic determinants in head and neck squamous cell carcinoma and their influence on global personalized medicine. *Genes Cancer*. 2016;7(5–6):182-200.
49. Satgunaseelan L, Virk SA, Lum T, Gao K, Clark JR, Gupta R. p16 expression independent of human papillomavirus is associated with lower stage and longer disease-free survival in oral cavity squamous cell carcinoma. *Pathology*. 2016;48(5):441-448.
50. Lopez S, Lim EL, Horswell S, et al. Interplay between whole-genome doubling and the accumulation of deleterious alterations in cancer evolution. *Nat Genet*. 2020;52(3):283-293.
51. Dorssers LCJ, Gillis AJM, Stoop H, et al. Molecular heterogeneity and early metastatic clone selection in testicular germ cell cancer development. *Br J Cancer*. 2019;120(4):444-452.
52. Lee JJ, Park S, Park H, et al. Tracing oncogene rearrangements in the mutational history of lung adenocarcinoma. *Cell*. 2019;177(7):1842-1857.e1821.
53. Calaminus G, Joffe J. Germ cell tumors in adolescents and young adults. *Progress Tumor Res*. 2016;43:115-127.
54. Luo W, Tian P, Wang Y, et al. Characteristics of genomic alterations of lung adenocarcinoma in young never-smokers. *Int J Cancer*. 2018;143(7):1696-1705.
55. Andreassen PR, Lohez OD, Lacroix FB, Margolis RL. Tetraploid state induces p53-dependent arrest of nontransformed mammalian cells in G1. *Mol Biol Cell*. 2001;12(5):1315-1328.

56. Pradhan MR, Siau JW, Kannan S, et al. Simulations of mutant p53 DNA binding domains reveal a novel druggable pocket. *Nucleic Acids Res*. 2019;47(4):1637-1652.
57. Wangsa D, Quintanilla I, Torabi K, et al. Near-tetraploid cancer cells show chromosome instability triggered by replication stress and exhibit enhanced invasiveness. *FASEB J*. 2018;32(7):3502-3517.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Satgunaseelan L, Strbenac D, Willet C, et al. Whole genome duplication in oral squamous cell carcinoma in patients younger than 50 years: implications for prognosis and adverse clinicopathological factors. *Genes Chromosomes Cancer*. 2022;61(9):561-571. doi:10.1002/gcc.23076