Article:
Marín-Márquez, C. orcid.org/0000-0002-9444-7852, Adisa, A.O., Niklander, S.E. et al. (2 more authors) (2025) Genomic and transcriptomic analysis of ameloblastoma reveals distinct molecularly aggressive phenotypes. Modern Pathology, 38 (3). 100682. ISSN 0893-3952

https://doi.org/10.1016/j.modpat.2024.100682

Genomic and transcriptomic analysis of ameloblastoma reveals distinct molecularly aggressive phenotypes

Constanza Marín-Márquez [a, b] DDS, MSc, PhD, Professor of Oral Pathology and Medicine.

Akinyele O. Adisa [c] BDS, FMCDS, Associate Professor of Oral Pathology

Sven E. Niklander [d] DDS, MDent, MSc, PhD, Professor of Oral Pathology and Medicine.

Janine Kirby [e] BSc, PhD, Professor of Neurogenetics

Keith D Hunter [f] PhD, FRCPath, Professor of Head and Neck Pathology.

[a] Unit of Oral and Maxillofacial Medicine, Pathology and Surgery, University of Sheffield, United Kingdom, 19 Claremont Crescent, Broomhall, Sheffield S10 2TA.

[b] Facultad de Odontología y Ciencias de la Rehabilitación, Universidad San Sebastián, Puerto Montt, Chile, constanza.marin@uss.cl

[c] Department of Oral Pathology, Faculty of Dentistry, University of Ibadan and University College Hospital Ibadan, Nigeria, aoadisa@com.ui.edu.ng

[d] Unit of Oral Pathology and Oral Medicine, Faculty of Dentistry, Universidad Andres Bello, Viña del Mar, Chile, sven.niklander@unab.cl

[e] Sheffield Institute for Translational Neuroscience, Division of Neuroscience, School of Medicine and Population Health, University of Sheffield, Sheffield, UK, 385a Glossop Road, Sheffield, S10 2HQ, UK, j.kirby@sheffield.ac.uk

[f] Liverpool Head and Neck Centre, Molecular and Clinical Cancer Medicine, University of Liverpool, 6 West Derby Street, Liverpool, L6 8TX, keith.hunter@liverpool.ac.uk

Corresponding author

Constanza Marín-Márquez
Faculty of Dentistry and Rehabilitation Sciences
Universidad San Sebastián, Lago Panguipulli #1390
Puerto Montt, Chile, +56 65 232 5500
constanza.marin@uss.cl

Disclosure

No conflict of interest is declared by the authors.

Abstract

Ameloblastoma (AM) is a benign but locally infiltrative epithelial odontogenic neoplasm of the jawbones that may reach grotesque proportions and be highly recurrent if inadequately removed. The $BRAF^{V600E}$ mutation has been demonstrated as key molecular event in its development, nevertheless, there are many queries about its aetiopathogenesis that are yet to be answered. In this study, we aimed to integrate results from whole-exome sequencing (WES) and RNA-sequencing in AM samples to identify novel candidate genes that may be relevant to its pathogenesis. Thirteen-matched tumors were subjected to WES and RNAseq, respectively, to detect gene mutations and gene expression profile, along to the presence of gene fusions. Mutations were validated with sanger sequencing, whereas transcriptome results were validated with qPCR. Results from both molecular techniques were merged in order to identify novel candidate genes, that were biological validated with immunohistochemistry. $BRAF^{V600E}$ mutation was present in 62% of the analyzed cases, and each AM presented at least two or three mutations affecting cancer-driver genes. RNAseq showed different molecular subgroups associated with an aggressive and cancer-related phenotype (epithelial-mesenchymal transition-EMT and KRAS gene sets). No gene fusions were detected among the cases. *CDH11* and *TGM2,* novel genes associated with EMT in AM, were selected and validated in tissue. Both WES and RNAseq results showed gene alterations related to proliferation, cell differentiation, and metabolic processes. These results show that AM share many of the hallmarks of cancer secondary to the presence of oncogenic mutations or activation of oncogenic signaling pathways.

Introduction

Ameloblastoma (AM) is a benign but locally infiltrative epithelial odontogenic neoplasm of the jawbones that derives from the dental lamina rests. If untreated, it may reach

grotesque proportions, become disfiguring and risk damage to adjacent vital structures. It usually recurs if inadequately removed, and the gold standard treatment is a complete excision with negative margins [1] . It corresponds to the most common odontogenic neoplasm in all ethnic groups, representing 1% of all head and neck neoplasms. Its frequency shows geographic variation, representing the second most common odontogenic tumor (following odontomas) in America and Europe, and the most common in some countries of Africa and Asia [2] .

The BRAF p.V600E mutation and the activation of mitogen-activated protein kinases/extracellular signal-regulated kinases (MAPK/ERK) signaling pathways have been demonstrated as key molecular events in the pathogenesis of AM, especially in mandibular AM [3-5]. Dysregulation of the hedgehog (Hh) signaling pathway may participate in maxillary AM, where mutation of *SMO* is the most frequent mutation [1]. Nevertheless, there are many queries about AM aetiopathogenesis that are yet to be answered. This is a tumor than can reach huge dimensions, it has a slow but persistent growth pattern, infiltrating medullary bone and, although there is no evidence of cytological atypia at the histopathological level, there is a small percentage of cases that metastasize (< 2%) [6].

In this study, we aimed to integrate results from whole-exome sequencing (WES) and RNA-sequencing (RNAseq) in matched-ameloblastoma samples to identify novel candidate genes that may be relevant to its pathogenesis. Efforts should be made to reduce the significant morbidity associated with this tumor's surgical management. Therefore, the study of AM at the -omics level, has the potential to identify new molecules that could be pharmacologically targeted.

Materials and Methods

Tissue samples and processing

Thirty-three tumor samples from Nigeria stored in RNAlater™ solution (Invitrogen, Thermo Fisher Scientific) were provided by AA. Approvals were obtained from the ethics committee at the College of Medicine, University of Ibadan, Nigeria (UI/EC/15/0164), whereas MTA2718 and 08/S0709/70 are from the University of Sheffield Ethics Committee. All samples were obtained in accordance with the Helsinki Declaration. The original diagnoses included four ameloblastic carcinomas (ACs). After the pathological review of serial sections of the cases made by CM and KH, there was a disagreement with the initial diagnosis of these four cases; thus, a third internal blind examination was made by an oral pathologist . The final review diagnosis concluded that all samples corresponded to AM, which was the criteria for the following analyzes. Tumor-only samples were included in the current project as this corresponded to a retrospective evaluation of the cases.

DNA and RNA extraction

Total DNA and RNA was extracted from tumors by manual macro-dissection using QIAamp DNA Mini Kit (QIAGEN, Germany) and RNeasy Mini Kit (QIAGEN), respectively, following manufacturer's instructions and quantified using a NanoDrop spectrophotometer. A total of 40µl of DNA and of 25µl of RNA (both with at least 20ng/µl) from each tumor were sent to Novogene (Cambridge, UK) to perform whole-exome and RNAseq, respectively.

Whole-exome sequencing and Data Processing

The library construction and sequencing processes were performed at Novogene (Cambridge, UK) (Supplementary methods). In brief, the exome-enriched libraries were sequenced on the Illumina (NovaSeq 6000 S4 platform) with 100x sequencing depth (equivalent to 12Gb data output per sample) and paired-end reads with an average of 150 base pairs. Details about the processing of the raw sequencing data are in Supplementary Figure S1A. In summary, raw files were subjected to quality assessment, trimming and mapped to the Human Genome Version

GRCh38. The following analysis was performed with cloud-based Galaxy Europe[7] and consisted on duplicates removal, variant calling which includes filtering unreliable/poor quality variants by working with a depth of coverage (DP) >30 and base quality score for the aligned read (QUAL) >40 [8]. Finally, annotations of the variants were made with Ensembl's Variant Effector Predictor (VEP) [9] in Rstudio, following a script that converted the variant calling format (VCF) file to a mutation annotation format (MAF) (Supplementary methods).

