CRISPR-CAS9 HSF4-TARGETED FOR CLEAR CELL RENAL CELL CARCINOMA PROLIFERATION CAPACITY, COLONY AND TUMORSPHERES FORMATION

MEI JUAN WONG

FACULTY OF SCIENCE
UNIVERSITI MALAYA
KUALA LUMPUR

2022

CRISPR-CAS9 HSF4-TARGETED FOR CLEAR CELL RENAL CELL CARCINOMA PROLIFERATION CAPACITY, COLONY AND TUMORSPHERES FORMATION

MEI JUAN WONG

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

INSTITUTE OF BIOLOGICAL SCIENCES
FACULTY OF SCIENCE
UNIVERSITI MALAYA
KUALA LUMPUR

2022

UNIVERSITI MALAYA
ORIGINAL LITERARY WORK DECLARATION

Name of Candidate: **MEI JUAN WONG**

Registration/Matric No: **S2020977/1**

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

Title of Dissertation ("this Work"):

CRISPR-CAS9 HSF4-TARGETED FOR CLEAR CELL RENAL CELL CARCINOMA PROLIFERATION CAPACITY, COLONY AND TUMORSPHERES FORMATION

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 right 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: 5 SEPTEMBER 2022

Subscribed and solemnly declared before,

Witness's Signature Date: 5 SEPTEMBER 2022

Name:
Designation:

CRISPR-CAS9 HSF4-TARGETED FOR CLEAR CELL RENAL CELL CARCINOMA PROLIFERATION CAPACITY, COLONY AND TUMORSPHERES FORMATION

ABSTRACT

Clear cell renal cell carcinoma (ccRCC) is the most prevalent form of kidney cancer, accounts for about 75% of all kidney cancer cases. Although anti-angiogenic drugs targeting the receptor tyrosine kinases and mTOR inhibitors were clinically approved for ccRCC, but the current available drugs are still having low efficacy and most of the patients can rapidly develop therapeutic resistance. Prior analysis of TCGA RNA-Seq data on the downstream analysis of KLF6-driven transcriptional programs in ccRCC revealed that HSF4 is among the genes those are significantly upregulated in ccRCC as compared to normal kidney tissues and high expression of HSF4 in ccRCC patients is associated with poor prognosis. To interrogate the HSF4 functional relevance in supporting kidney cancer cells growth and cellular fitness, we targeted HSF4 gene in ccRCC cell line (786-M1A) by using CRISPR-Cas9 (CRISPRko) system and carried out *in-vitro* functional assays to determine the phenotypic effects of targeting HSF4 in ccRCC cells. A point mutation of insertion was introduced into the HSF4-targeted 786-M1A cells and cause frameshift mutation and reduction of HSF4 protein synthesis. The *in-vitro* functional assays showed to decrease cell activities in terms of proliferation rate, colony formation, and ability to form tumorspheres in HSF4-targeted 786-M1A cells as compared with the non-targeted 786-M1A cells. The results indicated the HSF4 gene displayed functional role in maintain 786-M1A cells growth and cellular fitness. The further analysis will be focused on the analysis of the downstream pathways that are regulated by HSF4 gene. These findings are believed to be able to contribute to the foundation of the understanding on the molecular drivers in ccRCC pathogenesis and the development of effective therapeutic strategy for ccRCC patients.

Keywords: Kidney cancer, ccRCC, HSF4, CRISPR-Cas9.

