Original Research
Oral Pathology

Genome sequencing reveals molecular subgroups in oral epithelial dysplasia

Agustín MÁRQUEZ[a]
Isidora MUJICA[b]
Natalia JORDAN[c]
Pablo BAEZ[d]
Sandra TARQUINIO[e]
Jean NUNES[f]
Daniela ADORNO[g]
Benjamín MARTÍNEZ[h]
Sebastian MORALES-PISON[i]
Ricardo FERNANDEZ-RAMIRES[j]

[a]Universidad Autónoma, Faculty of Dentistry, Santiago, Chile.
[b]Universidad de Los Andes, Faculty of Dentistry, Santiago, Chile.
[c]Pontificia Universidad Católica de Chile, Faculty of Science, Santiago, Chile.
[d]Universidad de Chile, Faculty of Science, Santiago, Chile.
[e]Universidade Federal de Pelotas - UFPel, School of Dentistry, Pelotas, RS, Brazil.
[f]Universidade Federal da Bahia - UFBA, School of Dentistry, Salvador, BA, Brazil.
[g]Universidad de Chile, Faculty of Dentistry, Santiago, Chile.
[h]Universidad Mayor, Faculty of Dentistry, Santiago, Chile.
[i]Universidad Mayor, Precision Oncology Center, Santiago, Chile.
[j]Universidad Mayor, Faculty of Medicine and Heatlh Sciences, Santiago, Chile.

Declaration of Interests: The authors certify that they have no commercial or associative interest that represents a conflict of interest in connection with the manuscript.

Corresponding Author:
Ricardo Fernández-Ramires
E-mail: ricardo.ramires@umayor.cl

https://doi.org/10.1590/1807-3107bor-2023.vol37.0063

Submitted: April 26, 2022
Accepted for publication: November 21, 2022
Last revision: April 12, 2023

Abstract: This study aimed to analyze the molecular characteristics of oral epithelial dysplasia (OED), highlighting the pathways and variants of genes that are frequently mutated in oral squamous cell carcinoma (OSCC) and other cancers. Ten archival OED cases were retrieved for retrospective clinicopathological analysis and exome sequencing. Comparative genomic analysis was performed between high-grade dysplasia (HGD) and low-grade dysplasia (LGD), focusing on 57 well-known cancer genes, of which 10 were previously described as the most mutated in OSCC. HGD cases had significantly more variants; however, a similar mutational landscape to OSCC was observed in both groups. *CASP8+FAT1/HRAS, TP53,* and miscellaneous molecular signatures were also present. *FAT1* is the gene that is most affected by pathogenic variants. Hierarchical divisive clustering showed division between the two groups: "HGD-like cluster" with 4HGD and 2LGD and "LGD-like cluster" with 4 LGD. *MLL4* pathogenic variants were exclusively in the "LGD-like cluster". *TP53* was affected in one case of HGD; however, its pathway was usually altered. We describe new insights into the genetic basis of epithelial malignant transformation by genomic analysis, highlighting those associated with *FAT1* and *TP53*. Some LGDs presented a similar mutational landscape to HGD after cluster analysis. Perhaps molecular alterations have not yet been reflected in histomorphology. The relative risk of malignant transformation in this molecular subgroup should be addressed in future studies.

Keywords: Mouth Neoplasms; High-Throughput Nucleotide Sequencing; Precancerous Conditions; Molecular Sequence Annotation; Carcinoma *in Situ*; Mutation.

Introduction

The incidence of oral cancer in the United States (U.S.) is estimated at 35,130 new cases per year, most of which correspond to oral squamous cell carcinoma (OSCC).[1] Well-known risk factors for OSCC development include tobacco and alcohol consumption and immune system impairment.[2-5] Human papillomavirus (HPV) has a driving role in the oropharynx and cervical cancer. However, it is still a controversial risk factor for OSCC.[6, 7] Approximately 14% of OSCCs evolve from oral leukoplakia, a potentially premalignant oral lesion (PPOL) described as a "white plaque

of questionable risk having excluded known diseases or disorders that carry no increased risk for cancer".[8]

There are two clinical variants of oral leukoplakia: homogeneous and non-homogeneous leukoplakia, with the latter having a higher risk of malignancy (up to 20-25%).[4,9,10] The histological diagnosis of oral leukoplakia varies from hyperkeratosis to epithelial dysplasia and carcinoma. Dysplasia is a prognostic diagnosis, graded by the OMS as mild, moderate, and severe, depending on architectural and cytologic characteristics.[11] Severe dysplasia is associated with the highest risk of malignancy.[10,12] It has been proposed that dysplasia grades follow a continuum from normal mucosa to OSCC; however, this might not always be the norm.[13]

The risk of malignant transformation is difficult to assess and is unclear, especially in mild dysplasia.[14] Difficulties in grading dysplasia include observer subjectivity, lack of consensus classification, and the number of cytological atypia per field.[4,10,15] Kujan et al.[16] proposed a binary grading system that divides dysplasia into two categories (low-grade dysplasia (LGD) and high-grade dysplasia (HGD) after assessment of the epithelial and cytological characteristics proposed by the WHO Classification. Several studies have shown that the binary grading system reduces the differences between examiners and improves biopsy sensitivity.[11,16-18]

Additionally, molecular tests may help in assessing the risk of dysplasia due to malignant transformation. DNA mutations accumulate towards progression into OSCC, and loss of heterozygosity and microsatellite instability have been observed, which may account for the risk of malignant transformation assessment.[19] Molecular biology techniques assessing the latter are not widely available and are difficult to perform;[20,21] hence, the relevance of next-generation sequencing (NGS) systems provides broader laboratory techniques. According to existing panels, tools such as "*Hotspot Cancer Panel v2*" (*Illumina®, California, USA)* call variants of deleterious or driver gene regions for a series of cancers, including ovarian, lung, hepatic, renal, haematolymphoid, and melanoma.[22,23]

NGS studies have been performed on Head and Neck Squamous Cell Carcinoma (HNSCC) (including OSCC) to characterize its mutational landscapes.[7,19] The first OSCC NGS study was conducted in 2013 by Maitra et al. in an Indian population, reporting three mutational signatures: one characterized by *TP53* mutations, a second composed of CASP8±FAT1, and a third miscellaneous signature. The mutations mainly correspond to single nucleotide polymorphisms (SNP) and indels.[24] Only a few NGS studies have assessed PPOL, and none have mutations encountered across grades of dysplasia and cancer.[10,25] Hence, the present study compared the mutational landscape of ora epithelial displasya OED grades and known driver genes in OSCC and other cancers. Unique genomic characteristics may help to understand the progression from dysplasia to cancer and provide novel insights for future therapies.

Methodology

Samples

Ten patients from the study group with OED were retrieved. The histological slides were examined by two independent oral pathologists for a second diagnosis. A binary grading system for dysplasia was used for the histological diagnosis of the samples.[16] When no agreement was reached between the two pathologists, the diagnosis was settled by consensus. The patient information was codified to maintain anonymity. The subjects' rights were protected by the Institutional Review Board with ethical approval (FONDECYT #11140281), which complied with the Helsinki Declaration, and each subject signed a detailed informed consent form.

Exome sequencing and selection of focus gene list

Total DNA was obtained by an external laboratory (ALAMAK®) from formalin-fixed paraffin-embedded tissue blocks using the DNA extraction kit RecoverAll™ (Thermofisher™, Massachusetts, USA) following the manufacturer's instructions. The DNA yield ranged from 0.2-2.0 μg. After processing each sample, 20 μl of the solution was obtained, and an aliquot of 1.0 μl complemented with 99 μL Milli-Qâ water was analyzed using a spectrophotometer (DU-640, Beckman, Palo Alto, USA) to verify the

quantity and purity of each DNA sample. Genomic DNA was stored at -80 °C. DNA was prepared for whole-exome sequencing.

