Gene Sequence Analysis of Tumor Susceptibility Gene 101 (*TSG101*) in Nasopharyngeal Cancer Cell Lines

Michelle Yew Zhi Yan (70380)

Bachelor of Science with Honours
(Resource Biotechonology)
2022

Gene Sequence Analysis of Tumor Susceptibility Gene 101 (*TSG101*) in Nasopharyngeal Cancer Cell Lines

Michelle Yew Zhi Yan

A thesis submitted in partial fulfilment of the Requirement of The Degree Bachelor of Science with Honours

(Resource Biotechnology)

SUPERVISOR: Professor Dr Edmund Sim Ui Hang

Programme of Resource Biotechnology
Faculty of Resource Science and Technology
UNIVERSITI MALAYSIA SARAWAK

2022

UNIVERSITI MALAYSIA SARAWAK

Grade: _____

Please tick (√)

Final Year Project Report	☑
Masters	☐
PhD	☐

DECLARATION OF ORIGINAL WORK

This declaration is made on the ……15……day of…July……2022.

Student's Declaration:

I MICHELLE YEW ZHI YAN (70380) from Faculty of Resource Science and Technology (PLEASE INDICATE STUDENT'S NAME, MATRIC NO. AND FACULTY) hereby declare that the work entitled, Gene Sequence Analysis of Tumor Susceptibility Gene 101 (TSG101) in Nasopharyngeal Cancer Cell Lines is my original work. I have not copied from any other students' work or from any other sources except where due reference or acknowledgement is made explicitly in the text, nor has any part been written for me by another person.

15/7/2022
Date submitted

MICHELLE YEW ZHI YAN (70380)
Name of the student (Matric No.)

Supervisor's Declaration:

I, PROFESSOR DR. EDMUND SIM UI HANG (SUPERVISOR'S NAME), hereby certify that the work entitled, *As written in the student declaration section* (TITLE) was prepared by the above named student, and was submitted to the "FACULTY" as a * partial/full fulfillment for the conferment of BACHELOR OF SCIENCE WITH HONS (RESOURCE BIOTECHNOLOGY) (PLEASE INDICATE THE DEGREE), and the aforementioned work, to the best of my knowledge, is the said student's work

Received for examination by: Prof. Dr. Edmund Sim Ui Hang
(Name of the supervisor)

Date: 15 July 2022

I declare this Project/Thesis is classified as (Please tick (√)):

☐ **CONFIDENTIAL** (Contains confidential information under the Official Secret Act 1972)*** RESTRICTED** (Contains restricted information as specified by the organisation where research was done)*****

☑ **OPEN ACCESS**

Validation of Project/Thesis

I therefore duly affirmed with free consent and willingness declared that this said Project/Thesis shall be placed officially in the Centre for Academic Information Services with the abide interest and rights as follows:

- This Project/Thesis is the sole legal property of Universiti Malaysia Sarawak (UNIMAS).
- The Centre for Academic Information Services has the lawful right to make copies for the purpose of academic and research only and not for other purpose.
- The Centre for Academic Information Services has the lawful right to digitise the content to for the Local Content Database.
- The Centre for Academic Information Services has the lawful right to make copies of the Project/Thesis for academic exchange between Higher Learning Institute.
- No dispute or any claim shall arise from the student himself / herself neither third party on this Project/Thesis once it becomes sole property of UNIMAS.
- This Project/Thesis or any material, data and information related to it shall not be distributed, published or disclosed to any party by the student except with UNIMAS permission.

Student's signature michelle (15/7/2022)

Supervisor's signature: Edmund MH Sim (15/07/2022)

Current Address:

_____

_____

Notes: ***** If the Project/Thesis is **CONFIDENTIAL** or **RESTRICTED**, please attach together as annexure a letter from the organisation with the period and reasons of confidentiality and restriction.

[The instrument was duly prepared by The Centre for Academic Information Services]

ACKNOWLEDGEMENT

First and foremost, I would want to thank my supervisor, Professor Dr Edmund Sim Ui Hang, for supervising and advising me during this project. Furthermore, I would want to express my heartfelt gratitude to him for his assistance and engagement in the project; without his supervision, I would not be able to successfully complete the final year project.

