FUNCTIONAL INTERROGATION OF CRISPR-CAS9 MEDIATED EpCAM KNOCKOUT IN BREAST CANCER CELLS

AFIQAH FASIHAH BINTI ABDUL RAHIM

FACULTY OF SCIENCE

UNIVERSITI MALAYA

KUALA LUMPUR

2022

FUNCTIONAL INTERROGATION OF CRISPR- Cas9 MEDIATED EpCAM KNOCKOUT IN BREAST CANCER CELLS

AFIQAH FASIHAH BINTI ABDUL RAHIM

DISSERTATION SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF SCIENCE (BIOTECHNOLOGY)

INSTITUTE OF BIOLOGICAL SCIENCES
FACULTY OF SCIENCE
UNIVERSITI MALAYA
KUALA LUMPUR

2022

UNIVERSITI MALAYA

ORIGINAL LITERARY WORK DECLARATION

Name of Candidate: **AFIQAH FASIHAH BINTI ABDUL RAHIM**

Registration/Matric No: **S2020377**

Name of Degree: **MASTER OF SCIENCE (BIOTECHNOLOGY)**

Title of Dissertation ("this Work"):
FUNCTIONAL INTERROGATION OF CRISPR-CAS9 MEDIATED EPCAM KNOCKOUT IN BREAST CANCER CELLS

Field of Study: **BIOTECHNOLOGY**

I do solemnly and sincerely declare that:

(1) I am the sole author/writer of this Work;
(2) This Work is original;
(3) Any use of any work in which copyright exists was done by way of fair dealing and for permitted purposes and any excerpt or extract from, or reference to or reproduction of any copyright work has been disclosed expressly and sufficiently and the title of the Work and its authorship have been acknowledged in this Work;
(4) I do not have any actual knowledge nor do I ought reasonably to know that the making of this work constitutes an infringement of any copyright work;
(5) I hereby assign all and every rights in the copyright to this Work to the University of Malaya ("UM"), who henceforth shall be owner of the copyright in this Work and that any reproduction or use in any form or by any means whatsoever is prohibited without the written consent of UM having been first had and obtained;
(6) I am fully aware that if in the course of making this Work I have infringed any copyright whether intentionally or otherwise, I may be subject to legal action or any other action as may be determined by UM.

Candidate's Signature Date: 6 FEBRUARY 2022

Subscribed and solemnly declared before,

Witness's Signature Date: 6 FEBRUARY 2022

Name:
Designation:

FUNCTIONAL INTERROGATION OF CRISPR- Cas9 MEDIATED EpCAM KNOCKOUT IN BREAST CANCER CELLS

ABSTRACT

Breast cancer is the most common cancer, and it ranks among the top three in terms of new cases. Many factors contribute to the development of breast cancer, including overexpression of particular genes that promote tumourigenesis. EpCAM which encodes for a pro-metastasis protein, has been consistently associated with the developments of many epithelial cancers including breast cancer. This research aimed to determine the role of EpCAM in MCF-7 and T-47D breast cancer cell lines using CRISPR-Cas9 genome editing tool. Also, this study further investigated the phenotypic effects of knockout EpCAM using functional assays that measured cell migration, proliferation, adhesion and colony formation. The CRISPR-Cas9 genome editing approach was found to successfully perturbed EpCAM functions in both breast cancer cell lines. Similarly, observations on the functional assays showed that knocking out EpCAM reduced cell migration, inhibited cell proliferation while reducing the abilities of the cells to form colonies and cell adhesion. These findings indicate that knocking out EpCAM has a tumour suppressor effect, which suggests EpCAM as a potential target in breast cancer therapeutics.

Keywords: breast cancer, EpCAM gene, CRISPR-Cas9 genome editing technology, knockout, functional assays.

INTEROGASI BERFUNGSI CRISPR-Cas9 SEBAGAI PERANTARAAN NYAHSESAR EpCAM DALAM SEL KANSER PAYUDARA

ABSTRAK

Kanser payudara merupakan kanser yang paling biasa berlaku dan ia berada di antara tiga teratas dari segi kes baharu. Banyak faktor menyumbang kepada perkembangan kanser payudara, termasuk ekspresi berlebihan gen tertentu yang menggalakkan tumourigenesis. EpCAM merupakan salah satu gen yang lazim terlibat dalam kanser payudara dan ia juga dipercayai bertindak sebagai pro-metastasis yang banyak terlibat dalam kanser epitelium termasuk kanser payudara. Kajian ini dijalankan bertujuan untuk mengenalpasti fungsi EpCAM dalam sel kanser payudara; MCF-7 dan T-47D menggunakan teknologi penyuntingan genom CRISPR-Cas9 untuk menyasar dan nyahsesar EpCAM. Selain itu, kajian ini juga mengenalpasti kesan nyahsesar EpCAM ke atas fenotip sel menggunakan asai fungsian untuk mengukur kecergasan sel kanser payudara berdasarkan migrasi, proliferasi, lekatan dan pembetukan koloni. Hasil kajian mendapati bahawa pendekatan penyuntingan genom CRISPR-Cas9 berjaya mengganggu fungsi EpCAM dalam kedua-dua sel kanser payudara. Begitu juga, hasil kajian daripada asai fungsian menunjukkan keupayaan sel menurun dari segi migrasi, lekatan, proliferasi dan pembentukan koloni. Penemuan ini menunjukkan bahawa nyahsesar EpCAM mempunyai kesan penindas tumour dan EpCAM dicadangkan sebagai sasaran berpotensi untuk rawatan kanser payudara.

Kata kunci: kanser payudara, gen EpCAM, penyuntingan genom CRISPR-Cas9, nyahsesar, asai fungsian

ACKNOWLEDGEMENTS

Working on this dissertation has been a very enlightening, rewarding and gratifying experience for me. All praise to Allah S.W.T for the strength and courage that He has given me along the duration of conducting and completing this academic project.

First and foremost, I would like to extend my utmost appreciation to my supervisors Dr Nikman Adli and Dr Mohamad Aimanuddin, for their guidance and valuable suggestions. Their invaluable advice and constructive suggestions had helped greatly in the completion of the study. I sincerely appreciate the time they have taken patiently review my entire work. My sincere appreciation to Adib, Nadia, Hanani, Asmaa, Mei Juan, Su Min, Nisa, Farhan and others for all the help throughout the study. Also, thank you very much to Dr Saiful Effendi, all UMBI staff and people around me who have either directly or indirectly helped me with my research or study/work-related matters. I would also like to extend my sincere gratitude to Dato' Dr Norlizan and Dr Saifullizam for allowing me to finish my master study.

Also, thank you very much to my previous primary and secondary friends, my lecturers at Universiti Kebangsaan Malaysia and Universiti Malaya and all of my family members for making this journey possible and successful. Not forgetting, special thanks to my beloved parents Abdul Rahim and Surinan for their endless support, enormous sacrifice, tolerance and understanding. Their belief in me has never faltered and that at times was my only source of strength. Thank you to my siblings Ahmad Khairie, Amirah Najihah, Athirah Imanina and Aina Madihah. Without the support, patience and guidance of the following people, this study would have not been completed. This dissertation is dedicated to everyone who has made it possible. It is to them that I owe my deepest gratitude.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS AND ABBREVIATIONS

Ab	:	Antibody
Bp	:	Base pair
Cas9	:	CRISPR-associated protein 9
CK18	:	Cytokeratin 18
COAD	:	Colon adenocarcinoma
CRISPR	:	Clustered regularly interspaced short palindromic repeats
crRNA	:	Crispr RNA
CT	:	Computer tomography
DCIS	:	Ductal carcinoma in situ
DMEM	:	Dulbecco's modified eagle's media
DMSO	:	Dimethyl sulfoxide
DNA	:	Deoxyribonucleic acid
dNTP	:	Deoxyribonucleic triphosphate
DSB	:	Double stranded break
ECM	:	Extracellular matrix
ECL	:	Enhanced chemiluminescence reagent
EMT	:	Epithelial to mesenchymal transition
EpCAM	:	Epithelial cell adhesion molecule
EpEX	:	EpCAM cleaved extra-celullar domain
EpICD	:	EpCAM intra-celullar domain
ER	:	Estrogen receptor
FACS	:	Fluorescence activated cell sorting
FBS	:	Foetal bovine serum
FHL2	:	Four and a half LIM domains protein 2

GFP	:	Green fluorescent protein
HER2	:	Human epidermal growth factor receptor-2
IDC	:	Invasive ductal carcinoma
ILC	:	Invasive lobular carcinoma
kDa	:	Kilo Dalton
KO	:	Knock-out
Lef1	:	Lymphoid Enhancer Binding Factor 1
LUSC	:	Lung squamous cell carcinoma
mA	:	Milliampere
ml	:	Millilitre
mM	:	Millimolar
mRNA	:	Messenger RNA
ng/µl	:	Nanogram/Microlitre
NHEJ	:	Non homologous end joining
NTC	:	Non-targeting control
OD	:	Optical density
PAM	:	Protospacer adjacent motif
PBS	:	Phosphate- buffered saline
PCR	:	Polymerase chain reaction
pfu	:	Plaque-forming unit
pmol/µl	:	Picomole/microlitre
PR	:	Progesterone receptor
RIP	:	Regulated intramembrane proteolysis
RNA	:	Ribonucleic acid
SDS-PAGE	:	Sodium dodecyl sulphate-polyacrylamide gel electrophoresis
sgRNA	:	Single guide RNA

SMA	:	Smooth muscle actin
TACE/ADAM17	:	Tumour necrosis factor-α-converting enzyme/ disintegrin and metalloproteinase
TBS	:	Tris-buffered saline
TBST	:	Tris-buffered saline- Tween 20
TCGA	:	The Cancer Genome Atlas
TNBC	:	Triple negative breast cancer
tracrRNA	:	Trans-activating crispr RNA
WHO	:	World Health Organization
WT	:	Wild type
v/v	:	Volume/volume
w/v	:	Weight/volume
µg	:	Microgram
µl	:	Microlitre
µM	:	Micrometre

CHAPTER 1: INTRODUCTION

1.1 Background of Study

Cancer is one of the major causes of public health problems and fatality around the world. It affects people of all ages with approximately 14 million new cases in 2012 (Torre et al., 2015). The number of cases is expected to rise by about 70% over the next two decades and at some time in their lives, one out of every three people will be diagnosed with cancer. Cancer is a genetic disease caused by mutations in genes that control many of the cell's key processes. These functions are closely controlled in normal cells, by the genes that are expressed at the right time and place. In contrast, cancer cells do not stop growing and dividing; and this uncontrolled cell growth eventually results in the formation of tumours. Cancer has been around since 1500 BC (Lisa Fayed, 2020). Recently, the most common cancer worldwide is breast cancer (Breast Cancer Research Foundation, 2021a). A study by Yip et al. (2014), found that in Malaysia, the presentation of breast cancer varies significantly depending on the stage at presentation, tumour size and age. Scientists and researchers from all over the world are working on cancer research to unravel the mystery of this disease and discover the remedies. Curing this disease will be complex because numerous aspects must be considered throughout the decision-making process (Reyna et al., 2015).

Metastasis is a cellular mechanism whereby a group of cancer cells acquires the ability to migrate from the original tumour location to another non adjacent organ. During metastasis, the cancer cells travelling either in the bloodstream or lymphatics (Stitzenberg & Ridge, 2009). The process of cancer metastasis comprises a long series of several co-ordinated cascades. The fundamental steps in the metastatic process include invasion, intravasation and extravasation. Invasion of tumour cells into the surrounding host tissue

is the first step in metastasis. The tumour cells subsequently go through an intravasation process where they attempt to cross the wall of blood capillary. The tumour cells then travel in the blood capillary, exhibiting extravasation and spreading to new sites, eventually resulting in the formation of a new tumour (Lu et al., 2009). One of the important mechanisms in the metastatic process is epithelial-to-mesenchymal transition (EMT). According to Mohtar et al. (2020), for this process to occur, the invasive cancer cells first should modify the cell-to-cell interaction and cell attachment to the extracellular matrix (ECM).

Metastasis is an orchestrated show of different processes such as epithelial-to-mesenchymal transition (EMT) and mesenchymal-to-epithelial transition (MET). Many tumours demonstrate increased regulated intramembrane proteolysis (RIP) of EpCAM and decrease EpCAM expression under EMT circumstances (Martowicz et al., 2016). EMT is a crucial step in the metastatic process in which epithelial cancer cells acquire mesenchymal characteristics and lose their epithelial morphologies (Bullock et al., 2012). It assists invasion and intravasation of tumour cells into the bloodstream (Scully et al., 2012). In breast cancer, metastasis occurs at stage 3. However, 6-10% of patients already have *de novo* metastatic breast cancer which is a stage 4 breast cancer at the time of first diagnosis (Breast Cancer Research Foundation, 2021b). Thus, early detection of breast cancer is vital. However, most breast cancer patients realise their conditions at a late stage due to poor prognosis. Poor prognosis in breast cancer is attributed to its early metastatic behaviour (Abdullah et al., 2013).

EpCAM is overexpressed in varieties of human cancer of epithelial origin such as breast cancer (Mohtar et al., 2020). Therefore, it is important to study EpCAM as it is oncogenic; where it promotes tumourigenesis. In addition, overexpression of EpCAM is

also associated with several oncogenic signalling pathways including Wingless- related integration site/ Beta catenin (Wnt/β-catenin) pathways (Sankpal et al., 2021). Breast cancer is the first cancer discovered to be associated with WNT signalling. WNT signalling pathway; a well-established signalling network and EpCAM are important in embryogenesis and neonatal development (Yamashita et al., 2007; Xu et al., 2020).

Previous study by Zhou et al. (2015) showed that EpCAM is a part of the WNT pathway downstream reaction. When the EpCAM gene and protein expression were silenced in colon cancer cell lines, the expression level of β-catenin reduced as well thus preventing tumour formation. According to another study in Japan, there appeared to be a functional relationship between EpCAM expression and WNT/β-catenin activation (Yamashita et al., 2007). This is similar to a research by Mohtar et al. (2020) which stated that WNT/ β-catenin had a positive cascade effect in supporting tumourigenesis with EpCAM through cytoplasmic EpICD.