Whole exome capture was performed with 500 ng of DNA using Agilent SureSelect Human all Exon kit™ (Agilent Technologies™, Santa Clara, USA), following the manufacturer's instructions. Briefly, high-quality genomic DNA samples were sheared and randomly fragmented using the M220 Focused-ultrasonicator™ (Covaris®, Massachusetts, USA) with a base pared of 150–200bp. The ends were repaired, and 'A' bases were added at the 3' end. Adapters were ligated to both ends of the fragments. The adaptor-ligated library was amplified using ligation-mediated PCR, purified, and hybridized to the SureSelect Biotinylated DNA Library for enrichment. The captured library was amplified, purified, and tested using an adaptor-ligated library, amplified by ligation-mediated PCR, purified, and hybridized to the SureSelect Biotinylated DNA Library for enrichment. The captured library was amplified, purified, and tested using an Agilent 2100 Bioanalyzer™ (Agilent Technologies, Santa Clara, USA). The samples were analyzed by Whole Exome Sequencing, combining the most robust and efficient DNA capture with enrichment protocols.

A variant calling workflow for 57 genes was performed (47 genes of known impact for several cancers and 10 most mutated genes in OSCC) (Table 1).

Molecular analysis workflow

The first variant calling within the 57 genes region of interest was performed without restrictions based on their Minor Allele Frequency.

Variant classification

Variants were classified according to the impact ranking assigned by Variant Effect Predictor [26] from Ensembl version 91. The prediction was performed based on the type of genetic consequence of the mutation, classifying them as MODIFIER, LOW, MODERATE, and HIGH IMPACT.

Specific variant calling of HPV DNA sequences was performed for more than 170 HPV subtypes, including high-risk HPV strains.[27].

Variant characterization

The MODERATE and HIGH IMPACT variants were selected for further investigation of their pathogenicity. Divisive hierarchical clustering was performed based on SNPs presentation on samples to evaluate variant traits between samples. R programming environment v.3.6.0 was used for this purpose.[28] In a pool of 10 patients, we performed a cluster analysis to investigate whether certain variants occurred more frequently than others. Non-affected variants, heterozygosis, and homozygosis (0, 1, or 2) in 97 SNPs from each patient were used for clustering. We used the Manhattan distance formula to calculate the dissimilarity matrix of mutational profiles between SNPs and the hierarchical divisive clustering algorithm to identify broad clusters. We then performed a second cluster analysis to investigate whether groups of patients (n = 10) had similar mutational profiles in the 97 SNPs identified. We further evaluated whether these corresponded with certain physiological traits observed in patients. All cluster analyses were performed in R v.3.6.0, using the library cluster.[28]

Table 1. 57 selected genes for SNV analysis.

ABL1	CDH1	FAT1	GNAS	KIT	NRAS	SMARCB1	VHL
AKT1	CDKN2A	FBXW7	HNF1A	KRAS	PDGFRA	SMO	
ALK	CSF1R	FGFR1	HRAS	MET	PIK3CA	SRC	
APC	CTNNB1	FGFR2	IDH1	MLH1	PTEN	STK11	
ARID2	EGFR	FGFR3	IDH2	KMT2B	PTPN11	TP53	
ATM	ERBB2	FLT3	JAK2	MPL	RB1	TRPM3	
BRAF	ERBB4	GNA11	JAK3	NOTCH1	RET	UNC13C	
CASP8	EZH2	GNAQ	KDR	NPM1	SMAD4	USP9X	