Besides, I would like to thank the Postgraduates students at UNIMAS's Immunology Molecular Genetics Laboratory, Ms. Adrianne, Ms. Daphne, and MS. Alyaa Syafiqah. They provided me with valuable advice, information, and assistance in addressing every issue I had throughout the assignment.

Furthermore, I would want to thank my family members, particularly my parents, because I could not finish my report without their compassion, patience, and encouragement. They would provide me spiritual support when I met setbacks and challenges in completing this final report. Furthermore, I would like to thank my lab mates, Aliza Chung Yen Fen, Brennan Tang Yet Shen, Bonnie Tang Bao Chee, Emmanuella Nina Christina, Ling Zhao Wen, and Thanes Gunasekaran, for their compassion, support, and unity throughout the research. Last but not least, I would want to express my gratitude to all of my friends and classmates for their encouragement and support during the course of my project.

Gene Sequence Analysis of Tumor Susceptibility Gena (TSG101) in Nasopharyngeal Cancer Cell Lines

Michelle Yew Zhi Yan

Resource Biotechnology Programme
Faculty of Resource Science and Technology
Universiti Malaysia Sarawak

ABSTRACT

TSG101 acts as a tumor suppressor gene. Most research also indicates that *TSG101* may have a variety of functions in carcinogenesis as well as the development of tumors in diverse cell types. The recent findings proved that mutations of the *TSG101* gene were detected in many cancer diseases such as ovarian carcinomas, human papillary thyroid carcinomas, colorectal cancer, and others. Furthermore, a deletion known as *TSG101* Del 154-1054 was discovered in the NPC cell line (TW01). Despite this, research on nasopharyngeal carcinomas (NPC) cell line mutations such as insertions, duplication, and others were still restricted. The objectives of this study are to analyse and compare the *TSG101* gene sequence between the NCBI reference sequence and the NPC cell lines TW01 and EBV infected C666-1, as well as to observe the other types of mutation or genetic lesions in the *TSG101* gene in the context of NPC. The starting materials were total RNA from TW01 and C666-1, which were then transcribed into cDNA using M-MLV RT. The cDNA was then amplified using polymerase chain reaction and then proceed with sequence analysis. There were bands identified between 200 and 300bp, indicating that *TSG101* amplicon. The PCR product for these studies was supposed as a part of UEV. However, the result of this study are inconclusive, since the objectives of these studies cannot proceed due to lack of sequencing because the amplified *TSG101* was not sent for sequencing in time.

Key words: *TSG101*, Nasopharyngeal Carcinoma, gene sequence, TW01, C666-1

ABSTRAK

TSG101 *bertindak sebagai gen penindas tumor. Kebanyakan penyelidikan juga menunjuk bahawa* TSG101 *mungkin mempunyai pelbagai fungsi dalam karsinogenesis serta perkembangan tumor dalam pelbagai jenis sel. Dalam pembiayaan baru-baru ini, membuktikan bahawa mutasi gen* TSG101 *dikesan dalam banyak penyakit kanser seperti karsinoma ovari, karsinoma tiroid papillary manusia, kanser kolorektal dan lain-lain. Tambahan pula, pemadaman yang dikenali sebagai* TSG101 *Del 154-1054 ditemui dalam talian sel NPC (TW01). Walaupun begitu, penyelidikan mengenai mutasi sel karsinoma nasofaring (NPC) seperti sisipan, penduaan dan lain-lain masih terhad. bjektif kajian ini adalah untuk menganalisis dan membandingkan jujukan gen* TSG101 *antara jujukan rujukan NCBI dan talian sel NPC TW01 dan c666-1 yang dijangkiti EBV, serta memerhati jenis mutasi atau lesi genetik lain dalam gen* TSG101 *dalam konteks NPC. Bahan permulaan merupakan jumlah RNA daripada TW01 dan C666-1, yang kemudiannya ditranskripsikan ke dalam cDNA menggunakan M-MLV RT. cDNA kemudiannya dikuatkan menggunakan tindak balas rantai polimerase dan kemudian akan meneruskan dengan analisis jujukan. Terdapat jalur yang dikenal pasti antara 200 dan 300bp, menunjukkan bahawa amplikon* TSG101. *Produk PCR untuk kajian ini sepatutnya sebagai sebahagian daripada UEV. Walau bagaimanapun, hasil kajian ini tidak dapat disimpulkan, kerana objektif kajian ini tidak dapat diteruskan kerana kekurangan penjujukan kerana* TSG101 *yang dikuatkan tidak dihantar untuk penjujukan dalam masa.*