However, the role and the mechanisms underlying-EpCAM's action in the aetiology of breast cancer are relatively unclear, while the evidence to support previous studies remain scarce. Thus, this research aims to investigate the pro-oncogenic effects of EpCAM in breast cancer by using loss-of function or knockout EpCAM.

1.2 Problem Statement

Due to the overexpression of EpCAM in cancer and accessibility on the cell surface, EpCAM is a clinically-relevant biomarker for prognosis, therapeutic intervention and diagnosis for epithelial cancers including breast cancer. Previous studies have elucidated the role of EpCAM in cancers to some extents, but these studies mainly utilized the overexpression (gain-of-function) strategy while only a handful of studies utilized the

loss-of-function approaches. Here, the aim was to use the emerging CRISPR-Cas9 genome editing technology to target and/or completely knockout EpCAM in breast cancer cells. The resulting EpCAM isogenic cell models were further assessed to determine whether targeting EpCAM can reduce breast cancer cell fitness and/or intersect with previous molecular mechanisms.

1.3 Research Questions

1. Can EpCAM be targeted using CRISPR-Cas9 genome editing tools?
2. What are the phenotypic effects of targeting EpCAM in breast cancer cells? (i.e Will it reduce breast cancer cell fitness?)

1.4 Research Objectives

Main objective: To establish EpCAM-knockout breast cancer cells using CRISPR-Cas9 genome editing technology and understand the role of EpCAM in promoting breast cancer progression.

Specific objectives:

1. To target EpCAM in MCF-7 and T-47D breast cancer cells using CRISPR-Cas9 genome editing technology.
2. To assess phenotypic effects of targeting EpCAM in breast cancer cells using cell-based assays.

CHAPTER 2: LITERATURE REVIEW

2.1 Cancer

2.1.1 Overview of cancer

Cancer is a biological condition characterised by unrestrained and uncontrolled cell development in humans. It is generally a cell division disease in which failure of cell division control occurs. It has been noted in many studies that cancer is caused by mutations in genes that govern many basic aspects of cellular function. Many proteins are silenced or overexpressed in cells, resulting in changes in the overall activity of the cell (Bartee et al., 2017). The way genes are expressed is greatly influenced by the environment. Each cancer has a different age of onset, growth rate, invasiveness, prognosis and therapy response. Generally, cancer cells have two basic characteristics which are cell proliferation and metastasis. The former is an abnormal cell growth and division meanwhile the latter is an abnormality in the normal controls that keeps cells from propagating and invading other organs and systems of the body through lymphatic circulation.

In a healthy cell, normal cell growth and division are controlled by at least two types of genes which are proto-oncogenes that promote growth and tumour suppressor genes that inhibit growth. A normal cell becomes a tumour when one or both of these genes are altered. Tumours can be cancerous or noncancerous. Noncancerous tumour, also known as benign tumour is a non-invasive whereas a cancerous or malignant tumour is the one that invades the body and destroys normal body tissues and organs (metastasis). Metastatic cancers have been accounted to be the most predominant cause of deaths among cancer patients (Johnson, 2016).

Cancer is generally classified based on the tissues or organs from which the cells originate, as well as the specific type of cells from which they originate. Carcinoma, sarcoma, melanoma, lymphoma and leukaemia are the five major types of cancer (Pichardo, 2020). The most prevalent type of cancer is carcinoma. Carcinomas are malignancies that develop in the skin, lungs, breast, pancreas and other organs. Lymphomas are cancers of lymphocytes, leukaemia is a blood cancer, melanomas are cancers in cells that make the pigment in skin and sarcomas are malignant tumours that form in the muscle, bone, fat, blood vessel, cartilage or connective tissue of the body.

According to the World Health Organization (WHO), the top three most common types of cancers in terms of new cases in 2020 were breast (2.26 million cases) followed by lung (2.21 million cases) and colon and rectum (1.93 million cases) (World Health Organization, 2022a). Breast cancer represents one of the major malignancies and a leading cause of cancer-related death worldwide. According to WHO, breast cancer was the world's most prevalent cancer in 2020 with 2.3 million diagnoses. For cancer diagnosis, usually a histological examination through tissue biopsy or surgery specimen is done on the patient in order to determine the tumour's histological grade. Throughout the last decade, liquid biopsy which uses peripheral blood as a source of information on tumour status and treatment options has gained increasing interest in oncology (Keller et al., 2019). This examination is usually performed after the basic medical examinations such as X-rays, endoscopy, computer tomography (CT) scans and blood tests.

Good prognosis and various therapeutic options play important roles in the treatment of cancer. However, the prognosis for advanced-stage cancer remains poor particularly in developing countries (Wang et al., 2015). Nowadays, specific molecular markers have been developed and widely used to screen early stage of breast cancer and predict whether

it will progress or recur (Wu et al., 2009). In addition, researchers are also working on personalized medicine for cancer patients in order to predict the risk of the disease however for customised medicine to be successfully implemented, it requires stakeholder cooperation and technology to assay molecular analytes (Verma, 2012). Nonetheless, the incorporation of CRISPR-Cas9 system after nucleic acid amplification was found to improve the specificity of assays and minimized the cross-contamination during the assay (Feng et al., 2021). A review study by Katti et al. (2022) stated that nearly every element of cancer biology has been impacted by CRISPR, which has also sparked the production of enormous volumes functional data and unlocked a wealth of previously untapped knowledge about the disease. Additionally, this CRISPR system also offers a wide range of potential applications for determining gene function, analysing tumour heterogeneity as well as overcoming medication drug resistance in cancer (Chen & Zhang., 2018).

2.1.2 Hallmarks of cancer

'Hallmarks of cancer' is the observation on the metabolism of cancer cells that act in an atypical manner. Hanahan and Weinberg classified the complexity of cancer into six key properties which are (1) self-sufficiency in growth signals, (2) insensitive to anti-growth signals, (3) tissue invasion and metastasis, (4) limitless replicative potential, (5) sustained angiogenesis and (6) avoiding apoptosis (Hanahan & Weinberg, 2000). Additionally in 2011, a more recent version of hallmarks of cancer with the total of four hallmarks making ten hallmarks in total; (7) reprogramming energy metabolism, (8) circumventing the immune destruction (9) genome instability and (10) tumour promoting inflammation (Hanahan & Weinberg, 2011).

2.1.3 Breast

Breast is the tissue that covers the pectoral chest muscles. The breast is composed of skin, subcutaneous tissue and breast tissue. The base of female breast extends from the second to the sixth rib on the anterior thoracic wall. Breasts are made up of specialized tissue that produces milk as well as fatty tissue. The region of the breast that produces milk is divided into 15 to 20 lobes. Milk is generated in tiny structures called lobules within each lobe. The milk is transported by a network of small tubes known as ducts. The ducts join together to form bigger ducts that subsequently exit the skin in the nipple. The black patch of skin that surrounds the nipple is called the areola. Blood vessels, lymph nodes and lymph vessels can all be found in the breast (DerSarkissian, 2019).

According to Pandya & Moore (2011), breast is divided into four quadrants with the upper outer quadrant containing the majority of the breast volume and the most prevalent location of tumours of the breast. Glandular tissue is the most common tissue in the upper outer region of the breast and this area accounts for half of all breast malignancies. The main route of breast cancer metastatic spread is lymphatic drainage of the epithelial and mesenchymal parts of the breast. Majority of breast cancers begin in the terminal duct lobular units (American Joint Committee on Cancer, 2002).

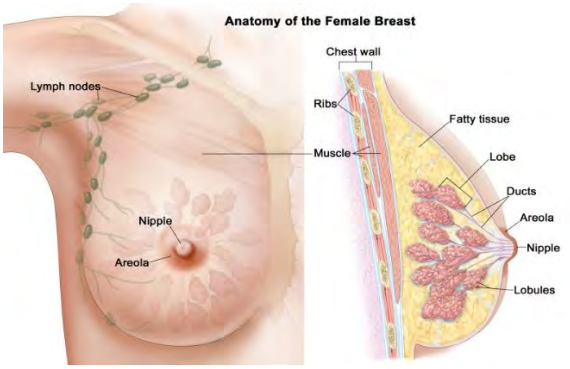


Figure 2.1: Breast anatomy – Breasts also known as mammary glands are made up of lobules, milk-producing glandular structures and a system of ducts that transports milk to the nipple. (Photo sourced from www.nationalbreastcancer.org/ breast-anatomy.)

2.1.4 Breast cancer overview

As aforementioned, breast cancer begins in the breast tissues and spreads primarily through the circulatory vessels. The lymphatic system is made up of veins and organs which serves two important functions in human lives. Excess fluids as well as any plasma proteins, are carried back into the circulatory system by the vessels, which effectively maintain interstitial fluid levels. The majority of the lymph vessels in the breast drain into auxiliary lymph nodes beneath the arm. Supraclavicular and infraclavicular lymph nodes are located around the collar bone, whereas internal mammary lymph nodes are found inside the chest around the breast bone (Rahman & Mohammed, 2015).

Breast tumours can be classified into two forms which are malignant that is cancerous and benign which is not cancerous. Benign breast tumour is an abnormal growth but it does not spread outside of the breast and is not life-threatening. Breast cancer is a malignant tumour that begins in the breast cells. Findings revealed that majority death-related breast cancer incidence are caused by distant organ metastasis instead of original tumour mass (Kamal et al., 2017). Lump in the breast or underarm is one of the symptoms of breast cancer. Cysts and fibrosis are the most common causes of lumps. Fibrosis is the production of scar-like tissue, whereas cysts are fluid-filled sacs. Breast swelling and soreness might result from these changes. Breast cancer is further divided into two categories based on where it starts in the breast. The first one is in-situ or ductal carcinoma in situ (DCIS) and the second one is invasive breast cancer, also known as invasive ductal carcinoma (IDC) or invasive lobular carcinoma (ILC) ("Types of Breast Cancer | Different Breast Cancer Types," 2019). IDC is a malignant cancer that has infiltrated the surrounding breast tissue, whereas DCIS is a non-invasive type of breast cancer or benign type of cancer that is regarded as the earliest form of breast cancer.

In Malaysia, the most common type of cancer affecting women is breast cancer (Meneka Kumaran, 2020). According to the Breast Cancer Foundation Malaysia, Chinese women appear to be at highest risk with an incidence of 59.7 per 100,000 followed by Indian women (55.8 per 100,000) and Malay women (33.9 per 100,000). The treatment options are determined on the type of breast cancer, its stage, grade, size, as well as if the cancer cells are hormone-sensitive. Breast cancer surgery is performed on the majority of the patient, and many of them require extra treatment afterward, for instance chemotherapy, hormone therapy or radiation. Surgery is the most comprehensive and effective treatment option for primary tumours (Mohtar, 2017).

The breast cancer incidence in Malaysia is lower than other countries such as Australia or the United States, however our survival rate is substantially lower (Meneka Kumaran, 2020). Dr Saunthari Somasundaram, the President of the National Cancer Society Malaysia (NCSM) said that the majority of Malaysian cancer patients are diagnosed at a later stage (Indramalar, 2019). She said that the size of the tumour and the extent to which it has spread are the most crucial aspects in determining a patient's outcome and the goal is to catch it as soon as possible. The cancer may exacerbate in rural areas due to lack of access to screening facilities. According to Schliemann et al. (2020), Malaysia's cancer incidence is expected to double by 2040.

Even though breast cancer cases worldwide keep increasing, there is still positive advancement in high income countries in which the survival rate 5 years after diagnosis exceeds 80% now. Cancer survival rates between countries are different due to delays in diagnosis and treatment. With regard to Malaysia, there are few studies on breast cancer survival rate and no current population-based survival studies have been reported (Abdullah et al., 2013). Global Breast Cancer Initiatives has been introduced by the WHO

to reduce breast cancer mortality worldwide by 2.5% per year until 2040, thus preventing an estimated 2.5 million deaths (World Health Organization, 2022b).

The survival rate of breast cancer cases is important for improving patient care and guiding physicians and health professionals in their methods in managing breast cancer cases (Movahedi et al., 2012). Survival rate of breast cancer patients depends on the stage of the disease at the time of diagnosis, tumour size, menstrual cycle and histology. There is also correlation between comorbidity and older age with mortality risk (Maskarinec et al., 2011). Cancer survival can be improved by better understanding of cancer biology, effective treatments as well as increasing the socioeconomic status, lifestyle and general health differences between populations. Another vital step in improving cancer survival is by advancement in gene editing technology. Recently, gene editing technology such as Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) has received enormous attention due to its precision in targeting cancer cells. Several studies and clinical trials using CRISPR/Cas9 approach are currently ongoing to assess biological functions of gene expression in cells that lead to cancer (Zhan et al., 2019). CRISPR-Cas9 has been used to knockout the EpCAM gene.

2.1.5 Breast cancer subtypes

Breast cancer subtypes are divided into several categories such as Luminal A, Luminal B, normal-like, HER2-enriched, basal-like/ triple negative breast cancer (TNBCs). The classification is determined by hormone receptors and proliferation rate. MCF-7 and T-47D breast cancer cell lines belong to the Luminal A molecular subtypes. Luminal A is the most prevalent subtype accounting for 28-31% of all breast cancer cases worldwide and 150,000 cases each year in the United States (Ciriello et al., 2013). Another

retrospective study conducted in India found that the prevalence of the luminal A breast cancer subtype was higher than that of other subtypes, with 762 individuals testing positive for luminal A and 157 for luminal B. Additionally, age was also found to be a factor in the high frequency of luminal A breast cancer subtype where 72% was the score for observation among patients aged 70 years and above (Pandit et al., 2020).

Luminal A subtype is a hormone-receptor positive cancer that expresses estrogen-receptor (ER) and/or progesterone-receptor (PR) and HER2 negative. ER-positive luminal A cell line will only develop a tumour in the presence of estrogen (Holliday & Speirs, 2011). In addition, this breast cancer subtype has the best prognosis and a modest growth rate. MCF-7 and T-47D cell lines exhibit pleural effusion metastatic behaviour, which is linked to a poor prognosis, breast cancer- related mortality and survival time is less than a year on average (Lee et al., 2015).