Variants filtering and assessment workflow

Variant Filtration process was conducted in Rstudio (v.4.1.1) using a Maftools package [10], and a Variant Validation process with visual inspection and direct sequencing. The former included the visualization on the Integrative Genome Viewer (IGV) [11] tool and the latter consisted of assessing the mutations with Sanger Sequencing.

In Phase 1 Stage 1, all the common variants were removed based on population frequency. Following the recommendations of the Association for Molecular Pathology (AMP), American Society of Clinical Oncology and College of American Pathologists [12] for eliminating polymorphic or benign variants in the absence of paired normal tissue, a modified cut-off of minor allele frequency (MAF) of 0.1% was applied. We defined variants with population frequency ≤ 0.001 in the African population from both 1000 Genomes [13], ExAC [14] and GnomAD exome v2.1 [15] databases. In Phase 1 Stage 2 functionally insignificant variants were removed. Only variants predicted to be damaging (D) and disease-causing automatic (A) based on four *in silico* prediction tools such as SIFT (prediction = D)[16], PolyPhen2 (prediction = D) [17], Mutation Taster (prediction = D and A) [18] and CADD phred score ≥ 25 [19] were included. Phase 2 Stage 1 consisted of excluding all false positive variants annotated in ClinVar [20] as "benign" or "likely benign" (Figure 1).

Identification of somatic mutations and cancer driver genes

Maftools package was used in Rstudio (v.4.1.1) to summarise, analyze and visualize the mutation set, including the summary of the somatic mutational landscape of ameloblastoma, mutational signatures and the oncoplot of the mutated cancer driver genes. Phase 2 Stage 2 consisted of identifying somatic mutations based on the annotations given by VEP's output. For the identification of cancer driver genes, a list of genomic variants obtained from Phase 2-stage 1, was uploaded to the open-web platform Cancer Genome Interpreter (CGI) (https://www.cancergenomeinterpreter.org/home) [21]. Variants present in at least three tumors and/or based on previously known involvement of the implicated genes as cancer driver genes were selected for further validation (Figure 1).

Mutation signature analysis

Maftools package bases the analysis of mutation signatures, characterized by a specific pattern of nucleotide substitutions, on Alexandrov et al analysis [22].

Pathway enrichment analysis

For the pathway enrichment analysis and visualization, the protocol stated by Reimand et al. was followed [23] (Supplementary methods). The statistical threshold used was $p < 0.05$, and the data sources for the analysis were: Gene ontology (GO molecular function and GO biological process) and biological pathways (Kyoto Encyclopedia of Genes and Genomes-KEGG, Reactome and WikiPathways). Clusters of nodes were created using AutoAnnotate Cytoscape application. The names of the clusters were annotated manually for clarity.

***In silico* protein-protein interactions**

In silico protein-protein interaction (PPI) from the lists of somatic mutations (obtained in Phase 2 Stage 2) was performed in Cytoscape 3.9.1 Desktop using STRING database [24]. Clustering

was performed using MCL cluster mode in clusterMaker2 app [25] with a granularity parameter (inflation value) of 3, array source set to stringdb::score and edge weight cut-off set to 0.4.

RNA-sequencing and Data Processing

The data processing was performed with cloud-based Galaxy Europe [7]. FASTQ files were subjected to FASTQC and Cut adapt tools, for the quality control and the trimming of the low-quality sequences, respectively. The alignment was performed using HISAT2, with the Human Genome Version GRCh38 as a reference. The feature Counts tool was used to count the number of reads per annotated gene and DESeq2 for the Differential Gene Expression analysis (Supplementary Figure S1B) For the visualization of the results, heatmaps of the differential expressed genes (DEGs) among different phenotypes were created with heatmap2 with Euclidean distance method and complete clustering method. A significance threshold of p-value <0.05 and a fold change of 1.2 ($1.2 \leq FC \leq -1.2$) were used. The unsupervised hierarchical analysis was conducted using Cluster 3.0 for Windows [26] between the tumor samples and visualized using Java TreeView 1.2.0 for Windows [27]. The ameloblastoma expression data were submitted to the Gene Expression Omnibus microarray database (accession number GSE263944).

Gene set enrichment analysis (GSEA)

Gene Set Enrichment Analysis (GSEA) was conducted using GSEA software v4.1.0 from the Broad Institute (Cambridge, MA, USA) [28] and the molecular data was assessed with the gene set databases “Hallmarks” (h.all.v7.5) and “all curated gene sets” (c2.all.v7.4); available via the Mutational Signatures Databases (MSigDB). All databases were analyzed with a threshold of FDR ≤ 0.25 and nominal $p < 0.05$.

Pathway enrichment analysis

In general terms, this analysis was similar to the one above described for WES. The protocol stated by Reimand et al. was followed [23], with mild changes such as the definition of the gene list of interest (a ranked list from all available genes downloaded from the DEGs analysis made in Galaxy EU was used). For the pathway enrichment analysis, the gene set databases assessed included "all curated gene sets" (c2.all.v7.5.1) was used.

Fusion transcripts

STAR-Fusion was applied to detect fusion transcripts [29]. As part of the STAR-Fusion software, Fusion Inspector was used for the IGV validation and annotation. The criteria determined by Chang et al. were followed, in which only fusion read counts > 10 were included. The sum of the "Junction Read Count" and "Spanning Frag Count" gave the fusion read number. If a fusion rearrangement was present in the same chromosome, the distance between the left and right breakpoints had to be > 10 kb [30].

Molecular Analysis

Polymerase chain reaction (PCR) and Sanger Sequencing

Methods about DNA extraction and polymerase chain reaction (PCR) are detailed in Supplementary methods. After purifying the DNA from the agarose gel or directly from the PCR, 5 µl of DNA plus 5 µl of forward and reverse primers, were sent for Sanger Sequencing to Eurofins Genomics (UK), and the sequencing results were sent in an ABI format. These results were observed as chromatograms using the software Finch TV 1.4.0 version. Then, the sequencing was analyzed with the nucleotide database BLAST (http://blast.ncbi.nlm.nih.gov/Blast.cgi) to compare the query sequencing with primary biological sequences.

Quantitative PCR (qPCR)

To validate the RNA-sequencing profiles, five genes were selected for qPCR. Gene expression was quantified with a Rotor-gene Q real-time PCR cycler (Qiagen, UK) using SYBR Green and TaqMan methods (Supplementary Table S2). Three technical replicates were completed for each condition. The method used to calculate the relative fold gene expression of samples was the $2^{-\Delta\Delta Ct}$, normalized to B2M. Expression validation was calculated using Pearson in GraphPad Prism 10.2.3 for Windows, GraphPad Software, Boston, Massachusetts USA, www.graphpad.com.

Immunohistochemistry

Paraffin-embedded tissue sections (5µm) were dewaxed, rehydrated, and endogenous peroxidase was quenched with 3% hydrogen peroxide for 20 minutes. An antigen retrieval step was undertaken using a steamer with 0.01M Sodium citrate buffer. Sections were blocked using protein-free blocking solution for 30 minutes at room temperature before incubation with primary antibody (diluted in serum) overnight at 4ºC. Specific antibodies against CDH11 (Invitrogen 32-1700) and TGM2 (HPA021019, Atlas Antibodies) were used. Secondary antibody and avidin-biotin complex (ABC) provided with Vectastain Elite ABC kit were used in accordance with the manufacturer's instructions. Finally, 3,3-diaminobenzidine tetrahydrochloride (DAB) was used to visualize peroxidase activity, and the sections were counterstained with hematoxylin, dehydrated and mounted in DPX and imaged by light microscopy. Results were considered positive when more than 40% of the total epithelial cells within a single sample were stained with CDH11 or TGM2 [31].