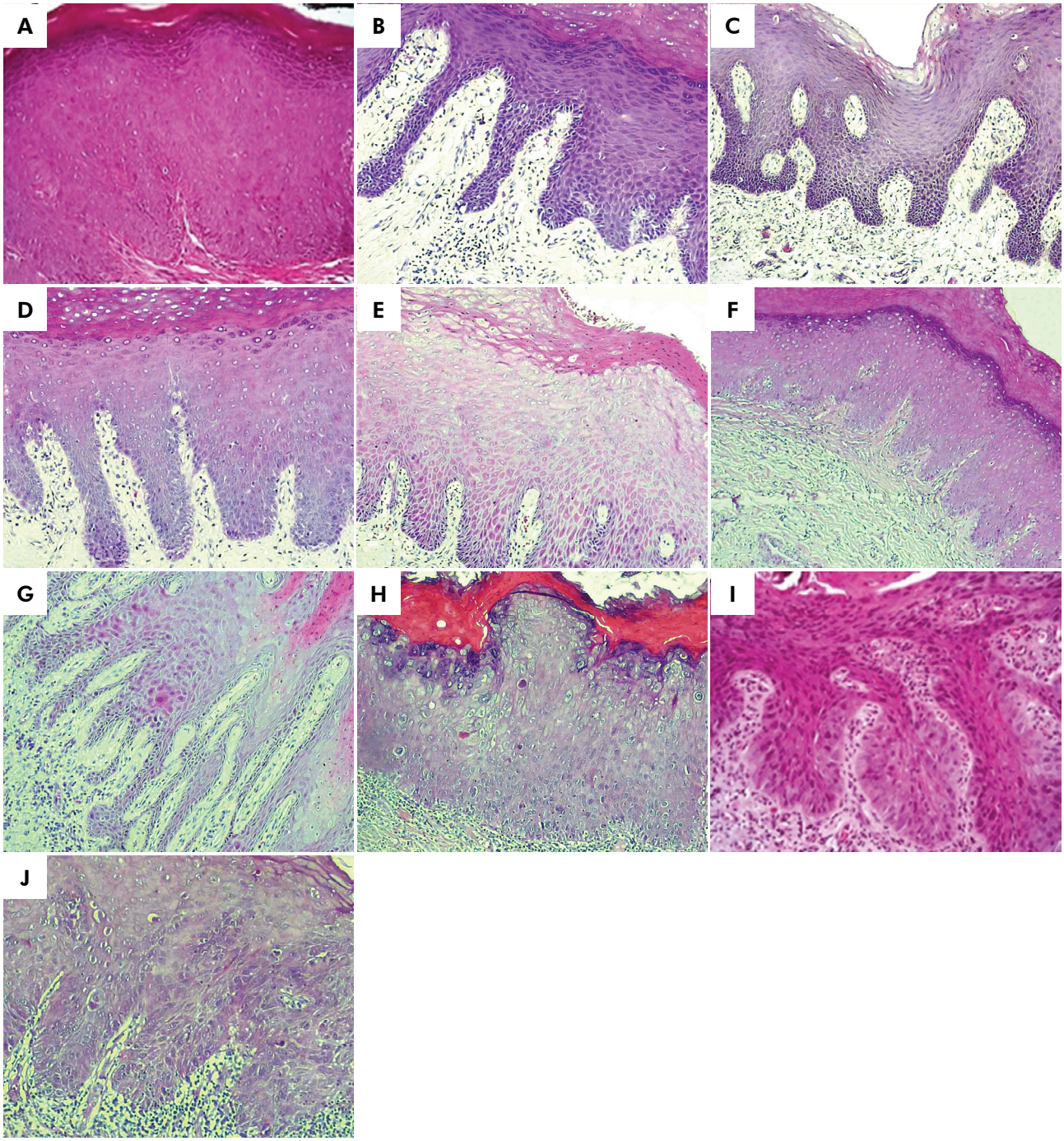

Figure 1. *Representative histopathological figures*. A. LGD in tongue sample with clinical diagnosis of erythroleukoplakia (ID 1); B. LGD in tongue sample with clinical diagnosis of homogeneous leukoplakia (ID 2); C. LGD in buccal mucosa sample with clinical diagnosis of verrucous leukoplakia (ID 3); D. LGD in palate sample with clinical diagnosis of erythroleukoplakia (ID 4); E. LGD in tongue sample with clinical diagnosis of homogeneous leukoplakia (ID 5); F. LGD in gingival ridge sample with clinical diagnosis of erythroleukoplakia (ID 6); G. HGD in gingival ridge sample with clinical diagnosis of homogeneous leukoplakia (ID 7); H. HGD in the floor of mouth sample with clinical diagnosis of homogeneous leucoplakia (ID 8); I. HGD in tongue sample with clinical diagnosis of erythroleukoplakia (ID 9); J. HGD in buccal mucosa sample with clinical diagnosis of erythroleukoplakia (ID 10).

SNPs with a variant allele frequency (VAF) ≥ 0.1 were excluded because of the lack of germline tissue for comparative analysis. Pathogenicity screening was performed with SIFT (Scale-Invariant Feature Transform), PolyPhen-2 (Polymorphism Phenotyping v2), CONsensus DELeteriousness score of non-synonymous single nucleotide variants (CONDEL), FATHMM, Combined Annotation Dependent Depletion (CADD), and LofTool. Variants with one or more pathogenic predictors or HIGH IMPACT were manually searched for in ClinVar and NCBI RefSeq to assess their clinical implications.

Panther Pathway analysis of genes affected by pathogenic variants with VAF≤0.1 was done at the functional and biological process levels to investigate molecular significance.[29] An overrepresentation test of the pathway analysis was performed for an extended review.

Variants from each sample were combined into a single gvcf file for comparative analysis between groups. The Case-Control routine included in SnpSift[30] was used to detect variants with differential occurrence between High- and Low-grade samples. P-values were obtained for different genetic models: dominant, recessive, allelic, genotypic/codominant, and Cochran-Armitage trend. Statistical significance was set at $p < 0.05$.

Pooled variants and their cumulative impacts were compared between LGD and HGD using a non-paired Student's *t*-test.

Results

Clinical outcomes

Six of the ten included cases were diagnosed as LGD and four as HGD (Figure 1). M: W ratio was 1:1, mean age was 58.8 years old. Four of the cases presented clinically as homogeneous leukoplakia (two HGD and two LGD), five as erythroleukoplakia (two HGD and three LGD), and one as verrucous leukoplakia (LGD). The sites of presentation were the tongue (four cases), buccal mucosa (two cases), gingival ridge (two cases), palatal mucosa (one case), and floor of the mouth (one case). A history of tobacco consumption was positive for all except one patient. No HPV infection was detected using sequencing.

Molecular outcomes

Variant calling for the coding exons of the genes of interest was performed. The mean depth coverage of SNPs in the coding area was 47x. The total number of variants was 7,283. Of these, 6,993 (96.23%) were noncoding variants grouped as MODIFIERS, and 274 (3.77%) were LOW/MODERATE/HIGH IMPACT, according to the ENSEMBL classification. Overall 177/274 (64.59%) were LOW IMPACT, and 136/274 (50%) were synonymous variants. Of the 87/274 (31.75%) variants of MODERATE IMPACT, 83 (95.4%) were missense, and 4 (4.6%) were in-frame indels. Finally, 10/274 (3.64%) variants were HIGH impact nonsense, stop gain/lost, and splice site variants. The variant counts and workflow are shown in Figure 2.

Nine of the ten patients were smokers or former smokers. C:G: A:T transversion was the most common alteration (64.16%). Variants tended to gather in a few chromosomes (CHR); 1,701 (23.3%) affected CHR 2; 1,357 (18.6%) affected CHR 9; and 1,541 (21.1%) affected CHR 4, 7, and 15 (mean of 513.6) (Figure 3a). The most affected genes were *ALK*, *ERBB4*, *TRPM3,* and *UNC13C,* with 830, 761, 678, and 487 SNPs, respectively.

Approximately 1,552 variants were detected in each HGD case. In comparison, about 1,254 variants per case were detected in LGD cases (Figure 3b). Unpaired Student's *t*-test of the mutational count revealed statistical significance between the groups ($p = 0.0347$). The normality assumption for each group was verified using the Shapiro-Wilk test (p-value = 0.4914). The results are shown in Figure 3.

Qualitative analysis of molecular alterations

After excluding MODIFIER and LOW IMPACT variants, the 97 remaining MODERATE and HIGH IMPACT variants were analyzed. The genomic variant burden for the OED cases is presented in Table 2. The mean number of SNPs for LGD and HGD cases was approximately 40 and 39, respectively, without significant differences between the groups.

Variants of the studied genes were examined in all cases, as shown in Table 2. The genes with