Kata kunci: TSG101, Karsinoma Nasofaring, jujukan gene, TW01, C666-1

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AGE	Agarose Gel Electrophoresis
BLAST	Basic Local Alignment Search Tool
bp	Base Pair
cDNA	Complementary Deoxyribonucleic Acid
DNA	Deoxyribonucleic Acid
dNTP	Deoxyribonucleotide
EBV	Epstein-Barr virus
EGF	Epidermal Growth Factor
GADPH	Glyceraldehyde-3-phosphate dehydrogenase
gDNA	Genomic DNA
IARC	International Agency for Research on Cancer
LMP1	Latent Membrane Protein 1
MAFFT	Multiple Alignment using Fast Fourier Transform
M-MLV RT	Moloney Murine Leukemia Virus Reverse Transcriptase
mRNA	Messenger Ribonucleic Acid
MVBs	Multivesicular bodies
NCBI	National Centre for Biotechnology

NPC	Nasopharyngeal Carcinoma
OD	Optical Density
PBS	Phosphate Buffered Saline
PCR	Polymerase Chain Reaction
PD	Primer Dimer
RT-PCR	Reverse Transcriptase- Polymerase Chain Reaction
SSTs	Similarity Searching Tools
Ta	Annealing Temperature
TAE	Tris-acetate EDTA
TE	Tris-EDTA
Tm	Melting Temperature
TSG101	Tumor susceptibility gene 101
UBC	Ubiquitin-conjugating Enzyme
UEV	Ubiquitin E2 Variant

CHAPTER 1

INTRODUCTION

TSG101 (Tumor Susceptibility Gene 101) is a human gene that maps on chromosome 11, p15.1 to p15.2. According to Li and Cohen (1996), the *TSG101* was found from mice NIH3T3 fibroblasts through an approach known as the homozygous knockout approach. When converted mice fibroblasts implant to nude mice, functional inactivation of such *TSG101* induces cell transformation in the laboratory as well as the metastasising tumor in animals. After using Cre-lox recombination to delete the genes which transactivated a nonsense promoter, the tumor characteristics will reverse in vitro. These findings revealed that *TSG101* deficiency caused tumorigenesis as well as transformation.

Besides, according to Xie, Li, and Cohen (1998), *TSG101* may alter the ½ life of other critical anti-oncogenes as well as cell cycle regulatory proteins. Because *TSG101* inactivation has been linked to aberrant effector centres but also nuclear aberrations, is therefore believed as *TSG101* protein loss or the lack of certain *TSG101* domains can result in chromosomal instability thus cancer development.

Nasopharyngeal cancer is also known as nasopharyngeal carcinoma or NPC. NPC is a cancer that originates at the upper part of the throat and behind the nostril. According to Chen et al. (2019), NPC is epithelial cancer that develops mostly from nasopharyngeal mucosa as well as was frequently observed in pharyngeal recess. As per the IARC, roughly 129000 new instances of nasopharyngeal carcinoma were detected in 2018, representing only 0.7 percent of all carcinoma detected in that year. Yet, the worldwide geographical allocation is very imbalanced, with even more than 70 percent of newly reported cases in East and Southeast Asia. Numerous variables, including host genetics, EBV infection, and environmental factors, are thought to contribute to the formation of nasopharyngeal

carcinomas. *TSG101* gene and its erroneously spliced isoform, *TSG101* Del154-1054, are strongly associated with carcinogenesis in a variety of cancer (Chua, Kameyama, Mayeda, & Yeh, 2019).