Results

Ameloblastoma demographics

From the original cohort composed of 33 tumors, thirteen were whole-exome sequenced after histopathological re-evaluation by pathologists (KH and CM) and confirming that over 90% of the tissue represented tumor [32]. Details about the demographic and clinical information of the cases are shown in Table 1. The mean age for the thirteen tumors was 36.5 years, from which a case affecting a 4-year-old and a 68-year-old male represented the youngest and the eldest, respectively. All thirteen cases of AM occurred in the mandible, from which four cases were reported as bilateral. Only one case corresponded to an unicystic AM, whreas the remaining were diagnosed as AM, conventional. The mean evolution time was 24 months. Regarding the radiological features, ten tumors were described as multilocular, and three as unilocular, ten tumors perforated the cortical bone, whereas 12 cases provoked root resorption of the adjacent tooth/teeth. Eight patients were treated with a segmental resection, three with a hemi-mandibulectomy and two with a mandible resection. Of the 13 cases, seven received an immediate graft reconstruction. Three patients were lost during the follow-up. Representative microphotographs of the tumors are shown in Figure 2.

Characteristics of the mutations

Variant filtering

The original recommendation of the Association for Molecular Pathology (AMP)[12] to eliminate polymorphic variants in the absence of paired normal tissue is a primary cut-off minor allele frequency (MAF) of 1% (0.01). However, when this was applied, the number of remaining variants was too high. Therefore, a stringent threshold was applied (MAF ≥0.001). Based on the population frequency, the WES pipeline was critical to remove variants that corresponded to polymorphisms or were benign/tolerated. Each sample typically began with ~ 26,000/27,000 variants before the filtering scheme (Supplementary Table S3). Phase 1 Stage 1 of the pipeline (Figure 1**)** removed approximately ~25,000/26,000 variants. Phase 1 Stage 2 removed

~1,100/1,200 variants that were predicted to have no functional significance. Around ~600 variants per sample were not annotated by these prediction tools; but they were not removed to avoid losing possible interesting and novel variations and for posterior analysis. Phase 2 Stage 1 removed ~120 variants reported as "benign/likely benign" in ClinVar or "synonymous". At this point, the mean number of annotated and non-annotated variants per sample was 21 and 472, respectively. From these, according to VEP's output a total of 202 annotated somatic variants were identified across the 13 tumors (Supplementary Table S4).

Missense mutations represent the most common mutation type in ameloblastoma

Following the previous analysis, only annotated somatic variants were included in this assessment. Missense mutations represented the most common type and a median of 13 somatic mutations per sample was identified (Supplementary Figure S2).

Mutational Signatures in ameloblastoma correlate with DNA mismatch repair

Signature analysis was performed in the Maftools package, and the three best-matched mutational signatures were estimated using non-negative matrix factorization (NMF) and compared against the updated version3 of 60 COSMIC signatures (single base substitution – SBS-signatures) (https://cancer.sanger.ac.uk/signatures/sbs/). Figure 3A shows the mutational signatures with their respective heatmaps of cosine similarities. Signature 1 found its match with SBS31, in which prior chemotherapy treatment with platinum drugs is proposed as an etiology. Signatures 2 and 3 found their best matches in SBS6, in which the proposed etiology is defective DNA mismatch repair.

Enriched biological pathways from mutated genes in ameloblastoma are mainly associated with cell proliferation, cell differentiation and regulation of cellular metabolic processes

The web-server g:Profiler was used to run the analysis from the list of genes obtained from Phase 2 Stage 1 of the Variant Filtration Scheme. The resultant file was uploaded into

Cytoscape software using the EnrichmentMap application. The obtained map was manually edited to facilitate its interpretation. Figure 3B shows the resulting enrichment map and the enriched Gene Ontology (GO) gene sets from the 13 whole-exome sequenced tumors. The final module included 25 nodes organized in 6 clusters, associated with the metabolic processes, apoptosis, cell differentiation, regulation of cell proliferation, immune response, and cell adhesion Supplementary Table S5 contains the biological process information of each GO gene set.

Protein-protein interaction (PPI) networks among somatic variants

Networks of PPIs give a framework for understanding the biological process and the molecular mechanisms occurring in the pathogenesis of a disease. Main KEGG pathways are clustered in light blue (MAPK pathway), blue (proteoglycans in cancer), red (osteoclast differentiation) and green (ErbB signaling pathway) (Figure 3C and Supplementary Table S6). Information about KEGG was retrieved from http://www.webgestalt.org/.

Each ameloblastoma carries an average of three mutations affecting cancer-driver genes

The list of genomic variants obtained from Phase 2 Stage 1, was uploaded to the open-web platform Cancer Genome Interpreter (CGI). The results were compared with the annotations from the Cancer Gene Census (CGC) of the Catalogue of Somatic Mutations in Cancer (COSMIC) [33] and OncoKB (https://www.oncokb.org). A total of 32 driver-cancer genes were identified across the thirteen tumors (Supplementary Table S7). From these, 7 genes (*ASPM, CLSPN, SERPINB3, SIRPA, SVEP1, PABPC3* and *MUC16*) were excluded because they were not annotated in the consulted *in silico* tools, or they were annotated as tolerated/benign. *SMO* was manually added because of its known involvement in the pathogenesis of AM. The *SMO* variant accomplished most of the criteria, however its CADD score of 24.9 eliminated it at the end of Phase 1 Stage 2. A total of 26 cancer driver genes were

identified across the 13 samples, including *BRAF*, *CTNNB1*, *KMT2D, BCOR, ERBB3, PIK3R1* and *ALK*, among others (Figure 3D). Information about the signaling/related pathways of the driver genes is available in Supplementary Table S8.

Variants Validation

Three genes (*BRAF, CTNNB1* and *KMT2D)* were validated because they were mutated in at least three tumors and they have a known involvement in tumor progression. Firstly, variants were inspected with the Integrative Genomics Viewer (IGV) [11] to assess the presence of strand bias (Figure 4 A, C and E). Supplementary Table S9 shows details of the total counts and the percentages of the reads of each variant and their genotype. Secondly, all variants were validated with Sanger Sequencing (SS) to confirm the mutations (Figure 4 B, D and F). All *BRAF* and *KMT2D* mutations were validated, showing a concordance between WES and SS of 100% (Figure 4 B and D). In the case of C*TNNB1,* 1/3 mutations was not validated with SS, thus represented a false positive (tumor 13) (Figure 4F).

RNA-sequencing

Despite an average RNA Integrity Number (RIN) number of 2.4 ± 0.98 SD, only one tumor failed the quality control test performed by the company (T31). Table 1 shows the demographics, clinicopathological features and *BRAF* status of the included samples. To validate the RNA-seq profiles five genes were selected for qPCR. Overall, a high correlation between RNA-seq and qPCR results was found ($R^2 = 0.7805$; Supplementary Figure S3).

Ameloblastoma shows distinct transcriptome profiles

The relationship among the 12 tumor samples was first studied via unsupervised hierarchical cluster analysis based on normalized transcript intensities and by applying correlation as a distance method and average linkage method for clustering. Two distinct

molecular clusters were found in AM. Cluster 1 included five tumors (T9- T13), whereas cluster 2 included seven tumors (T8, T16, T17, T19, T26, T29, T30). (Figure 5A).

