Genetic Profiling of Mucoepidermoid Carcinoma of Minor Salivary Glands in Thai Patients: A Preliminary Study Using Whole-Exome Sequencing

Pyae Phyo Thaw[1,*], Poramaporn Klanrit[1], Ajiravudh Subarnbhesaj[1] and Poramate Klanrit[2,3]

[1]Department of Oral Biomedical Sciences, Research Group of Chronic Inflammatory Oral Diseases and Systemic Diseases Associated with Oral Health, Faculty of Dentistry, Khon Kaen University, Khon Kaen 40002, Thailand
[2]Department of Systems Biosciences and Computational Medicine, Faculty of Medicine, Khon Kaen University, Khon Kaen 40002, Thailand
[3]Cholangiocarcinoma Research Institute, Khon Kaen University, Khon Kaen 40002, Thailand
[*]Corresponding author, E-mail: pyaephyothaw.ac@gmail.com

Abstract

Mucoepidermoid carcinoma (MEC) is the most prevalent malignant salivary gland tumor. However, the genomic landscape of minor salivary gland MECs in Southeast Asian populations remains largely under-characterized compared to Western populations. This study aimed to identify genetic alterations in MECs from Thai patients to uncover their mutational profile and potential therapeutic targets. Whole-exome sequencing (WES) was performed on three formalin-fixed paraffin-embedded (FFPE) samples of minor salivary gland MECs using a next-generation sequencing (NGS) platform. All cases were diagnosed according to the 5th edition of the World Health Organization (WHO) Classification of Head and Neck Tumors and graded using the Armed Forces Institute of Pathology (AFIP) grading system. The sequencing data were processed using the Genome Analysis Toolkit (GATK) bioinformatics pipeline to identify single-nucleotide variants (SNVs) and insertions/deletions (INDELs), followed by functional pathway enrichment and druggability analysis. The results revealed driver mutations in genes involved in major oncogenic pathways, including DNA repair, MAPK, Hedgehog. Notch signaling, with Notch pathway-related alterations consistently observed across all cases. Pathway-level analysis indicated shared biological features despite intertumor diversity, and druggability assessment identified several candidate targets. As an exploratory study, these findings are limited by a small sample size (n = 3) and the absence of matched normal controls, which would allow definitive distinction between somatic mutations and germline variants. Overall, this study provides a preliminary genomic overview of minor salivary gland MECs in Thai patients and demonstrates the advantage of pathway-based analysis in improving our understanding of MEC biology and establishing a basis for future expanded genomic investigations.

Keywords: *mucoepidermoid carcinoma, minor salivary glands, whole-exome sequencing, genetic alterations*

1. Introduction

Salivary gland carcinomas are rare and heterogeneous tumors, accounting for less than 1% of all malignancies and 2-6% of head and neck cancers (Schvartsman et al., 2019). The parotid gland is the most common site of salivary gland neoplasms, with approximately 25% being malignant tumors, whereas about half of tumors arising from the minor salivary gland tumors are malignant. The sublingual glands are the exception, in which all tumors are malignant (Speight & Barrett, 2002). In Thailand, 68.8% of salivary gland tumors are benign and 31.2% are malignant tumors, with a mean patient age of 47.1 years. The most common malignant and benign tumors were mucoepidermoid carcinoma and pleomorphic adenoma, respectively. The most common sites of tumors were the parotid gland (62.7%), submandibular gland (18.8%), and the intraoral minor salivary glands (18.0%) (Juengsomjit et al., 2015).

The World Health Organization (WHO) defines mucoepidermoid carcinoma as “a malignant neoplasm of the salivary gland that is composed of cells of mucous, intermediate, and epidermoid (squamoid) types and has both cystic and solid growth patterns” (WHO Classification of Tumours Editorial Board, 2023). Beyond its histopathological definition, mucoepidermoid carcinoma demonstrates distinct epidemiological characteristics across different populations. Various articles have reported different male- to- femaleratios, with the majority reporting that MECs show a greater female predominance. The reported mean age at diagnosis ranges from 47 to 55 years, and together with adenoid cystic carcinoma, it is one of the most

common malignant neoplasms of the adolescent salivary glands (Peraza et al., 2020). The prevalence of mucoepidermoid carcinoma in Thailand is reported at 8.9% of all salivary gland tumors (Juengsomjit et al., 2015).

In addition to its epidemiological features, mucoepidermoid carcinoma is notable for its significant cellular heterogeneity, including intermediate, epidermoid, and mucin-producing cells. There are multiple patterns in which the cells are grouped, including glandular patterns, compact solid clusters, and cystic appearance (Chen et al., 2021). Mucin-producing cells with cystic structures are common in low-grade tumors, whereas high-grade tumors are mainly composed of solid sheets of epidermoid cells. Perineural invasion and necrosis are also features of high-grade MECs (Peraza et al., 2020).

