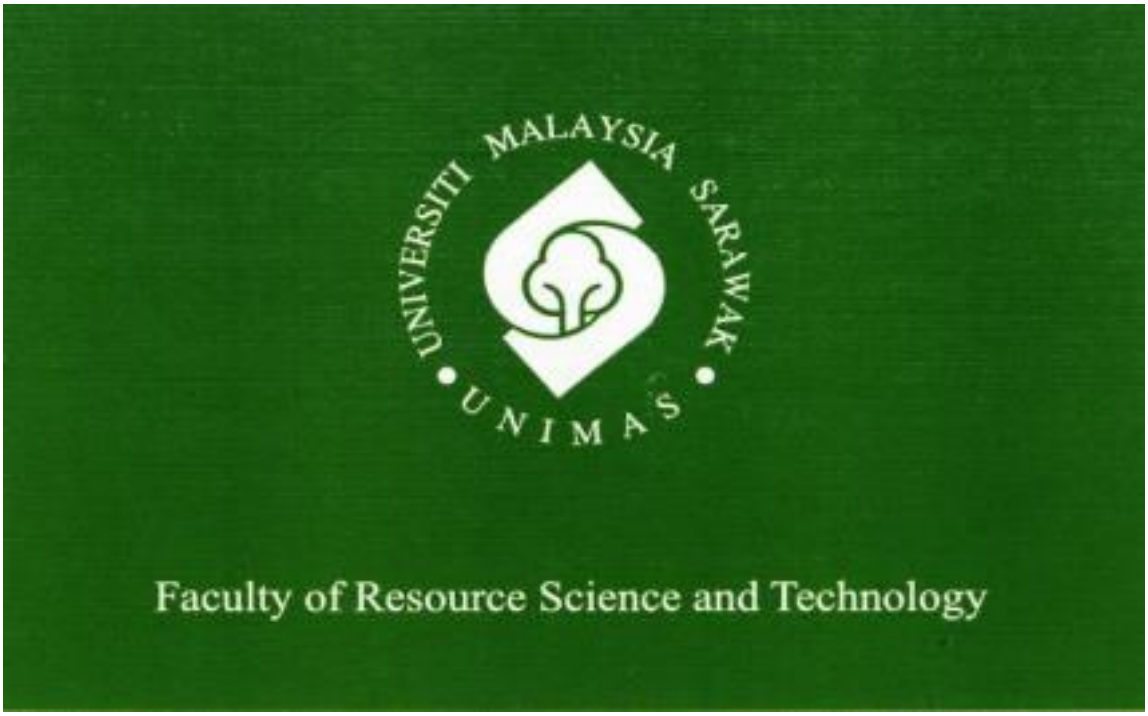

Gene Sequence Analysis of *Tumour Susceptibility Gene (TSG101*) in Colorectal Carcinoma Cell Lines

Bonnie Tang Bao Chee (69238)

Bachelor of Science with Honour

(Biotechnology Resource)

2022

DECLARATION

UNIVERSITI MALAYSIA SARAWAK

Grade: _____

Please tick (√)

Final Year Project Report	√
Masters	
PhD	

DECLARATION OF ORIGINAL WORK

This declaration is made on the 27th day of May 2022

Student's Declaration:

I Bonnie Tang Bao Chee (69238), Faulty Resource Science and Technology

(PLEASE INDICATE STUDENT'S NAME, MATRIC NO. AND FACULTY) hereby declare that the work entitled, Gene Sequencing of Tumour Susceptibility Gene 101 (TSG 101) in Colorectal Carcinoma 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.

27th May 2022
Date submitted

Bonnie Tang Bao Chee (69238)
Name of the student (Matric No.)

Supervisor's Declaration:

I, PROF DR EDMUND SIM UI HANG (SUPERVISOR'S NAME), hereby certify that the work entitled, *As written above (refer student declaration)* (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 Honours (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: 1 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 _____ (27th May 2022)

Supervisor's signature: _____ (Date) 1 July 2022

Current Address:
503, Lorong Unicentral 5A1B, Kota Samarahan, Sarawak.