A molecular sub-group of ameloblastoma shows enrichment of cancer-related gene sets

GSEA conducted between cluster 2 and cluster 1 showed that 2939 out of the 4754 gene sets in the "c2: curated gene sets" (c2.all.v7.5.1) database were enriched in cluster 2. One hundred and twelve gene sets were significantly enriched at FDR < 25% and two hundred and seventy-nine at a nominal p- value <5. In general, cluster 2 showed enrichment of the gene sets associated with regulation of DNA replication, protein synthesis and oxidative phosphorylation (Figure 5B). On the other hand, in cluster 1, one hundred and thirty-seven gene sets were significantly enriched at FDR < 25% and three hundred and eighty-four at a nominal p value <5 , mostly related to carcinogenesis (Figure 5C) and EMT (Figure 5D and Supplementary Table S10).

BRAF mutated vs. and BRAF WT tumors

A DEG analysis was performed comparing BRAF mutated (BRAFm) and BRAF WT tumors. From a total of 28,394 annotated genes, 586 were differentially expressed at a p-value <0.05 and a fold change 1.2. To visualize the relationship between the most differentially expressed genes in AM with and without BRAF p.V600E mutation, a heatmap clustering of the top 100 genes is shown in Figure 5E.

BRAF WT ameloblastomas show enrichment of the epithelial-mesenchymal transition and KRAS-signaling gene sets

Gene set enrichment analysis (GSEA) conducted between BRAF WT and BRAFm showed that 31 out of the 50 gene sets in the Hallmarks database were enriched in BRAF WT. Twelve gene sets were significantly enriched at FDR < 25% and at nominal p value<5%, including epithelial-mesenchymal transition (EMT) and KRAS-signaling gene sets

(Supplementary Table S11). Figure 5F shows representative hallmarks enrichment plots in BRAF WT tumors. Conversely, 19 out of the 50 gene sets in the Hallmarks database were enriched in BRAFm. However, zero gene sets were significantly enriched at FDR < 25% and only two gene sets were significantly enriched at nominal p <5% (data not shown).

Absence of Gene Fusions

None of the 12 tumors subjected to RNA-sequence passed the recommended criteria of Chang et al. [30] for fusion reads >10 and intrachromosomal rearrangement distance cutoff of >10kb. Therefore, no fusion transcripts were identified in our AM cohort (Supplementary Table S12).

Candidate genes identification and validation

As the EMT-gene set appeared enriched when comparing the tumors according to their *BRAF* status and the unsupervised clustering analysis, the focus was analyzing the overexpressed genes among this gene set. Within these genes, cadherin-11 (*CDH11*) and transglutaminase II (*TGM2*) were chosen for biological validation because of their known role in tumorigenesis (breast, pancreatic, gastric cancer and thyroid, hepatocellular and ovarian cancer, respectively) and unknown role in AM pathogenesis.

A validation cohort of AM samples from the Department of Oral Pathology of the University of Sheffield was used. Immunohistochemical analyzes were carried out to examine the expression of CDH11 and TGM2 in human AM specimens. Results were considered positive when more than 40% of the total epithelial cells within a single sample were stained with against CDH11 or TGM2 [31]. Cytoplasmic/membrane staining for CDH11 was observed to be positive in 10/15 tumors (67%), whereas it was hardly detected in adjacent oral non-tumorous stratified squamous cell regions. The monoclonal antibody showed specific cytoplasmic immunoreactivity for CDH11 in the neoplastic cells. Regarding the

immunostaining pattern, positive staining was seen in both the periphery and central cells of the islands and cords of the neoplastic epithelium (Figure 6 A-F). On the other hand, cytoplasmic staining for TGM2 was observed in 5/11 tumors (46%), with no expression in the normal oral mucosa. Both weak, moderate, and strong cytoplasmic positivity was observed in the neoplastic and stromal cells (Figure 6 G-K).

Discussion

Understanding the biology and behavior of AM is a huge challenge; however, with the advancement of molecular biology techniques, there is a better understanding of the molecular basis of this tumor. Nevertheless, many queries remain unknown, such as what drives the pathogenesis in *BRAF* wild-type cases? Given the higher incidence in African populations [34, 35], are they molecularly different?

This study focused on characterizing the coding mutations and the transcriptome of AM to identify novel genes and molecular pathways that may be relevant to its aetiopathogenesis. To our knowledge, this is the first time that a series of African AMs have been subjected to next-generation sequencing (NGS) techniques to characterize their genome. No other whole-exome and RNA-sequencing in matched-samples have been published before. Moreover, no transcriptome comparison between AM harboring BRAF p.V600E and AM BRAF WT has yet been reported.

Criteria for including tumors and candidate gene selection

Different methods have been used when consulting literature about NGS in AM, mainly working with targeted NGS panels, and less frequently with WES [36]. The two reports that have worked with WES, did not validate the variants [37, 38]. On the other hand, most of the studies that have worked with targeted panels, validated the mutations with direct sequencing or TaqMan allele-specific qPCR [36].

There is no unique method for the variant filtration when working with NGS. Although there are computational approaches for raw data processing, filtering thousands of genes to identify a group of candidates relevant to a specific study is a big challenge. Therefore, the workflows for prioritizing genetic variants differ from one study to another. In addition, when working with only-tumor samples, careful interpretation must be done in case novel variants are identified. This is because, in the absence of normal tissue, the complete certainty of excluding germline variants cannot be assumed.

In the current study we stated that variants or genes present in at least three tumors and/or based on previously known involvement of the implicated genes as driver cancer genes will be further validated. Sweeney et al. selected candidate genes based on their presence on both tumor samples and/or involvement in gene/pathway implicated in tissue proliferation, differentiation, or neoplasia [5]. On the other hand, Shi et al. established a categorization method in which they classified mutations as "rare" when they passed the population frequency filter and as "severe" if they were predicted to be damaging by at least two *in silico* tools. They included mutations categorized as rare and severe in the final list [38]. Similarly, Guan et al. filtered variants by population frequency and only included variants predicted as deleterious by five of the twelve prediction tools available [37].

Ameloblastoma coding mutations with a focus on cancer driver genes

Consistent with earlier reports [3, 5, 32, 39], the current investigation showed that BRAF p.V600E was the most frequent mutation, affecting 62% of the cases, followed by other less frequent mutations that occurred in *BRAF* background. Earlier reports have shown mutations affecting other genes from the MAPK pathway (*KRAS, NRAS, HRAS,* and *FGFR2*) with a tendency to be mutually exclusive with *BRAF* [3, 5]. Nevertheless, our research did not find mutations affecting other genes from the RAS family or *FGFR2*.

After a strict variant filtration workflow, we determined that each AM showed a mean number of three mutations affecting cancer driver genes. This allowed us to suggest that the mutation load of cancer driver genes in AM is small and that the mutational profile of AM is relatively stable. Shi et al. reported an unbalanced distribution of somatic mutations in four AMs that ranged from 5 to 37 mutations per tumor [38]. These differences may be because of the variant filtration workflow used in their study and its focus on somatic mutations in general. A recent article aimed to remove drivers from germline mutations [40]. Working with a comprehensive genomic profiler (CGP), the authors observed, from a total of fourteen samples, 6 BRAF p.V600E (42%), 5 *PIK3CA* (36%), 5 *SMO* (36%), 4 *FGFR2* (29%), 1 *EGFR* and *ROS1* mutations (7% each) [40]. BRAF p.V600E mutations were mutually exclusive with *SMO*, *FRGR2*, *KRAS*, and *NRAS* mutations. The samples assessed in Gates et al. study corresponded to seven primary AMs and seven "complex AMs" described as clinically advanced, unresectable, or metastatic. In addition, the median age was 64 years old, compared to our results, which were 36.5 years old. The median number of mutations per sample was four. Their results may suggest that even in long-term tumors, AM's genomic landscape remains relatively stable. An important difference between Gates's findings and our results is that the six BRAF WT tumors harbored a median number of six mutations per sample, compared to two mutations per BRAF WT AMs in our study. Differences could be related to the methodology since we worked with WES; thus, our analysis was subjected to a variant filtration process, and the variants were validated. On the other hand, they worked with a panel of cancer genes in which variants were not validated. Therefore, larger cohorts subjected to the same methodology are needed to compare the results.