CRISPR-CAS9 HSF4-SASARAN UNTUK KAPASITI PEMBIAKAN, FORMASI KOLONI DAN SFERA TUMOR KARSINOMA SEL RENAL SEL JELAS

ABSTRAK

Karsinoma sel renal sel jelas (ccRCC) ialah bentuk kanser buah pinggang yang paling lazim, menyumbang kira-kira 75% daripada semua kes kanser buah pinggang. Walaupun ubat anti-angiogenik yang menyasarkan reseptor tyrosine kinase dan perencat mTOR telah diluluskan secara klinikal untuk ccRCC, tetapi ubat-ubatan yang ada sekarang masih mempunyai keberkesanan yang rendah dan kebanyakan pesakit boleh menunjukkan rintangan terapeutik dengan cepat. Analisis terdahulu data TCGA RNA-Seq mengenai analisis hiliran program transkrip dipacu KLF6 dalam ccRCC telah mendedahkan bahawa HSF4 adalah antara gen yang diekspresi tinggi dalam ccRCC berbanding dengan tisu buah pinggang biasa dan ekspresi tinggi HSF4 dalam pesakit ccRCC dikaitkan dengan prognosis yang buruk. Untuk menyoal siasat kaitan fungsi HSF4 dalam menyokong pertumbuhan sel kanser buah pinggang dan kecergasan selular, kami menyasarkan gen HSF4 dalam jenis sel ccRCC (786-M1A) dengan menggunakan sistem CRISPR-Cas9 (CRISPRko) dan menjalankan ujian fungsi *in-vitro* untuk menentukan fenotip bawah kesan penyasaran HSF4 dalam sel ccRCC. Mutasi titik tunggal sisipan yang diperkenalkan ke dalam sel 786-M1A yang disasarkan HSF4 telah menyebabkan mutasi anjakan bingkai dan pengurangan sintesis protein HSF4. Ujian fungsi *in-vitro* telah menunjukkan penurunan aktiviti sel dari segi kadar percambahan, pembentukan koloni, dan keupayaan untuk membentuk tumorspheres dalam sel 786-M1A yang disasarkan HSF4 berbanding dengan sel 786-M1A yang tidak disasarkan. Keputusan menunjukkan gen HSF4 mempunyai peranan berfungsi dalam mengekalkan pertumbuhan sel 786-M1A dan kecergasan selular. Analisis selanjutnya akan ditumpukan kepada analisis laluan hiliran yang dikawal oleh gen HSF4. Penemuan ini dipercayai dapat menyumbang

kepada asas pemahaman tentang pemacu molekul dalam patogenesis ccRCC dan pembangunan strategi terapeutik yang berkesan untuk pesakit ccRCC.

Kata Kunci: Kanser buah pinggang, ccRCC, HSF4, CRISPR-Cas9.

ACKNOWLEDGEMENT

I cannot express enough thanks to my supervisors for their continued guidance and encouragement: Dr. Nikman and Dr. Effendi. I offer my sincere appreciation for the learning opportunities provided by my respectful supervisors. Their detailed and constructive comments as well as great patience have been the keys to the completion of this research.

My completion of this research project could not have been accomplished without the support from the lab members, Asmaa, Nadia, and Farhan. To Dr. Aiman, Hanani, and Afiqah, thank you for the sharing and discussion about the experimental procedures and administration information. My gratitude also goes to the administrative and teaching staff of Universiti Malaya (UM) and UKM Medical Molecular Biology Institute (UMBI). It is great to be the part of this wonderful group.

I would like to thank my family members who sponsored me financial supports, and comfortable accommodations throughout my Master Degree study. Finally, I would also like to express my special thanks to my loving and supportive friends, especially Wei Jie, Xun Chan, Brian, Vivian, and So Kian for them accompanies and mentally support, bringing me a joyful and stress less life along my study.

Ultimately, I would like to acknowledge the National University of Malaysia fro funding via Young Lecturers Incentive Grant (GGPM) [UKM-SPKP-CRIM-PK01-GP06].

TABLE OF CONTENT

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS AND ABBREVIATIONS

~	:	approximately
%	:	percentage
bp	:	base pair
U	:	unit
mL	:	millilitre
µ	:	micro
℃	:	degree celsius
min	:	minute
rpm	:	rotation per minute
M	:	Molarity
g	:	gram
sec	:	second
v	:	volt
h	:	hour
CRISPR-Cas9	:	clustered regularly interspaced short palindromic repeats and CRISPR-associated protein 9
HSF	:	heat shock factor
ccRCC	:	clear cell renal cell carcinoma
mTOR	:	mammalian target of rapamycin
TCGA	:	The Cancer Genome Atlas
CRISPRko	:	CRISPR-Cas9 or CRISPR knockout
VHL	:	*von-Hippel-Lindau*
HIF2α	:	hypoxia-inducible factor 2 alpha
ZFN	:	zinc finger nuclease
TALEN	:	transcription-activator-like effector nuclease
dCas9	:	Cas9 Endonuclease Dead
KLF6	:	Kruppel-like factors