MEC's marked histological heterogeneity suggests an underlying molecular complexity that cannot be fully explained by morphology alone. The most frequent genetic alteration in MECs is a peculiar chromosomal translocation t(11;19)(q21;p13), which produces a unique fusion. This translocation was first reported by Tonon et al. (2003) and results in a *CRTC1-MAML2* fusion. This fusion occurs when exons 2 through 5 of the *MAML2* gene combine with the first exon of the *CRTC1* gene, genetic material is fused, creating the characteristic *CRTC1-MAML2* fusion product (O'Neill, 2009). A recent study identified the *CRTC1-MAML2* fusion in all four cell lines and, fortuitously, identified and confirmed the breakpoint structures between *MYBL1* and *MAML2* and between *SGK3* and *CRTC1*. These findings suggest a indicated the potential role for the *MYBL1* gene in the pathogenesis of MECs. They also identified and confirmed *TERT* (Telomerase Reverse Transcriptase) rearrangements and amplifications in MECs through whole-exome sequencing and Sanger sequencing; this was the first report identifying *TERT* as a potential novel driver in MECs (Gensterblum-Miller et al., 2024).

Whole-exome sequencing (WES) is the sequencing of the exome, the protein-coding regions of the genome where most of the variants that cause disease arise. WES is also a more cost-effective methodology for identifying rare diseases, cancer genetics, and population genetics than whole-genome sequencing. There are a number of commercial kits available for exome capture that are compatible with the Illumina next-generation sequencing platform. After exome enrichment, high-throughput sequencing is performed, generating large volumes of sequencing data that are subsequently analyzed using bioinformatics software in a manner similar to whole-genome sequencing (Satam et al., 2023). Illumina, a second-generation sequencing technology, is widely used to perform various sequencing protocols. These include genomic sequencing, exome sequencing, metagenomics, and RNA sequencing. Different technologies provide different throughput levels (Slatko et al., 2018).

Current studies of genetic landscapes, including gene fusions, in mucoepidermoid carcinoma mostly use techniques such as fluorescence in situ hybridization (FISH) and next-generation sequencing (NGS). However, most studies focus on non-Asian populations and major salivary glands. This lack of diversity creates a gap in understanding how the disease behaves and responds to treatment in Asian populations, especially in Southeast Asia. To fill that gap by providing valuable insights into the unique genetic alterations in MECs in underrepresented ethnic groups and other types of salivary glands, ensuring a more comprehensive understanding of the disease, this study aims to perform whole-exome sequencing (Illumina) to explore genetic profiles of mucoepidermoid carcinomas of minor salivary glands to identify tumor-specific alterations.

2. Objectives

1) To explore exonic variants in mucoepidermoid carcinoma of minor salivary glands in Thai patients.
2) To compare the identified exonic variants and affected pathways with previously reported genomic data for mucoepidermoid carcinoma from public databases.

3. Materials and Methods

3.1 Study design

The present study was an exploratory cross-sectional study, in which whole-exome sequencing was performed on three formalin-fixed, paraffin-embedded tissue specimens (n=3) diagnosed with

mucoepidermoid carcinoma of the minor salivary glands. The specimens were obtained from the archive of the Oral Pathology Unit, Faculty of Dentistry, Khon Kaen University. This study was approved by the Khon Kaen University Ethics Committee for Human Research (HE682176).

3.2 Sample preparation

Three formalin-fixed paraffin-embedded (FFPE) tissue specimens diagnosed with mucoepidermoid carcinoma were included in the study. Subsequently, patient identifiers were removed and the cases were designated as MEC1, MEC2, and MEC3. Histopathological re-evaluation was performed according to the 5th edition of the World Health Organization Classification of Head and Neck Tumors (WHO Classification of Tumours Editorial Board, 2023). Tumor grading was performed using the Armed Forces Institute of Pathology (AFIP) grading system (Auclair et al., 1992). Of the three cases, two were classified as low-grade mucoepidermoid carcinoma and one as intermediate-grade. A digitally scanned slide of each case was used to identify a corresponding section and select a tumor area containing more than 60% tumor cells in each sample (Kang et al., 2017).