Both cell lines used in this study are epithelial-like cells which have a polygonal form with more regular dimensions in close contact with one another and grow in discrete patches attached to a substrate (Figure 2.2). MCF-7 is a non-invasive adenocarcinoma cell line that is typically considered to have a low metastatic potential (Comşa et al., 2015). T-47D, on the other hand, is a ductal carcinoma and highly progressive. According to Juliette Adjo Aka (2012), proteins involved in cell proliferation and cancer formation are more abundantly expressed in T-47D cells than in MCF-7 cells.

As aforementioned, therapeutic treatment varies depending on the subtypes of breast cancer. For all types of breast cancer, therapy response and metastases are important clinical concerns (Roarty & Echeverria, 2021). MCF-7 cells and T-47D were shown to express epithelial markers including E-cadherin, β-catenin and cytokeratin 18 (CK18) but

not mesenchymal markers for example vimentin and smooth muscle actin (SMA) (Comşa et al., 2015). Thus, it is important to study this breast cancer subtype in order to further understand the tumour metastatic fitness and its biology.

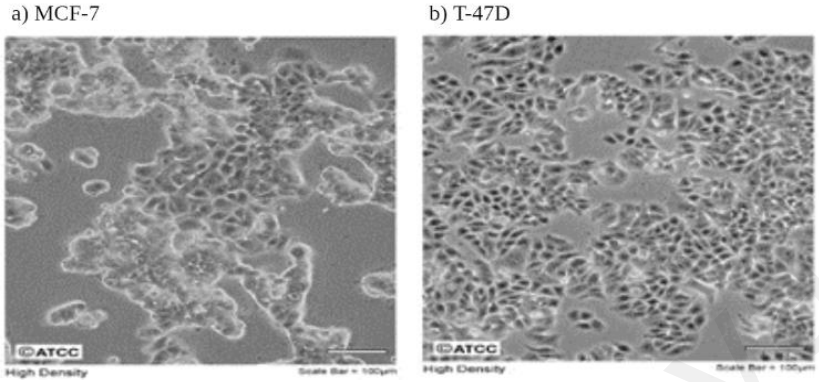


Figure 2.2: (a) Structure of MCF-7 cell lines and (b) Structure of T-47D cell lines. (Photo sourced from ATCC Resources.)

2.2 Epithelial cell adhesion molecule (EpCAM)

Epithelial cell adhesion molecule (EpCAM), also known as CD326, is located on the short arm of chromosome 2 (2p21) that spans about 42kb in the genomic region. EpCAM-201, EpCAM-202, EpCAM-203, EpCAM-204, EpCAM-205 and EpCAM-206 are six transcript variations of this type-1 transmembrane glycoprotein. EpCAM-201 is the most well-studied transcript and commonly used as the reference sequence, as confirmed by the TCGA large scale cancer transcriptomic results hence, EpCAM is the name referred to this isoform.

EpCAM is abundantly expressed at basal levels in normal epithelial cells and is predominantly localized in the basolateral membrane. Breast, colorectal and ovarian malignancies have all been found to have hypomethylation in the EpCAM promoter region. In humans, EpCAM protein can be found in the cells that line the surfaces and cavities of the body. It helps the cells to stick to one another. Human EpCAM protein molecular weight ~40kDa and comprises 314 amino acids residues (Mohtar et al., 2020).

The full-length EpCAM protein has many cleavage sites that are required for its biological activity as well as protein expression regulation (Keller et al., 2019). A large N-terminal extracellular domain (EpEX), a single-spanning transmembrane domain (TM), a short C-terminal cytoplasmic domain and an intracellular domain (EpICD) are the four critical elements of this protein.

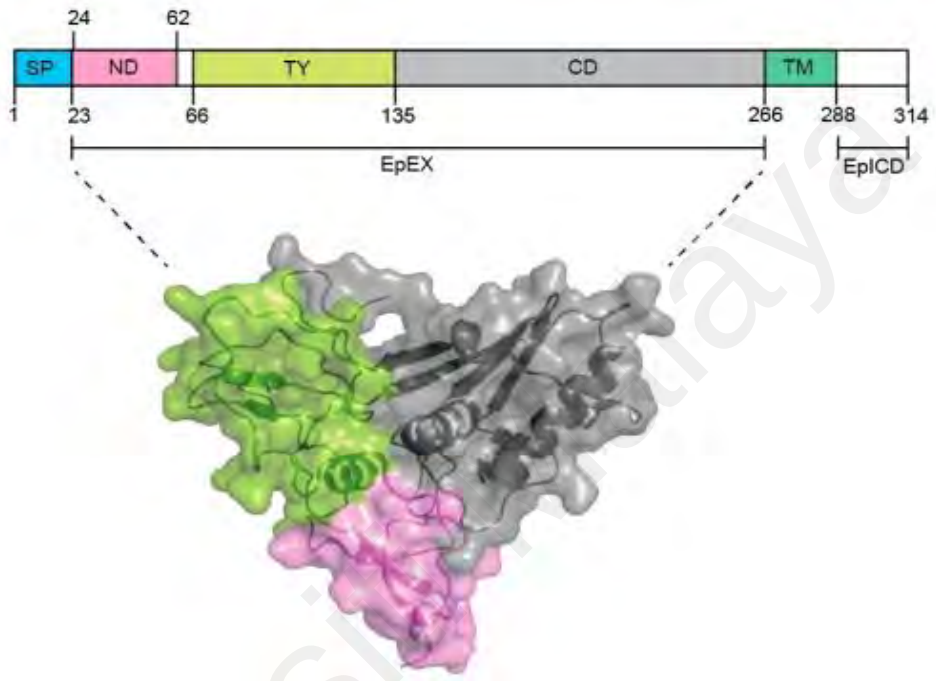


Figure 2.3: Epithelial cell adhesion molecule (EpCAM) protein structure. EpCAM secondary structure consists of signal peptide (SP, blue), thyroglobulin type-1 domain (TY, lime green), N-domain (ND, pink), transmembrane domain (TM, grey), C-domain (CD, grey), and intracellular domain (EpIC, white). The EpCAM cleaved extracellular domain is depicted in three dimensions. Source: Adapted from Mohtar et al., (2020).

2.2.1 Signalling of EpCAM

EpCAM controls cell cycle progression and specialization through regulated intramembrane proteolysis (RIP). Keller et al. (2019) described RIP as the process of transmembrane proteins cleavage inside the membrane plane to release cytosolic fragments that are physiologically active to enter the nucleus and regulate transcription process of the gene. There are two essential steps in RIP; firstly, removal of EpCAM's ectodomain by the action of TACE/ADAM17 enzyme and the second step is the cleavage

of C-terminal intracellular domain (EpICD) by presenilin 2 (PS-2), a protease that is part of the γ-secretase complex (Mohtar et al., 2020). Following RIP, EpICD along with FHL2, beta-catenin and Lef1 translocate from the cytoplasm enter the nucleus and bind to promoter regions of cell division regulators, pluripotency genes and genes implicated in the control of EMT-related processes that are responsible in many cell activities including proliferation, tight junction, maturation, adhesion as well as cell migration (Gires et al., 2020). In addition, according to Mohtar et al. (2020), EpCAM's roles and activity are strongly reliant on proteolytic processing.

2.2.2 Functions of EpCAM

As previously stated, most epithelial cells have a high level of EpCAM expression. In terms of EpCAM functions, according to Mohtar et al. (2020), this transmembrane glycoprotein can exist in multimers, allowing it to perform different functions under different physiological conditions. There are several studies which showed that cell adhesion, migration, cellular proliferation and differentiation are all aided by EpCAM. Li Huang et al., (2018) reported other roles of EpCAM in normal tissues for example EpCAM proteins contribute in regulating cell signalling as well as the maintenance and formation of organ morphology. Other than that, EpCAM is also involved in several signalling pathways and epithelial-to-mesenchymal transition (EMT) (Gaber et al., 2020). To fulfil these functions, EpCAM must follow RIP to generate functionally active extracellular domain, EpEX and intracellular fragments, EpICD (Liang et al., 2018). According to Schnell et al. (2013), a cytosolic signalling molecule, EpICD, leads to degradation and inactivation of transmembrane protein. Overall, downstream signalling has been reported as a result of regulated intramembrane proteolysis action of EpCAM.

2.2.3 EpCAM in cancer

Human epithelial cell adhesion molecule (EpCAM) is considered as the most studied tumour-associated antigen. It is commonly found in carcinomas and stimulates proliferation after regulated intramembrane proteolysis (Tsaktanis et al., 2015). In epithelial cancer, high EpCAM expression is linked to tumour progression (Schnell et al., 2013). Research has shown that DNA hypomethylation in the EpCAM promoter region has often been found in various cancer types for example colorectal, ovarian and breast cancer (Mohtar et al., 2020).

EpCAM is the most commonly used epithelial marker in capturing circulating tumour cells (CTCs) in the blood circulation of carcinomas patients. CTCs are the focus of liquid biopsy, which allows for the collection of a wide range of data at the DNA, RNA and protein levels (Keller et al., 2019). EpCAM abnormal expression levels have been linked to a variety of illnesses including cancer metastasis and cancer stem cells. The elevated EpCAM expression in tumour tissues often induced by tumour tissues losing their sticky structure and cell polarity. Despite all the research done, many molecular details about role of EpCAM in cancer development are unknown (Gaber et al., 2020).

Gain-of-function and loss-of-function are two types of studies that have recently emerged as fascinating approaches due to their sophisticated and beneficial characteristics; thus, filling a critical gap in cancer biology (Li et al., 2019). EpCAM in-vitro and in-vivo studies have been conducted to identify gain-of-function of EpCAM expression (Gostner et al., 2018). But for loss-of-function of EpCAM, the studies are limited especially in breast. So far, research has been done to knockout zebrafish and mouse models to discover more about the functional roles of EpCAM in disease like cancer. Meanwhile EpCAM knockdown in Xenopus models supported its role in

adherens junction and E-cadherin integrity regulation (Mohtar et al., 2020). The roles of EpCAM in supporting tumourigenesis were reported by Mohtar et al. (2020) as in Figure 2.3.

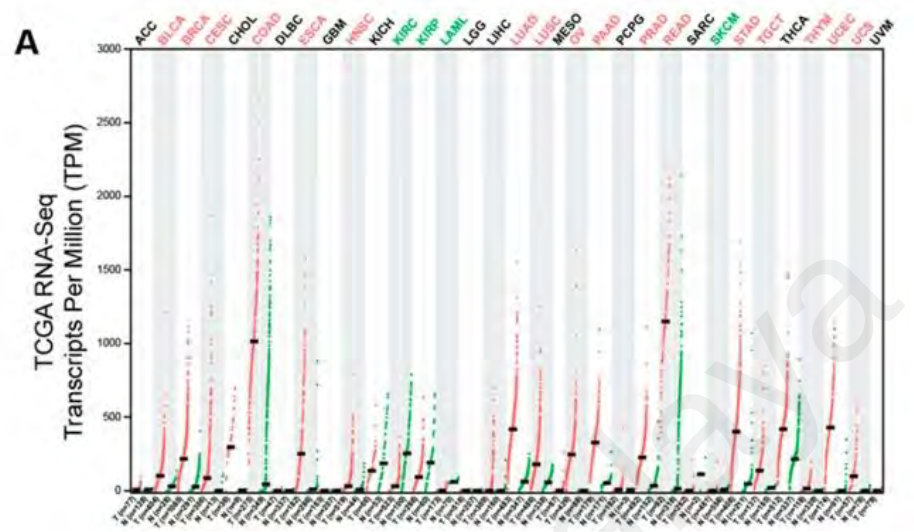


Figure 2.4: EpCAM expression in tumour and normal tissues samples is represented by dot plots. Cancer types that have higher EpCAM expression compared to the normal tissues (red) meanwhile cancer types that have reduced EpCAM expression than their normal counterparts (green). Source: Adapted from Mohtar et al., (2020).

CHAPTER 3: METHODOLOGY

3.0 Cell culture and handling

All cell culture work was carried out in a biosafety cabinet using standard aseptic tissue culture techniques. Only molecular biology grade chemicals and plasticware were used, otherwise sterilisation was done via autoclaving (TOMY SX300). For autoclave, water was added into the autoclave machine until it covered the basket holes and default setting was used. The consumables and reagents were then transferred to an oven to dry before use.

3.1 Cell recovery

To recover the cell from liquid nitrogen, the cryovial containing frozen cell lines were removed from the liquid nitrogen tank and swiftly thawed in a 37°C water bath until a tiny amount of ice remained before being transferred into a 60-mm culture plate containing fresh complete culture media. The cells were then incubated in a humidified incubator 37°C with 50% carbon dioxide. On the next day, the media were discarded in order to remove dimethyl sulfoxide (DMSO).

3.2 Cell lines and maintenance

All cell MCF-7 and T-47D (Table 3.1) were cultured in Dulbecco's Modified Eagle Medium (DMEM) (Nacalai Tesque, Japan) supplemented with 10% fetal bovine serum (FBS) (Sigma Life Science) and penicillin-streptomycin mixed solution (Nacalai Tesque, Japan). The cells were maintained in 60-mm tissue culture dishes (Trueline) and kept in a humidified cell incubator at 37°C with 95% humidity and 5% carbon dioxide.

The cells were sub-cultured 1 to 3 times per week. For sub-culturing, the existing media were discarded and the cells were rinsed once with sterile 1X phosphate-buffered saline (PBS). The cells were then trypsinized with 0.5 ml of pre-warmed 1X trypsin-EDTA 0.5%, and incubated in the incubator for approximately 5 minutes until they detached. The cell's detachment was observed under a microscope.

The detached cells were transferred into a new tissue culture dish containing 2.5 ml of the warmed medium supplement specified below, which was then added up to a total amount of 3 ml. Trypsinized cells were mixed 1:1 with Trypan Blue Stain 0.4% to stain for dead cells, placed onto a haemocytometer and counted under microscope before sub-culturing process.

Table 3.1: Cell lines and growth conditions and media.