mTORC1	:	mammalian target of rapamycin complex 1
KIRC	:	kidney renal clear cell carcinoma
PHD	:	plant homeodomain
HRE	:	HIF-responsive element
VEGFA	:	vascular endothelial growth factor A
EGFR	:	estimated glomerular filtration rate
CXCR4	:	C-X-C chemokine receptor type 4
GLUT1	:	glucose transporter 1
EPO	:	erythropoietin
DNA	:	deoxyribonucleic acid
VEGFR	:	vascular endothelial growth factor receptor
FDA	:	Food and Drug Administration
UMPP	:	ubiquitin-mediated proteolysis pathway
PBRM1	:	protein polybromo-1
PBAF SWI/SNF	:	Polybromo BRG1 (Brahma Related Gene 1) Associated Factor SWItch/Sucrose Non-Fermentable
BAP1	:	BRCA1 (Breast Cancer gene 1) associated protein 1
SETD2	:	SET Domain Containing 1A, Histone Lysine Methyltransferase
JARID1C	:	[Histone H3]-trimethyl-L-lysine(4) demethylase
UTX	:	ubiquitously transcribed tetratricopeptide repeat on chromosome X
DBD	:	DNA binding domain
HR-A/B	:	leucine zipper oligomerization domain
RD	:	regulatory domain
TAD	:	transactivation domain
HR-C	:	oligomerization repressor site
PTM	:	post-translational modifications

PDSM	:	phosphorylation-dependent sumoylation motif
CNS	:	central nervous system
MAPK	:	mitogen-activated protein kinase
RNA	:	ribonucleic acid
E. coli	:	*Escherichia coli*
crRNA	:	CRISPR RNA
tracrRNA	:	trans-activating crRNA
PAM	:	palindromic repeat
sgRNA	:	single guide RNA
gRNA	:	guide RNA
DSB	:	double-strand break
NHEJ	:	non-homologous end joining
HDR	:	homology directed repair
CRISPRa	:	CRISPR activation
CRISPRi	:	CRISPR interference
STR	:	short tandem repeat
RPMI	:	Roswell Park Memorial Institute Medium
FBS	:	fetal bovine serum
DMEM	:	Dulbecco's Modified Eagle Medium
SDS-PAGE	:	sodium dodecyl-sulfate polyacrylamide gel electrophoresis
TBS-T	:	tris-buffered saline-tween
LB media	:	Luria Broth
786-M1A LB	:	786-M1A lentiCas9-blasticidin
mRNA	:	messenger ribonucleic acid
2D	:	two-dimensional
3D	:	three-demensional

LIST OF APPENDIXES

CHAPTER 1: INTRODUCTION

1.1 Background of Study

Clear cell renal cell carcinoma (ccRCC) is the most prevalent form of kidney cancer, accounts for ~75% of all kidney cancer cases (Hsieh et al., 2017). The key event in ccRCC pathogenesis is bi-allelic inactivation of the *von-Hippel-Lindau* (VHL) gene and subsequent accumulation of hypoxia-inducible factor 2 alpha (HIF2), which collectively account for 90% of sporadic ccRCC cases (Turajlic et al., 2018). The aberrant activation of hypoxic-response transcriptional programs leads to the upregulation of angiogenesis and glucose metabolism rewiring that would favour the ccRCC growth (Kaelin, 2017). Although the roles of VHL-HIF2α axis in ccRCC initiation have been well established, other molecular determinants that drive ccRCC progression remain elusive. Recent multi-region high-throughput sequencing efforts in ccRCC revealed significant intra-tumoral heterogeneity, with the ubiquitous VHL mutations projected to represent the truncal event in ccRCC pathogenesis, while the majority of the predicted driver mutations are subclones (Gerlinger et al., 2014; Turajlic et al., 2018; Turajlic et al., 2018). These widespread heterogeneities along with lack of complete understanding on ccRCC pathogenesis have hampered the development of efficient therapeutic strategy for ccRCC.

Disease recurrence has often been seen in ccRCC patients who had undergone the curative surgery and about one-third of ccRCC patients have already had metastatic disease upon diagnosis (Dabestani et al., 2016; Hsieh et al., 2017). Anti-angiogenic drugs targeting the receptor tyrosine kinases and mTOR inhibitors have been clinically approved for ccRCC. However, these treatment options are unsatisfactory whereby significant fractions of patients rapidly develop therapeutic resistance (Rini & Atkins, 2009). Combinatorial treatment with anti-angiogenic and mTOR inhibitors have also been tested, however increased adverse effects with no added efficacy were observed in

majority of patients (Choueiri & Motzer, 2017). Small molecule inhibitors targeting HIF2a (W. Chen et al., 2016; Cho et al., 2016) as well as the immune checkpoint inhibitors (Mazza et al., 2017) have also been developed. Despite the promising results in some ccRCC patients, low efficacy and rapid development of resistance are still the major concerns. As a result, a greater knowledge of the molecular requirements of VHL mutant ccRCC is required to improve ccRCC therapy options.