Similar to Gates et al., as shown in Figure 3C and D, our findings highlight the relevance of taking into account that each AM may be at least double or triple-mutated, and

these mutations affect cancer driver genes that are targets of Food and Drug Administration (FDA) approved drugs or of small molecule compounds.

KMT2D: A novel gene with unknown participation in AM pathogenesis

Our investigation identified that 23% (3/13) of the AMs harbored likely loss-of-function mutations in *KMT2D*, corresponding to a tumor suppressor gene (TSG). Three articles have reported mutations affecting this gene in 2/10 [37], 1/4 [38], and 5/14 [40] AMs. There may be a detection bias, and this mutation could be underestimated in AM pathogenesis. The identification of the mutations in *KMT2D* in this study and the other reports was made using WES or CGP, in contrast to other publications that have worked with targeted panels (such as those with "hot spot" regions that are frequently mutated in human cancer genes) in which *KMT2D* is not included within the arrays [5, 32, 41-44].

KMT2D represented the most heterogeneous mutated gene in the current series of cases since it presented three variants p.Q3293*, p.P648Tfs*2, p.Y389* (two nonsense mutations affecting tumors 9 and 16, and one frameshift insertion in tumor 30) (Supplementary Table S7), leading to a truncated protein. Only KMT2D p.Q3293* was annotated by Mutation Taster as disease-causing with a CADD score of 38. However, the remaining variants were not annotated by any consulted tools. As the gene corresponds to a TSG, the mutation effect is predicted to be a loss of function. Previous reports about mutations affecting *KMT2D* in AM have indicated that these mutations usually co-occur with others, in most cases in *BRAF* background. In our study, 2/3 cases with mutations affecting this gene occurred in *BRAF* background. It is important to consider that *KMT2D* is a large gene, and as such, its fragments may map to the reference sequence with less accordance than shorter reads, leading to false positive results [45]; therefore, it is important to validate the mutations. Of the four studies that have reported *KMT2D* mutations in AM, our study is the only one that has validated the

variants. Combining our results with the previous articles, we suggest *KMT2D* may be an interesting candidate gene with an unknown role in AM pathogenesis that needs to be further characterized in cohorts from different ethnic backgrounds. Further research is required to confirm its pathogenicity in the tumorigenesis of AM.

Our results suggest no molecular differences regarding mutations in African AMs. Published literature has worked with cohorts from the US, Brazil, Finland, European countries, Japan, China, and Singapore, among others [3, 5, 32, 39]. All of them have demonstrated the same trend: BRAF p.V600E is the most frequent driver mutation in this tumor, regardless of ethnic background.

Ameloblastoma has distinct transcriptome phenotypes associated with tumorigenesis

Cluster 1 and BRAF WT molecular subgroups were associated with an aggressive and cancer-related phenotype. Results showed that the EMT gene set was activated in both groups, and *KRAS* gene set was activated in the BRAF WT tumors. Distinct phenotypes in AM have been reported before by Hu et al., who identified two distinct phenotypes associated with pre-secretory ameloblast and odontoblast. [46]; and Heikinheimo et al. reported two distinct sub-clusters in AM related to dental epithelial genes [47].

The EMT is a process in which polarised epithelial cells assume a mesenchymal cell phenotype, acquiring enhanced migratory capacity, invasiveness, resistance to apoptosis, and increased extracellular matrix (ECM) component production. This switch allows them to move away from their epithelial context and to integrate into surrounding tissue and remote locations, a critical process in (epithelial) cancer progression [48]. EMT in AM has been suggested by immunohistochemistry before [49-52]; however, the mechanisms that activate this process are still poorly defined. Experimentally, both IL-8 (by activating β-catenin) and $FOSL_1$, a component of the AP-1 transcriptional complex, induce EMT in AM [53, 54]. On the other hand, one study

concluded that the neoplastic epithelial cells in AM, showed higher expression of mesenchymal markers without evident morphological mesenchymal phenotype, suggesting partial EMT in this tumor [51].

Cluster 2 and BRAFm tumors showed enrichment of the genes related to oxidative phosphorylation (OXPHOS). This process seems to be the preferred way these molecular clusters regulate energy supply. OXPHOS is the final stage in cellular aerobic respiration, in which ATP is generated within the mitochondria in the presence of oxygen [55]. Classically, it has been postulated that in a cancer context, cells prefer to generate energy through "aerobic glycolysis" rather than mitochondrial respiration, although it provides much less energy (Warburg effect)[55]. This switch is considered an early event in oncogenesis, such as an immediate consequence of an initial *BRAF* oncogenic mutation in melanoma [56].

Nevertheless, during the last decades, increased expression of mitochondrial genes involved in OXPHOS has been observed in many malignant neoplasms and metastasis [57]. Our findings allow us to hypothesize, for the first time, that enhanced mitochondrial respiration may have a role in AM's tumorigenesis. These results need further validation since certain cancers, such as melanomas resistant to the BRAF kinase inhibitor therapy, have shown a good response at the *in vitro* level when treated with OXPHOS inhibition therapy [58].

Overall, WES and RNAseq findings raise the question of the exact role of the (oncogenic) activation of the MAPK signaling in the etiopathogenesis of AM. On one hand, it is known that the implications of oncogenic mutations are context-dependent and that the MAPK/ERK pathway is dynamic and shows complex cross-talk with other cellular regulatory pathways [59]. On the other hand, the activation of the MAPK/ERK pathway is linked to cell senescence in other tumorigenic processes, which may inhibit or contribute to oncogenesis [60].

Further *in vitro* and *in vivo* studies may help elucidate these biological events' role in AM pathogenesis.

CDH11 and TGM2 are expressed in human ameloblastoma tissue

One of our aims was to identify candidate genes relevant to AM aetiopathogenesis. Consequently, we selected *CDH11* and *TGM2,* novel genes associated with EMT in AM. Both have been previously associated with the acquisition of EMT in cancer [61, 62].

Cadherin-11 (*CDH11*), or osteoblast cadherin (OB-Cadherin), is a type II transmembrane protein, a member of the cadherin adhesion family, that mediates Ca^{2+} dependent homophilic interactions between cells through the formation of intercellular connections and allowing cell migration. Its expression is usually restricted to mesenchymal cells such as smooth muscle cells, stromal cells, osteoblasts, and endothelial cells [63]. In triple negative breast cancer and pancreatic cancer, high expression is associated with poor prognosis, favoring metastasis [61]. Conflicting results regarding CDH11 transcript expression in AM exist, and no investigation about its role has been performed. Heikinheimo et al. found that mRNA CDH11 was underexpressed in AM compared to fetal teeth [64]. Lim et al. observed that mRNA CDH11 was not differentially expressed when comparing the expression pattern of AM versus two dentigerous cysts [65]. On the other hand, Kondo et al. reported an overexpression of mRNA CDH11 in AM, compared to normal oral tissue from gingiva [66]. These three studies worked with a cDNA microarray; thus, the differences between these results are probably due to using different control specimens.

The second candidate chosen in the current study was Transglutaminase II (*TGM2*). *TGM2* is a calcium-dependent cross-linking enzyme, a member of the transglutaminase family, that catalyzes protein modification (post-translational modification). It is involved in many physiological and pathological processes, such as extracellular matrix (ECM) stabilization, cell

differentiation, signal transduction, apoptosis, maintenance of oral cancer stem cell survival, and invasive and metastatic behavior[62]. In AM, only one article has reported under-expression of mRNA *TGM2* in a cohort of tumors compared with tooth germs [46].