According to Chua et al. (2019), *TSG101* Del 154-1054 mRNA generated an increase of *TSG101* protein in human nasopharyngeal carcinoma (NPC) biopsies, which was replicated by overexpression of *TSG101* Del 154-1054 in the NPC cell line TW01. However, the research on mutations such as insertions, duplication, and others of the NPC cell line is still limited. Thus, further studies on the gene sequence analysis of *TSG101* between NPC cell lines and reference sequence of *TSG101* gene are required to understand the relationship between the *TSG101* gene in NPC cell lines and reference sequence of *TSG101* gene, as well as to observe the other types of mutation or genetic lesions in the *TSG101* gene in the context of NPC. Thus, in this research, *TSG101* was derived from the TW01 (NPC cell line), and EVB infected C666-1 (NPC cell line) and being analyse and compare with reference sequence of *TSG101* gene (accession number: NM 006292.4) at NCBI website. *TSG101* was also found a mutation in human papillary thyroid carcinomas (Liu et al., 2002). Thus, the gene sequence of *TSG101* is expected to be different between the reference sequence of *TSG101* gene at NCBI website and NPC cell line (TW01 and EVB infected C666-1).

The objectives of this study are:

1. To analyse and compare the gene sequence of *TSG101* between reference sequence of *TSG101* gene (accession number: NM 006292.4) at National centre for Biotechnology (NCBI) website and NPC cell line TW01 and EBV infected C666-1.
2. To observe the other types of mutation or genetic lesions in the *TSG101* gene in the context of NPC.

The hypothesis statement of this study is:

The gene sequence of *TSG101* is expected to be different between the reference sequence of *TSG101* gene at NCBI website and NPC cell line (TW01 and EVB infected C666-1).

CHAPTER 2

LITERATURE REVIEW

2.1 *TSG101*

TSG101 is a human gene that maps on chromosome 11. *TSG101*, formerly called CC2, has been discovered inside a *saccharomyces cerevisiae* two-hybrid screens as one coiled-coil domain-containing protein which binds to stamin. *TSG101* encodes the multi-domain protein with one potential DNA binding protein at its COOH terminus that could act as a transcriptional cofactor for inhibiting nor stimulating nuclear hormone receptor-mediated transactivation. In contrast, *TSG101*'s NH_2-terminal region shares considerable sequence similarity with the UBC domain of the ubiquitin-conjugating enzyme, yet lacking a crucial active site cysteine required for enzyme production, indicating that *TSG101* is destroyed by ubiquitin-mediated protein (Young et al., 2007).

According to Lu, Hope, Brasch, Reinhard, and Cohen (2003), *TSG101* may be implicated in the negative control of receptor tyrosine kinase signalling because it blocks the transport of activated EGF receptors to the endosome of the lysosome, resulting in prolonged stimulation of the downstream signal cascade. As a consequence, much research is providing light on the molecular mechanism through which *TSG101* acts as a tumor suppressor gene. On the other hand, Zhu et al. (2004) stated that *TSG101* has been shown to serve as a tumor-promoting gene in certain epithelial tumor cells. This indicates that *TSG101* may have a variety of functions in carcinogenesis as well as the development of tumors in diverse cell types.

In terms of cancer, *TSG101* suppresses the tumor suppressor protein p21 and is involved in ovarian carcinomas (Young et al., 2007). This *TSG101* has also been linked to

more aggressive activity in colorectal cancer (Gheytanchi et al., 2021). In addition, *TSG101* overexpression seems to enhance the vulnerability of mammary epithelia to malignant transformation in aged women. *TSG101* is therefore expected to have a more prominent role in the advancement of a subset of spontaneously occurring breast cancers rather than cancer start (Oh, Stanton, West, Todd, & Wanger, 2007). While in the NPC context, *TSG101* Del 154-1054 mRNA increased TSG101 protein in human NPC biopsies, that was replicated by *TSG101* Del 154-1054 overexpression in the NPC cell line TW01 (Chua, Kameyama, Mayeda, & Yeh, 2019). The research above showed that *TSG101* is related to carcinogenesis.