Cell lines	Description	Growth Conditions	Media
MCF-7	Luminal A Adherent cells	5% CO_2, 37°C	DMEM, 10% FBS, 1% Penicillin-Streptomycin
T-47D	Luminal A Adherent cells	5% CO_2, 37°C	DMEM, 10% FBS, 1% Penicillin-Streptomycin

3.3 Cell storage

Cell lines were grown to 90% confluence in 10 cm dishes and then trypsinized before being diluted in a final volume of 10 ml. In sterile 15 ml falcon tubes, the cells were pelleted by centrifuging at 1000 rpm for 5 minutes. The cell pellet was then resuspended in freezing medium containing 50% FBS, 20% DMSO and 30% serum-free media. Before being transferred to liquid nitrogen containers for long-term storage, cells were aliquoted

into 1 ml cryotubes (Nunc) and placed in gradient freezing containers (Nalgene) for 24 hours at -80°C.

3.4 Harvesting cell

The highly confluent cells (80- 100% confluency) were harvested on ice. The cells were rinsed with cold 1X PBS after the existing media were discarded. 1ml of cold 1X PBS was pipetted into the plate containing cells. Then, cell scraper (BioMedia) was used for detachment of the cells. Cells detachment were observed under inverted microscope before the cells were transferred into a centrifuge tube. The cells were centrifuged at 3000 rpm for 5 mins at 4°C in order to get the cell pellets. The supernatant was discarded and cell pellets were subjected to cell lysis or stored at -80°C until further use.

3.5 Cell lysis

A box containing ice was prepared beforehand, the cell pellets were removed from -80°C and placed on ice. 50µl of urea lysis buffer or RIPA buffer was added into each tube containing pellet, mixed thoroughly and incubated on ice for 30 minutes. The mixture of cell pellets and lysis buffer was then centrifuged at 10000 rpm for 15 minutes at 4°C. After that, the cell pellets were discarded and the supernatant were transferred into a new 1.5 ml Eppendorf tube and subjected to BCA assay or stored at -80°C until further use.

3.6 To establish EpCAM knockout breast cancer cells using CRISPR-Cas9 gene editing technology

CRISPR-Cas9 gene editing was used to generate EpCAM knockout in two types of breast cancer cell lines (MCF-7 and T-47D). The experimental pipeline to generate

EpCAM-knockout breast cancer cells using CRISPR-Cas9 is described in Figure 3.1 below.

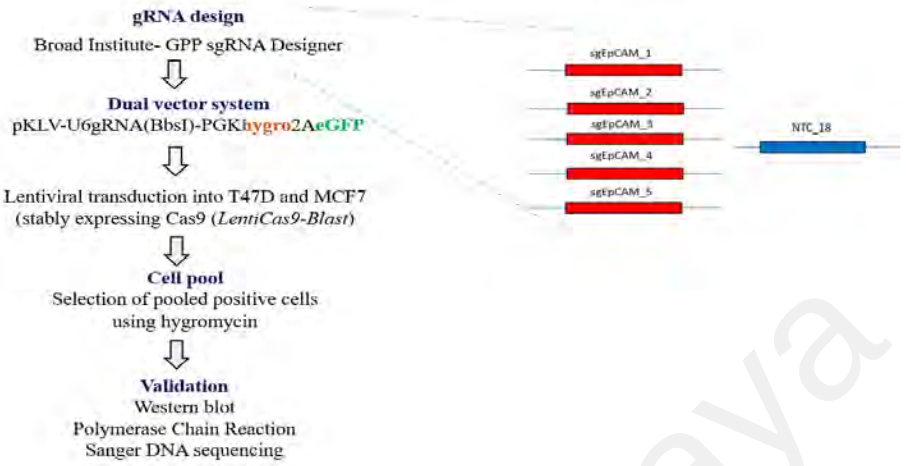
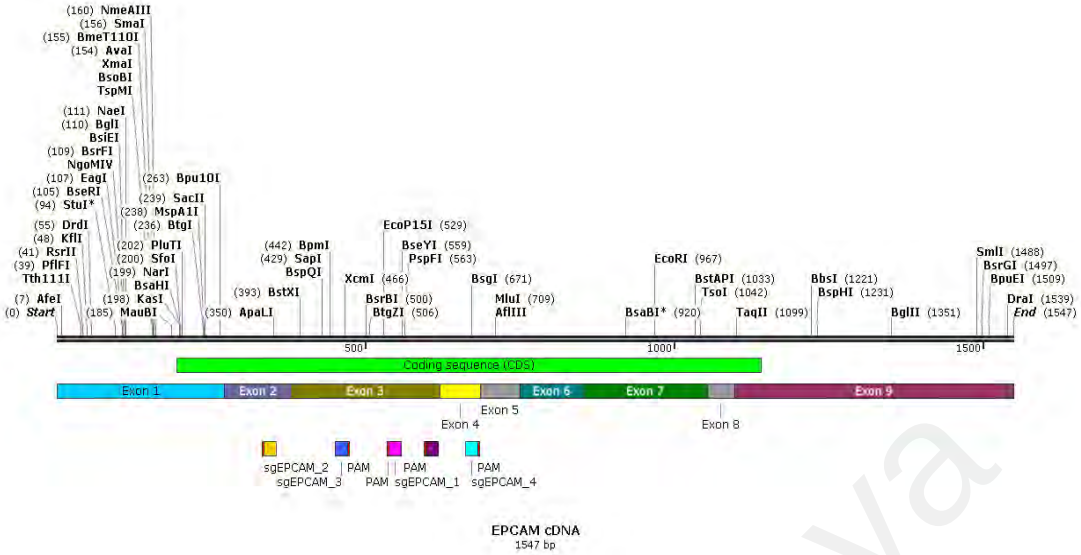


Figure 3.1: Experimental pipeline to generate EpCAM-knockout breast cancer cells using CRISPR-Cas9.

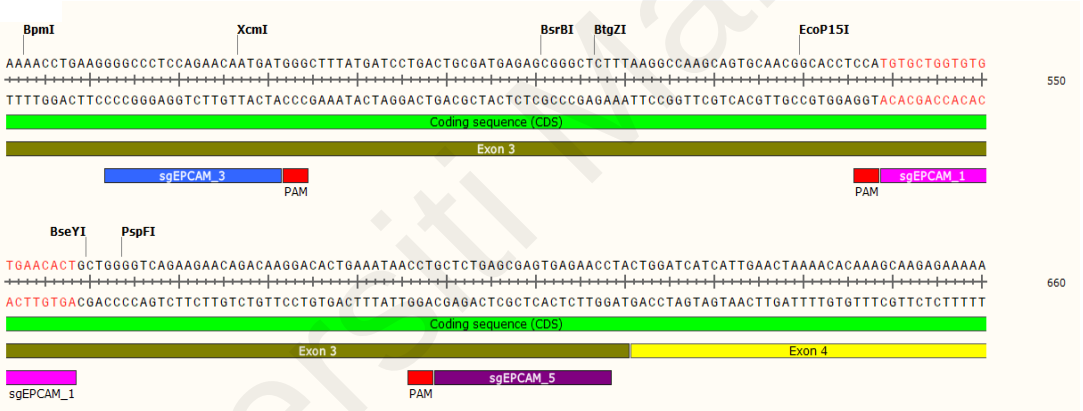
3.6.1 Guide RNA Design

Five single guide-RNAs (sgRNAs) targeting EpCAM and one sgRNA of non-targeting control (NTC) were designed for each breast cancer cell line using the design tool at the Broad Institute sgRNA (https://portals.broadinstitute.org/gpp/public/analysis-tools/sgrna-design). In this study, the top two selected sgRNAs, sgEpCAM 1 and sgEpCAM 2 were used. These two sgRNAs present on exon 3 and exon 2, respectively (Figure 3.2). Non-targeting control (NTC) served as the negative control. All sgRNA constructs used in this study were purchased from Addgene.

a)



b)



c)

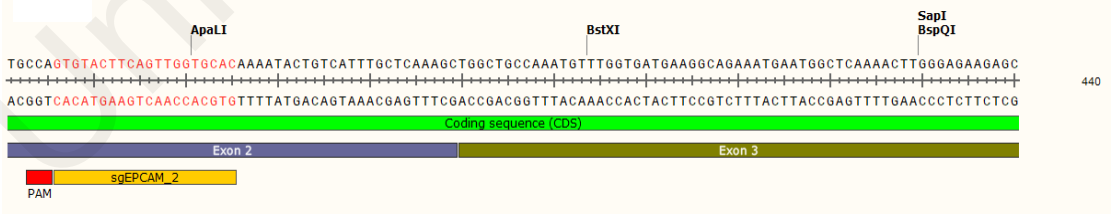


Figure 3.2: a) Coding sequence (CDS) of EpCAM, b) CRISPR-Cas9 mediated EpCAM targeting. Region targeted by CRISPR-Cas9 in exon 3 named as sgEpCAM_1, c) CRISPR-Cas9 mediated EpCAM targeting. Region targeted by CRISPR-Cas9 in exon 2 named as sgEpCAM_2.

3.6.2 Generation of stable Cas9-expressing breast cancer cell lines

In this study, a dual vector system was used with one vector containing sgRNA expression (pKLV-U6-gRNA (Bbsl)-PGKhygro2AeGFP), and with another vector containing Cas9 enzyme (lentiCas9-Blast). The positively transduced cells were selected with antibiotic. The strategy employed to generate stable Cas9-expressing breast cancer cell lines is simplified in Figure 3.3. The sgRNA plasmids were delivered into the breast cancer cell lines by lentiviral transduction as it allows for stable expression of the sgRNAs and genomic integration. The cells were harvested as pooled knockout cells and subjected to Western blot after selection with antibiotic.

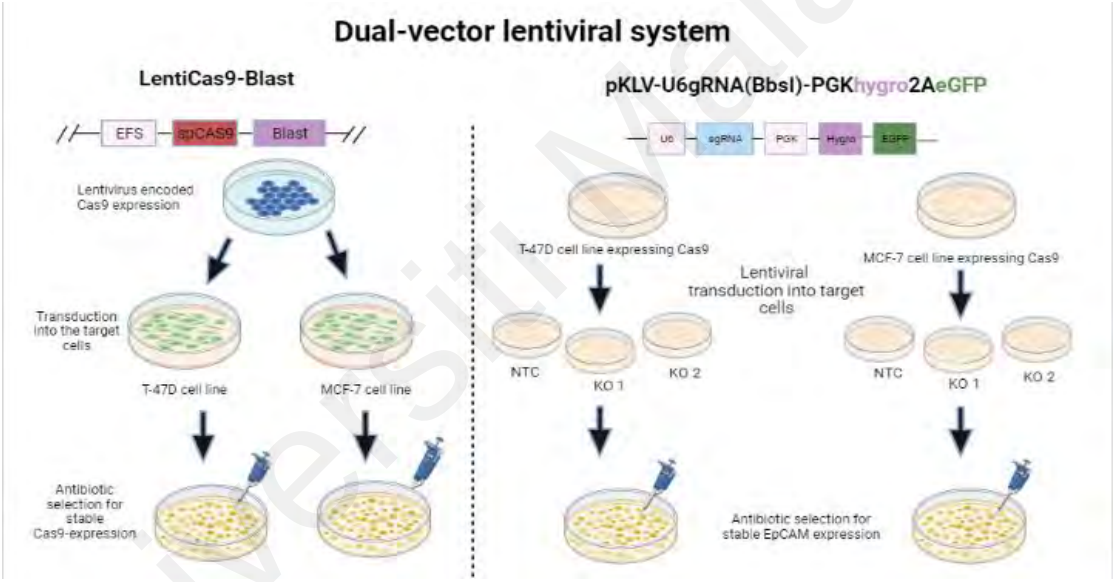


Figure 3.3: Strategy employed to generate stable Cas9-expressing breast cancer cell lines.

3.6.3 Cloning sgRNA construct into the sgRNA expression vector

The quantity and purity of sgRNA expression plasmid were determined using Nanodrop before cloning the sgRNA construct into the expression vector. The sgEpCAM and sgNTC constructs were cloned into the pKLV-U6-gRNA(BbsI)-PGKhygro2AeGFP sgRNA expression vector, followed by transduction into the HEK293T cells expressing

the Cas9. The sgRNA expression plasmid, pKLV-U6-gRNA(BbsI)-PGKhygro2AeGFP was digested with 20 unit/µg *BbSI-HF* (NEB R3539), the reaction was incubated for 2 hours at 37°C and 65°C for 20 minutes in order to inactivate the enzyme. After that, the digested plasmids were treated with 5 unit/µg Antarctic phosphatase (AnP) (NEB M0289) or stored in -20°C freezer until further use.

T4 Polynucleotide Kinase (NEB M0201) and T4 ligation buffer (NEB) were used to anneal and phosphorylate both top and bottom strands of sgRNAs (Sigma-Aldrich) (NEB). The reaction was incubated for 30 minutes at 37°C before being heat-inactivated for 5 minutes at 95°C to denature the enzyme. Then, at a rate of 5°C per minute, the temperature was gradually reduced to 25°C. The ramping down step is crucial because it allows the top and bottom of the sgRNA strands to anneal. The annealed sgRNA construct was diluted in 1:200 in DNAse/RNAse-free water and ready for ligation or stored in a -20°C freezer until further use.

T4 ligase (NEB M0202) was used to ligate the sgRNA strands into the sgRNA expression vector. The reaction was incubated at 25°C for 1 hour before being inactivated for 10 minutes at 65°C. Instead of annealed sgRNA, water was used as a ligation negative control. The cells were then transformed into chemically competent DH5α cells. Then, 10- 20 µl of ligated plasmid were added to the Eppendorf tube and was gently mixed by flicking the tube 4-5 times. The competent cells were incubated on ice for 30 minutes before being heat-shocked at 42°C for 30 seconds. Then, the reaction was incubated on ice for another 5 minutes. SOC medium or LB broth were added and incubated at 37°C while shaking for 200-250 rpm. The broth was cultured overnight at 37°C on the LB + Ampicillin (100 µg/ml) agar plate in a bacteria incubator.

On the following day, the colony were picked and cultured overnight in 100 µg/ml LB + Ampicillin broth before the plasmid extraction. Hygromycin was used to select cells that had been transduced with the sgRNA expression vector.

3.6.4 Lentiviral production

HEK293T cells were plated in a 6-well plate and incubated overnight until around 70-80% confluency for lentiviral production. The transfection mix was prepared, and 10 µl Attractene was added dropwise into the mixture. After that, the mixture was incubated for 30 minutes at room temperature. 1.8 ml fresh DMEM media was added to the HEK293T cells media. The transfection mixture was then added dropwise into the HEK293T cells. The total volume of media including the transfection mix was 2 ml. The HEK293T cells were cultured and incubated for 48-72 hours until 70-80% confluency before harvesting the lentivirus particle.