3.3 DNA extraction

Tissue DNA was extracted using a QIAamp DNA FFPE tissue kit (QIAGEN, Hilden, Germany). Using a microtome, FFPE tissues were sectioned into 15 sections of 10 micrometers thickness and placed directly into a 1.5 mL microcentrifuge tube. Next, 160 microliters of deparaffinization solution (QIAGEN, Hilden, Germany) was added, vortexed vigorously for 10 seconds, and briefly centrifuged. After this step, 180 microliters of lysis buffer (Buffer ATL) was added, followed by vortexing and centrifuged for 1 minute at 11,000 x g (10,000 rpm). Twenty microliters of proteinase K were added to the lower, clear phase, and the mixture was gently mixed by pipetting up and down. The mixture was incubated at 56°C for two hours to digest the tissue completely. After the above steps, the DNA was purified through binding and washing steps. First, the lower, clear phase was transferred into a new 2 mL microcentrifuge tube, then 200 µL of Buffer AL and 500 µL of 100% ethanol were added, and the mixture was thoroughly vortexed. The samples were then transferred to a QIAamp MinElute column and centrifuged for 1 minute at 8,000 rpm. Then, the column was transferred to a clean 2 mL collection tube. Next, 500 µL of Buffer AW1 was added to the spin column, followed by centrifuged at 8,000 rpm for 1 minute to wash contaminants. The spin column was transferred to a clean microcentrifuge tube, and 500 µL of Buffer AW2 was added, followed by centrifugation at 8,000 rpm for 1 minute. Then, 30 µL of elution buffer (Buffer ATE) was added to the column The column was incubated at room temperature for 3 minutes, then centrifuged for 1 minute.

3.4 DNA quantity and quality assessment

The extracted genomic DNA was quantified for concentration and purity using a NanoDrop spectrophotometer. DNA integrity was further evaluated by determining DNA Integrity Number (DIN) values using the Genomic DNA ScreenTape Assay on the TapeStation system (Agilent Technologies), ensuring suitability for downstream whole-exome sequencing.

3.5 Library construction and whole-exome sequencing

Extracted genomic DNA was sent to NovogeneAIT Genomics Singapore Pte Ltd. for whole-exome sequencing (WES). Genomic DNA was randomly fragmented to 180-280 bp, followed by end repair, A-tailing, and ligation of Illumina sequencing adapters. Libraries were size-selected, PCR-amplified, and enriched for exonic regions using biotin-labeled capture probes. Library quality and concentration were assessed using Qubit fluorometer, real-time PCR, and a bioanalyzer. Qualified libraries were pooled and sequenced on an Illumina sequencing platform using paired-end 150-bp reads (PE150).

3.6 Bioinformatic analysis

Raw sequencing reads were quality-filtered and trimmed using fastp (v0.23.1), and then aligned to the human reference genome (hg38) using BWA (v0.7.17). BAM files were sorted, and duplicate reads were

marked using Sambamba (v1.0.0) and Picard (v2.18.9). Single-nucleotide polymorphisms (SNVs) and insertions and deletions (INDELs) were identified using the Genome Analysis Toolkit (GATK, v4.3.0). The variant calling was performed against the human reference genome (hg38) and results were reported in VCF file. The variants were annotated using ANNOVAR, focused on exonic function, amino acid change, ClinVar ID, COSMIC ID, gnomAD variant allele frequency, 1000 genomes variant allele frequency, Mutation Taster Annotation, and Combined Annotation Dependent Depletion (CADD) scores (Wang et al., 2010).

To identify the potential variants of mucoepidermoid carcinomas of minor salivary glands, all samples (MEC1, MEC2, and MEC3) were consensually prioritized and summarized into a new file. As matched normal controls were not available, the datasets were filtered using a CADD Phred score (≥20 cut-off, Top 1% most deleterious variants), minimize the inclusion of germline background variants. Next, the variants were summarized by gene name and subjected to pathway enrichment analysis and druggable gene targets.

3.7 Statistical analysis and data visualization

All analyses were conducted in using Google Colab with standardized package versions to ensure reproducibility. The complete analysis pipeline, including data preprocessing, pathway analysis, druggability assessment, and visualization generation, was implemented as an integrated Python script. All statistical analyses were performed using Python's scientific computing ecosystem, including pandas (version 2.0.3), Numpy (version 1.24.3), and Scipy (version 1.10.1). Data visualization was implemented with matplotlib (version 3.7.1) and seaborn (version 0.12.2).

4. Results and Discussion

This preliminary study on whole-exome sequencing (WES) of three formalin-fixed, paraffin-embedded (FFPE) archival tissues samples provided a valuable opportunity to explore the genomic alterations in minor salivary gland mucoepidermoid carcinoma (MEC) in a small group of Thai patients. The results identified genomic variants and further included functional pathway enrichment analysis and evaluation of potential druggable targets. The findings were presented in a structured manner to facilitate interpretation of molecular alterations and associated biological pathways involved in the pathogenesis of MEC.