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]

ACKNOWLEDGEMENTS

The project would have not been successfully completed without the assistance of a few people that have been such a great help to me. First and foremost, I thank God for his showers of blessings and never forsaking me in times of need. This project would not have been possible without his blessings. Furthermore, deepest appreciation to my supervisor, Prof Dr. Edmund Sim Ui Hang for his patient guidance and consultation throughout this project. Besides, I would like to extend my gratitude toward the postgraduate students, Adrianne Vong Dien Yu, Allya Ayafiqah binti Shamsuri, and Daphne Olivia Jawai Anak Sadai for their help along with this project. They had provided me with lots of guidance and shared their experience with me regarding all the research work despite of her own busy schedule. In addition, I also wish to thank my colleagues, Brennan Tang Yet Shen and Michelle Yew that had advised and helped me a lot in this project. Lastly, my family members are my great support and I thank them for their encouragement throughout this project. They had given me endless support although they were not by my side.

Gene Sequence Analysis of *Tumour Susceptibility Gene (TSG101)* in Colereactal Carcinoma Cell Lines

Bonnie Tang Bao Chee

Resource Biotechnology
Faculty of Resource Science and Technology
University Malaysia Sarawak

ABSTRACT

Colorectal carcinoma, or colorectal cancer (CRC), is a type of cancer that affects the colon or rectal area and it is likely to be related with *TSG101* gene expression. *TSG101* is a gene that encodes a multi-domain protein that is homologous to an inactive ubiquitin-conjugating enzyme and it may act as a positive enhancer in carcinogenesis and CRC progression. In the recent finding on the *TSG101* has shown to be high expression in CRC, but there are no studies that have been done in determining and analyzing the gene sequences of *TSG101* in CRC cell lines. The objectives of this study are to determine the existence of *TSG101* in CRC cell lines and to perform *In Silico* Characterization of *TSG101* in CRC cell lines. The starting materials of this study were the total RNA that was isolated from HCT116 and HT29 cell lines, and they were reverse transcribed into cDNA to perform PCR. Lastly, the amplified products proceeded for sequencing and analysis. However, due to some unavoidable errors, we could not send our PCR product on time. The PCR products showed expression in the expected size range (200-300bp), hence the products were likely to be *TSG101* in HCT116 and HT29 cell lines.Therefore, we subjected the expected amplified sequence (NM_006292.4) on CD search and we found out that the predicted amplified product is located at the nucleotide 499 to 739 in *TSG101* and it is a region of UEV domain.

Keywords: *Colorectal carcinoma (CRC), HCT116, Tumour susceptibility gene 101 (TSG101), Sequence analysis*

Analisis Urutan Tumour Susceptibility Gene (TSG101) dalam Garisan Sel Kanser Kolorektal

Karsinoma Kolorektal ialah sejenis barah yang mempengaruhi kawasan kolon dan rektal. Karsinoma kolorektal berkemungkinan berkaitan dengan ekspresi gen TSG101. TSG101 merupakan gen yang mengkodekan protein berbilang domain yang homolog kepada enzim ubiquitin-conjugating yang tidak aktif dan ia boleh bertindak sebagai penambah positif di karsinogenesis dan perkembangan CRC. Penyelidikan terkini melaporkan TSG101 mempunyai espresi tinggi dalam karsinoma kolorektal, tetapi tiada kajian dilakukan untuk menentukan dan menganalisis gen urutan TSG101 dalam karsinoma kolorektal. Objektik kajian ini adalah untuk menetukan kewujudan TSG101 dalam karsinoma kolorektal dan pencarian "in silico" TSG101 dalam karsinoma kolorektal. Bahan permulaan dalam kajian ini merupakan jumlah RNA daripada HCT116 dan HT29, dan ia akan ditranskripsi kepada cDNA untuk menjalankan PCR. Akhirnya, produk PCR akan diteruskan untuk penganalisasi urutan. Walau bagaimanapun, disebabkan ralat yang tidak dapat dielakkan, kami tidak dapat menghantar produk PCR kami tepat pada masanya. Produk PCR menunjukkan ekspresi dalam julat saiz yang dijangkakan (200-300bp), maka produk tersebut berkemungkinan TSG101 dalam garisan sel HCT116 dan HT29. Oleh itu, kami tertakluk kepada urutan yang diamplifikasi (NM_006292.4) pada carian CD dan kami mendapati bahawa produk yang diamplifikasi terletak pada nucleotida 499 hingga 739 dalam TSG101 dan ia merupakan sebahagian daripada domain UEV.