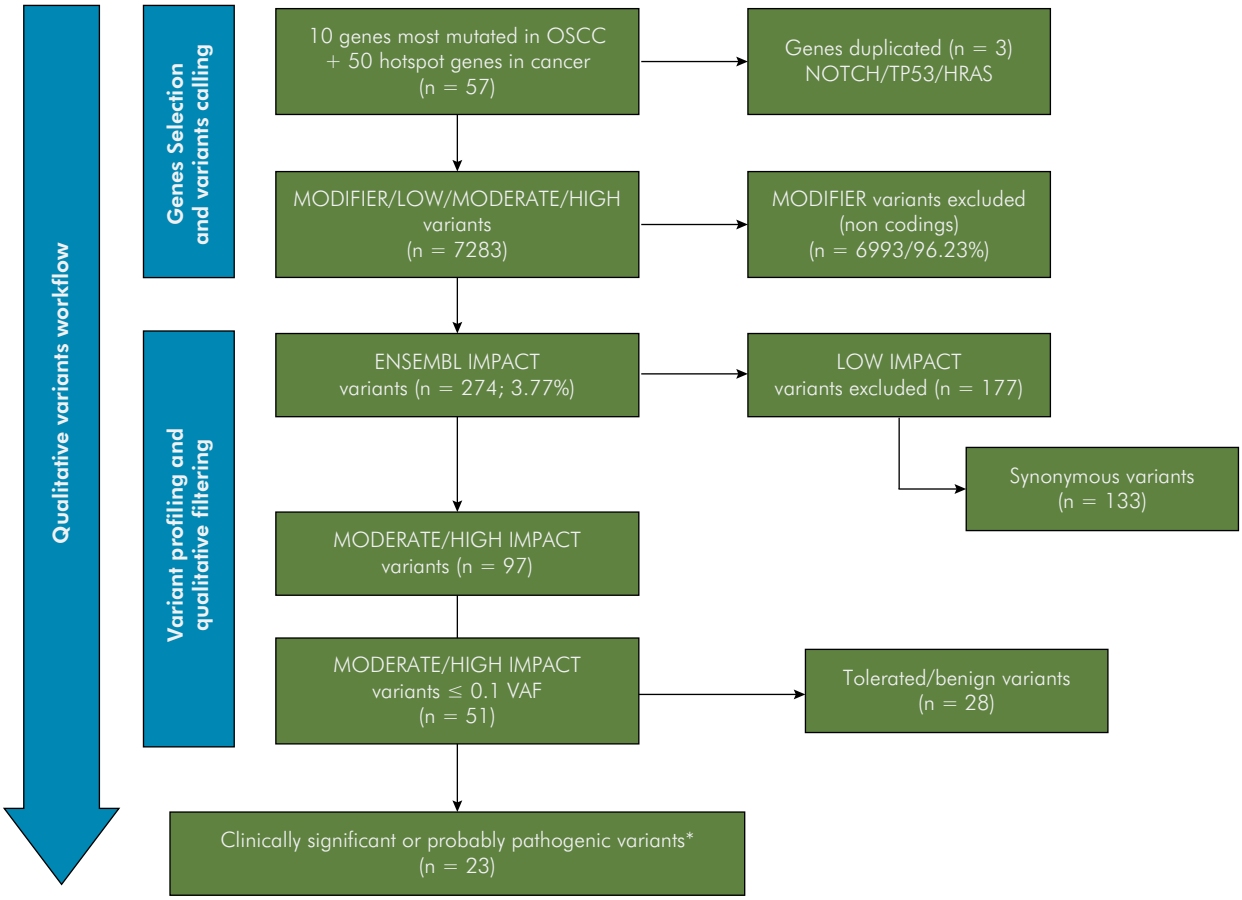

Figure 2. *Qualitative variant analysis:* 51 specific HIGH/MODERATE probable somatic variants are shown (VAF ≤0.1). 23 out of 51 had possible pathological consequences in proteins, according to software predictions.

the most significant number of variants were *FAT1* ($\bar{X}$ ~ 11.2), *HNF1A* ($\bar{X}$ ~ 4.4), *ALK* ($\bar{X}$ ~ 4.4), and *TRPM3* ($\bar{X}$ ~ 4.4). Interestingly, only two variants were identified for *TP53*: rs1042522 (a relatively common variant associated with hereditary cancer syndrome in Clinvar), which was present in all cases of homozygosis, and rs28934575, a heterozygous variant present in only one HGD case. Seven new variants were assessed, and 3 of them had HIGH IMPACT (frameshift/splice donor/start loss). The number of *FLT3* gene variants was significantly higher in HGD than in LGD (p = 0.0112) (Figure 3c).

Hierarchical clustering

Two non-supervised dendrograms were obtained using "R" software, as shown in Figure 4 (97 variants distributed across cases are shown). The first dendrogram is shown at the top and comprises two broad clusters: LGD-like clusters (green, including 4/6 LGD) and HGD-like clusters (brown, including 2/6 LGD and 4/4 HGD). The second dendrogram is shown on the right and is divided into two main clusters: the "common variants between cases" (shown in pink) and the "rare variants between cases" (shown in blue). The orange frame depicts three *FAT1* missense variants. Interestingly, they appear together in homozygosity in the HGD-like cluster, but in the LGD-like cluster, they appear in heterozygosity or are not mutated (rs11939575; rs1877731; rs367863). The purple frame depicts a heterogeneous group of variants that tend to appear more in the HGD-like cluster and is composed of missense variants of *FAT1*, *PIK3CA*, *KDR*, *HNF1A*, and *GNAS*. Variants with a VAF≤ 0.1 were part of the "rare variants between cases" cluster, except for *ALK* variant rs1569156, a HIGH IMPACT deletion with a

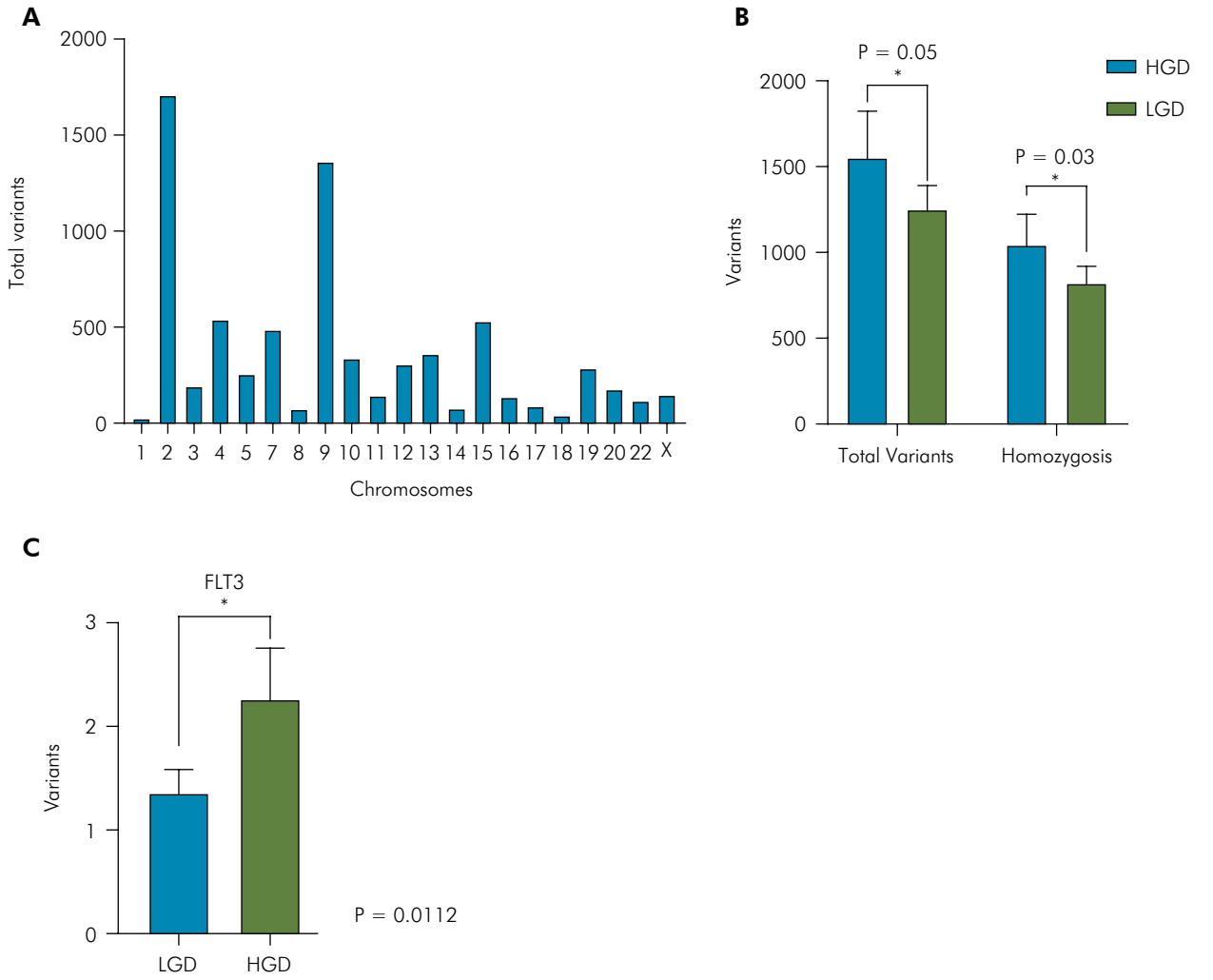

Figure 3. *A. Chromosome variants distribution*: CHR 2 and 9 genetic burdens can be seen; *B. Total mean variants comparison between HGD and LGD:* significantly different affected groups in the total variant count and homozygotic variant burden grouped; C. *FTL3 gene variants burden:* the only significantly affected gene between LGD and HGD in the gross variant analysis.

stop gain consequence (identified in homozygosity in 9/10 cases).