4.1 Sequencing quality and mapping statistics

The utilization of formalin-fixed, paraffin-embedded (FFPE) tissue for high-throughput sequencing presents well-documented challenges, primarily due to DNA fragmentation and cytosine deamination artifacts induced by fixation processes (Srinivasan et al., 2002). Sequencing quality assessment identified variability in read mapping, sequencing depth, and target coverage among MEC1, MEC2, and MEC3. Base quality across the samples was high overall, with low sequencing error rates (0.01%) and Q30 values >95%. The Q30 score indicates that the probability of an incorrect base call is less than 1 in 1,000, ensuring that the sequencing data have high accuracy for the reads that were mapped (Yauy et al., 2023). Among the three samples, MEC1 showed the best sequencing performance. This sample achieved a 99.39% mapping rate to the reference genome, a mean sequencing depth of 136.52x, and 96.33% of target regions with ≥20x depth. This level of coverage exceeded the standard recommendations for cancer exome sequencing, which typically require mean depths of approximately 100x to ensure reliable detection of somatic variants with lower variant allele frequencies (VAF) (Satam et al., 2023). MEC2 showed acceptable sequencing quality but had a lower mapping rate (91.11%) and lower target coverage than MEC1. Although it had a mean sequencing depth comparable to MEC1 (123.73x), only 30.07% of target regions achieved ≥20x coverage. On the other hand, MEC3 had the lowest sequencing performance. It had fewer reads mapped to the reference genome (77.50%), a much lower mean sequencing depth (34.59x), and limited target coverage. Only 16.07% of target regions achieved ≥20x coverage. This degradation in quality is characteristic of more aggressively fixed FFPE blocks, where DNA crosslinking and fragmentation impede library preparation and alignment (Srinivasan et al., 2002). Despite this, the inclusion of MEC3 in the analysis was justified by the preservation of essential quality

metrics, such as base quality scores (Q30 > 95%) and the transition-to-transversion (Ts/Tv) ratio. The variability in coverage underscores the need to use specialized variant callers (such as GATK) that are calibrated to handle uneven coverage distributions typical of FFPE-derived libraries (McKenna et al., 2010). Despite the differences in sequencing performance, all samples met the minimum quality requirements for downstream variant detection and comparative genomic analysis (Table 1).

Table 1 Sequencing quality, mapping performance, and coverage metrics of MEC samples

Sample	Total Reads	Mapped (%)	Total Unique Reads	Uniquely Mapped (%)	Error (%)	Q30 (%)	Mean Depth (x)	Coverage ≥20x (%)
MEC1	89,435,120	99.39	88,382,790	98.82	0.01	96.69	136.52	96.33
MEC2	89,289,474	91.11	79,701,296	89.26	0.01	96.31	123.73	30.07
MEC3	31,730,748	77.50	23,130,852	72.90	0.01	95.50	34.59	16.07

4.2 Summary of INDEL and SNV variants in MEC samples

The majority of INDELs and SNVs identified in MEC samples before filtering were located in the introns and intergenic regions of the genes, with only a small proportion affecting coding regions. However, there was a substantially higher frequency of SNVs than of INDELs in coding sequences of genes. This distribution is consistent with the mutational landscapes observed in most solid tumors, where point mutations are the primary drivers of oncogene activation and tumor suppressor inactivation (Vogelstein et al., 2013). Additionally, most variants appear to be homozygous rather than heterozygous for both SNVs and INDELs. A critical quality control metric in whole-exome sequencing is the transition-to-transversion (Ts/Tv) ratio, which measures the frequency of purine-to-purine or pyrimidine-to-pyrimidine changes (transitions) versus purine-to-pyrimidine changes (transversions) (Guo et al., 2012). In the human exome, this ratio is empirically expected to fall between 2.8 and 3.0 due to the biochemical properties of spontaneous mutation rates, particularly CpG deamination (Wang et al., 2015). In the present study, the Ts/Tv ratios for SNVs were 2.50 for MEC1, 3.03 for MEC2, and 2.92 for MEC3. These values are remarkably consistent with the expected range for high-quality human exome data, providing strong confidence in the validity of the called variants (Wang et al., 2013). Notably, MEC3, despite its lower coverage, maintained a Ts/Tv ratio of 2.92, suggesting that reduced read depth did not introduce a substantial bias toward random sequencing errors, which are typically associated with a marked reduction in the Ts/Tv ratio (Bainbridge et al., 2011). Furthermore, although most variants had been previously documented in dbSNP, many novel variants were identified across the three samples (Table 2).

One of the most striking findings of this study was the substantial number of novel variants identified that were absent in the dbSNP database. Specifically, MEC2 harbored 88,987 novel SNVs and 10,910 novel INDELs, while MEC3 contained 44,913 novel SNVs. This finding suggests that a significant portion of the somatic landscape in underrepresented populations may actually reflect rare germline variants, highlighting the urgent need for population-specific reference genomes to improve cancer precision medicine in Thailand (Karczewski et al., 2020).