Cancer progression are complex evolutionary process, facilitated by accumulation of various cellular alterations and rewiring that confer for selective growth advantage and acquisition of aggressive traits (Vanharanta & Massagué, 2013). The transformation of normal cells into malignant, neoplastic cells is accompanied by the acquisition of several important traits which are described as the hallmarks of cancer (Hanahan & Weinberg, 2011). Cancer cells could also develop high dependencies towards the molecular drivers that support cancer growth and progression as well as adaptation to various tumorigenesis-associated cellular stresses (Luo et al., 2009). The understanding of the molecular pathways and mechanisms hence become the most fundamental cancer research concern.

There are various of genetic tools such as zinc finger nucleases (ZFNs) or transcription-activator-like effector nucleases (TALENs) have been deployed to examine the molecular details of cancer. Compared to these two popular tools, the recent developed genetic modification tool, CRISPR-Cas9 is more rapid and simple to insert the target gene and desired gene modification such as activation, repress, purify, or multiple gene screening. Therefore, CRISPR is quickly arise as a handy genetic modification tool. The CRISPR-Cas9 was originally identified as the adaptive immune system in prokaryotes to protect against invading bacteriophage (Barrangou & Marraffini, 2014). Since its discovery, this CRISPR-Cas9 system has been successfully repurposed for highly efficient gene knockout in eukaryotic cells (Lander, 2016). There are wide variety

of CRISPR-Cas9/dCas9 applications developed to date that include gene expression modulations, epigenetic modifications, precise base editing, chromatin imaging and genetic screens (Adli, 2018).

CRISPR genetic screens have been widely utilized to uncover essential genes, molecular drivers or druggable targets in various cancer types (Behan et al., 2019; Dempster et al., 2019; Hart et al., 2015; Tzelepis et al., 2016). This approach has been proven to be a powerful and robust tool that allows systematic, large scale functional interrogation of genes even at the genome-wide level. The previous study has performed by Syafruddin and and colleagues at the University of Cambridge *in-vitro* and *in-vivo* CRISPR-Cas9 genetic screen in ccRCC cells and identified several of candidate hits. One of the candidate hits, Kruppel-like factors (KLF6), has been functionally validated where they found that null KLF6 ccRCC cells manifested impaired growth *in-vitro* and *in vivo* as well as having significantly reduced capability for lung colonisation (Saiful E. Syafruddin et al., 2019). A comprehensive downstream analysis revealed that KLF6 modulate the lipid homeostasis and mTORC1 signalling pathways, two important pathways implicated in ccRCC pathogenesis (Saiful E. Syafruddin et al., 2019).

Prior analysis of ccRCC TCGA RNA-Seq data has revealed that HSF4 gene is one of the genes that are significantly upregulated in ccRCC as compared to normal kidney tissues and high expression of HSF4 in ccRCC patients is associated with poor prognosis (unpublished data). Motivated by this finding, in this proposed study we embark on studying the phenotypic effects of heat shock factor 4 (HSF4) in kidney cancer cells. Functionally, HSF4 inactivation has been shown to reduce cells growth, induce cellular senescence and suppress the development of lymphoma in the p53/Arf-deficient mice (Jin et al., 2012). However, the roles of HSF4 in ccRCC pathogenesis remains to be elucidated. The upregulation of HSF4 and its association with poor prognosis in ccRCC

may indicate that HSF4 could play important roles in supporting pathogenesis of this kidney cancer subtype.