Kata kunci: Karsinoma Kolorektal (CRC), HCT116, HT29, Tumour susceptibility gene 101 (TSG101), Analysis Urutan

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AGE	Agarose Gel Electrophoresis
BLAST	Basic Local Alignment Search Tool
CC	Coiled Coil
cDNA	Complementary Deoxyribonucleic Acid
CDS	Coding DNA Sequence
CD Search	Conserved Domain Search
CRC	Colorectal Carcinoma
ddH20	Double Distilled Water
DNA	Deoxyribonucleic Acid
ESCRT	Endosomal Sorting Complexes Required for Transportation
EtBr	Ethidium Bromine
FBS	Fetal Bovine Serum
GADPH	Glyceraldehyde-3-Phosphate Dehydrogenase
GISTs	Gastrointestinal Stromal Tumours
MDM2	Mouse double minute 2 homolog
MEME	Multiple Expectation maximizations for Motif Elicitation
miRNA	Micro Ribonucleic Acid
M-MLVRT	Moloney Murine Leukemia Virus Reverse Transcriptase
MVB	Multivesicular Bodies
NETs	Neuroendocrine Tumours
NCBI	National Centre for Biotechnology
NPC	Nasopharyngeal Carcinoma
TSG101	*Tumour Susceptibility Gene 101*
PCR	Polymerase Chain Reaction

PRR	Proline-rich Region
qPCR	Quantitative Polymerase Chain Reaction
RNA	Ribonucleic Acid
RT-PCR	Reverse Transcriptase Polymerase Chain Reaction
S-box	Steadiness Box
SDS	Sodium dodecyl Sulfate
UBC	Ubiquitin-Conjugating
UEV	Ubiquitin E2 Variant
UV	Ultraviolet

CHAPTER 1

INTRODUCTION

1.1 Research Background

Colorectal carcinoma, or colorectal cancer (CRC), is a type of cancer that affects the colon or rectal area. The reason these two cancers are often grouped together is that they share numerous characteristics in common. According to the International Agency for Cancer Research (2020), CRC is the second most common type of cancer in Malaysia, affecting 19.6 percent of Malaysians of all sexes and ages. CRCs frequently begin as growths on the epithelial cells of the colorectal mucosa, which are referred to as polyps. Indeed, not all polyps are malignant and these benign polyps usually do not invade nearby tissue; hence, they can be removed by using colonoscopy. However, if the polyps are not removed, they could turn into a malignant tumour, also known as cancer. Adenomas are the most prevalent type of neoplastic polyps that may develop into cancer. Adenocarcinomas account for almost 90% of CRC cases, whereas other uncommon forms include spindle cells, undifferentiated carcinoma, squamous cells, and adenosquamous cells (Fleming, Ravula, Tatishchev & Wang, 2012).

In actuality, around 70% of CRCs are moderately differentiated or grade 2 tumours in a two-tiered grading system. It is apparent that determining the tumour grade is critical because it correlates with the survival rate of patients. CRC can be diagnosed through screening, which involves performing an endoscopic biopsy and polypectomy to determine whether the polyps are malignant. Early identification of CRC is crucial, as it is highly curable. This is because cancer cells remain localized to a particular region and do not spread to other tissues in the early stages. The process of cancer cells invading neighbouring tissues is known as metastasis. It is facilitated by an extracellular vesicle called an exosome that is secreted by cancer cells during cancer progression. Studies had shown that exosomes are critical for tumour

growth because they carry specific miRNA to neighbouring cells, and *tumour susceptibility gene 101 (TSG101)* is one of the exosome markers that aid in miRNA transport (Azmi, Bao, & Sarkar, 2013). As a result, the main reason for choosing *TSG101* in this study as the upregulation of *TSG101* and its aberrant form have been strongly related to cancer development.