Table 2 Summary of genomic variants in MEC samples

Variant Category	MEC1	MEC2	MEC3
Total SNVs	843,821	209,344	97,477
Total INDELs	177,866	46,545	20,620
Coding (CDS) Variants	27,254	14,535	9,606
Ts/Tv Ratio (SNVs)	2.50	3.03	2.92
Known Variants (dbSNP %)	73.80%	57.49%	53.92%
Novel SNVs	21,102	88,987	44,913
Novel INDELs	31,090	10,910	5,634

4.3 Data analysis and filtering by exonic functions

Exonic function filtering was used to focus on genetic variants most likely to affect protein-coding regions (e.g., frameshift, stop-gain, and splice-site variants), under the assumption that they are more functionally relevant to cancer phenotypes. The total number of variants decreased substantially after filtering based on annotated exonic function. As a result, many variants located outside coding regions were removed, thereby enriching for variants more likely to be biologically relevant. Although the filtering process reduced the frequency of SNVs relative to INDELs across all MEC samples, SNVs remained much more abundant than INDELs, consistent with typical whole-exome sequencing patterns. A total of 9,111 unique genes (identified by combining SNV and INDEL variants) had at least one variant; most of these genes had only SNVs, and fewer genes had only INDELs and/or combinations of both variant types. These results indicate that SNVs are the primary source of exonic variation in MEC, while INDELs contribute to a smaller subset of affected genes, although they are also biologically significant. In addition, the fact that some genes carry both SNVs and INDELs emphasizes the complexity and genetic heterogeneity of the mutational spectrum observed in mucoepidermoid carcinoma (Table 3).

Table 3 Data Analysis and filtering by exonic functions

Filtering all variants with annotated exonic functions						
	INDELs			**SNVs**		
Filtered Variants	**MEC1**	**MEC2**	**MEC3**	**MEC1**	**MEC2**	**MEC3**
	1,106	1,131	686	3,068	6,918	4,719
Total Filtered Variants	2,923 Variants			14,705 Variants		
Combined INDEL and SNV						
Total unique genes	9,111 genes					
Genes with SNV only	6,883 genes (75.55 %)					
Genes with INDEL only	826 genes (9.07 %)					
Genes with both SNV and INDEL	1,402 genes (15.38 %)					

4.4 Comparative genomic analysis and heterogeneity

Analysis of gene overlap among MEC1, MEC2, and MEC3 demonstrated pronounced intertumoral heterogeneity. Only 411 genes were shared across all three samples, indicating a relatively small common set of altered genes. In contrast, a large number of genes were unique to individual cases, with MEC2 showing the highest number of sample-specific genes (3,235), followed by MEC3 (1,928) and MEC1 (1,349). Pairwise overlaps were also observed, with 708 genes shared between MEC1 and MEC2, 353 genes shared between MEC1 and MEC3, and 1,127 genes shared between MEC2 and MEC3 (Figure 1). This finding supports the concept of salivary gland carcinomas as “malignant snowflakes,” tumors that are histologically similar but genetically unique (Kurzrock and Giles, 2015).

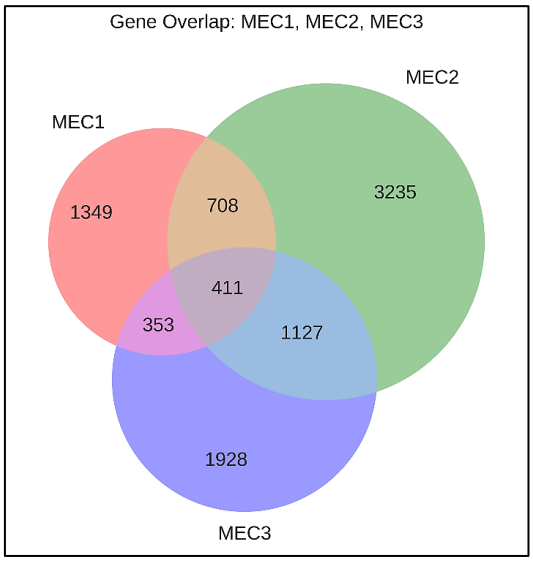

Figure 1 Number of unique genes presented with alterations (SNV only, INDEL only, or both) across 3 patients

A total of 411 overlapping genes encompassing 6,418 variants were initially identified after excluding likely germline polymorphisms using multiple population databases (e.g., gnomAD, 1000 Genomes, and ClinVar). Given the absence of matched normal controls, these alterations were considered putative somatic variants. To further prioritize variants with potential functional relevance, we applied an additional filter based on predicted deleteriousness, retaining only variants with CADD Phred scores ≥20 (corresponding to the top 1% most deleterious substitutions genome-wide). This refinement reduced the dataset to 381 genes harboring 799 variants. While this stepwise filtering strategy enriches for variants with predicted functional impact, rare germline variants could not be completely excluded, and therefore the results should be interpreted as exploratory candidate alterations rather than definitive somatic mutations.