1.2 Problem Statement

ccRCC is the most majority form of kidney cancer, diagnosed annually in ~400,000 individuals with reported deaths of ~144,000 worldwide. Bi-allelic inactivation of *von-Hippel-Lindau* (*VHL*) tumor suppressor gene and subsequent accumulation of the pro-oncogenic hypoxia-inducible factor 2 alpha (HIF2α) is the gatekeeper event in ccRCC pathogenesis, contributing to ~90% of sporadic ccRCC cases. Angiogenesis and mTOR signalling pathway inhibitors are the clinically approved treatments for ccRCC. Nevertheless, the overall patients' response rates are still unsatisfactory whereby rapid acquisition of therapeutics resistance and disease relapse pose further challenges in treating this disease efficiently. Despite ccRCC is not among the commonly diagnosed cancers worldwide, this cancer type especially the advanced stage ccRCC usually irresponsive towards treatments and is highly lethal. Unfortunately, a significant fraction of ccRCC patients already have the disease progressed to the late stages at the time of diagnosis. Therefore, there is a pressing need for efficient diagnostic and therapeutic strategies for ccRCC. However, limited understanding on ccRCC pathogenesis have thus far been hampering the development of these much-needed novel strategies. The underlying mechanisms that support ccRCC progression is not completely understood despite the establishment of VHL-HIF2α axis as ccRCC pathogenesis gatekeeper and identification of several ccRCC molecular drivers from recent high-throughput sequencing efforts. Cancer cells regularly undergo cellular and metabolic rewiring and would become highly addicted to certain genes and/or pathways that confer for selective growth advantage. These are also to maintain the cancer phenotypes and adaptation to various cellular stresses. Prior to this research, my co-supervisor and his team at

University of Cambridge performed both *in-vitro* and *in-vivo* CRISPR-Cas9 genetic screening in ccRCC cells in order to discover novel molecular drivers and pathways in which ccRCC cells are highly dependent on. Among the identified candidate hits, several of the hits are yet to be functionally validated in the previous study.