Filtered variants analysis

To describe probable somatic variants among the 97 already selected variants, we retained variants with VAF≤0.1 in the general population. 51 MODERATE and HIGH impact variants were identified as probable somatic mutations (Table 3 accompanied by clinical information for every case.

The *FAT1* gene was affected by different variants in out of 9/10 cases. Most *FAT1*+ cases were *CASP8*+ (6/9 cases), *NOTCH1*+ (3/9 cases) or *MLL4*+ (4/9 cases). FAT1 mutations in the same cases were accompanied by a specific nonsense variant in *UNC13C* (3/9 cases) and infrequently by a *TRPM3* heterozygotic variant (1/9). In the LGD cluster, case #5 did not show the same mutational signature and had *PIK3CA/ATM/FGFR2/JAK3*/MLH1+ variants without *FAT1/NOTCH1/CASP8* signature or *TP53*+. Only one HGD case had a *FAT1* mutation accompanied by a TP53 mutation (10%). Furthermore, the *ALK* SNP rs1569156 was present in all cases, and 9 out of 10 cases presented with homozygosis, predicting a STOP GAIN mutation. The mean number of probably pathogenic somatic variants per case was

9.9, and no significant differences between groups were observed.

Angiogenesis and TP53 pathways were mainly affected by pathogenic variants in the Panther analysis, while cellular and biological regulatory processes were the most affected in the biological process analysis (Table 4 and 5). Over-representation analysis showed that TP53 feedback loop 2 and

Table 2. Mutational burden of coding variants. 97 HIGH/MODERATE variants distribution between cases.

Genes	LGD 1	LGD 2	LGD 3	LGD 4	LGD 5	LGD 6	HGD 7	HGD 8	HGD 9	HGD 10	Mean variants
FAT1	10	16	12	4	10	13	14	16	10	7	11.2
ALK	5	5	5	5	5	5	3	5	2	4	4.4
HNF1A	0	5	7	6	6	3	4	3	6	4	4.4
TRPM3	2	4	2	2	1	2	1	3	2	3	2.2
UNC13C	1	3	3	1	1	1	1	1	2	1	1.5
CASP8	2	2	1	2	0	1	2	1	1	2	1.4
ERBB2	2	3	0	2	1	1	2	1	2	0	1.4
FLT3	0	2	1	1	1	0	2	2	2	3	1.4
MLL4	2	2	2	3	0	0	1	0	1	1	1.2
MET	2	0	1	1	1	1	1	1	2	1	1.1
TP53	1	1	1	1	1	1	1	1	2	1	1.1
RET	0	2	2	1	1	1	0	2	1	0	1
APC	1	1	1	1	1	1	1	1	1	1	1
KDR	0	0	3	0	1	1	1	0	0	2	0.8
EGFR	0	1	0	1	1	1	0	1	0	1	0.6
PDGFRA	1	1	0	0	1	1	1	0	1	0	0.6
GNAS	0	0	0	1	0	1	1	0	0	2	0.5
MLH1	1	0	0	1	1	0	1	0	0	1	0.5
ATM	0	0	1	0	1	0	0	0	0	1	0.3
CSF1R	0	0	0	1	0	1	0	0	0	1	0.3
NOTCH1	0	0	0	1	0	0	0	1	0	0	0.2
PIK3CA	0	0	0	0	1	1	0	0	0	0	0.2
SMO	0	1	0	0	1	0	0	0	0	1	0.3
CDKN2A	0	0	1	0	0	0	0	0	1	0	0.2
FGFR2	0	0	2	0	0	0	0	0	0	0	0.2
JAK3	0	0	0	0	1	0	0	0	0	1	0.2
MPL	1	0	0	0	0	0	0	0	0	0	0.1
SRC	0	0	0	0	1	0	0	0	0	1	0.2
ABL1	1	0	0	0	0	0	0	0	0	0	0.1
EZH2	0	0	0	1	0	0	0	0	0	0	0.1
FGFR1	0	0	0	0	0	0	0	1	0	0	0.1
NDUFB2	0	0	0	0	0	1	0	0	0	0	0.1
Genes affected	14	15	16	19	18	18	16	15	15	20	16.6
Total variants	39	49	44	36	36	37	40	40	36	39	39.6

Genes are shown in rows and cases are depicted in columns.