2.2 Nasopharyngeal Carcinoma (NPC)

Nasopharyngeal carcinoma (NPC) is a cancer that originates at the upper part of the throat and behind the nostril. According to Wei, Xu, Liu, Zhang, and Liang (2011), primary nasopharyngeal carcinoma is a tumor that develops inside the mucosal epithelium and nasopharyngeal salivary glands. Clinical manifestations of NPC include headache, neck lumps, difficult breathing or discharge, epistaxis, and other non-specific indicators. Besides, since the carcinoma is placed in a quiet anatomical region as well as NPC has a high metastatic rate (Lo, To, & Huang, 2004), NPC is frequently detected at an advanced stage. Tang et al. (2015) found that about 70 percent of patients were found to be at an advanced stage at the time of diagnosis. Patients with nasopharyngeal carcinoma had a 10-year survival rate of 98 percent in stage I and 60 percent in stage II (Chuan, Sham, Kwong, & Au, 2003), with a median survival period of 3 years in advanced patients (Al-sarraf et al., 1998).

There were approximately 129000 cases reported with nasopharyngeal cancer in 2018. Furthermore, males have a greater incidence of nasopharyngeal cancer compared to females (Chen et al., 2019). The incidence of NPC is highest among Chinese residing in

southeast China, and the age-standardized ratio of men is 21 per 100,000. The prevalence of indigenous peoples in North African nations, Southeast Asian nations, as well as Arctic areas, is moderately high. Instead, the rate is less than 1 per 100,000 (Devi, Pisani, Tang, & Parkin, 2004). Nasopharyngeal carcinoma has become one of Malaysia's top 5 carcinomas since 1966. Furthermore, nasopharyngeal carcinoma was indeed the second most prevalent carcinoma among men in Sarawak between 1981 and 1982. Between January 1981 until December 1982, the Land Dayak had a crude rate of nasopharyngeal carcinoma of 9.0 per 100,000 people per year, followed by the Sarawak Chinese at 5.8 per 100,000 (Linton et al., 2021).

A variety of factors, including host genetics, EBV infection, and environmental factors, are thought to play an impact on the development of nasopharyngeal carcinomas. Linton et al., (2021) found Epstein-Barr virus (EBV), which can cause Hodgkin and some no Hodgkin lymphomas, NPC, and stomach carcinoma is one of the cancer-causing viruses. More than 95 percent of NPC cases are linked with EBV in areas where the disease is prevalent. According to Chua et al., (2020) Rta, an instant protein encrypted by the EBV, interacts with *TSG101* in increasing the expression for late EBV genes. Anti-TSG101 antibody enhanced *TSG101* levels in EBV-negative NPC-TW01-pZip cells as well as NA when treated with TPA and SB. Thus, either Rta being activated intracellularly with chemical manipulation nor produced extracellularly through ectopic transfection, Rta protein can successfully interact with *TSG101* protein without the assistance of other viral products. These findings showed that overexpression of *TSG101* will cause tumorigenesis and can be found in NPC cell lines.

2.3 Epstein-Barr Virus (EBV)

According to Sarwari, Khoury, and Hernandez (2016), the Epstein-Barr virus (EBV) is a double-stranded DNA virus of the Herpes family that causes infectious mononucleosis (IM). It is the highest prevalent infectious disease in the world with a prevalence of 90 percent. EBV was initially found in African Burkitt's lymphoma isolated cells, but it was later revealed to be widespread around the world (Epstein, Achong, Barr, 1964). As with other herpesviruses, following a primary infection, EBV undergoes a latency phase in which it invades epithelial cells, penetrates circulating B lymphocytes, and survives for life (Young & Rickinson, 2004).

EBV infection can also produce neurological and hematological problems, however, this is less common. EBV have been related to autoimmune diseases like multiple sclerosis, arthritis, as well as systemic lupus erythematosus (Lossius, Johansen, Torkildsen, Vartdal, & Holmoy, 2012). Besides, Tzellos and Farrell (2012) stated that there are numerous malignancies associated with this virus, including nasopharyngeal carcinomas (NPCs), stomach cancers, posttransplant lymphoproliferative disorders (PTLDs), as well as Hodgkin's lymphomas. EBV is transmitted primarily via the oral route. One of the most common EBV infections is mononucleosis, which is spread through kissing (Kaye, 2021). Furthermore, organ transplantation and blood transfusion have been linked to EBV transmission (Gerber, Walsh, Rosenblum, & Purcell, 1969; Alfieri, 1996).