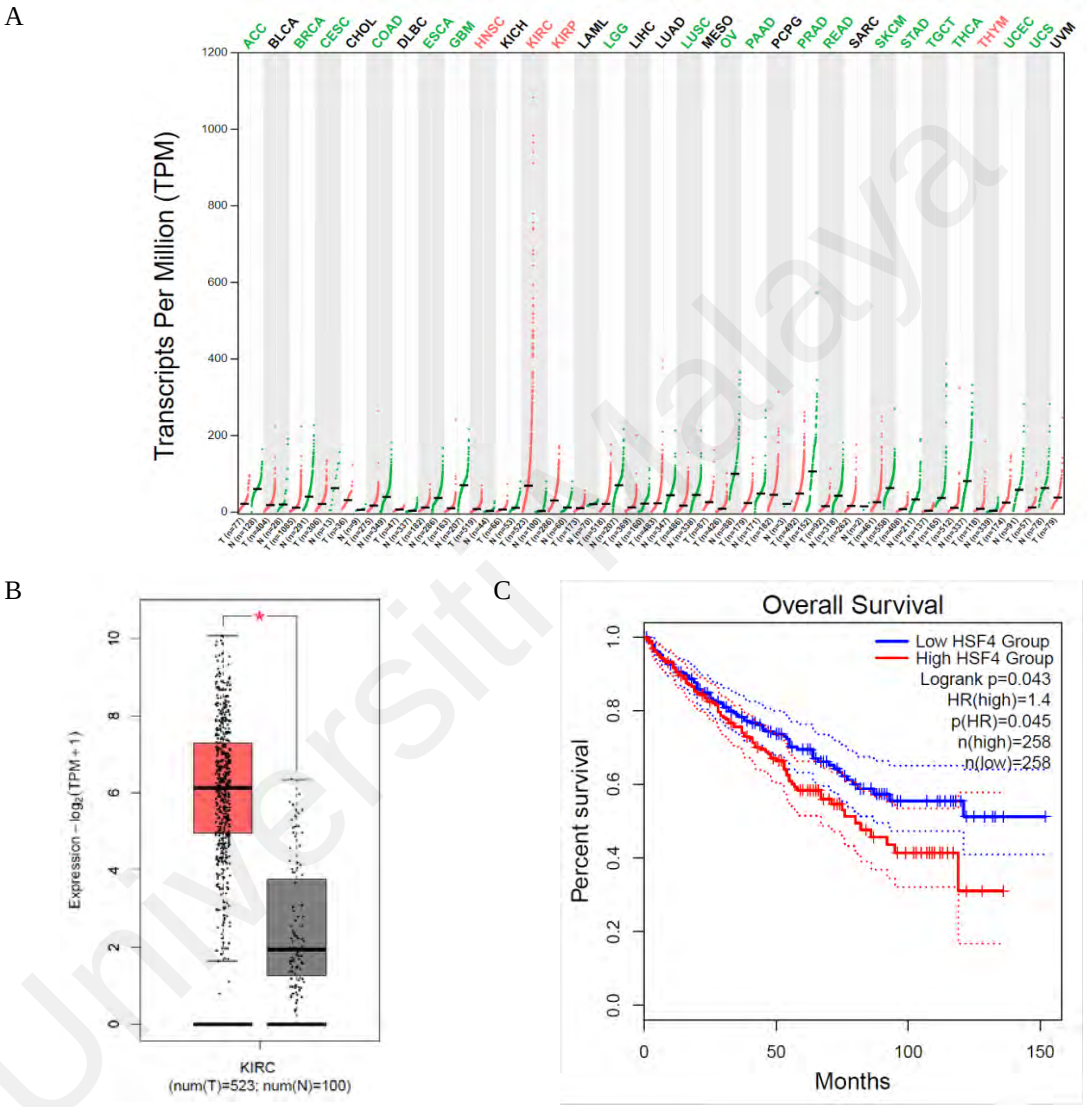


Figure 1.1: HSF4 is a novel candidate driver supporting ccRCC pathogenesis. A) HSF4 gene expression profile across different tumor samples and their respective normal tissues. Red and green labels indicate upregulation and downregulation of HSF4 expression in tumor samples, respectively, in comparison to their respective normal counterparts. B) HSF4 is highly expressed in ccRCC (red boxplot) as compared to the normal kidney tissues (grey boxplot). C) High expression of HSF4 is significantly associated with poor overall survival in kidney cancer patients. (Queried and extracted from GEPIA2 (Gene Expression Profiling Interactive Analysis) web application (http://gepia2.cancer-pku.cn/))

Therefore, in this proposed study, we are specifically interested in investigating the roles of candidate hits HSF4 in ccRCC pathogenesis. We believe that HSF4 functions are important in supporting ccRCC progression because sgRNAs targeting HSF4 were among the top significantly depleted sgRNAs in the CRISPR-Cas9 genetic screening (unpublished data). Also, HSF4 expression is found to be upregulated in ccRCC clinical samples and patients with high expression of HSF4 had worse prognosis (Figure 1.1), further signifying HSF4 potential roles in driving ccRCC pathogenesis. Moreover, to our knowledge, HSF4 roles in ccRCC are yet to be reported in the previous studies.

1.3 Research Questions

1. Would the CRISPR-Cas9 system be able to target HSF4 in ccRCC cells effectively?
2. What are the phenotypic effects of targeting HSF4 in ccRCC cells using the CRISPR-Cas9 system?

1.4 Main Aim and Objectives

1.4.1 Main Aim

- To determine the phenotypic effects of targeting HSF4 in ccRCC cells by performing *in-vitro* functional assays.

1.4.2 Specific objectives

1. To design and construct CRISPR-Cas9 system targeting HSF4 in ccRCC cells
2. To deliver the CRISPR-Cas9 system into the ccRCC cells via lentiviral transduction approach and select positively transduced cells
3. To assess the HSF4 targeting efficiency in ccRCC cells using Western blotting and Sanger sequencing
4. To determine the preliminary effects of CRISPR-Cas9 HSF4-targeted for ccRCC cells' proliferation capacity, colony and tumuorspheres formation.

1.5 Hypothesis

1. Targeting HSF4 using CRISPR-Cas9 system would completely knockout or reduce the expression of HSF4 expression in ccRCC cells.
2. CRISPR-Cas9-mediated HSF4 targeting impairs ccRCC cells growth and the formation of tumorsphere *in-vitro*.

CHAPTER 2: LITERATURE REVIEW

2.1 Clear Cell Renal Cell Carcinoma (ccRCC)

Kidney cancer is among the top ten cancers in the world. This research emphasises on the clear cell RCC subtype, also known as clear cell renal cell carcinoma (ccRCC), which is the most common kind of kidney cancer and the cause of the majority of kidney cancer-related deaths. The ccRCC is the most frequent RCC subtype, accounting for 75 % of all kidney cancer cases. Papillary renal cell carcinoma (pRCC) and chromophobe renal cell carcinoma (crRCC) are two more RCC subtypes that account for around 10 % and 5 % of all kidney cancer cases respectively (Rini et al., 2009). The classification of RCC subtypes continuously updating but overall, these three subtypes account for more than 90% of all RCCs (Athanazio et al., 2021; Cai et al., 2020; Lopez-Beltran et al., 2009). Renal cell carcinoma (RCC) is a heterogeneous group of cancer that arise from the lining of tubules in the kidney (Figure 2.1) (PDQ Adult Treatment Editorial Board, 2002).