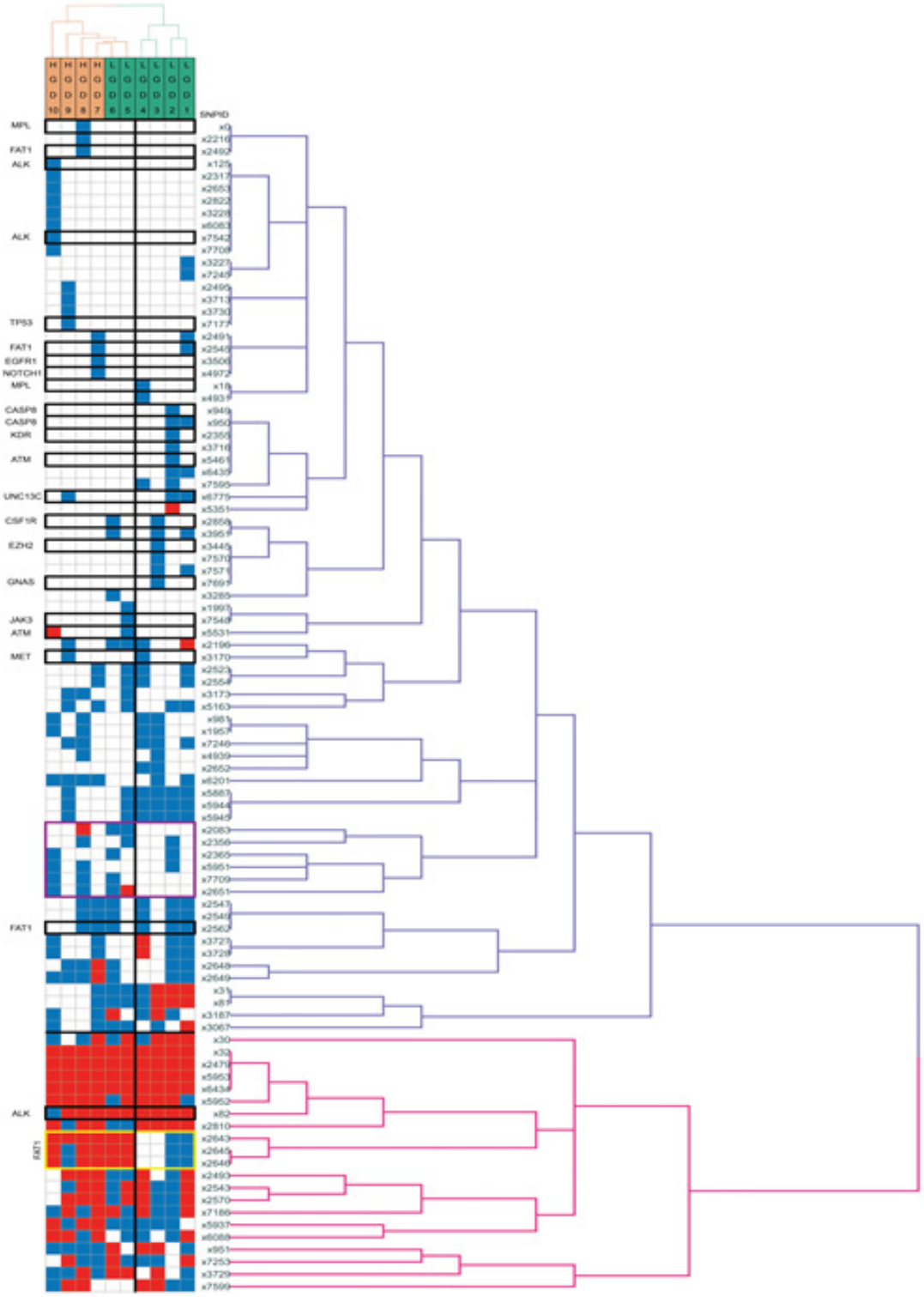

Figure 4. *Hierarchical clustering of HDG and LGD cases:* top dendrogram shows "HGD- like cluster" in brown and "LGD-like cluster" in green. Cases are shown in the top row. The right dendrogram shows the "*common variants between cases*" cluster in pink and the "*rare variants between cases*" cluster in blue. Variants are shown in rows and are identified by the SNPID number given in the study. Blue boxes are heterozygous variants, and red boxes are homozygous variants. The left side of the heatmap shows genes symbols, and black frames indicate specific variants with VAR ≥ 0.1. The green frame highlights *FAT1* variants that are mainly in HGD-like cluster. The purple frame highlights *PKI3CA, KDR, HNF1A, GNAS,* and *FAT1*, rare variants that are primarily present in the HGD-like cluster.

Table 3. Mutational landscape of HGD and LGD.

			LGD						HGD				
			1	2	3	4	5	6	7	8	9	10	
	Alcohol												
	Tobacco												
	Clinical diagnosis												
	Lesion site												
		%											% in OSCC*
OSCC genes	FAT1	90	1<	1<	1<			1<	1<	1<			40%
	CASP8	60			1<								34%
	MLL4	40				1<							12%
	NOTCH1	30											16%
	UNC13C	30											12%
	TP53	10											62%
	TRPM3	10											10%
Common mutated cancer genes	ALK	100											
	MET	40									1<		
	PIK3CA	30											
	CSF1R	30											
	ATM	30											
	HNF1A	30											
	JAK3	20											
	KDR	20											
	MPL	20											
	MLH1	20											
	SMO	20											
	GNAS	20											
	CDKN2A	20											
	PDGFRA	10											
	ABL1	10											
	FLT3	10											
	EZH2	10											
	NDUFB2	10											
	FGFR1	10											
	FGFR2	10											
	ERBB2	10											
	SRC	10											
	Total mutations		10	11	13	9	6	6	12	8	8	12	
	Genes affected		7	7	10	10	6	5	9	7	7	11	

	Alcohol
	Homogeneous Leukoplakia
	Erythroleukoplakia
	Verrucous Leukoplakia
	Tobacco (+)
	not known
	Tongue
	Buccal Mucosa

	Palate
	Gingival ridge
	Floor of mouth

	Homozygosis	Nonsense/frameshift/ splice site
	Heterozygosis	
	Homozygosis	Missense/inframe insertion-deletion
	Heterozygosis	
1<	Affected more than once	

LGD and HGD cases are depicted in the first line, and red numbers correspond to the HGD-like cluster cases. The first section presents the clinical diagnosis and history of alcohol and tobacco consumption. The second section describes gene alterations based on mutations with a VAR ≤0.1, accompanied by a percentage compared to OSCC[24]. RED/pink box: nonsense/frameshift and splice-site variants. Blue/Light blue box: "Missense/in-frame indel variants". "1<" define a gene affected more than once by different variants. The total mutation accounts for the lowest number of genes affected.

TP53 pathways were enriched 41 and 31 folds, respectively, followed by angiogenesis and apoptosis signaling pathways with 24 and 18 folds, respectively (Table 6).

Variants predicted to be pathogenic

Of the 51 variants shown in Table 3, 23 were predicted by software analysis or ClinVar deleterious annotations to have a greater chance of being

Table 4. Panther Biological process.

Biological process name	Genes	Total genes (%)	Hits (%)
cellular process (GO:0009987)	19	63.3	17.4
biological regulation (GO:0065007)	16	53.3	14.7
response to stimulus (GO:0050896)	11	36.7	10.1
signaling (GO:0023052)	11	36.7	10.1
developmental process (GO:0032502)	11	36.7	10.1
multicellular organismal process (GO:0032501)	10	33.3	9.2
metabolic process (GO:0008152)	10	33.3	9.2
immune system process (GO:0002376)	5	16.7	4.6
cellular component organization or biogenesis (GO:0071840)	4	13.3	3.7
localization (GO:0051179)	4	13.3	3.7
locomotion (GO:0040011)	4	13.3	3.7
cell population proliferation (GO:0008283)	3	10.0	2.8
biological adhesion (GO:0022610)	1	3.3	0.9

Table 5. Panther pathway. Only the 15 most represented pathways of 69 hits are shown.

Pathways	Genes	Gene hit against total # genes (%)	Gene hit against total # Pathway hits (%)
Angiogenesis (P00005)	6	20	8,7
p53 pathway (P00059)	4	13,3	5,8
Apoptosis signaling pathway (P00006)	3	10	4,3
Integrin signaling pathway (P00034)	3	10	4,3
p53 pathway feedback loops 2 (P04398)	3	10	4,3
FGF signaling pathway (P00021)	3	10	4,3
PDGF signaling pathway (P00047)	3	10	4,3
Cadherin signaling pathway (P00012)	3	0	4,3
Interleukin signaling pathway (P00036)	2	6,7	2,9
Huntington's disease (P00029)	2	6,7	2,9
Wnt signaling pathway (P00057)	2	6,7	2,9
VEGF signaling pathway (P00056)	2	6,7	2,9
Endothelin signaling pathway (P00019)	2	6,7	2,9
EGF receptor signaling pathway (P00018)	2	6,7	2,9
Gonadotropin-releasing hormone receptor pathway (P06664)	2	6,7	2,9

Table 6. Panther Overrepresentation Test (Released 20200407).

Variable	Homo sapiens (REF)					
Panther pathways	#	#	Expected	Fold Enrichment	+/-	Rawp-value
p53 pathway feedback loops 2	50	3	.07	41.70	+	5.97E-05
p53 pathway	87	4	.13	31.96	+	8.48E-06
Angiogenesis	172	6	.25	24.25	+	1.77E-07
Apoptosis signaling pathway	115	3	.17	18.13	+	6.39E-04
FGF signaling pathway	121	3	.17	17.23	+	7.38E-04
PDGF signaling pathway	145	3	.21	14.38	+	1.23E-03
Cadherin signaling pathway	160	3	.23	13.03	+	1.62E-03
Integrin signaling pathway	191	3	.27	10.92	+	2.67E-03
Unclassified	18243	12	26.25	.46	-	1.12E-09

Fisher exact test with false discovery rate correction data are shown.

pathogenic. The 23 variants are highlighted in the dendrogram (Figure 4) and are part of the "rare variant cluster," except for *ALK* variant rs1569156, which has already been described. Most of the studies did not show a specific pattern between the groups. These variants were reviewed manually for specific implications during data retrieval.

Discussion

We described the mutational landscape of OED based on OSCC-affected genes and common cancers. This is the first study to compare the OED genetic landscape with the histopathological diagnosis of dysplasia using a binary grading system. It is evident that the LGD and HGD exhibit similar mutational landscapes. A significantly higher total number of variants in HGD has been described already[12]. The exclusion of passenger variants provides insights into the molecular signature of these cases. The ≈7.9 affected genes per sample were less than those reported by Villa et al.[10] (13 genes per patient). They characterized histopathological "Keratosis of Unknown Significance" (KUS) and moderate to high-grade dysplasia, demonstrating similar landscapes but different risks and times of progression to OSCC.