2.4 Cell Lines

According to Kaur and Dufour (2012), in most cases, cell lines have been utilised instead of original cells to research biological processes. In addition, it had modified scientific studies and is now used for vaccine manufacturing, antibodies creation, medication metabolic and cytotoxic tests, gene function study, the synthesis of biological substances, as

well as the formation of artificial tissues. Besides, Borell (2010) states that most of the mutations could be found by comparing tumor cell sequences with normal tissue from the same person.

It has numerous benefits, such as being easy to use, unlimited in the supply of material, cost-effective, and forwarding on ethical issues related to the use of experimental animals and humans. Cell lines can provide a purified population of cells, which is advantageous as it ensures a constant sample as well as repeatable findings (Kaur & Dufour, 2012). However, not all cancer types can be continually cultured in the laboratory, which is a drawback of cell lines (Borell, 2010). As a result, we must exercise caution when substituting cell lines for primary cells. The cell lines should have functional properties that are as near to primary cells as feasible.

Gullo, Low, and Teoh (2008) state that more than twenty NPC cell lines have been identified since 1975. The NPC cell line is commonly used in NPC research. Continuous cell lines, in particular, are useful tools for studying the genetic pathways and environmental variables that contribute to NPC, both in terms of tumorigenicity as well as metastatic potential. HONE-1, HK1, SUNE-1, C666-1, TW04, and TW01 are some examples of NPC-derived cell lines. In this research, the nasopharyngeal carcinoma cell lines selected were TW01 and EVB infected C666-1.

2.4.1 TW01 Cell Line

TW01 is one of the examples of nasopharyngeal carcinoma cell lines. Human NPC cell lines included NPC-TW01, NPC-TW03, and NPC-TW04 (Wang et al., 2019). This cell line is often used to study nasopharyngeal carcinoma to have a better understanding of this cancer. For example, TW01 has been utilised to investigate how fibronectin enhances nasopharyngeal cancer cell migration as well as proliferation. EBV LMP1 was found to be

a causative facilitator of fibronectin formation in NPC cell lines, as stimulated fibronectin adds towards LMP1-stimulated cell migration (Wang et al., 2019). In addition, *TSG101* Del 154-1054 in the NPC cell line TW01 demonstrated that it did produce *TSG101* protein increase in human NPC biopsy tissue. They also indicate TSG154-1054-induced protein stability of *TSG101* is required for tumor cell metastasis (Chua, Kameyama, Mayeda, & Yeh, 2019). The following figure 1 shows the image of TW01 under inverted microscope.