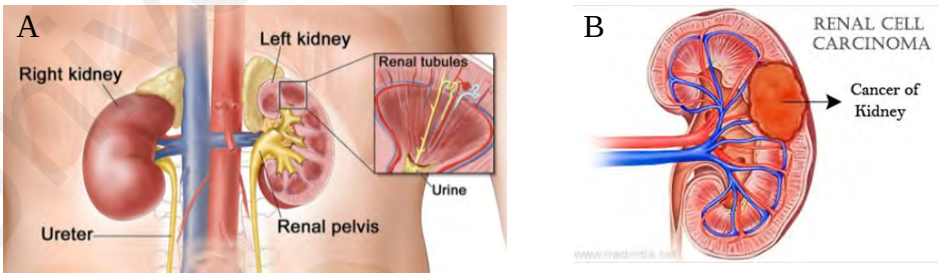


Figure 2.1: A) Structure of kidney which showing the position of renal tubules in the kidney. B) renal cell carcinoma in kidney. (Photo sourced from PDQ Adult Treatment Editorial Board, 2002; Reeja, 2013)

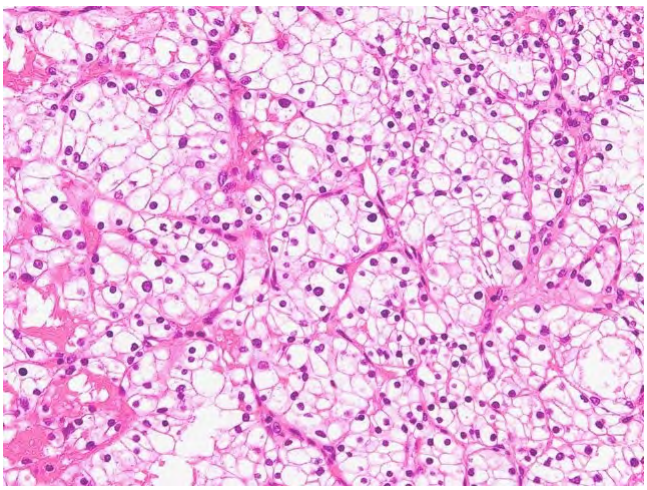


Figure 2.2: High magnification micrograph of ccRCC (H&E stain, ×200), showing cells with clear cytoplasm due to the accumulation of cytoplasmic and lipid and glycogen. (Photo sourced from Sakai et al., 2010)

The ccRCC develops when the number of cells in the kidney grows rapidly, producing in a lump (mass). The term "clear cell" refers to clear, glass-like cytoplasm due to the accumulation of cytoplasmic lipid and glycogen in ccRCC cells (Figure 2.2). However, the cause of the accumulation of these lipid and glycogen remains to elusive, even though metabolic rewiring and how the ccRCC cells consume and metabolize glucose, lipid and other nutrients have been proposed to contribute to this clear cells histology (Gebhard et al., 1987; Sundelin et al., 2012).

According to previous reports, the ccRCC tumorigenesis is initiated by the VHL-HIF alpha axis. Hypoxic conditions in the tumor site eventually promote cancer cell proliferation and maintenance. During normoxia, HIFα is hydroxylated by PHD proteins via an oxygen-dependent enzymatic mechanism. The VHL complex binds to the hydroxylated HIFα and leading to degradation of HIFα. However, in hypoxia, hydroxylation step by PHDs does not happen, resulting in HIFα stabilization. HIFα will then accumulates and form dimers with its constitutively active partner HIF-β. The HIF complex translocates to the nucleus and binds to HIF-responsive element (HRE), leading to the transcription of hypoxia-associate genes, and activation of various downstream pathways, mainly to adapt to the hypoxic condition. These include angiogenesis (e.g.,

VEGFA), proliferation (e.g., EGFR), cell migration and invasion (e.g., CXCR4), metabolic shift towards glycolysis (e.g., GLUT1), survival (e.g., survivin), erythropoiesis (e.g., EPO) that will ultimately contribute to ccRCC tumorigenesis. The HIF pathway also interacts and contribute to the genome-wide epigenetic alterations, which alter the apoptotic, cell cycle regulatory, immune evasion and DNA mismatch repair pathways. All these interactions and mechanisms play important roles in the formation and progression of ccRCC (Figure 2.3).