TSG101 is a gene that encodes a multi-domain protein that is homologous to an inactive ubiquitin-conjugating enzyme. The previous study by Liu et al. (1997) reported that the functional inactivation of *TSG 101* in mice will lead to tumorigenesis, especially in metastasis tumour. Therefore, we aimed to study the existence of *TSG101* in CRC cell lines and characterize the genomic structure of TSG101 and compare their differences with wild type of *TSG101* (NM_006292.4) in National Center for Biotechnology Information (NCBI). There is evidence indicate that *TSG101* has shown to be high expression in CRC, but no studies that have been done in determining and analyzing the gene sequences of *TSG101* in CRC cell lines. Therefore, in this study, two CRC cell lines which are HT29 and HCT116 were screened by using polymerase chain reaction (PCR) and the products were outsourced for sequencing by commercial vendors. In addition, bioinformatics analysis was carried out for sequence alignment and comparison.

1.2 Objectives

The main purpose of this study is to characterize the *TSG101* in HT29 and HCT116 and compare them between the wild type of *TSG101*. The objectives of this study are:

1. To isolate gene sequences of *TSG101* in CRC cell lines

2. To perform comparative characterization of *TSG101* in CRC cell lines.

CHAPTER 2

LITERATURE REVIEW

2.1 Colorectal Cancer

CRC is a kind of cancer that can affect either the colon or the rectum and is the most frequent type of gastrointestinal cancer. The majority of CRCs originate as polyps, which are referred to as colorectal polyps. Colorectal polyps are usually benign, meaning that the polyps are not cancerous or invading to other tissues. However, if the polyps are not removed, they have the potential to develop into a malignant tumour. According to Bujanda et al., (2010), around 90% of CRCs are adenocarcinomas, a kind of neoplastic polyps capable of developing into malignant tumours. Additionally, polyp size is associated with an increased risk of the polyp developing into cancer. The larger the polyps, the more malignant they are. Apart from adenocarcinomas, other uncommon forms of colorectal carcinoma include neuroendocrine tumours (NETs), gastrointestinal stromal tumours (GISTs), and lymphoma.

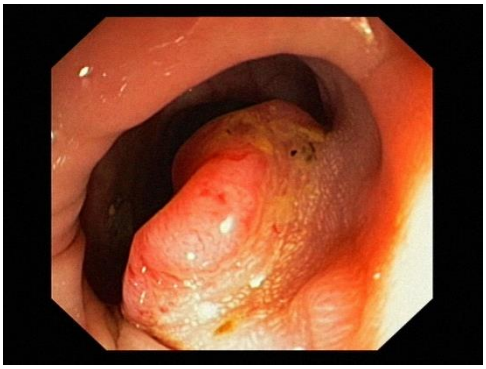

Figure 1: Adenocarcinoma inside colon. From "The colonoscopic view of adenocarcinoma of the colon" by Minhhuyen, N., 2021, *MSD Manual*. (https://www.msdmanuals.com/professional/gastrointestinal-disorders/tumors-of-the-gastrointestinal-tract/colorectal-cancer)

2.1.1 Staging of Colorectal Cancer

There are two primary staging systems used to describe the size and progression of cancer: TNM staging and number staging. The TNM staging method uses the term "tumour" to refer to the size and invasion of cancer; "N" refers to the number of invaded lymph nodes; and "M" refers to the spread of disease to surrounding tissues. For numbering system, it is categorized into five stages, which is stage 0, I, II, III and IV. Stage 0 cancers are in situ, meaning they have not spread to adjacent tissues, whereas stage I or II cancers have grown but have not invaded adjacent tissues. In general, the higher the staging system's number, the larger the cancer tumours and invasion area. Stage III cancer, sometimes called locally advanced cancer, has begun to spread to other tissues and may potentially infect adjacent lymph nodes. Finally, stage IV tumours have spread to other organs, a process known as metastasis.