Lentivirus particles were harvested the next day by passing the media through a 0.45 µM syringe filter with a 1 ml/ 10 ml syringe and the filtrate was collected in a 15 ml falcon tube. The lentivirus-containing medium was aliquoted into several cryovial tubes and kept at -80°C or used directly for cell transduction.

For transduction, the harvested lentivirus was introduced to the cells in the presence of polybrene transfection reagent (Merck Millipore #TR-1003-G) and thoroughly mixed by swirling the plate. The mixture was incubated overnight and the media containing lentiviral supernatant was discarded the next day. The cells were then rinsed with 1X PBS and fresh medium was added to the cells. The positively transduced cells were incubated for another 24 hours or proceeded to either antibiotic selection.

3.6.5 Gene knockout validation

3.6.5.1 Western blot

For Western blotting, the harvested cells were subjected to cell lysis using 1X cell lysis buffer. The protein quantification was evaluated by determining the protein concentration using BCA reagent as a standard and total proteins were separated on a 12% SDS-PAGE gel. Samples were transferred to nitrocellulose membranes (Amersham, Protran, Germany) electrophoretically. The membranes were incubated with primary antibodies (EpCAM antibody) at 4°C overnight after blocking with TBST containing 5% fat-free dry milk for 1 hour. On the following day, the membranes were incubated with secondary antibodies (anti-mouse) before being observed in the ChemiDoc™ MP Imaging system (Bio-Rad, USA) using a chemiluminescence (ECL) detection system (Thermo Scientific). For control, incubation with housekeeping gene, β-actin was also performed. The density of each band for the target protein was quantified using ImageLab 5.0 software. The size of the protein was quantified by comparing it to a molecular weight marker, PM2610 ExcelBand™ (SMOBIO).

(a) Protein quantification

The lysate was vortexed for 15 seconds before being spun down at 4°C at 14000 rpm for 20 minutes. Then, the protein concentration was determined using the Pierce ™ BCA Protein Assay Kit (Thermo Scientific) according to the manufacturer's procedure, by measuring the absorbance at 562nm using a microplate reader (Thermo Scientific). The readings were compared to a linear curve calculated from BCA standard values using SkanIt software 2.4.3.

(b) SDS-PAGE

The stacking and resolving gels were prepared as described in (Table 3.2) and allowed to polymerise at room temperature. After adding the stacking gel, a 10-well comb was used and this depends on the number of samples.

Table 3.2: Components of resolving and stacking gel in preparing SDS-PAGE gel.

12% resolving gel		5 ml stacking gel	
Component	**Volume (ml)**	**Component**	**Volume (ml)**
Water	6.6	Water	3.4
30% acrylamide mix	8.0	30% acrylamide mix	0.83
1.5 M Tris (pH 8.8)	5.0	1.5 M Tris (pH 6.8)	0.63
10% SDS	0.2	10% SDS	0.05
10% ammonium persulfate	0.2	10% ammonium persulfate	0.05
Tetramethyl-ethylenediamine (TEMED)	0.008	Tetramethyl-ethylenediamine (TEMED)	0.005

Generally, 50 µg protein lysate was mixed 1:1 with 4X reducing sample buffer (Table 3.4) and boiled at 95°C for 5 minutes before loading onto the gels. 5 µl of PM2610 ExcelBand™ (SMOBIO) was loaded as a marker. Proteins were separated by electrophoresis at 180-200V using a 1X running buffer (Table 3.3) for 90 minutes.

The PM2610 ExcelBand™ (SMOBIO) was used as the molecular weight marker in order to measure the protein separation throughout SDS-polyacrylamide gel electrophoresis, evaluation of Western transfer efficiency on nitrocellulose membrane and estimation of protein size.

Table 3.3: 1X running buffer and 4X reducing sample buffer.

1X running buffer	4X reducing sample buffer
192 mM Glycine	250 mM Tris pH 6.8
25 mM Tris-base	50% Glycerol
0.1% SDS (w/v)	5% SDS (w/v)
	Bromophenol blue to the desired colour
	10% β-Mercaptoethanol (added fresh prior to use)

(c) Transferring gel to membrane

The separated proteins were transferred onto 0.2 μM pore nitrocellulose membrane using Western blot buffers (Table 3.4). The transfer was carried out electrophoretically at 300 mA for 1 hour 30 minutes on a hotplate stirrer. It is important to ensure that the protein is transferred properly. The membrane was then blocked with 5% milk in 1X TBS-Tween on a shaker before being incubated overnight at 4°C with EpCAM primary antibody diluted in 5% fat-free dry milk (1:1000).

Table 3.4: Western blot buffers.

Transfer buffer	TBS-Tween 20 buffer	Blocking buffer
24 mM Tris	1x TBS	1x TBS
191 mM glycine	0.1% Tween 20	0.1% Tween 20
20% (v/v) methanol		5% Dried Skimmed Milk

(d) Blotting with primary and secondary antibodies

For blotting, EpCAM mouse monoclonal (Invitrogen) and β-actin (Santa Cruz Biotechnology) were used as primary antibodies while rabbit anti-mouse was used as the secondary antibodies for this experiment (Table 3.5).

After blotting with primary antibodies, the membrane was washed three times with TBST for 5 minutes each on a shaker (Heidolph Rotamax 120). The membrane was then subjected for incubation with secondary antibodies for 1 hour at room temperature. After that, the membrane was then washed three times with TBST for 5 minutes each on a shaker. Target protein bands were identified by overlaying with enhanced chemiluminescence (ECL) reagents (Thermo Scientific) at a 1:1 ratio for 2 minutes. Excess ECL solution will be removed with tissue paper before the blot was visualized under a gel imaging system. The band densities were observed using the ChemiDoc™ MP Imaging system. The membrane was then further blotted against β-actin mouse monoclonal (Santa Cruz Biotechnology) as control.

Table 3.5: Primary antibodies and secondary antibodies.

Primary Antibody	Type	Dilution	Source	Secondary Antibody
EpCAM	Mouse monoclonal	1:1000	Invitrogen	Rabbit anti-mouse 1: 1000
β-actin	Mouse monoclonal	1:1000	Santa Cruz Biotechnology	Rabbit anti-mouse 1: 1000

3.7 Functional assays to assess the phenotypic effects of EpCAM-knockout in breast cancer cells

For this objective, EpCAM-knockout breast cancer models were subjected for further analysis. The EpCAM-knockout expressions were characterized by assaying the activity; adhesion assay, clonogenic assay, MTT assay and wound healing assay. The functional assays were carried out to determine cell fitness in terms of migration, adhesion, proliferation and ability of a single cell to grow into a colony. The migration of cells was assessed by wound healing assay to evaluate cell motility, gap distance and determine wound-closing rate. MTT assay was carried out to identify proliferation and growth rate

of cells. All of these assays allowed us to elucidate the role of EpCAM using a loss-of-function approach. Each sample was prepared in triplicates for analysis and the cell activity in EpCAM-knockout cells was compared to non-targeting control (NTC) cells.

3.7.1 Wound healing assay

Cells were seeded in triplicate 6-well plates and cultured monolayer in a humidified incubator until they reached 90% confluence. For the wound healing experiment, a scratch was made across the whole diameter of each well to induce the wound using the P-200 sterile pipette tip. The cells were then being imaged under an inverted microscope (Olympus, Japan) at 0, 24, 48, 72 hours. The gap distance and wound-closing rate were measured over time to determine cell migration. The gap distance was quantitatively evaluated using ImageJ software (National Institutes of Health). The following is the equation for calculating percentage of wound closure:

$$\textit{Wound closure (\%)} = ((W_0 - W_t)/ (W_0) \times 100 \quad (3.1)$$

W_0 = Wound area at 0 hours (μm^2)

W_t = Wound area at Δh (μm^2)

3.7.2 Adhesion assay

For adhesion assay, cells were trypsinized using 200 µl of pre-warmed 1X Trypsin-EDTA 0.5% (Gibco, Life Technologies) and incubated in an incubator for 5 minutes for cell detachment. 50 µl of trypsinized cells were mixed with Trypan Blue Stain 0.5% (Biowest) and subjected to cell counting. Then, the cells (5 x 10^5 ml) were seeded in triplicate 6-well plates. The plates were incubated for various lengths of time (4, 8 and 24

hours) to allow for cell adherence to the well of the plates. At each time point, the existing media were discarded, rinsed with 1X PBS and 500 µl fixative solution was added before staining with 500 µl crystal violet. The stained cells were visualized (with normal camera or Bio-Rad viewer) then 10% acetic acid was used to lyse the cells. The lysed cells were transferred to a 96-well plate, then the absorbance was measured at 600 nm using a microplate reader (Bio-Rad, California, USA). At each time intervals, a percentage of the total number of cells adhering to the total number of cells were calculated.

3.7.3 Clonogenic assay

Approximately 2 x 10^3 cells/ well were seeded in triplicate 6-well plates and allowed to adhere for 3 weeks. Cells were subjected to fixation and crystal violet stain on day 21 and percent of colony formation was calculated for each cell in comparison to non-targeting control cells. The observations were recorded in three independent experiments, and the results were presented with a mean percent colony formation. The percentage of colony formation was calculated according to the following equation:

$$\textit{Percent of colony formation (PE)} = \frac{\textit{No. of colonies formed}}{\textit{No. of cells seeded}} \times 100\% \qquad (3.2)$$

3.7.4 MTT assay

Cells (2x 10^3/100 µl and 3x 10^3/100 µl) were seeded in triplicate 96-well plates and allowed to grow for 24 hours until about 80% confluence. On the next day, 10 µl MTT reagent was added into each well, mixed and incubated for 4 hours in a cell culture incubator. After that, 100 µl of detergent were added and incubated overnight. The following day, cells were transferred to another 96-well plate and the absorbance at 570

nm was measured in a microplate reader (Bio-Rad, California, USA). The percentage of cell viability was calculated according to the following equation:

$$\text{Cell viability (\%)} = \frac{\text{OD570nm sample} - \text{mean (OD570nm sample-blank)}}{\text{Mean (OD570nm control)}} \times 100\% \quad (3.3)$$

3.8 Statistical analysis

GraphPad Prism (Version 9.3.1) was used to conduct the statistical analysis. P-values less than 0.05 were regarded statistically significant, either a two-way ANOVA with Tukey's range or a one-way ANOVA with Dunnett's test was employed depending on the assays.

CHAPTER 4: RESULTS

4.1 Establishing and screening of EpCAM knockout breast cancer cells

4.1.1 Gene knockout validation

Validation of EpCAM-knockout in breast cancer cell lines was required prior to performing functional assays. As highlighted in the introduction section 1.1, several studies have reported EpCAM is highly expressed in varieties of human cancer of epithelial origin including breast cancer. In this study, the effects of EpCAM knockout were tested on two different breast cancer cell lines; MCF-7 and T-47D.

CRISPR-Cas9 gene editing approach was employed in this study to target and disrupt EpCAM function in both breast cancer cell lines. The EpCAM-targeting sgRNAs and non-targeting control construct were designed using the Broad Institute sgRNA design tool. The top two sgRNAs; sgRNA_1 and sgRNA_2 were chosen based on the prediction algorithm. These two sgRNAs targeted two independent regions in EpCAM exon 2 and exon 3, respectively (Figure 3.2). Additionally, for this study, a stable expression vector using lentivirus was used to deliver the CRISPR components into the cell lines.

To knockout the EpCAM gene, the CRISPR-Cas9 system recognized the PAM sequence of the targeted site. The Cas9 protein then cleaved at these sgRNA-targeted sites and caused non-homologous end joining (NHEJ). The double strand break was repaired by the error-prone NHEJ in the absence of a repair template, resulting in the introduction of insertions and/or deletions (INDELS) that caused frameshift or nonsense mutations in the gene. Hence, perturb the EpCAM function in the cell lines. In addition, a dual vector system was used, with two plasmids encoding Cas9 and sgRNA in order to ensure that

the cell lines had stably expressed the Cas9 protein before being introduced to the respective cell lines. Then, selection of pooled positive EpCAM knockout using hygromycin was done. EpCAM expression in these cell lines were confirmed by Western blotting. In each condition, the sgEpCAM were compared to the control group, the non-targeting control. Figure 4.1 shows the Western blot analysis result of EpCAM knockout experiment in MCF-7 cell line.

Western blot using mouse monoclonal EpCAM showed that protein expression was reduced in the knockout group as compared to the control group. There was a slight band seen in sgEpCAM_1, probably because of the pooled knockout selection and it showed that sgEpCAM_2 had a better knockout than that in sgEpCAM_1. The same trend also could be seen in the Western blot analysis of EpCAM knockout in T-47D cell line. Generally, our finding demonstrated that EpCAM expression was significantly reduced in the knockout group of both cell lines, MCF-7 and T-47D as compared to their respective controls. This shows that CRISPR-Cas9 gene editing tool was effective at targeting EpCAM in both breast cancer cell lines; MCF-7 and T-47D. The generation of EpCAM knockout using CRISPR-Cas9 gene editing tool allowed us to understand the function of EpCAM in promoting breast cancer progression.

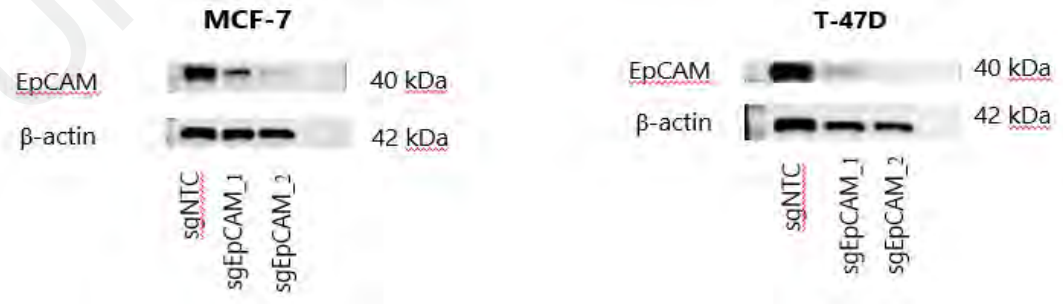


Figure 4.1: Characterisation of EpCAM expression in breast cancer cell lines. Western blot analysis of MCF-7 and T-47D EpCAM knockout. EpCAM antibody was used to analyse EpCAM protein expression in MCF-7 and T-47D cells. EpCAM (above) and β-Actin (below) antibodies were separated using 12% SDS polyacrylamide gel electrophoresis. Proteins in sgEpCAM_1 and sgEpCAM_2 were depleted as compared to sgNTC in both cell lines.