OED is strongly associated with tobacco consumption, which induces mutations associated with nucleotide transversions, as revealed in a series of OSCC.[24] In our study, the variants tended to accumulate in CHR 2, 9, 4, 7, and 15. Zhang et al. showed that these CHR presented a loss of heterozygosity and microsatellite instability when evaluating OED and HNSCC.[19,31]

In the present study, a significantly higher number of variants in the *FLT3* gene were observed in the HGD group. *FLT3* is the most commonly mutated gene in acute myeloid leukemia and represents an important potential target in acute myeloid leukemia.[32] Mutations in *FLT3* have been found in OSCC[33]; however, there is little evidence. *FLT3* is a transmembrane ligand-activated receptor tyrosine kinase that plays an important role in the expansion of multipotent progenitor cells in the bone marrow.[32]

Hierarchical clustering showed well-delineated groups of "HGD-like cluster" and "LGD-like cluster" LGD-like clusters by divisive clustering of the imputed variants. As we can observe, there are 2 LGD cases classified in the "HGD-like cluster". This might be explained by the fact that the two cases already had molecular alterations consistent with HGD. However, these molecular characteristics do not affect the morphological architecture of dysplasia. This observation needs to be confirmed in a larger case series.

Specific variant distribution could not be demonstrated between "common variants between

cases" and "rare variants between cases" clusters due to a lack of a more representative sample size. Most of the homozygote variants in the "common variants between cases" cluster were missense variants that might probably be germline demographic polymorphisms unrelated to the development of OED. For example, most of the *HNF1A* gene variants (seven variants detected) are associated with diabetes mellitus (rs1169288, rs56348580, rs2464196, rs2464195, rs55834942, rs1169304, and rs1169305), but are still annotated in COSMIC related to cancer.

This study identified that *FAT1* was frequently mutated in OED, and *TP53* was relatively unaltered, although its pathway was significantly affected. *FAT1* is a known tumor suppressor gene that plays an essential role in cellular differentiation and cell adhesion.[34] Accordingly, *FAT1* had been described to be frequently mutated in an OSCC series by Maitra et al.[24] Our study found that *FAT1* was greatly affected by several mutations (9 of 10 cases), in contrast to earlier studies that found lower mutation frequencies in OED[10]. Regarding OSCC, it has been observed that *FAT1* mutation is present in 44% of all patients[24] and is considered a prognostic factor for OSCC progression and invasion capabilities.[35] The prognosis of PPOL for malignant transformation.[34] In addition, *FAT1* mutations are usually associated with other OSCC mutations.[24]. In the present study, positive *FAT1* cases also had *CASP8* (6 of 9), *MLL4* (4 of 9), *NOTCH1* (3 of 9), and *UNC13C* (3 of 9) mutations. Maitra et al. also described a specific subgroup of OSCC characterized by the presence of mutated *CASP8* with or without *FAT1,* representing 34% of OSCC cases, and had a better disease-free survival.[24] It seems that *FAT1* alone is insufficient to explain the progression to OSCC and might justify differences in signature percentage between studies.

TP53 has an average somatic mutation count of 62-84%, depending on the study.[19,24] Villa et al.[10] reported that *TP53* was mutated in 35% of KUS/severe dysplastic cases. Our series showed that *TP53* was mutated in 1 of 10 cases (10%). These differences could be explained by the fact that 6 of 10 cases were LGD, which is known to have a low malignant transformation rate compared to KUS and moderate/severe dysplasia. According to the results of a previous study, *CASP8/FAT1* positive samples might be less likely to transform into OSCC than those with *TP53* mutations.[24] These findings open a field of opportunities in diagnostic, prognostic, and treatment investigations, raising alarms for those cases that are *TP53*+. The few cases found to be *TP53*+ in this study might be explained by the fact that *TP53* could be a late driver mutation, only present when the risk of malignant transformation is high. In addition, TP53 might be indirectly affected by this pathway, as demonstrated by the Panther analysis.

ALK is another relevant gene that was found to be mutated in the whole series (90% homozygosis, 10% heterozygosis), presenting a specific stop-gain variant. This gene has not been reported in any OED study. However, it has been reported as a hypomethylated gene in advanced OSCC, and it has been described that its co-inhibition enhances the anti-tumor action of EGFR inhibitors.[36]

Future studies should consider that most of the limitations here derive from a lack of germline sequencing, and the inclusion of the latter is recommended. The manual filtering of SNPs may be an important source of bias. No copy number variation (CNV) analysis was performed, which limited the deeper characterization of the samples. RNA sequencing may be useful for better understanding protein translation patterns during malignant transformation.

Acknowledgments

We are grateful to members of the Diagnostic Service of the Dentistry School at Universidad Mayor, Chile, for their helpful discussions and reading of the text. The sequencing workflow and guidelines were performed by BioinfoGP, Spain. This research was funded by ANID FONDEQUIP grant number EQY220014, PEP I-2022016, and ANID SUBVENCIÓN A INSTALACIÓN EN LA ACADEMIA CONVOCATORIA AÑO 2022 grant number 85220068.