For CRC, a specific stage method known as the Duke staging system is used to determine the depth of invasion. Duke's staging system is divided into four stages, denoted by the letters A, B, C, and D. According to Jass and Morson (1987), Dukes A are tumours that have spread beyond the mucosa but have not entered the intestinal wall; Duke B are tumours that have spread beyond the wall but have remained submucosal and have not invaded lymph nodes. Additionally, Duke C tumours involve lymph nodes but do not exceed the intestinal wall, whereas Duke D tumours exceed the intestinal wall and have widespread metastases. However, the Duke staging system has been mainly replaced by the TNM staging system.

2.1.2 Colorectal Cancer Symptoms

Typically, there are no symptoms in the early stages of CRC until the tumour has grown or spread. As a result, it is preferable to have routine screening to detect precancerous polyps and remove them before they develop into cancer. CRC is most frequently manifested by changes in bowel habits such as constipation or diarrhoea, abdominal pain, and blood in the stool

(Majumdar, Fletcher, & Evans, 1999). However, these symptoms may be perplexing because gastrointestinal symptoms do not always signal the presence of CRC; therefore, it is preferable to undergo a colonoscopy to detect cancer or precancerous polyps.

2.1.3 HCT116 Cell Lines

The HCT116 cell line is a human colorectal cell line derived from an adult male patient and it usually used as a common model in CRC cell line. Due to its strong growth rate properties and around 18-hour-old doubling time, confluency can reach 80-90 % after a day of sub-culturing (Yeung, Gandhi, Wilding, Muschel, & Bodmer, 2010). The HCT116 cell line has a mutation in the ras proto-13 oncogene's codon, and it possesses epithelial shape and adherent growth capabilities. It is derived from a metastasis cells found in ascending colon (Sim *et al*., 2018a). Figure 2 demonstrated that HCT116 cell line was grown in 89% McCoy's 5A medium with 1% of Penicillin-Streptomycin and 10% of fetal bovine serum (FBS).

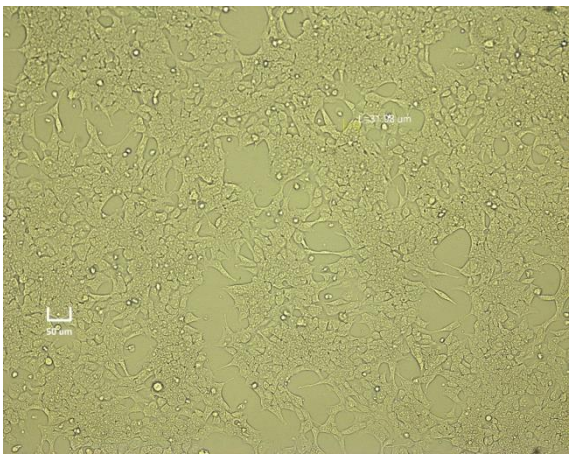

Figure 2: Diagram of HCT116 cell lines under microscope.

2.1.4 HT29 Cell Lines

HT 29 cell line is also a human colorectal cell line but it was isolated from a 44 years old female patient in 1964. It has the same morphology and growth properties as HCT116, the cells are adherent with epithelial morphology. Because it can express the features of mature gut cells, it is widely used in tumour and toxicity studies (Martnez-Maqueda, Miralles, & Recio, 2015). In serum medium, the doubling time of HT29 is 24 hours. Figure 3 showed that HT29 was sub-cultured in 89% of McCoy's 5A medium with 1% of Penicillin-Streptomycin and 10% of fetal bovine serum (FBS).