4.2 Assessing phenotypic effects of EpCAM knockout in breast cancer cells through functional assays

To functionally test the link between protein expression and breast cancer cells fitness, four functional assays were carried out including wound healing assay, adhesion assay, MTT assay and clonogenic assay. The phenotypic effects of EpCAM knockout cells were evaluated in terms of migration, proliferation, adhesion and ability of single cells to form colonies. These functional assays are important to study the characteristic of tumour progression/tumour development as they provide a measure of tendency/potential of tumour to metastasize. The advanced stage of breast cancer development is highly aggressive which increases in metastatic activity involving epithelial-mesenchymal transitions.

4.2.1 Effect of EpCAM knockout on cell migration

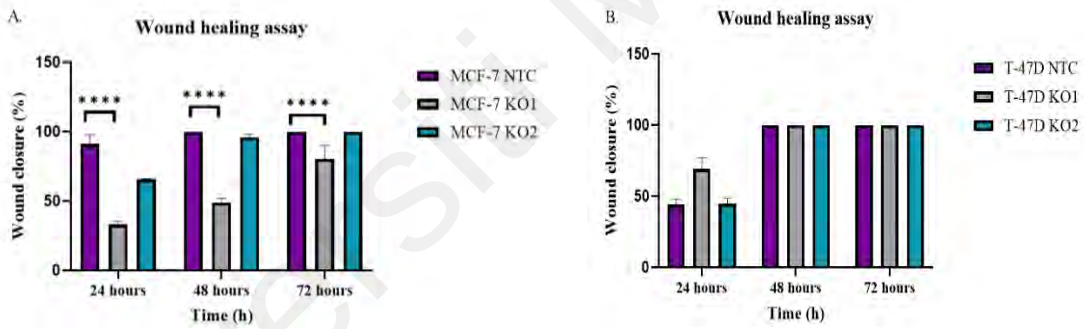
Cell migration was assessed by measuring the surface area occupied by cells over time after creating scratch assay. For migration analysis, three replicates of each cell were plated out in a 6-well plate. A sterile P-200 tip was used to create a linear wound devoid of cells on the confluent monolayer.

After conducting a scratch experiment, the wound closure was measured at different time intervals; 0, 24, 48 and 72 hours by capturing the image. ImageJ software was used to carry out automated wound area measurements by detecting the wound edges. Then, the percentage of wound closure was calculated. The relative wound closing rate at 24 hours were significantly reduced in both knockout cells; MCF-7 and T-47D as compared to that in NTC (Figure 4.2a). There was a significant difference in MCF-7 cell line during

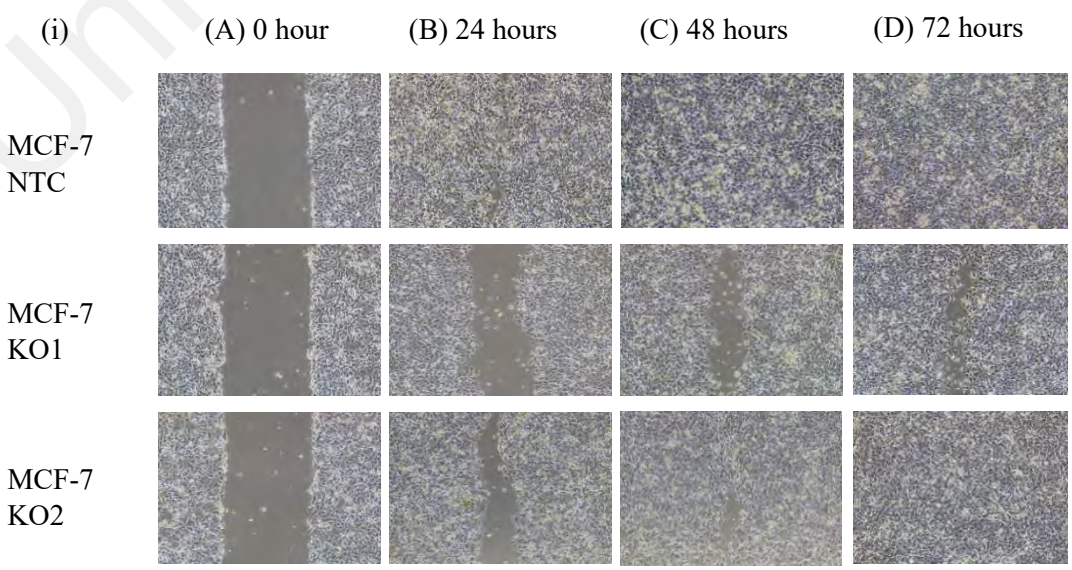
48 and 72 hours however the effect was not quite significant in T-47D cell line. For T-47D at 48 hours, the difference of T-47D wound closure could be seen under microscope but there was no statistically significant change in T-47D cell line however at 72 hours, all of the cells were at stationary phase or plateau as the cell population becoming confluent (Figure 4.2b).

Generally, this finding showed that the ability to migrate differed between breast cancer cell lines and percentage of migration in both breast cancer cell lines were significantly higher in NTC than in knockout cell lines which indicated that EpCAM was one of the migration-related proteins involved in metastasis of tumour development.

a)



b)



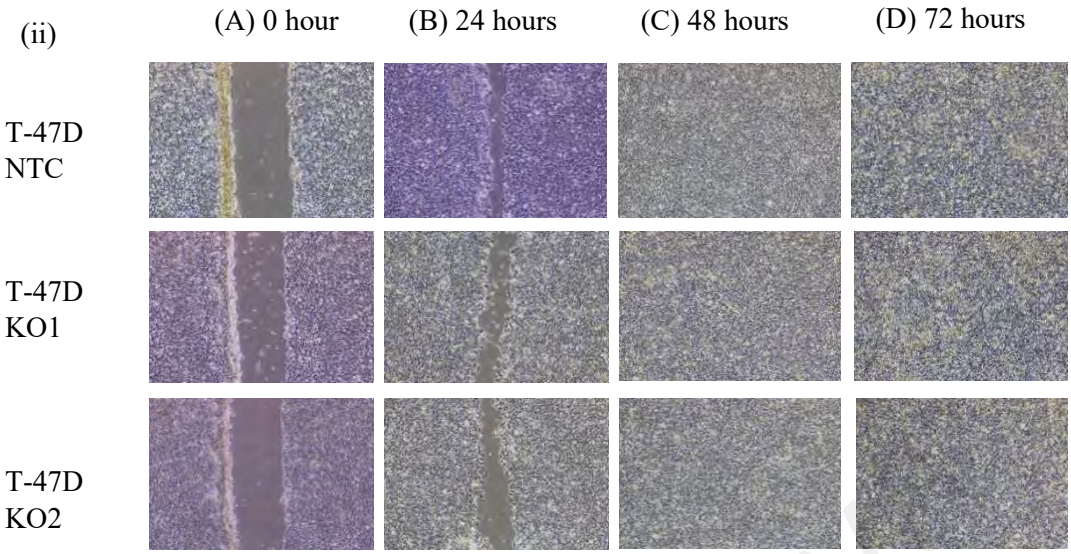


Figure 4.2: a) Data for wound healing assay for MCF-7 (A) and T-47D (B). Data presented as mean ± SEM percent wound closing rate of cells (* $p < 0.05$), b) Representative images of scratch of (i) MCF-7 and (ii) T-47D cells. Wound healing assay was carried out in knockout EpCAM cells and parent cell lines to detect migration after a time period of 24 hours. (A) 0 hour (just after inducing the wound) (B) after 24 hours (C) after 48 hours and (D) after 72 hours. ImageJ software was used to calculate the wound of each cell line.

4.2.2 Effect of EpCAM knockout on cell proliferation

The cell proliferation was quantified by assessing metabolic activity of the cells. MTT assay works by converting a water-soluble dye into an insoluble formazan in metabolically active cells. For the MTT assay, three replicates of each cell were plated in a 96-well plate. This experiment was conducted for 4 days and optical density was read using a microplate reader for each day. The optical density and the number of cells seeded were directly proportional to the range of cell concentration. Figure 4.3 showed that the percentage of cell proliferation in the EpCAM knockout and the control group differed significantly. This signified that the EpCAM knockout group had a low metabolic activity hence had a low reduction of MTT. The reduction of tetrazolium salts was visible 24 hours after cell seeding, as the metabolically active cell converted the yellow solution to

dark purple. These results indicated that EpCAM knockout altered cell metabolism, resulting in a decrease in the proliferation rate of both MCF-7 and T-47D cell lines.

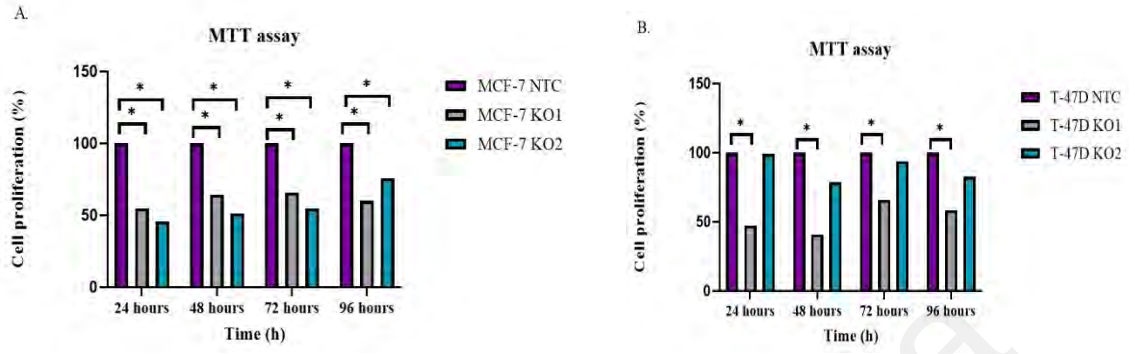
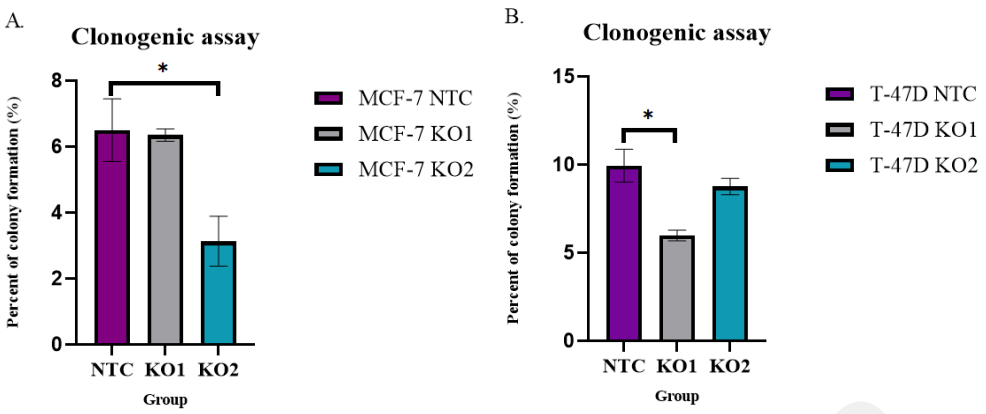


Figure 4.3: Data on effect of EpCAM knockout on the cell proliferation at 24, 48, 72 and 96 hours of MCF-7 (A) and T-47D (B) cells. Data presented as mean percent of cell proliferation (* $p < 0.05$).

4.2.3 Effect of EpCAM knockout on ability of single cell to form colonies

Additionally, a clonogenic assay also was conducted to determine the potential of a single cell to form a colony. The colonies were fixed with fixative solution, dyed with crystal violet and counted using ImageJ software on the last day of the experiment, which lasted three weeks. Proportion of colonies developed from the seeded cells were calculated according to the equation in section 3.7.3. A colony was defined as a group of 50 or more cells. When comparing the size of colony formed in T-47D and MCF-7, the colony form in MCF-7 was considerably larger than that in T-47D (Figure 4.4b). The absence of EpCAM gene reduced the cell's ability to form colonies, resulting in a lower clonogenic potential of breast cancer cells as compared to wild-type MCF-7 and T-47D cells (Figure 4.4a).

a)



b)

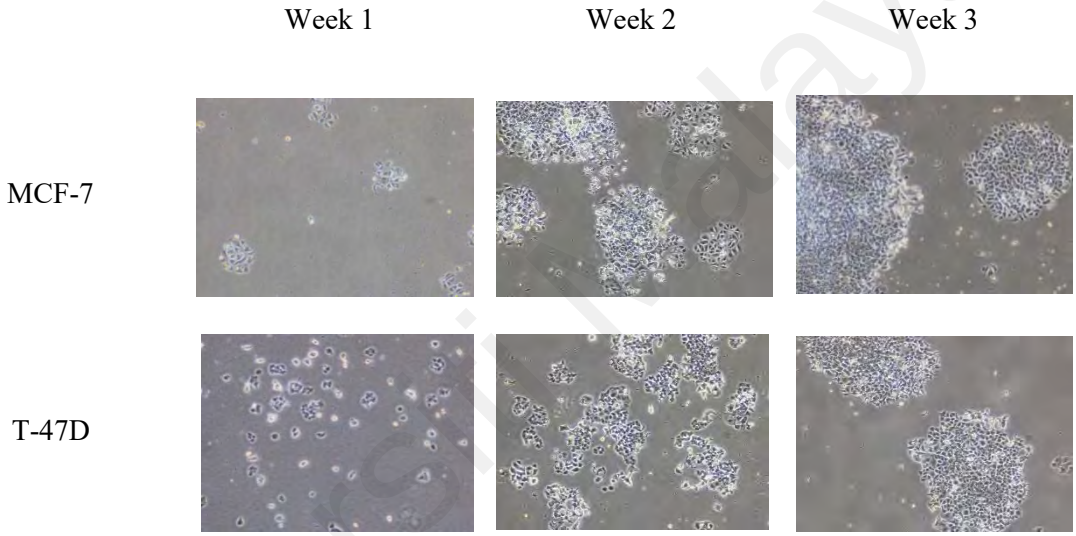


Figure 4.4: a) Data for colony formation for MCF-7 (A) and T-47D (B). Data presented as mean ± SEM percent of colony formation, b) Representative images clonogenic assay of MCF-7 and T-47D cells on week 1, week 2 and week 3.