References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. CA Cancer J Clin. 2019 Jan;69(1):7-34. https://doi.org/10.3322/caac.21551
2. Personal habits and indoor combustions. Volume 100 E: A review of human carcinogens.Lyon: IARC Monogr Eval Carcinog Risks Hum; 2012.
3. Porter S, Gueiros LA, Leão JC, Fedele S. Risk factors and etiopathogenesis of potentially premalignant oral epithelial lesions. Oral Surg Oral Med Oral Pathol Oral Radiol. 2018 Jun;125(6):603-11. https://doi.org/10.1016/j.oooo.2018.03.008
4. Napier SS, Speight PM. Natural history of potentially malignant oral lesions and conditions: an overview of the literature. J Oral Pathol Med. 2008 Jan;37(1):1-10. https://doi.org/10.1111/j.1600-0714.2007.00579.x
5. Ng JH, Iyer NG, Tan MH, Edgren G. Changing epidemiology of oral squamous cell carcinoma of the tongue: A global study. Head Neck. 2017 Feb;39(2):297-304. https://doi.org/10.1002/hed.24589
6. Agrawal N, Frederick MJ, Pickering CR, Bettegowda C, Chang K, Li RJ, et al. Exome sequencing of head and neck squamous cell carcinoma reveals inactivating mutations in NOTCH1. Science. 2011 Aug;333(6046):1154-7. https://doi.org/10.1126/science.1206923
7. Stransky N, Egloff AM, Tward AD, Kostic AD, Cibulskis K, Sivachenko A, et al. The mutational landscape of head and neck squamous cell carcinoma. Science. 2011 Aug;333(6046):1157-60. https://doi.org/10.1126/science.1208130
8. Warnakulasuriya S. Clinical features and presentation of oral potentially malignant disorders. Oral Surg Oral Med Oral Pathol Oral Radiol. 2018 Jun;125(6):582-90. https://doi.org/10.1016/j.oooo.2018.03.011
9. Waal I. Potentially malignant disorders of the oral and oropharyngeal mucosa; terminology, classification and present concepts of management. Oral Oncol. 2009;45(4-5):317-23. https://doi.org/10.1016/j.oraloncology.2008.05.016
10. Villa A, Hanna GJ, Kacew A, Frustino J, Hammerman PS, Woo SB. Oral keratosis of unknown significance shares genomic overlap with oral dysplasia. Oral Dis. 2019 Oct;25(7):1707-14. https://doi.org/10.1111/odi.13155
11. Katabi N, Lewis JS. Update from the 4th Edition of the World Health Organization Classification of Head and Neck Tumours: What Is New in the 2017 WHO Blue Book for Tumors and Tumor-Like Lesions of the Neck and Lymph Nodes. Head Neck Pathol. 2017;11(1):48-54. https://doi.org/10.1007/s12105-017-0796-
12. Warnakulasuriya S, Ariyawardana A. Malignant transformation of oral leukoplakia: a systematic review of observational studies. J Oral Pathol Med. 2016 Mar;45(3):155-66. https://doi.org/10.1111/jop.12339
13. Neville BW, Day TA. Oral cancer and precancerous lesions. CA Cancer J Clin. 2002;52(4):195-215. https://doi.org/10.3322/canjclin.52.4.195
14. Warnakulasuriya S. Histological grading of oral epithelial dysplasia: revisited. J Pathol. 2001 Jul;194(3):294-7. https://doi.org/10.1002/1096-9896(200107)194:3<294::AID-PATH911>3.0.CO;2-Q
15. Woo SB, Grammer RL, Lerman MA. Keratosis of unknown significance and leukoplakia: a preliminary study. Oral Surg Oral Med Oral Pathol Oral Radiol. 2014 Dec;118(6):713-24. https://doi.org/10.1016/j.oooo.2014.09.016
16. Kujan O, Oliver RJ, Khattab A, Roberts SA, Thakker N, Sloan P. Evaluation of a new binary system of grading oral epithelial dysplasia for prediction of malignant transformation. Oral Oncol. 2006 Nov;42(10):987-93. https://doi.org/10.1016/j.oraloncology.2005.12.014
17. Nankivell P, Williams H, Matthews P, Suortamo S, Snead D, McConkey C, et al. The binary oral dysplasia grading system: validity testing and suggested improvement. Oral Surg Oral Med Oral Pathol Oral Radiol. 2013 Jan;115(1):87-94. https://doi.org/10.1016/j.oooo.2012.10.015
18. Speight PM, Khurram SA, Kujan O. Oral potentially malignant disorders: risk of progression to malignancy. Oral Surg Oral Med Oral Pathol Oral Radiol. 2018 Jun;125(6):612-27. https://doi.org/10.1016/j.oooo.2017.12.011
19. Cancer Genome Atlas Network. Comprehensive genomic characterization of head and neck squamous cell carcinomas. Nature. 2015 Jan;517(7536):576-82. https://doi.org/10.1038/nature14129
20. Cromwell I, Regier DA, Peacock SJ, Poh CF. Cost-effectiveness analysis of using loss of heterozygosity to manage premalignant oral dysplasia in British Columbia, Canada. Oncologist. 2016 Sep;21(9):1099-106. https://doi.org/10.1634/theoncologist.2015-0433
21. Rock LD, Rosin MP, Zhang L, Chan B, Shariati B, Laronde DM. Characterization of epithelial oral dysplasia in non-smokers: first steps towards precision medicine. Oral Oncol. 2018 Mar;78:119-25. https://doi.org/10.1016/j.oraloncology.2018.01.028
22. Santos JN, Sousa Neto ES, França JA, Diniz MG, Moreira RG, Castro WH, et al. Next-generation sequencing of oncogenes and tumor suppressor genes in odontogenic myxomas. J Oral Pathol Med. 2017 Nov;46(10):1036-9. https://doi.org/10.1111/jop.12598
23. Simen BB, Yin L, Goswami CP, Davis KO, Bajaj R, Gong JZ, et al. Validation of a next-generation-sequencing cancer panel for use in the clinical laboratory. Arch Pathol Lab Med. 2015 Apr;139(4):508-17. https://doi.org/10.5858/arpa.2013-0710-OA
24. Maitra A, Biswas NK, Amin K, Kowtal P, Kumar S, Das S, et al. Mutational landscape of gingivo-buccal oral squamous cell carcinoma reveals new recurrently-mutated genes and molecular subgroups. Nat Commun. 2013;4(1):2873. https://doi.org/10.1038/ncomms3873

25. Wood HM, Daly C, Chalkley R, Senguven B, Ross L, Egan P, et al. The genomic road to invasion-examining the similarities and differences in the genomes of associated oral pre-cancer and cancer samples. Genome Med. 2017 Jun;9(1):53. https://doi.org/10.1186/s13073-017-0442-0
26. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GR, Thormann A, et al. The Ensembl Variant Effect Predictor. Genome Biol. 2016 Jun;17(1):122. https://doi.org/10.1186/s13059-016-0974-4
27. Chandrani P, Kulkarni V, Iyer P, Upadhyay P, Chaubal R, Das P, et al. NGS-based approach to determine the presence of HPV and their sites of integration in human cancer genome. Br J Cancer. 2015 Jun;112(12):1958-65. https://doi.org/10.1038/bjc.2015.121
28. Team RC. A language and environment for statistical computing. 2018 [cited 2022 Mar 3]. Available from: https://www.R-project.org/
29. Mi H, Muruganujan A, Casagrande JT, Thomas PD. Large-scale gene function analysis with the PANTHER classification system. Nat Protoc. 2013 Aug;8(8):1551-66. https://doi.org/10.1038/nprot.2013.092
30. Cingolani P, Patel VM, Coon M, Nguyen T, Land SJ, Ruden DM, et al. Using Drosophila melanogaster as a Model for Genotoxic Chemical Mutational Studies with a New Program, SnpSift. Front Genet. 2012 Mar;3:35. https://doi.org/10.3389/fgene.2012.00035
31. Zhang L, Poh CF, Williams M, Laronde DM, Berean K, Gardner PJ, et al. Loss of heterozygosity (LOH) profiles: validated risk predictors for progression to oral cancer. Cancer Prev Res (Phila). 2012 Sep;5(9):1081-9. https://doi.org/10.1158/1940-6207.CAPR-12-0173
32. Nitika, Wei J, Hui AM. Role of Biomarkers in FLT3 AML. Cancers (Basel). 2022 Feb;14(5):1164. https://doi.org/10.3390/cancers14051164
33. Oikawa Y, Morita KI, Kayamori K, Tanimoto K, Sakamoto K, Katoh H, et al. Receptor tyrosine kinase amplification is predictive of distant metastasis in patients with oral squamous cell carcinoma. Cancer Sci. 2017 Feb;108(2):256-66. https://doi.org/10.1111/cas.13126
34. Chung CM, Hung CC, Lee CH, Lee CP, Lee KW, Chen MK, et al. Variants in FAT1 and COL9A1 genes in male population with or without substance use to assess the risk factors for oral malignancy. PLoS One. 2019 Jan;14(1):e0210901. https://doi.org/10.1371/journal.pone.0210901
35. Nishikawa Y, Miyazaki T, Nakashiro K, Yamagata H, Isokane M, Goda H, et al. Human FAT1 cadherin controls cell migration and invasion of oral squamous cell carcinoma through the localization of β-catenin. Oncol Rep. 2011 Sep;26(3):587-92.
36. Gonzales CB, De La Chapa JJ, Saikumar P, Singha PK, Dybdal-Hargreaves NF, Chavez J, et al. Co-targeting ALK and EGFR parallel signaling in oral squamous cell carcinoma. Oral Oncol. 2016 Aug;59:12-9. https://doi.org/10.1016/j.oraloncology.2016.05.007