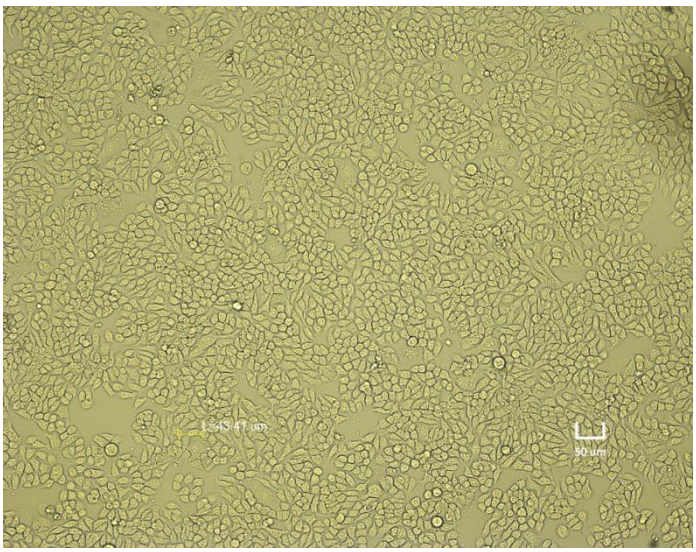

Figure 3: Diagram of HT29 cell lines under microscope.

2.2 *Tumour Susceptibility 101 (TSG101) Gene*

TSG101 is a gene that encodes a multidomain protein that acts as an inactive ubiquitin-conjugating enzyme homolog. It is classified as a critical component of the endosomal sorting complexes required for transportation (ESCRT). ESCRT plays a significant role in biological processes such as retroviral budding, multivesicular bodies (MVB) biogenesis and cytokinesis. The ESCRT machinery is subdivided into three subcomplexes: ESCRT-I, ESCRT-II, and

ESCRT-III, with *TSG101* serving as a component of ESCRT-I. *TSG101*, in general, is essential for the sorting of complexes required for ESCRT-I transport, which is required for membrane receptor degradation and cytokinesis (Luyet et al., 2008; Ferraiuolo, 2020). *TSG101* was initially found in Li and Cohen's (1996) immortalised mouse embryonic fibroblasts model, and their study indicated that depletion or overexpression of *TSG101* could result in metastasis in murine fibroblasts. *TSG101* is comprised of four known structural motifs which are the N-terminal UEV domain, proline-rich region (PRR), a coiled coil (CC) region and a C-terminal steadiness box (SB) domain. TSG101's N-terminal UEV domain is an ubiquitin-conjugating enzyme E2 variant domain; its sequences are homologous to those of the E2 family of ubiquitin conjugation proteins, but it is catalytically inactive due to the absence of an active site (Luyet et al., 2008; Ferraiuolo, 2020). However, *TSG101* retains its ability to bind to ubiquitin, which enables it to function as a mediator of intracellular mobility (Luyet et al., 2008).

2.2.1 Role of *TSG101* in Colorectal Cancer

Even though the role of *TSG101* in cancer remains a contention, multiple studies have identified *TSG101*'s participation in a variety of malignancies. In general, oncogene activation and tumour suppressor gene inactivation are strongly related to the development of malignant tumours, particularly with cancer invasion. Nonetheless, the topic of whether *TSG101* lack or increased expression will result in neoplastic development remains controversial. According to the hypothesis given by Ma et al. (2008), *TSG101* may act as a positive enhancer in carcinogenesis and CRC progression, rather than as a tumour suppressor. This data corroborates the findings of Gheytanchi et al. (2021) who showed that *TSG101* expression is increased in CRC tumour tissues relative to normal tissues. Several investigations, however, indicated that *TSG101* loss is strongly related to p53/MDM2 feedback control, which results in carcinogenesis. MDM2 is a negative regulator of the p53 tumour suppressor gene. *TSG101* plays a vital function in p53/MDM2 feedback control by inhibiting MDM2 degradation when

its UBC domain interferes with MDM2 ubiquitination (Li et al., 2001). MDM2 overexpression results in p53 downregulation, which results in tumorigenesis.