4.2.4 Effect of EpCAM knockout on cell adhesion

Cell adhesion is important in cell communication and regulation. In this study, it showed that the knockout cell's ability to adhere were reduced in both cell lines; MCF-7 and T-47D as compared to respective controls. Cell attachment is a time-dependent process. This experiment was carried out in three different time intervals. The cells were allowed to adhere for 4, 8 and 24 hours. The cells were stained with crystal violet for each time point and subjected to absorbance reading using a microplate reader.

Cells in the knockout group have reduced stickiness as compared to the control group. Under the microscope, the number of floating cells decreased as time passed. In both cell lines; MCF-7 and T-47D, there was a significant difference as compared to the respective control group (Figure 4.5). It showed that the ability of breast cancer cells to adhere have reduced in the absence of EpCAM protein.

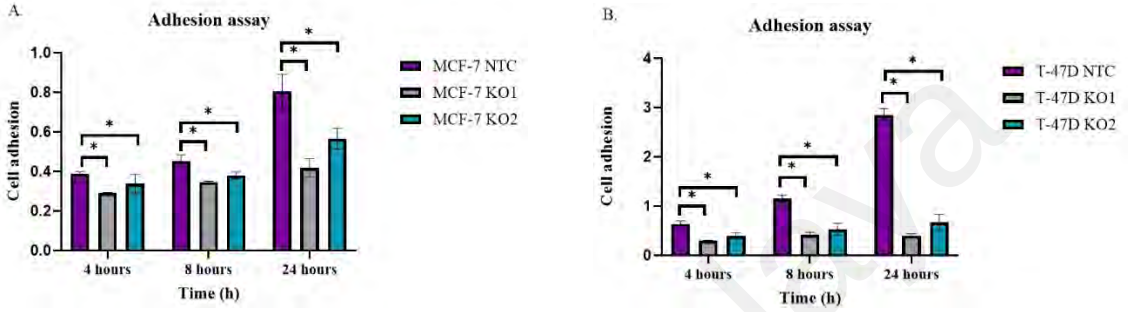


Figure 4.5: Data on effect of EpCAM knockout on the cell adhesion at 4, 8 and 24 hours of MCF-7 (A) and T-47D (B) cells. Data presented as mean ± SEM cell adhesion.

CHAPTER 5: DISCUSSION

5.1 CRISPR-Cas9

In this study, it is found that knockout of EpCAM significantly decreased the fitness of both cell lines, MCF-7 and T-47D. This was caused by the activity of Cas9 protein that cleaved the gene sequence and induced double stranded break. NHEJ, as previously stated, is a type of error-prone repair mechanism that can result in nucleotide deletion.

CRISPR system consists of three major components which are crRNA, tracrRNA and Cas9 protein. crRNA-tracrRNA form duplex fused into a chimeric single guide RNA (sgRNA). All sgRNAs must be adjacent to a protospacer adjacent motif (PAM) site (NGG). Cas9 protein acts as a nuclease that will cause double stranded break (DSB). This system is a natural process that has long functioned as a bacterial immune system and it is programmable. There are three types of CRISPR systems commonly found in bacterial and archaeal host. The type II CRISPR system, which is derived from *Streptococcus pyogenes*, is one of the best and widely employed in research.

The CRISPR system will transcribe and translate creating short crRNA molecules and those repeating palindromic sequences will form hair-pin structure (gRNA) helping the Cas enzyme to cut exactly where it is supposed to. CRISPR is a precise tool, no observable off-target effects detected as compared to others. The discovery of CRISPR-Cas9 in the year 2016 triggered a huge amount of innovative scientific inquiry to treat diseases including cancer, Coronavirus disease (COVID-19) and others.

In the past two decades, a number of researchers sought to determine the function of genes by doing knock-out, knock-in, CRISPRi, CRISPRa. The link between CRISPR-Cas9 and diseases including cancer has been the centre of attention. Metastasis of breast cancer cells is the leading cause of death among women around the world. One of the genes that is upregulated in breast cancer is EpCAM. EpCAM is upregulated in various epithelial cancers including lung squamous cell carcinoma (LUSC) and colon adenocarcinoma (COAD) as well as breast cancer as illustrated in dot plots EpCAM expression (figure 2.4) by Mohtar et al. (2020) This gene has a pro-metastatic activity that leads to the worst case of cancer. In this study, EpCAM knockout breast cancer was generated using CRISPR-Cas9 gene editing approach in order to study the function of this pro-metastatic gene.

Western blot analysis revealed that EpCAM protein expression was reduced in breast cancer cell lines; MCF-7 and T-47D, respectively. The pooled-knockout EpCAM gene reduced the chance to obtain a complete knockout because the editing outcomes were random; however, antibiotic selection were further conducted to select the knockout cells. Additionally, the variation could potentially be resulted from different exons being targeted. As for sgEpCAM_1 the editing was at exon 3 whereas sgEpCAM_2 was on exon 2. However, it is still premature to speculate about the therapeutic value of the knockout in the earlier exon but at least based on the functional assay, it is possible to observe the difference in the fitness of edited cells for knockout groups in both cell lines; MCF7 and T47D. Also, reports on the differences in the deletion of location genomic locus are scarce. However, report by Canver et al. (2014), stated that individual sgRNAs construct may exhibit variations in editing efficacy in the presence of Cas9 due to deletion frequency and deletion size. The CRISPR-Cas9 approach in this study was accurate in

targeting EpCAM in both cell lines. Indeed, the loss-of-function study is important to determine gene function and its effects on cancer development.

5.2 FUNCTIONAL ASSAYS

The purpose of this study was to identify the phenotypic effects of targeting EpCAM, a gene that supports breast cancer tumourigenesis in the hopes that the gained knowledge and information could be useful for prognosis, diagnostic and therapeutic intervention for epithelial cancers. As mentioned in Chapter 1, there were several studies aimed to elucidate the role of EpCAM but the studies mainly used overexpression strategy and thorough understanding on the underlying mechanisms that support its progression is needed in order to provide a better treatment for cancer patients.

In recent years, there has been a growing academic interest in gene editing technology such as CRISPR-Cas9 and functional assays particularly in oncology to have a better understanding on cancer pathogenesis. As aforementioned, CRISPR-Cas9 was employed in order to knockout the specific protein that is involved in breast cancer metastasis with high specificity and low error. We used a stable expression using viral particles-mediated delivery to ensure the continuous gene editing, which is considered the most efficient method for evaluating cancer cells activity.

In this study, the functional assays were carried out in order to assess the phenotypic effects of EpCAM knock-out in breast cancer cells, determine cell fitness and track cancer cells metastatic potential. *In vitro* functional assays showed that this pro-metastatic gene was functionally important in supporting breast cancer cells growth. The findings showed that EpCAM knock-out resulted in reduced breast cancer cells migration, proliferation,

adhesion and ability to form colonies. This was in line with the reports demonstrating loss-of-function studies by silencing EpCAM expression particularly caused decrease in cell proliferation, migration and invasiveness (Gostner et al., 2011).

To assess cell migration, wound-healing assay was carried out to evaluate the cell's migratory behaviour. According to several studies, cancer cells use different movement strategies, they can migrate individually or collectively. For this study, we observed that both MCF-7 and T-47D cells were collectively migrated. The cell migration occurs as a result of cell-to-cell interactions with neighbouring cells. Migration is the main mechanism involved in metastasis of cancer cells. Different extent of wound healing closing rate was found that represented cancer cell metastasis among different cell types.

The experiment was assessed at different time points; 0, 24, 48 and 72 hours. Finding showed that the knockout cells exhibited lower metastasis capability compared to respective controls. ImageJ software was used to measure the wound area. Then, the percentage of wound closure was calculated using the equation in section 3.7.1. The statistical data demonstrated that EpCAM perturbation reduced the MCF-7 and T-47D cells migratory properties *in vitro* experiments. However, the cells were at stationary phase during 48 and 72 hours, hence a lower time-point should be considered after this. These results indicate that knockout breast cancer cell lines reduced migration in breast cancer cells by regulating the expression of migration-related proteins.

In addition, to assess the phenotypic effect of EpCAM knock-out in cell proliferation, MTT assay was performed. MTT is a tetrazolium salt that is often used in cell analysis to identify reductive metabolism. As aforementioned, the MTT assay involves the activity

of mitochondrial reductase, which converts the MTT reagent to a crystal formazan by the activity of NAD(P)H-dependent oxidoreductase enzymes in live cells. The darker the solution indicated the more viable and metabolically active cells. In this study, the proliferation of the cells was monitored for four continuous days. The finding showed that lack of EpCAM in the knockout group decreased the ability of the cells to convert the dye resulting in low absorbance readings as compared to the respective controls. This showed that knocking out EpCAM resulted in a reduction in cell metabolic activity.

Another assay, clonogenic assay was conducted to evaluate the ability of a single cell to form colony. This test was often used to determine a single cell's capacity to withstand treatments and develop into a colony (Guzman et al., 2014). In this study, the cells in the control group multiplied faster than the knockout group. Cell colonies were determined in week 1, where cells began to divide. The size of colonies formed in MCF-7 were relatively big, containing 60 to 70 cells whereas the colonies formed in T-47D were smaller but more scattered and attached to one another. Additionally, it was found that the time required for breast cancer cell lines to double were different and it was also directly linked to hormone receptors, patient age and histological grade (Andersson et al., 2015). This finding indicated that different cell lines reflect to different response to treatment.

Adhesion assay on the other hand was carried out to observe the adherence of cells for 4, 8 and 24 hours. In tumour metastasis, alteration in the cell adherence enables the tumour cells to migrate through the tissue and intravasate into arteries, where the cells are carried through the circulation system and extravasate into other organs (Mayurika & Martin, 2009). EpCAM is a molecule that is commonly found in intercellular adherens junctions. It stimulates epithelial cell migration and proliferation by modulating cadherin-

mediated cell adhesion. Same pattern as other assays can be detected in this adhesion assay as knockout groups significantly reduced in the ability to adhere as compared to their corresponding control groups.

According to the findings of this study, the CRISPR-Cas9 gene editing technology approach have successfully knocking out the right gene candidate; EpCAM, in both MCF-7 and T-47D breast cancer cells. Significantly, this study also revealed that the fitness of edited cells in both cell lines is affected by the different cell lines used resulting in differences in cell activity disruption; cell migration, adhesion and proliferation were reduced, and the ability of single cells to form colonies was also significantly reduced when compared to the respective controls. This is in line with a report on functional proteomic analyses that demonstrated that a large number (at least 164) of proteins, including proteins involved in breast cancer cell growth regulation, are differentially expressed in MCF7 and T47D breast cancer cell lines. The proteomic results showed that T47D expresses more protein than MCF7 (Adjo Aka & Lin, 2012).

These findings are solely based on pooled-knockout experiments, thus further experiments about this gene could possibly lead to the identification of machinery that targeting would perturb this gene. Also, it is appealing to test this pro-metastatic gene using mass-spectrometry based proteomics in the future. The traditional methods of functional assays that were carried out in this study are favourable because of their low cost and easy to conduct, with fewer steps and a lower demand for specialized equipment. There are more advanced techniques that require costly reagents available (Venter & Niesler, 2019), which could be considered in future. *In-vitro* assays are useful for extrapolating to *in-vivo* situations and studying the behaviour of live cells. The results of these assays showed that the most important indicator of cancer recurrence is metastasis,

which is strongly linked to a low survival rate. Hence, more research is required to fill the oncology gap.

CHAPTER 6: CONCLUSION AND FUTURE RECOMMENDATIONS

In this study, disruption of EpCAM expression resulted in a decrease in cell motility, proliferation, adhesion and colony formation. The induction of double-strand break by CRISPR-Cas9 approach puts the cells under stress and consequently reduces their metastatic efficiency. Tumour development is dependent on the processes of migration, proliferation and adhesion, all of which are critical for tumour development. The functional assays were performed to determine the effects of loss of function of EpCAM protein on the fitness of breast cancer cell models; MCF-7 and T-47D. These *in-vitro* studies have provided further insights on the involvement of EpCAM in the aggressiveness of breast cancer cells, which is linked to a higher risk of metastasis.

CRISPR-Cas9 gene editing technology can specifically knockout proteins of interest that are involved in development of cancer metastasis. The stable expression of knockout in both breast cancer cell lines models, resulting in a good correlation with minimal error in experimental data. This study provides a more practical, cost-effective and high accuracy method. The establishment of this approach allows researchers to study the function of genes in cancer development and identify the best therapeutic efficacy for different cancer cells using CRISPR-Cas9 gene editing technology approach.

In a wide-ranging perspective, this research potentially benefits breast cancer patients by paving the way in developing new targeted therapies and offering a possible cure for the disease. It could be a source of reference for other related studies, especially studies on breast cancer in Malaysia and acts as a benchmark of EpCAM study specifically in its functional interrogation. On a more specific perspective, this research provided critical insights into the mechanisms underlying EpCAM and demonstrated the putative role of

EpCAM in controlling breast cancer progression. This study may assist other researchers in elucidating the underlying mechanisms of EpCAM during breast development and may provide new or alternative ways and strategies in our collective effort to fight cancer, while providing the researchers with an in-depth knowledge pertaining to breast cancer especially on the functional interrogation of EpCAM knockout.

This study provides the necessary information in order to have a better understanding of the biology of EpCAM in cancer cells. However, there are several limitations for this research. First, time constraint due to Malaysian Government Movement Control Order (MCO). A longer period of time can enable for more exploration especially in terms of functional assays, such as adjusting the time-point and employing the proper seeding number for each cells in each assays. Secondly, employing a single cell cloning instead of pooled knockout can be done to obtain a fully complete gene knockout, allowing for more distinguish robust difference between the non-targeting (NTC) and knockout cell group. Continuous research and tremendous efforts on genetic engineering or more advanced approaches in the future will open doors for improvements in all living organisms in this world including human beings, animals as well as plants.