This is the first time these genes have been further explored and validated in AM. We demonstrate that CDH11 and TGM2 are variably expressed in AM tissue. With CDH11, no differences in the expression of the epithelial cells from the periphery or the center of the neoplastic islands or cords were observed. Nevertheless, it has been postulated that the expression level of cadherins may vary during various cellular processes [67]. Therefore, we can not rule out the possibility of CDH11 differential expression during AM tumorigenesis. Regarding TGM2, the immunostaining suggested a higher immunoexpression at the peripheral cells of the epithelial islands or cords. However, a larger cohort is needed for both proteins to confirm these results.

Limitations

One limitation of this study was the degradation status of the RNA. The average RIN of the tumors sequenced in the current study is similar to the RIN of FFPE samples (<2.5) [68].A very high correlation between protein-coding transcripts has been reported when comparing fresh, frozen, and FFPE tissue [69]. Library results prepared with a low RIN can be used; however, they should be carefully interpreted, and validation at the individual gene level is recommended. Another limitation of the current investigation was the lack of matched-normal samples to compare with AM's mutational landscape and gene expression profile. Even when AM does not have a matched-normal oral tissue, the most similar controls are tooth germs; however, these are very difficult to obtain. Other controls that have been used are normal oral mucosa and other odontogenic lesions, including cysts and tumors. Ideally, when performing WES, a blood sample or NOM from the patient should be collected, allowing the discrimination

of germline mutations from somatic mutations. For the RNAseq, comparing the gene expression profile from the lesion with a normal-matched sample is not the only way, and the selection of the optimum control sample depends on the experimental design and research questions that are intended to be answered.

Conclusions

Our results confirm that AM has a stable mutational profile, consistent with what is expected for a benign neoplasm. However, both WES and RNAseq results showed gene alterations related to proliferation, cell differentiation, and metabolic processes. These results support the statement that benign tumors share many of the hallmarks of cancer secondary to the presence of oncogenic mutations or activation of oncogenic signaling pathways.

To our knowledge, this is the first time that the transcriptome of AM with BRAF p.V600E and BRAF WT tumors have been compared. Therefore, our results confirmed that genes from the MAPK pathway are overexpressed in AM, regardless of the *BRAF* mutation status. Furthermore, from the transcriptome point of view, it can be suggested that BRAF WT tumors possess an aggressive profile compared with their *BRAF* mutated counterpart (with activation of the EMT-related genes mainly); however, larger cohorts are needed to confirm these findings. The molecular heterogeneity of AM should be further characterized and correlated with clinical features to determine the impact on the treatment response and the prognosis of the disease. Overall, our findings suggest that molecular screening needs to be considered, at least for aggressive and advanced cases, when planning a neoadjuvant targeted therapy, and due to the molecular heterogeneity of AM and BRAF WT cases, targeted-tailored treatment should be considered to diminish morbidities in certain cases.

Ethics Approval / Consent to Participate

Approvals were obtained from the ethics committee at the College of Medicine, University of Ibadan, Nigeria (UI/EC/15/0164), whereas MTA2718 and 08/S0709/70 are from the University of Sheffield Ethics Committee.

Author Contribution Statement

CMM Conceptualization, laboratory experimentation and data analysis, writing original draft, writing review and editing. KH and JK Conceptualization, writing review and editing, supervision. AA: collation of clinical cohort, ethical approval and writing review. SEN: initial laboratory experimentation and writing review.

Funding Statement

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. CMM was funded by the National Agency for Research and Development (ANID)/Doctorado Becas Chile/ 2019, Grant/ Award Number 72200035.

Data Availability Statement

Most of the datasets analyzed during the current study are included in this published article and its supplementary information files. The ameloblastoma expression data is available at the Gene Expression Omnibus microarray database (accession number GSE263944). Raw data from whole-exome sequencing is available from the corresponding author upon reasonable request.

References

1. WHO Classification of Tumors Editorial Board. Head and neck tumors. Lyon (France): International Agency for Research on Cancer; 2022. (WHO classification of tumors series te.
2. Siriwardena B, Crane H, O'Neill N, Abdelkarim R, Brierley DJ, Franklin CD, et al. Odontogenic tumors and lesions treated in a single specialist oral and maxillofacial pathology unit in the United Kingdom in 1992-2016. Oral Surg Oral Med Oral Pathol Oral Radiol. 2019;127(2):151-66.

3. Brown NA, Rolland D, McHugh JB, Weigelin HC, Zhao LL, Lim MS, et al. Activating FGFR2-RAS-BRAF Mutations in Ameloblastoma. Clinical Cancer Research. 2014;20(21):5517-26.
4. Kurppa KJ, Caton J, Morgan PR, Ristimaki A, Ruhin B, Kellokoski J, et al. High frequency of BRAF V600E mutations in ameloblastoma. J Pathol. 2014;232(5):492-8.
5. Sweeney RT, McClary AC, Myers BR, Biscocho J, Neahring L, Kwei KA, et al. Identification of recurrent SMO and BRAF mutations in ameloblastomas. Nat Genet. 2014;46(7):722-5.
6. Hosalkar R, Saluja TS, Swain N, Singh SK. Prognostic evaluation of metastasizing ameloblastoma: A systematic review of reported cases in literature. Journal of Stomatology, Oral and Maxillofacial Surgery. 2021;122(2):192-8.
7. Afgan E, Baker D, Batut B, van den Beek M, Bouvier D, Čech M, et al. The Galaxy platform for accessible, reproducible and collaborative biomedical analyzes: 2018 update. Nucleic Acids Res. 2018;46(W1):W537-W44.
8. Blighe K. Next Generation Sequencing in the National Health Service England: A Pipeline that Completely Agrees with Sanger. Journal of Cancer Science & Therapy. 2014;06.
9. Cunningham F, Allen JE, Allen J, Alvarez-Jarreta J, Amode M R, Armean Irina M, et al. Ensembl 2022. Nucleic Acids Res. 2021;50(D1):D988-D95.
10. 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):1747-56.
11. Robinson JT, Thorvaldsdóttir H, Wenger AM, Zehir A, Mesirov JP. Variant Review with the Integrative Genomics Viewer. Cancer Research. 2017;77(21):e31-e4.
12. Li MM, Datto M, Duncavage EJ, Kulkarni S, Lindeman NI, Roy S, et al. Standards and Guidelines for the Interpretation and Reporting of Sequence Variants in Cancer: A Joint Consensus Recommendation of the Association for Molecular Pathology, American Society of Clinical Oncology, and College of American Pathologists. J Mol Diagn. 2017;19(1):4-23.
13. Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, et al. A global reference for human genetic variation. Nature. 2015;526(7571):68-74.
14. Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, et al. Analysis of protein-coding genetic variation in 60,706 humans. Nature. 2016;536(7616):285-91.
15. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. Nature. 2020;581(7809):434-43.
16. Sim NL, Kumar P, Hu J, Henikoff S, Schneider G, Ng PC. SIFT web server: predicting effects of amino acid substitutions on proteins. Nucleic Acids Res. 2012;40(Web Server issue):W452-7.
17. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, et al. A method and server for predicting damaging missense mutations. Nat Methods. 2010;7(4):248-9.
18. Schwarz JM, Rödelsperger C, Schuelke M, Seelow D. MutationTaster evaluates disease-causing potential of sequence alterations. Nat Methods. 2010;7(8):575-6.
19. Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. Nat Genet. 2014;46(3):310-5.
20. Landrum MJ, Lee JM, Benson M, Brown GR, Chao C, Chitipiralla S, et al. ClinVar: improving access to variant interpretations and supporting evidence. Nucleic Acids Res. 2018;46(D1):D1062-D7.
21. Tamborero D, Rubio-Perez C, Deu-Pons J, Schroeder MP, Vivancos A, Rovira A, et al. Cancer Genome Interpreter annotates the biological and clinical relevance of tumor alterations. Genome Medicine. 2018;10(1):25.

22. Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SAJR, Behjati S, Biankin AV, et al. Signatures of mutational processes in human cancer. Nature. 2013;500(7463):415-21.
23. Reimand J, Isserlin R, Voisin V, Kucera M, Tannus-Lopes C, Rostamianfar A, et al. Pathway enrichment analysis and visualization of omics data using g:Profiler, GSEA, Cytoscape and EnrichmentMap. Nat Protoc. 2019;14(2):482-517.
24. Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, et al. STRING v10: protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Res. 2015;43(Database issue):D447-52.
25. Morris JH, Apeltsin L, Newman AM, Baumbach J, Wittkop T, Su G, et al. clusterMaker: a multi-algorithm clustering plugin for Cytoscape. BMC Bioinformatics. 2011;12:436.
26. de Hoon MJ, Imoto S, Nolan J, Miyano S. Open source clustering software. Bioinformatics. 2004;20(9):1453-4.
27. Saldanha AJ. Java Treeview--extensible visualization of microarray data. Bioinformatics. 2004;20(17):3246-8.
28. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proceedings of the National Academy of Sciences. 2005;102(43):15545-50.
29. Haas BJ, Dobin A, Li B, Stransky N, Pochet N, Regev A. Accuracy assessment of fusion transcript detection via read-mapping and de novo fusion transcript assembly-based methods. Genome Biology. 2019;20(1):213.
30. Chang Y-S, Tu S-J, Yen J-C, Lee Y-T, Fang H-Y, Chang J-G. The Fusion Gene Landscape in Taiwanese Patients with Non-Small Cell Lung Cancer. Cancers (Basel). 2021;13(6):1343.
31. Fujii S, Ishibashi T, Kokura M, Fujimoto T, Matsumoto S, Shidara S, et al. RAF1–MEK/ERK pathway-dependent ARL4C expression promotes ameloblastoma cell proliferation and osteoclast formation. The Journal of Pathology. 2022;256(1):119-33.
32. Heikinheimo K, Huhtala JM, Thiel A, Kurppa KJ, Heikinheimo H, Kovac M, et al. The Mutational Profile of Unicystic Ameloblastoma. J Dent Res. 2019;98(1):54-60.
33. Sondka Z, Bamford S, Cole CG, Ward SA, Dunham I, Forbes SA. The COSMIC Cancer Gene Census: describing genetic dysfunction across all human cancers. Nat Rev Cancer. 2018;18(11):696-705.
34. Regezi JA, Kerr DA, Courtney RM. Odontogenic tumors: analysis of 706 cases. J Oral Surg. 1978;36(10):771-8.
35. Shear M, Singh S. Age-standardized incidence rates of ameloblastoma and dentigerous cyst on the Witwatersrand, South Africa. Community Dent Oral Epidemiol. 1978;6(4):195-9.
36. Marín-Márquez C, Kirby J, Hunter KD. Molecular pathogenesis of ameloblastoma. Journal of Oral Pathology & Medicine. 2024;53(5):277-93.
37. Guan P, Wong SF, Lim JQ, Ng CCY, Soong PL, Sim CQX, et al. Mutational Signatures in Mandibular Ameloblastoma Correlate with Smoking. J Dent Res. 2019;98(6):652-8.
38. Shi Y, Li M, Yu Y, Zhou Y, Wang S. Whole exome sequencing and system biology analysis support the "two-hit" mechanism in the onset of Ameloblastoma. Med Oral Patol Oral Cir Bucal. 2021;26(4):e510-e7.
39. Diniz MG, Gomes CC, Guimaraes BVA, Castro WH, Lacerda JCT, Cardoso SV, et al. Assessment of BRAFV600E and SMOF412E mutations in epithelial odontogenic tumors. Tumor Biol. 2015;36(7):5649-53.

40. Gates JC, Clark AP, Cherkas E, Shreenivas AV, Kraus D, Danzinger N, et al. Genomic profiling and precision medicine in complex ameloblastoma. Head Neck. 2023;45(4):816-26.
41. Shimura M, Nakashiro KI, Sawatani Y, Hasegawa T, Kamimura R, Izumi S, et al. Whole Exome Sequencing of SMO, BRAF, PTCH1 and GNAS in Odontogenic Diseases. In Vivo. 2020;34(6):3233-40.
42. Bartels S, Adisa A, Aladelusi T, Lemound J, Stucki-Koch A, Hussein S, et al. Molecular defects in BRAF wild-type ameloblastomas and craniopharyngiomas-differences in mutation profiles in epithelial-derived oropharyngeal neoplasms. Virchows Archiv. 2018;472(6):1055-9.
43. Gultekin SE, Aziz R, Heydt C, Senguven B, Zoller J, Safi AF, et al. The landscape of genetic alterations in ameloblastomas relates to clinical features. Virchows Archiv. 2018;472(5):807-14.
44. Yukimori A, Oikawa Y, Morita K, Nguyen CTK, Harada H, Yamaguchi S, et al. Genetic basis of calcifying cystic odontogenic tumors. Plos One. 2017;12(6).
45. Lin L-H, Chou C-H, Cheng H-W, Chang K-W, Liu C-J. Precise Identification of Recurrent Somatic Mutations in Oral Cancer Through Whole-Exome Sequencing Using Multiple Mutation Calling Pipelines. Frontiers in Oncology. 2021;11.
46. Hu SJ, Parker J, Divaris K, Padilla R, Murrah V, Wright JT. Ameloblastoma Phenotypes Reflected in Distinct Transcriptome Profiles. Scientific Reports. 2016;6.
47. Heikinheimo K, Kurppa KJ, Laiho A, Peltonen S, Berdal A, Bouattour A, et al. Early dental epithelial transcription factors distinguish ameloblastoma from keratocystic odontogenic tumor. J Dent Res. 2015;94(1):101-11.
48. Kalluri R, Weinberg RA. The basics of epithelial-mesenchymal transition. J Clin Invest. 2009;119(6):1420-8.
49. Aborisade A, Akinyele A, Aregbesola B, Adesina O, Ladeji A. Immunohistochemical expression of E-cadherin, N-cadherin and Snail/slug in ameloblastoma. Journal of Stomatology, Oral and Maxillofacial Surgery. 2022;123(6):e801-e7.
50. Kurioka K, Wato M, Iseki T, Tanaka A, Morita S. Differential expression of the epithelial mesenchymal transition factors Snail, Slug, Twist, TGF-β, and E-cadherin in ameloblastoma. Medical Molecular Morphology. 2017;50(2):68-75.
51. Siar CH, Ng KH. Epithelial-to-mesenchymal transition in ameloblastoma: focus on morphologically evident mesenchymal phenotypic transition. Pathology. 2019;51(5):494-501.
52. Hao F, Liu J, Zhong M, Wang J, Liu J. Expression of E-cadherin, vimentin and β-catenin in ameloblastoma and association with clinicopathological characteristics of ameloblastoma. Int J Clin Exp Pathol. 2018;11(1):199-207.
53. Xiong G, Ouyang S, Xie N, Xie J, Wang W, Yi C, et al. FOSL1 promotes tumor growth and invasion in ameloblastoma. Frontiers in Oncology. 2022;12.
54. Zhang J, Wang Y, Fan C, Xiao X, Zhang Q, Xu T, et al. Interleukin-8/β-catenin mediates epithelial-mesenchymal transition in ameloblastoma. Oral Dis. 2019;25(8):1964-71.
55. Saraste M. Oxidative Phosphorylation at the fin de siècle. Science. 1999;283(5407):1488-93.
56. Liberti MV, Locasale JW. The Warburg Effect: How Does it Benefit Cancer Cells? Trends Biochem Sci. 2016;41(3):211-8.
57. Uslu C, Kapan E, Lyakhovich A. Cancer resistance and metastasis are maintained through oxidative phosphorylation. Cancer Letters. 2024;587:216705.
58. Cheng G, Hardy M, Zielonka J, Weh K, Zielonka M, Boyle KA, et al. Mitochondria-targeted magnolol inhibits OXPHOS, proliferation, and tumor growth via modulation of energetics and autophagy in melanoma cells. Cancer Treat Res Commun. 2020;25:100210.