2.2.2 UEV Domain

The UEV domain is one of the *TSG101* protein domains, consisting of around 145 amino acids with an alpha/beta fold that is similar to canonical E2 ligases. However, because the active site cysteine is missing, it would be unable to synthesize ubiquitin enzymes. The *TSG101* UEV domain's 3D structure is illustrated in the diagram below. The UEV domain adopts the structural characteristics of a twisted antiparallel beta-sheet and four alpha-helices, with an additional N-terminal helix (Sundquist, Schubert, Hill, Holton & Hill, 2004).

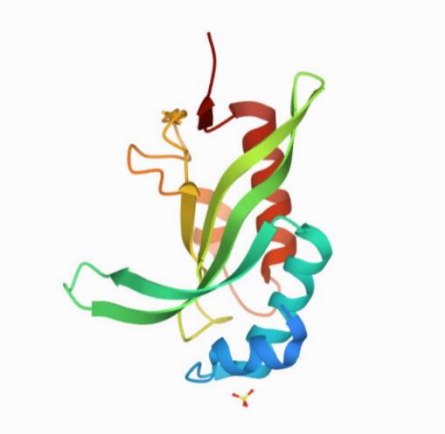

Figure 4. The 3D structure of UEV domain in *TSG101*. (**Red**= UEV domain of *TSG101* bound to ubiquitin, **Green**= Twisted antiparallel beta-sheet, **Blue**= Alpha-helices) From "Domain Binding and Function", Sundquist et al. (2004), *Molecular Cell*.
(https://www.cellsignal.com/learn-and-support/protein-domains-and-interactions/uev-protein-domain#:~:text=The%20Ubiquitin%20E2,the%20two%20C%2Dterminal%20helices.)

2.2.3 Housekeeping gene

The housekeeping gene is a cellular maintenance gene that is constantly expressed in all cells for regulating basic and extensive cellular functions (Turabelidze et al.,2010). The example of housekeeping genes includes Glyceraldehyde-3-Phosphate Dehydrogenase (GADPH), beta-

actin and ubiquitin. Housekeeping genes usually serve as an internal positive control during PCR to improve the reliability of the PCR experiment. Since the housekeeping genes are expected to maintain a constant expression throughout different cells, it is always used for data normalization in gene expression. In our study, *GADPH* served as our internal positive control to ensure that our PCR experiment was successfully performed.

2.3 Polymerase Chain Reaction

Polymerase chain reaction (PCR) is a laboratory technique that is widely used nowadays due to its sensitivity, speed and specificity. It was first invented by biochemist, Kary Mullis in 1998 to amplify a specific target DNA sequence by enzymatic method and cycling condition in a short time. Before being subjected to PCR, the starting materials could be RNA or genomic DNA. In this study we were using RNA as our starting material. Therefore, RNA extraction was carried out to isolate the RNA from CRC cell lines by using GENEzol approach, which is one of the examples of Gianidium-acid-phenol extraction to isolate RNA from CRC cell lines because it can ensure high yielding of good quality RNA and less costly.

PCR has been used in a variety of applications including sequencing, cloning and phenotyping. PCR amplification can be achieved by 3 simple steps which are denaturation, annealing and extension. During denaturation, a double-stranded DNA template into single strands by using heat, followed by annealing of primers in which the designed primer will bind to the flaking region of target DNA for new strand synthesis. Lastly, DNA polymerase will bind to the annealed primer and extend at 3’ end along the template. There are different types of PCR, including reverse transcriptase PCR (RT-PCR), Hot Start PCR, Quantitative PCR (qPCR) and Nested PCR. In this study, RT-PCR were employed to detect the presence of TSG 101 in CRC cell lines.

Reverse transcriptase PCR is variety of PCR that widely used in medical science and biomedical field to study the targeted gene expression. In this study, RT-PCR was used to detect the presence and expression of *TSG101* in CRC cell lines. In general, RT-PCR uses RNA template as a starting material to catalyse DNA synthesis and transcribe it into complementary DNA (cDNA). RT-PCR can be performed either in one step or two-step process. In one step approach, the reverse transcription reaction and PCR will be performed at the same time by mixing the reverse transcriptase and DNA polymerase into same tube. In contrast, two steps process will perform reverse transcription first, followed by PCR. RT-PCR is a very simple and cost-effective technique for quantifying mRNA in small samples. Besides, it also has been used for molecular detection of circulating tumour cells (Skrondra et al., 2014). After that, PCR products of the normal *TSG101* can be analysed by using agarose gel electrophoresis (AGE) to determine the specific nucleotides that we desired.