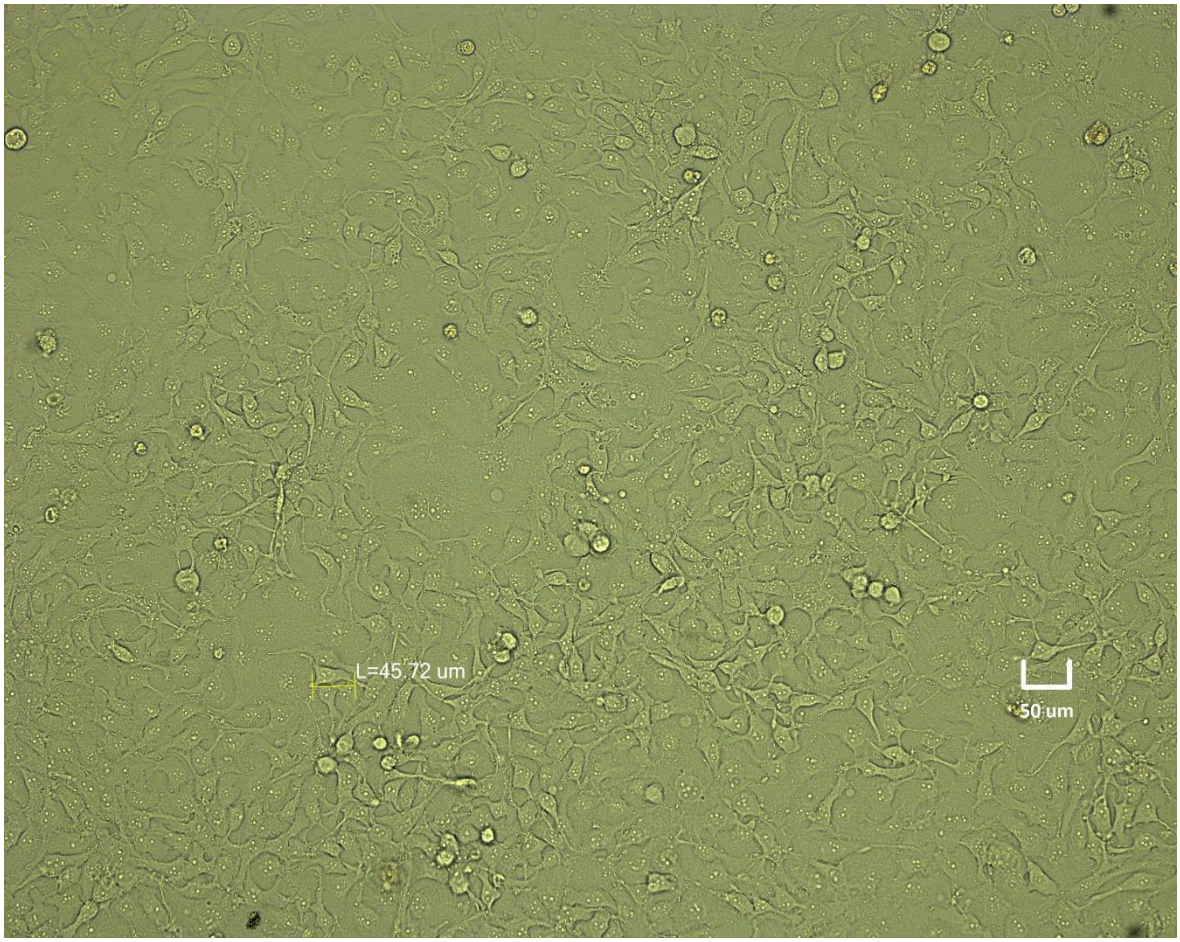

Figure 1. Image of TW01 under inverted microscope

2.4.2 C666-1 Cell Line

C666-1 is one of the examples of nasopharyngeal carcinoma cell lines. Nasopharyngeal carcinoma (NPC) is uncommon in almost all of the globe, although it is common in South China and Southeast Asia. One of epithelial cancer's distinguishing traits is its continuous relationship with Epstein–Barr virus (EBV) infection (Chan et al., 2008).

EBV has been established in research to play a major function in the formation and maintenance of the tumor phenotype in this cancer (Lo, To, & Huang, 2004). To date, just a few NPC cell lines consistently carrying the EBV genome, such as C666-1, were employed as EBV-positive NPC models (Chan et al., 2008). According to Cheung et al. (1999), C666-1 is a sub clone of C666 (its original cell line), which was obtained from a xenograft of a southern Chinese NPC. It grows as an adhering colony with no inhibition to touch. Furthermore, it also induces tumors among athymic naked rats. Because the cells continuously produce EBV-encoded RNAs, they were positive for the epithelial marker cytoskeletal proteins. C666-1 may be useful as a research tool since the genotype as well as virus latent patterns are found inside the most of original NPC biopsies among patients diagnosed. The figure 2 below shows the image of EBV infected C666-1 under microscope inverted.