59. Guimarães LM, Coura BP, Gomez RS, Gomes CC. The Molecular Pathology of Odontogenic Tumors: Expanding the Spectrum of MAPK Pathway Driven Tumors. Frontiers in Oral Health. 2021;2.
60. Yang J, Liu M, Hong D, Zeng M, Zhang X. The Paradoxical Role of Cellular Senescence in Cancer. Front Cell Dev Biol. 2021;9:722205.
61. Satriyo PB, Bamodu OA, Chen JH, Aryandono T, Haryana SM, Yeh CT, et al. Cadherin 11 Inhibition Downregulates β-catenin, Deactivates the Canonical WNT Signaling Pathway and Suppresses the Cancer Stem Cell-Like Phenotype of Triple Negative Breast Cancer. J Clin Med. 2019;8(2).
62. Sun W, Qin Y, Wang Z, Dong W, He L, Zhang T, et al. The NEAT1_2/miR-491 Axis Modulates Papillary Thyroid Cancer Invasion and Metastasis Through TGM2/NFκb/FN1 Signaling. Front Oncol. 2021;11:610547.
63. Chang SK, Noss EH, Chen M, Gu Z, Townsend K, Grenha R, et al. Cadherin-11 regulates fibroblast inflammation. Proc Natl Acad Sci U S A. 2011;108(20):8402-7.
64. Heikinheimo K, Jee KJ, Niini T, Aalto Y, Happonen RP, Leivo I, et al. Gene expression profiling of ameloblastoma and human tooth germ by means of a cDNA microarray. J Dent Res. 2002;81(8):525-30.
65. Lim J, Ahn H, Min S, Hong SD, Lee JI, Hong SP. Oligonucleotide microarray analysis of ameloblastoma compared with dentigerous cyst. Journal of Oral Pathology & Medicine. 2006;35(5):278-85.
66. Kondo S, Ota A, Ono T, Karnan S, Wahiduzzaman M, Hyodo T, et al. Discovery of novel molecular characteristics and cellular biological properties in ameloblastoma. Cancer Med-Us. 2020.
67. Alimperti S, Andreadis ST. CDH2 and CDH11 act as regulators of stem cell fate decisions. Stem Cell Res. 2015;14(3):270-82.
68. Lu W, Zhou Q, Chen Y. Impact of RNA degradation on next-generation sequencing transcriptome data. Genomics. 2022;114(4):110429.
69. Jacobsen SB, Tfelt-Hansen J, Smerup MH, Andersen JD, Morling N. Comparison of whole transcriptome sequencing of fresh, frozen, and formalin-fixed, paraffin-embedded cardiac tissue. PLoS One. 2023;18(3):e0283159.

Figure legends

Figure 1. Summary of Variants Filtering and Validation Workflow

The first part of the pipeline included the bioinformatics pre-processing of the data mainly performed in Galaxy EU and ended with annotations of the variants by the VEP. The second part corresponded to the Variant Filtration and Validation and included two phases. In phase 1, common variants were removed based on population frequency and functional non-significant variants. Phase 2 further narrowed down the number of variants, firstly, by removing false positive variants shown as benign/likely benign, secondly by identifying somatic mutations; and ultimately, by identifying driver cancer genes and including those present in at least three tumors and/or based on previously known involvement of the implicated genes as driver cancer genes. The variants that passed all the filters were validated by visualization on IGV and direct sequencing.

Figure 2: Representative sections of the whole-exome sequenced tumors

From A to F, six representative microphotographs of the included tumors with hematoxylin-eosin stains.

Figure 3: Analysis of somatic variants in ameloblastoma

In A the mutational signatures in ameloblastoma. In A.1, de novo mutational signatures were compared against the 60 COSMIC signatures and in A.2, heatmaps of cosine similarities between de novo mutational signatures and COSMIC signatures for A.1. In B, pathway enrichment analysis in ameloblastoma. Nodes represent GO gene sets, and edges represent overlap (similarity) between the gene sets. The statistical threshold used was p-val < 0.05. In C, a visualization of the predicted results of protein-protein interactions (PPI) between somatic variants in ameloblastoma. Each node represents a protein, and each edge refers to an interaction. Edge thickness reflects the strength of PPIs. Only proteins with more than three interactions were included. Yellow nodes correspond to targets of Food and Drug Administration (FDA) approved drugs, and purple nodes correspond to targets of small molecule compounds. In D, a total of 26 driver cancer genes were included across the 13 samples. *BRAF* is the most frequent (62%), followed by *KMT2D* (23%) and *CTNNB1* (23%). The remaining 23 genes were present in 8%, meaning they were mutated in one sample. In addition, the oncoplot shows the relation between the presence of the *BRAF* mutated tumors and the *BRAF* wild type. Colored boxes indicate the presence of different types of mutations in the indicated genes (rows) and samples (columns). T8 represents a case of an unicystic ameloblastoma (UA) whereas the remaining cases are conventional.

Figure 4. Inspection and validation of the variants

In A, C and E, representative IGV visualization of the variants affecting, *BRAF*, *KMT2D* and *CTNNB1*, in which red reads are forward strands and blue reads are reverse strands. In B, D and F, representative chromatograms showing the nucleotides responsible for the identified mutations in representative tumors.

Figure 5: Ameloblastoma shows inter-tumor heterogeneity

In A, a heatmap from the unsupervised hierarchical cluster analysis, showing the twelve tumor samples with 2158 genes grouped in two molecular clusters, not explained by the presence/absence of $BRAF^{V600E}$ mutation, location or histopathological features (* indicates WT tumors). Cluster 1 in orange dendrograms and cluster 2 in blue dendrograms. In B, enrichment map of all curated genes from human MSigDB Collections. The network was manually rearranged to improve the layout. Clusters of nodes were labeled using AutoAnnotated Cytoscape Application and manually edited to better comprehend their functional significance. Clusters with ≤ 2 nodes were removed. In C, a table showing the top 15 enriched gene sets in the molecular cluster 1, most of them associated with carcinogenesis. The size column indicates the number of genes represented in each gene set. In D, an enrichment plot showing the activation of the hallmark epithelial-mesenchymal transition set in the molecular cluster 1. In E, differentially expressed genes according to *BRAF* mutation status in ameloblastoma. Red and blue colors represent up- and downregulated expression in AM, respectively. P-value <0.05, and 1.2≤FC≤-1.2. T8 corresponds to a unicystic ameloblastoma, the remaining cases are conventional. In F, representative enrichment plots and heatmaps showing the epithelial-mesenchymal transition and the KRAS-signaling sets in BRAF WT tumors. Both enriched sets were significantly enriched at FDR < 25%. GSEA was conducted using GSEA v4.1.0 software and Molecular Signatures Database (Broad Institute). All raw data were applied to "hallmarks" (h.all.v7.5) database.

Figure 6: CDH11 and TGM2 are expressed in ameloblastoma

Panels A to F are representative photomicrographs of CDH11 in different cases of AM. Moderate to strong cytoplasmic immunolabeling in the neoplastic cells is present, while the mesenchymal stroma shows no reactivity. No differences in the expression of the epithelial cells from the periphery or the center of the neoplastic islands or cords were observed. Panels J to K are representative photomicrographs of TGM2 in different cases of AM. Moderate to strong cytoplasmic immunolabeling in the neoplastic cells and mesenchymal stroma are present. In I, higher immunoexpression was seen in the tumor cells localized along the basal cell layers at the periphery of the islands and cords of the neoplastic epithelium. Scale bar: 50µm.