REFERENCES

Abdullah, N. A., Mahiyuddin, W. R. W., Muhammad, N. A., Mohamad Ali, Z., Ibrahim, L., Tamim, N. S. I., Mustafa, A. N., & Kamaluddin, M. A. (2013). Survival rate of breast cancer patients in Malaysia: A population-based study. *Asian Pacific Journal of Cancer Prevention*, *14*(8), 4591–4594.

Adjo Aka, J., & Lin, S. X. (2012). Comparison of functional proteomic analyses of human breast cancer cell lines T47D and MCF7. *PloS ONE*, *7*(2), Article#e31532.

American Joint Committee on Cancer. (2002). Breast. In *AJCC cancer staging manual*.

Andersson, I., Dustler, M., Borgquist, S., Timberg, P., Fo, D., La, K., Physics, M. R., Medicine, T., Physics, R., Radiology, D., Medicine, T., & Lund, C. S. (2015). Estimates of Breast Cancer Growth Rate From Mammograms and Its Relation To Tumour Characteristics. *Radiation Protection Dosimetry*, *169*(1–4), 151–157.

Bartee, L., Shriner, W., & Creech, C. (2017). *Cancer and Gene Regulation*. Open Oregon Educational Resources.

Breast Cancer Research Foundation. (2021a, n.d.). *Breast Cancer Statistics & Resources | Breast Cancer Research Foundation*. Retrieved on 19 November 2021 from https://www.bcrf.org/breast-cancer-statistics-and-resources

Breast Cancer Research Foundation. (2021b, May 19). *Metastatic Breast Cancer: Symptoms, Treatment, Research | BCRF*. Retrieved on 19 November 2021 from https://www.bcrf.org/metastatic-breast-cancer-symptoms-treatment

Bullock, M. D., Sayan, A. E., Packham, G. K., & Mirnezami, A. H. (2012). MicroRNAs: critical regulators of epithelial to mesenchymal (EMT) and mesenchymal to epithelial transition (MET) in cancer progression. *Biology of the Cell*, *104*(1), 3-12.

Canver, M. C., Bauer, D. E., Dass, A., Yien, Y. Y., Chung, J., Masuda, T., ... & Orkin, S. H. (2014). Characterization of Genomic Deletion Efficiency Mediated by Clustered Regularly Interspaced Palindromic Repeats (CRISPR)/Cas9 Nuclease System in Mammalian Cells*♦. *Journal of Biological Chemistry*, *289*(31), 21312-21324.

Chen, Y., & Zhang, Y. (2018). Application of the CRISPR/Cas9 system to drug resistance in breast cancer. *Advanced Science*, *5*(6), Article#1700964.

Ciriello, G., Sinha, R., Hoadley, K. A., Jacobsen, A. S., Reva, B., Perou, C. M., Sander, C., & Schultz, N. (2013). The molecular diversity of Luminal A breast tumours. *Breast Cancer Research and Treatment*, *141*(3), 409–420.

Comşa, Ş., Cîmpean, A. M., & Raica, M. (2015). The story of MCF-7 breast cancer cell line: 40 Years of experience in research. *Anticancer Research*, *35*(6), 3147–3154.

DerSarkissian, C. (2019, May 18). *The Breast (Human Anatomy): Picture, Function, Conditions, & More*. Retrieved on 18 October 2021 from https://www.webmd.com/women/picture-of-the-breasts

Feng, W., Newbigging, A. M., Tao, J., Cao, Y., Peng, H., Le, C., ... & Le, X. C. (2021). CRISPR technology incorporating amplification strategies: molecular assays for nucleic acids, proteins, and small molecules. *Chemical science*, *12*(13), 4683-4698.

Gaber, A., Lenarčič, B., & Pavšič, M. (2020). Current View on EpCAM Structural Biology. *Cells*, *9*(6), Article#136.

Gires, O., Pan, M., Schinke, H., Canis, M., & Baeuerle, P. A. (2020). Expression and function of epithelial cell adhesion molecule EpCAM: where are we after 40 years?. *Cancer and Metastasis Reviews*, *39*(3), 969-987.

Gostner, J. M., Fong, D., Wrulich, O. A., Lehne, F., Zitt, M., Hermann, M., ... & Spizzo, G. (2011). Effects of EpCAM overexpression on human breast cancer cell lines. *BMC cancer*, *11*(1), 1-14.

Guzman, C., Bagga, M., Kaur, A., Westermarck, J., & Abankwa, D. (2014). ColonyArea: an ImageJ plugin to automatically quantify colony formation in clonogenic assays. *PloS ONE*, *9*(3), Article#e92444.

Hanahan, D., & Weinberg, R. A. (2000). The Hallmarks of Cancer. *Cell*, *100*(1), 57–70.

Hanahan, D., & Weinberg, R. A. (2011). Hallmarks of cancer: The next generation. *Cell*, *144*(5), 646–674.

Holliday, D. L., & Speirs, V. (2011). Choosing correct breast cancer cell line for breast cancer research. *Breast Cancer Research*, *13*(4), 1–7.

Indramalar, S. (2019, October 16). 1 in 30 Malaysian women will have breast cancer, so get checked now | *The Star*. Retrieved on 23 March 2021 from https://www.thestar.com.my/lifestyle/family/2019/10/16/breast-cancer-2/

Johnson, N. A. (2016). Evolutionary Medicine I. An Overview and Applications to Cancer.

Kamal, M., Razaq, W., Leslie, M., Adhikari, S., & Tanak, T. (2017). Circulating Tumour Cells in Breast Cancer: A Potential Liquid Biopsy. *Breast Cancer - From Biology to Medicine*.

Keller, L., Werner, S., & Pantel, K. (2019). Biology and clinical relevance of EpCAM. *Nature*, *3*(6), 165–180.

Lee, A. V., Oesterreich, S., & Davidson, N. E. (2015). MCF-7 Cells - Changing the Course of Breast Cancer Research and Care for 45 Years. *Journal of the National Cancer Institute*, *107*(7), 1–4.

Li Huang, Yanhong Yang, Fei Yang, Shaomin Liu, Ziqin Zhu, Zili Lei, J. G. (2018). Functions of EpCAM in physiological processes and diseases (Review). *International Journal of Molecular Medicine*, 1771–1785.

Li, Y., Zhang, Y., Li, X., Yi, S., & Xu, J. (2019). Gain-of-Function Mutations: An Emerging Advantage for Cancer Biology. *Trends in Biochemical Sciences*, *44*(8), 659–674.

Liang, K. H., Tso, H. C., Hung, S. H., Kuan, I. I., Lai, J. K., Ke, F. Y., Chuang, Y. T., Liu, I. J., Wang, Y. P., Chen, R. H., & Wu, H. C. (2018). Extracellular domain of EpCAM enhances tumour progression through EGFR signaling in colon cancer cells. *Cancer Letters*, *433*, 165–175.

Lisa Fayed. (2020, April 22). *The History and Discovery of Cancer*. Retrieved on 10 February 2021 from https://www.verywellhealth.com/the-history-of-cancer-514101

Lu, J., Steeg, P. S., Price, J. E., Krishnamurthy, S., Mani, S. A., Reuben, J., ... & Yu, D. (2009). *Breast Cancer Metastasis: Challenges and Opportunities*.

Martowicz, A., Seeber, A., & Untergasser, G. (2016). The role of EpCAM in physiology and pathology of the epithelium. *Histology and Histopathology*, *31*(4), 349–355.

Maskarinec, G., Pagano, I., Lurie, G., Bantum, E., Gotay, C. C., & Issell, B. F. (2011). *Factors Affecting Survival Among Women with Breast Cancer in Hawaii*.

Mayurika Lahiri, J. H. M. (2009). Nitric oxide decreases motility and increases adhesion in human breast cancer cells. *Research Gate*, 275–281.

Meneka Kumaran. (2020, August 28). *Breast Cancer Awareness 2020 | DTAP Clinics Malaysia*. Retrieved on 4 June 2021 from https://drtanandpartners.com.my/breast-cancer-awareness-malaysia/

Mohtar, M. A., Syafruddin, S. E., Nasir, S. N., & Low, T. Y. (2020). Revisiting the roles of pro-metastatic EpCAM in cancer. *Biomolecules*, *10*(2), Article#255.

Mohtar, M. A. (2017). *Novel role of an ER-resident chaperone pathway in cancer signalling* (Doctoral dissertation, University of Edinburgh).

Movahedi, M., Haghighat, S., Khayamzadeh, M., Moradi, A., Ghanbari-Motlagh, A., Mirzaei, H., & Esmail-Akbari, M. (2012). Survival Rate of Breast Cancer Based on Geographical Variation in Iran, a Na-tional Study. *Iranian Red Crescent Medical Journal*.

Pandit, P., Patil, R., Palwe, V., Gandhe, S., Patil, R., & Nagarkar, R. (2020). Prevalence of molecular subtypes of breast cancer: a single institutional experience of 2062 patients. *European journal of breast health*, *16*(1), Article#39.

Pandya, S., & Moore, R. G. (2011). Breast development and anatomy. *Clinical Obstetrics and Gynecology*, *54*(1), 91–95.

Pichardo, G., (2020, January 20). *Cancer: Sarcoma, Carcinoma, Lymphoma, and Leukemia*. Retrieved on 26 April 2021 from https://www.webmd.com/cancer/guide/understanding-cancer-basics

Rahman, M., & Mohammed, S. (2015). Breast cancer metastasis and the lymphatic system (Review). *Oncology Letters*, *10*(3), 1233–1239.

Reyna, V. F., Nelson, W. L., Han, P. K., & Pignone, M. P. (2015). Decision making and cancer. *American Psychologist*, *70*(2), Article#105.

Roarty, K., & Echeverria, G. V. (2021). Laboratory Models for Investigating Breast Cancer Therapy Resistance and Metastasis. *Frontiers in Oncology*, *11*(March), 1–17.

Sankpal, N. V., Brown, T. C., Fleming, T. P., Herndon, J. M., Amaravati, A. A., Loynd, A. N., & Gillanders, W. E. (2021). Cancer-associated mutations reveal a novel role for EpCAM as an inhibitor of cathepsin-L and tumour cell invasion. *BMC Cancer*, *21*(1), 1–13.

Schliemann, D., Ismail, R., Donnelly, M., Cardwell, C. R., & Su, T. T. (2020). Cancer symptom and risk factor awareness in Malaysia: Findings from a nationwide cross-sectional study. *BMC Public Health*, *20*(1), 1–10.

Schnell, U., Kuipers, J., & Giepmans, B. N. G. (2013). EpCAM proteolysis: new fragments with distinct functions? *Biosci. Rep*, *33*(2), Article#30.

Scully, O. J., Bay, B. H., Yip, G., & Yu, Y. (2012). Breast Cancer Metastasis. *Cancer Genomics & Proteomics*, *9*(5), 311–320.

Stitzenberg, K., & Ridge, J. A. (2009). What Is Cancer? In *Abernathy's Surgical Secrets*. Elsevier Inc.

Torre, L. A., Bray, F., Siegel, R. L., Ferlay, J., Lortet-Tieulent, J., & Jemal, A. (2015). Global cancer statistics, 2012. *CA: A Cancer Journal for Clinicians*, *65*(2), 87–108.

Tsaktanis, T., Kremling, H., Pavšič, M., Von Stackelberg, R., Mack, B., Fukumori, A., Steiner, H., Vielmuth, F., Spindler, V., Huang, Z., Jakubowski, J., Stoecklein, N. H., Luxenburger, E., Lauber, K., Lenarčič, B., & Gires, O. (2015). Cleavage and cell adhesion properties of human epithelial cell adhesion molecule (HEPCAM). *Journal of Biological Chemistry*, *290*(40), 24574–24591.

Venter, C., & Niesler, C. U. (2019). Rapid quantification of cellular proliferation and migration using ImageJ. *BioTechniques*, *66*(2), 99–102.

Verma, M. (2012). Personalized Medicine and Cancer. *Journal Of Personalized Medicine*, *2*(1), 1-14.

Wang, Y., Zhang, Y., Pan, C., Ma, F., & Zhang, S. (2015). Prediction of poor prognosis in breast cancer patients based on microRNA-21 expression: a meta-analysis. *PloS ONE*, *10*(2), Article#e0118647.

World Health Organization. (2022a, February 3). *Cancer*. Retrieved on 20 April 2021 from https://www.who.int/news-room/fact-sheets/detail/cancer/

World Health Organization. (2022b, June 30). *The Global Breast Cancer Initiative (GBCI)*. Retrieved on 23 April 2021 from https://www.who.int/publications/m/item/the-global-breast-cancer-initiative-gbci

Wu, J. M., Flynn, J. F., & Wong, C. (2009). Anti-EGFR therapy: Mechanism and advances in clinical efficacy in breast cancer. *Journal of Oncology*, *2009*(i).

Xu, X., Zhang, M., Xu, F., & Jiang, S. (2020). Wnt signaling in breast cancer: biological mechanisms, challenges and opportunities. *Molecular Cancer*, *19*(1), 1–35.

Yamashita, T., Budhu, A., Forgues, M., & Xin, W. W. (2007). Activation of hepatic stem cell marker EpCAM by Wnt-β-catenin signaling in hepatocellular carcinoma. *Cancer Research*, *67*(22), 10831–10839.

Yip, C. H., Pathy, N. B., & Teo, S. H. (2014). A review of breast cancer research in Malaysia. *Medical Journal of Malaysia*, *69*(August), 8–22.

Zhan, T., Rindtorff, N., Betge, J., Ebert, M. P., & Boutros, M. (2019). CRISPR/Cas9 for cancer research and therapy. *Seminars in Cancer Biology*, *55*(April 2018), 106–119.

Zhou, F. Q., Qi, Y. M., Xu, H., Wang, Q. Y., Gao, X. S., & Guo, H. G. (2015). Expression of EpCAM and wnt/ β-catenin in human colon cancer. *Genetics and Molecular Research*, *14*(2), 4485–4494.