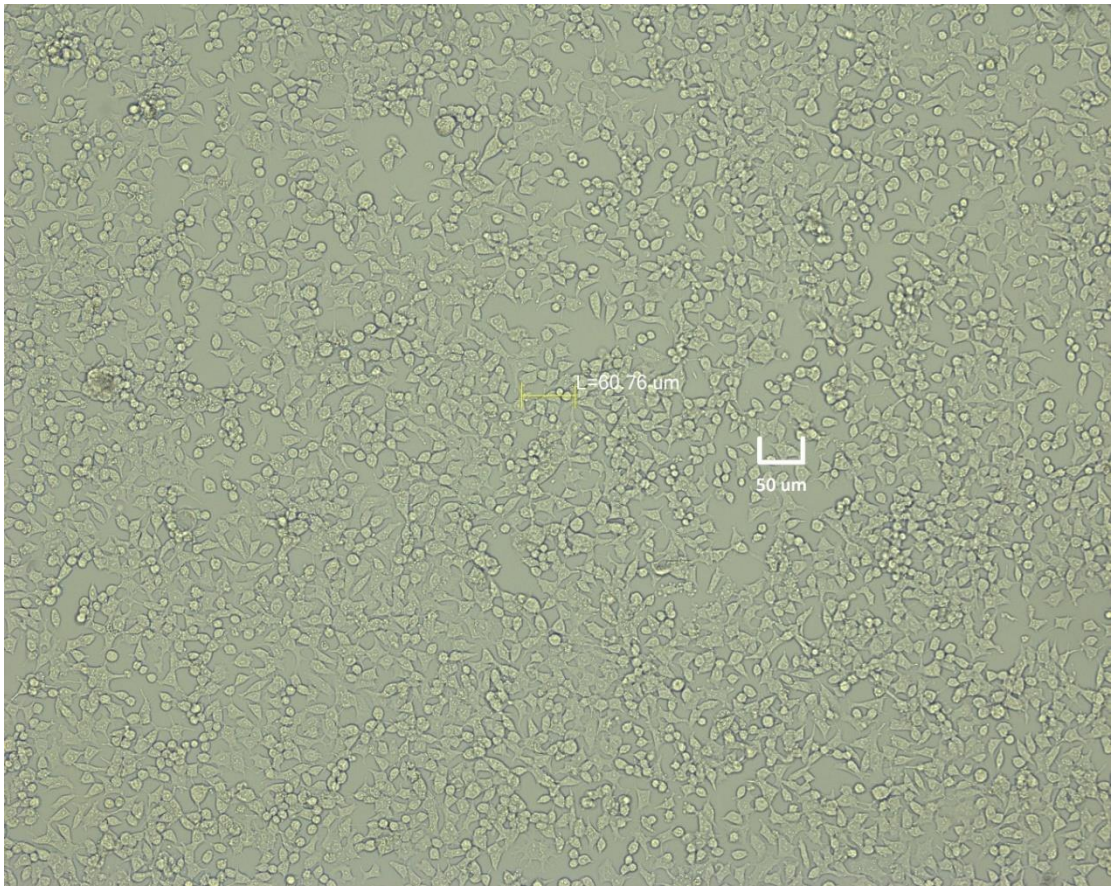

Figure 2. Image of EVB infected C666-1 under inverted microscope

2.5 Housekeeping Genes

The housekeeping gene was employed as a positive control in this study. Housekeeping genes seem to be cellular maintenance genes that govern fundamental and pervasive cellular activities (Turabelidze, Guo, & DiPierro, 2010). The expression levels of housekeeping genes vary slightly nor not at all between individual samples or experimental settings inside a broad range of cell and tissues types (Reboucas et al., 2013). The housekeeping gene used in this study was glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*). The *GADPH* enzyme catalyses the conversion of glucose to pyruvate in the sixth stage of glycolysis. It is frequently used as a housekeeping gene. Its engagement in glycolysis is not its only function, it also helps in transcriptional control (Sirover 2005) as well as DNA repair (Meyer-Siegler et al. 1991).

2.6 DNA Sequence Analysis

DNA sequence identification, as well as similarity searching tools (SSTs), are the first and most essential programmed in biological research, assisting scientists in making proper decisions regarding species identity or categorisation by offering information regarding closely related organisms as a consequence. Bioinformatics uses statistics, mathematical models, and computer science to analyze biological data and compare sequences to find similarities in two or more sequences at the time of a similarity search (Dwivedi, Bharadwaj, Mohanty, & Gupta, 2018).

BLAST is a powerful similarity search tool with a plethora of options for looking for sequence similarity in lengthy DNA sequences (Altschul, Gish, Miller, Myers, & Lipman, 1990). In addition to conducting alignments, BLAST gives the statistical information about an alignment, highlighting mismatches between comparable sequences and identifying where the query was masked for low-complexity sequences (Ye, McGinnis, & Madden,