2.4 Sequence Analysis

Bioinformatics is generally known as an application that is using computer programming or software tools to analyse and interpret biological data. It is a combination of biology, computer science and information technology (Kumar & Chordia, 2017). In this era of technology, bioinformatics is a significant player in nearly all aspects of biology fields including biotechnology, biomedicine and cancer genetics. With the exponentially growing of biological data, bioinformatics is crucial to store, manage and interpret these data, especially for the sequencing data. In bioinformatics, sequence analysis is a definitive technique that involves computational analysis of DNA, RNA or peptide sequence to study its features, functions and structures. The establishment of sequence analysing has contributed to many individuals with different scientific backgrounds.

Before subjecting the amplified gene to bioinformatics analysis, the gene has to undergo DNA sequencing to determine nucleotide sequence. DNA sequencing is a very common laboratory technique that is used to determine the specific nucleotide sequence of the desired gene. The nucleotide sequence of a gene carries every information of the cell that might be associated with certain diseases or genetic variations; hence, it is crucial for researchers to determine the nucleotide sequence of their desired gene in order to investigate the role of the gene.

2.4.1 CD Search

CD search (Marchler-Bauer and Bryant, 2004), is a free web-based tool that is provided by NCBI for the detection of structural and functional domains in query sequence. CD search employed the BLAST statistics to identify the conserved domain(s) present in a sequence by matching the sequence with domain and domain family models. It provides a concise, standard and complete result. The standard result demonstrated in CD search provides information on specific domain hits, non-specific domain hits, superfamily and each details on domain hit and superfamily.

2.4.2 MEME Suite

Multiple Em for Motif Elicitation Suite, also known as MEME Suite (Bailey et al., 2009), is a free software toolkit that allows us to identify the novel motifs in unaligned nucleotides or sequences. Motifs are the short conserved sequences that represent a small region within the three-dimensional structure of protein. The output of MEME provides information on the discovered motifs and its location.

2.4.3 BLAST

BLAST (Altschul et al., 1990) is a free tool hosted by NCBI that enables researchers to search and aligned the nucleotide sequences by segment and enable to detect the identical sequences

in sequence analysis. In addition, it also can calculate the statistical significance of matches. The output of BLAST provides information on percent identity, expected value (E-value), query cover, accession number, total score and maximum score of target sequence.

CHAPTER 3

MATERIALS AND METHODS

3.1 Cell lines and cultured cells

In this study, human colorectal carcinoma (CRC) cell line, HCT116 and HT29 were used to examine the presence of *TSG101* and characterize the genes. These HCT116 and HT29 cell lines were provided by UNIMAS postgraduate students in cell form and cultured in the T25 flasks.

3.2 Extraction of RNA

RNA extraction, also known as RNA isolation is a purification process of RNA from biological samples such as cells or tissues. The principle of RNA extraction is to separate the RNA with DNA and protein. RNA isolation is more difficult compared to DNA isolation due to the abundantly present RNases in the environment, which can degrade RNA. Therefore, good aseptic techniques need to be strictly applied during the extraction.

RNA of HCT116 and HT29 cell lines were isolated by using the GENEzol method (Chomczynski, 1993). First, the used media inside the T25 flask was removed by using a pipette, and then 1mL of PBS was added to wash the cells. After that, 1mL of GENEzol reagent was added to each flask and the cells were lysed using a cell scraper. The cell pellets were re-suspended again by using the mechanical method, which is pipetting the cells in GENzol and vortex. The cells in GENEzol reagents were transferred into new tubes, followed by 5 minutes of incubation on ice. Two hundred microliters of chloroform was added into the tube later to promote phase separation. Next, the mixture was centrifuged using a micro-centrifuge at 12,000 rpm for 15 minutes at 4 °C.