4.5 Functional pathway enrichment analysis

To assess the functional consequences of the mutations across these case series, the candidate genes (n = 381) from all three samples were mapped to primary oncogenic signaling pathways (Figure 2). Mapping to the primary oncogenic pathways demonstrated a consistent pattern of pathway involvement across the samples. The pathway analysis suggests a convergence on DNA repair and Notch signaling pathways rather than the *TP53*-centric profile often observed in major gland MECs. Previous large-scale sequencing studies of MEC have predominantly focused on tumors arising in the major salivary glands (parotid) (Kang et al., 2017). These studies have established *TP53* (mutated in ~28% of cases) and *POU6F2* as the most common non-fusion drivers, particularly in intermediate- and high-grade tumors. In contrast, the minor salivary gland tumors in this study exhibited a distinct mutational burden and diversity. The next most commonly involved processes were alterations to the MAPK, Hedgehog, and Notch signaling pathways. The Notch signaling pathway was consistently altered across all samples in this study. This finding is particularly intriguing given the known biology of MEC. The canonical *CRTC1-MAML2* fusion, present in ~50-80% of MECs, acts as a functional mimic of the Notch co-activator *MAML2* but functions primarily to activate CREB signaling (Tonon et al., 2003; Wu et al., 2005). However, the fusion protein also interacts with the Notch complex, and its presence has been linked to aberrant Notch target expression (e.g., *HES1*, *HES5*) (O'Neill, 2009). Recurrent alterations in *NOTCH2* have been observed in MEC cases irrespective of *CRTC1-MAML2* fusion status, indicating a potential Notch-driven subtype beyond the canonical fusion mechanism (Yin & Ha, 2016).

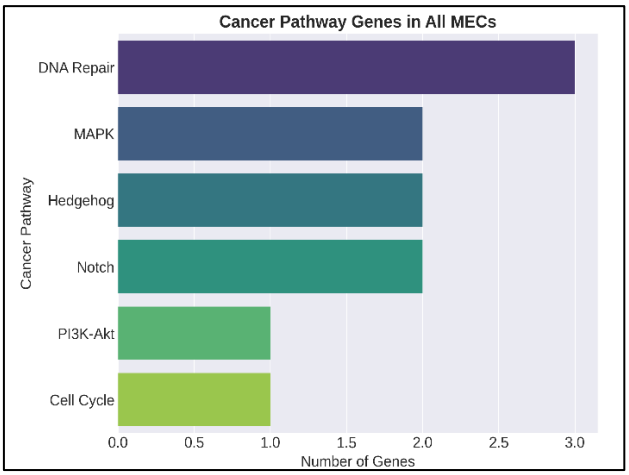

Figure 2 Functional pathway enrichment in MEC samples, based on 381 candidate genes

4.6 Targeted therapy analysis

The potential clinical utility of the genomic information derived from this study was assessed by evaluating the potential for drug treatment. This analysis identified multiple gene families whose products are targeted by established therapeutic classes, particularly agents involved in DNA repair and kinase signaling. The top-ranked categories of druggable genes included those whose products are targeted by FDA-approved drugs or agents currently under clinical investigation, including PARP inhibitors and specific kinase inhibitors. In addition to identifying high-prevalence targets shared across all three MEC samples, the analysis also revealed a high degree of inter-tumor heterogeneity; however, common pathway-level vulnerabilities may provide exploratory avenues for the application of personalized treatments (Figure 3).

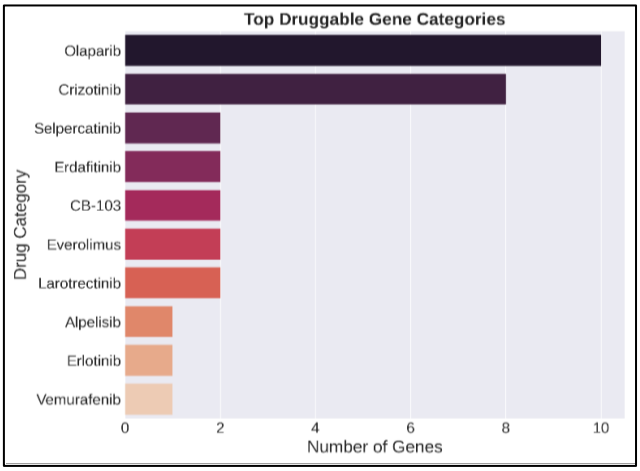

Figure 3 Top druggable gene categories

4.7 Comparison the present findings with major salivary gland studies

Biologically, minor salivary glands differ from major glands in their embryonic development, the absence of encapsulation, and their direct exposure to the oral cavity environment (mucosal carcinogens). The observed high variant counts in MEC1 and MEC2 suggest that minor gland MECs may be more genomically unstable or driven by a wider array of somatic mutations than the "fusion-driven, genomically quiet" model often applied to low-grade parotid MECs (Seethala et al., 2010). This finding supports clinical observations that minor gland MECs, particularly those occurring at the palate or retromolar trigone, can

exhibit more aggressive behavior than parotid tumors (Speight & Barrett, 2002). However, we acknowledge that our filtering approach cannot definitively isolate somatic mutations.

