I. INTRODUCTION

1.1 Background

The head and neck cancer cells are considered one of the top 10 common cancer deaths worldwide which is capable of affecting a massive amount of individuals per year. In the head and neck cancer cells, the Oral Squamous Cell Carcinoma (OSCC) represents 95% of all head and neck cancer cells. In addition, Oral Squamous Cell Carcinoma (OSCC) is also considered a disease that still becomes preventable due to possible early detection and its treatment (Rivera & Venegas, 2014). Oral Squamous Cell Carcinoma (OSCC) could develop due to several risk factors available. Since Oral Squamous Cell Carcinoma (OSCC) is preventable, the major risk factors to occur in most of the OSCC cases are smoking (tobacco exposure) and alcohol consumption. People with smoking habits often develop oral cancer 3 times more compared to those who are not smoking. On the other hand, alcohol consumption also increases the risk of developing oral cancer by the dissociation of the lipids on the surface of the epithelium which plagues the DNA synthesis as well as repair mechanisms (Rivera, 2015). Regarding its epidemiology, according to the World Health Organization (WHO) (2020), Oral Squamous Cell Carcinoma cases are higher in the continent of Asia compared to any other continent which accounts for 566,583 cases in 2020. In the continent of Asia specifically in Taiwan itself, According to IHME (2019), OSCC death cases were reported about 2,239 cases in 2019.

There are several causes of cancer development. One of the causes of the development of cancer is an aberrant expression of an oncogenic long non-coding RNA (lncRNA). In addition, numerous studies have stated the involvement of an oncogenic gene specifically in the Oral Squamous Cell Carcinoma (OSCC). The oncogenic long non-coding RNA (lncRNA) is considered a non-coding type of RNA which is unable to code protein and consists of more than 200 nucleotides (Yang et al., 2014). The oncogenic long non-coding RNA (lncRNA) itself appears to play an important role in the regulation of cell division, apoptosis, migration as well as invasion. Recently, the series of oncogenic lncRNA, especially the lncX, have been reported to have an association with OSCC progression (Zhang et al., 2019). However, the detailed mechanism of the lncX is still not fully understood.

In silico studies which are in the area of bioinformatics have been used recently for aiding the research investigation. The in silico analysis utilizes a computer-based method which is capable of providing precise predictions for biological studies. Hence, the analysis could aid by mitigating the pressure of laboratory work such as the cost as well as sacrificial animals (Moradi et al., 2022). For assisting the in silico studies, high throughput platforms such as the bioinformatic tools have been utilized for the purpose of in silico analysis (Wang et al., 2021).

In this present study, we would like to utilize the GEO NCBI bioinformatic tools with the reference series of GSE6631 (overexpressed genes), GSE3524 (overexpressed genes) and GSE144752 (lncX knock-downed) for the purpose of gene screening and will further analyze its gene by the expression level of overexpressed and lncX knock-downed dataset, survivability analysis of each gene, the pathway identification and the protein-protein interactions between the genes. The tools we utilize include cBioPortal (to determine correlation analysis), UALCAN (to determine expression analysis), Kaplan-Meier Plotter (to determine overall survivability analysis), DAVID (to determine pathway enrichment analysis), Metascape (to determine pathway enrichment analysis), and STRING bioinformatic tools (to determine protein-protein interactions analysis).

1.2 Objective

The aim of the study is to identify different co-expressed genes and key pathways of lncX in Oral Squamous Cell Carcinoma (OSCC) through in silico studies.

1.3 Research Scope

The scope of the study is as follows:

1. Gene screening utilizing GEO NCBI database
2. Expression analysis utilizing UALCAN & GEO NCBI database
3. Survivability analysis utilizing Kaplan Meier-Plotter bioinformatic tool
4. Pathway identification utilizing DAVID & Metascape bioinformatic tool
5. Protein-protein interaction analysis using STRING bioinformatic tool

1.4 Research Question

Various research questions were made the objective of the studies. The research questions are:

1. What are the potential genes that are considered co-expressed toward the lncX?
2. What are the potential pathways of lncX based on the bioinformatic analysis?

1.5 Research Hypothesis

For this research, the following hypothesis was made:

H0: Several of the gene candidates will not indicate co-expression as well as protein-protein interaction toward the lncX

H1: Several of the gene candidates will indicate co-expression as well as protein-protein interaction toward the lncX