4.8 Limitations of study

While this study provides valuable initial insights, the small sample size limits broader statistical generalizations. Consequently, the interpretation of these genomic profiles remains exploratory at this stage. Future investigations incorporating more samples (e.g., N ≥ 30 per group) and matched normal tissues will be essential to definitively distinguish somatic mutations from inherited germline variations with statistical power. Because MEC3 had a significantly lower sequencing depth, some mutations in this sample may have gone undetected. Therefore, the true mutational overlap between the three samples might be higher than observed; however, it is partially masked by these technical limitations. Finally, we acknowledge that this approach cannot definitively distinguish somatic mutations. As a result, our findings represent exploratory genomic profiling rather than a definitive analysis of somatic mutations.

5. Conclusion

This study presented a preliminary whole-exome sequencing analysis of three mucoepidermoid carcinomas (MECs) of the minor salivary glands, providing an initial overview of their genomic characteristics. Comparative analysis with publicly available genomic data highlighted the involvement of key oncogenic signaling pathways. DNA repair, MAPK, Hedgehog, and Notch pathways were among the most frequently affected, with Notch pathway-related alterations consistently observed across all MEC samples. Pathway-level analysis suggested shared biological vulnerabilities despite inter-tumor genetic diversity, and druggability assessment identified multiple candidate target-associated drugs, particularly those related to DNA repair and kinase signaling. These findings represent an initial exploratory analysis and will be further investigated in more samples with extended analysis, refined variant interpretation, and deeper pathway-level investigation to confirm the biological and translational relevance of the results. Future studies should include matched normal tissues for all cases, perform functional validation of key altered genes and pathways, integrate multi-omics data, and correlate genomic findings with clinicopathological features and therapeutic responses.

6. Acknowledgement

This work was supported by the Fundamental Fund of Khon Kaen University for the fiscal year 2025 and the National Science, Research, and Innovation Fund (NSRF) of Thailand. Dr. Pyae Phyo Thaw was supported by the KKU ASEAN-GMS Scholarship. The authors are grateful to Ms. Thitima Khonrak and Mrs. Supaporn Singhara for their generous help in the pathology laboratory.

7. References

Auclair, P. L., Goode, R. K., & Ellis, G. L. (1992). Mucoepidermoid carcinoma of intraoral salivary glands evaluation and application of grading criteria in 143 cases. *Cancer*, *69*(8), 2021-2030. https://doi.org/10.1002/1097-0142(19920415)69:8%3C2021::AID-CNCR2820690803%3E3.0.CO;2-7

Bainbridge, M. N., Wiszniewski, W., Murdock, D. R., Friedman, J., Gonzaga-Jauregui, C., Newsham, I., ... & Gibbs, R. A. (2011). Whole-genome sequencing for optimized patient management. *Science translational medicine*, *3*(87), 87re3-87re3.

Chen, Z., Ni, W., Li, J. L., Lin, S., Zhou, X., Sun, Y., ... & Wu, L. (2021). The CRTC1-MAML2 fusion is the major oncogenic driver in mucoepidermoid carcinoma. *JCI insight*, *6*(7), Article e139497. https://doi.org/10.1172/jci.insight.139497

Gensterblum-Miller, E., Bhangale, A., Al Majid, D., Pienkowski, V. M., Rydzanicz, M., Janiszewska, J., ... & Brenner, J. C. (2024). Long read sequencing identifies complex structural variant landscape and recurrent TERT rearrangements in mucoepidermoid carcinoma. *Oral oncology*, *159*, Article 107108. https://doi.org/10.1016/j.oraloncology.2024.107108

Guo, Y., Long, J., He, J., Li, C. I., Cai, Q., Shu, X. O., ... & Li, C. (2012). Exome sequencing generates high quality data in non-target regions. *BMC genomics*, *13*(1), Article 194. https://doi.org/10.1186/1471-2164-13-194

Juengsomjit, R., Lapthanasupkul, P., Poomsawat, S., & Larbcharoensub, N. (2015). A clinicopathologic study of 1,047 cases of salivary gland tumors in Thailand. *Quintessence International*, *46*(8). 707–716. https://doi.org/10.3290/j.qi.a32922

Kang, H., Tan, M., Bishop, J. A., Jones, S., Sausen, M., Ha, P. K., & Agrawal, N. (2017). Whole-exome sequencing of salivary gland mucoepidermoid carcinoma. *Clinical Cancer Research*, *23*(1), 283-288. https://doi.org/10.1158/1078-0432.CCR-16-0720

Karczewski, K. J., Francioli, L. C., Tiao, G., Cummings, B. B., Alföldi, J., Wang, Q., ... & MacArthur, D. G. (2020). The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*, *581*(7809), 434-443. https://doi.org/10.1038/s41586-020-2308-7

Kurzrock, R., & Giles, F. J. (2015). Precision oncology for patients with advanced cancer: the challenges of malignant snowflakes. *Cell Cycle*, *14*(14), 2219-2221. https://doi.org/10.1080/15384101.2015.1041695

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., ... & DePristo, M. A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome research*, *20*(9), 1297-1303. https://doi.org/10.1101/gr.107524.110

O'Neill, I. D. (2009). t (11; 19) translocation and CRTC1-MAML2 fusion oncogene in mucoepidermoid carcinoma. *Oral oncology*, *45*(1), 2-9. https://doi.org/10.1016/j.oraloncology.2008.03.012

Peraza, A., Gómez, R., Beltran, J., & Amarista, F. J. (2020). Mucoepidermoid carcinoma. An update and review of the literature. *Journal of stomatology, oral and maxillofacial surgery*, *121*(6), 713-720. https://doi.org/10.1016/j.jormas.2020.06.003

Satam, H., Joshi, K., Mangrolia, U., Waghoo, S., Zaidi, G., Rawool, S., ... & Malonia, S. K. (2023). Next-generation sequencing technology: current trends and advancements. *Biology*, *12*(7), Article 997. https://doi.org/10.3390/biology12070997

Schvartsman, G., Pinto, N. A., Bell, D., & Ferrarotto, R. (2019). Salivary gland tumors: Molecular characterization and therapeutic advances for metastatic disease. *Head & neck*, *41*(1), 239-247. https://doi.org/10.1002/hed.25468

Seethala, R. R., Dacic, S., Cieply, K., Kelly, L. M., & Nikiforova, M. N. (2010). A reappraisal of the MECT1/MAML2 translocation in salivary mucoepidermoid carcinomas. *The American journal of surgical pathology*, *34*(8), 1106-1121. https://doi.org/10.1097/PAS.0b013e3181de3021

Slatko, B. E., Gardner, A. F., & Ausubel, F. M. (2018). Overview of next-generation sequencing technologies. *Current protocols in molecular biology*, *122*(1), Article e59. https://doi.org/10.1002/cpmb.59

Speight, P. M., & Barrett, A. W. (2002). Salivary gland tumours. *Oral diseases*, *8*(5), 229-240. https://doi.org/10.1034/j.1601-0825.2002.02870.x

Srinivasan, M., Sedmak, D., & Jewell, S. (2002). Effect of fixatives and tissue processing on the content and integrity of nucleic acids. *The American journal of pathology*, *161*(6), 1961-1971. https://doi.org/10.1016/S0002-9440(10)64472-0

Tonon, G., Modi, S., Wu, L., Kubo, A., Coxon, A. B., Komiya, T., ... & Kaye, F. J. (2003). t (11; 19)(q21; p13) translocation in mucoepidermoid carcinoma creates a novel fusion product that disrupts a Notch signaling pathway. *Nature genetics*, *33*(2), 208-213. https://doi.org/10.1038/ng1083

Vogelstein, B., Papadopoulos, N., Velculescu, V. E., Zhou, S., Diaz Jr, L. A., & Kinzler, K. W. (2013). Cancer genome landscapes. *science*, *339*(6127), 1546-1558. https://doi.org/10.1126/science.1235122

Wang, J., Raskin, L., Samuels, D. C., Shyr, Y., & Guo, Y. (2015). Genome measures used for quality control are dependent on gene function and ancestry. *Bioinformatics*, *31*(3), 318-323.

Wang, K., Li, M., & Hakonarson, H. (2010). ANNOVAR: Functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Research*, *38*(16), Article e164. https://doi.org/10.1093/nar/gkq603

Wang, Q., Jia, P., Li, F., Chen, H., Ji, H., Hucks, D., ... & Zhao, Z. (2013). Detecting somatic point mutations in cancer genome sequencing data: a comparison of mutation callers. *Genome medicine*, *5*(10), Article 91. https://doi.org/10.1186/gm495

WHO Classification of Tumours Editorial Board. (2023). *WHO classification of tumours: Head and neck tumours* (5th ed., Vol. 9). Lyon, France: International Agency for Research on Cancer.

Wu, L., Liu, J., Gao, P., Nakamura, M., Cao, Y., Shen, H., & Griffin, J. D. (2005). Transforming activity of MECT1-MAML2 fusion oncoprotein is mediated by constitutive CREB activation. *The EMBO journal*, *24*(13), 2391–2402. https://doi.org/10.1038/sj.emboj.7600719

Yauy, K., Van Goethem, C., Pégeot, H., Baux, D., Guignard, T., Thèze, C., ... & Cossée, M. (2023). Evaluating the transition from targeted to exome sequencing: a guide for clinical laboratories. *International Journal of Molecular Sciences*, *24*(8), Article 7330. https://doi.org/10.3390/ijms24087330

Yin, L. X., & Ha, P. K. (2016). Genetic alterations in salivary gland cancers. *Cancer*, *122*(12), 1822-1831. https://doi.org/10.1002/cncr.29890