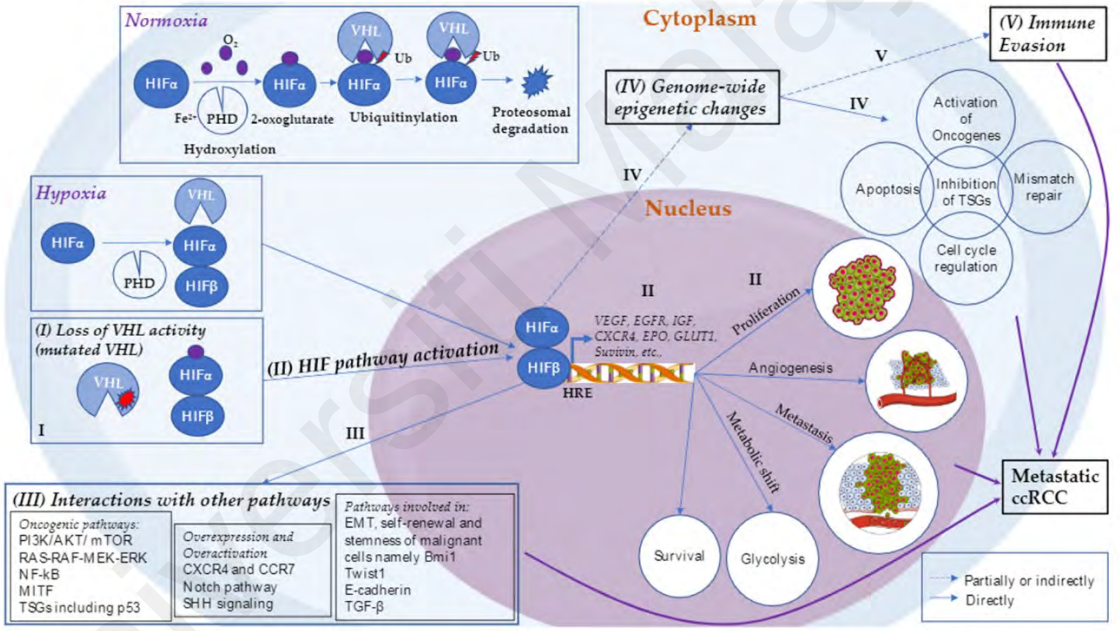


Figure 2.3: Physiology of the VHL-HIF pathway along with proposed etiopathogenesis that leading to metastatic ccRCC. During normoxia, the VHL complex binds to the hydroxylated HIF-1a and leading to degradation of HIF-1a. However, in hypoxia, the inactivation of the von-Hippel-Lindau (VHL) gene and subsequent hypoxia-inducible factor alpha (HIFα) accumulation lead to upregulation of cell growth control genes (II), triger of oncogenic and other pathways (III), genome wide changes (IV) and immune invasion (V). All these connections have roles in ccRCC tumorigenesis. (Photo sourced from Mehdi & Riazalhosseini, 2017)

There are also variety of treatments available for ccRCC patients, included surgery, radiation therapy, chemotherapy, immunotherapy and targeted therapy. However, the conventional treatments are normally toxic to normal cells, often less effective when dealing with recurrent ccRCC and high post treatment mortality chance. On top of that, tumor heterogeneity can account for a wide range of clinical outcomes in RCC, options to enhance clinical outcomes based on individual tumor characteristics (so-called precision medicine) hence becomes a growing paradigm. Therefore, data import to ccRCC molecular characterization will enhance the development of precision medicine.

Advanced renal cancer is primarily treated with antiangiogenic drugs and mTOR inhibitors. Antiangiogenic drugs such as vascular endothelial growth factor receptor (VEGFR) prevents new blood vessels from developing in the tumor, causing it to be cut off from receiving enough supply of oxygen and nutrients, hence will lead the tumors cells to stop proliferating and die (PDQ Adult Treatment Editorial Board, 2002). The medications such as sunitinib, temsirolimus, bevacizumab, pazopanib, sorafenib, nivolumab, cabozantinib and lenvatinib have been approved by the Food and Drug Administration (FDA) for the treatment of metastatic ccRCC (Hsieh et al., 2017).

Although the specific aetiology of ccRCC is unknown, smoking, overuse of certain drugs, and a number of genetic susceptibility factors may all play a role in its development. In most cases of ccRCC, the malfunction of the tumor suppressor gene *von Hippel–Lindau* (VHL) causes constitutive activation of hypoxia-inducible factor HIFα, production of hypoxia-induced genes in normoxic conditions (Creighton et al., 2013; Sundelin et al., 2012). The alterations in ubiquitin-mediated proteolysis pathway (UMPP); *PBRM1*, a subunit of the PBAF SWI/SNF chromatin remodelling complex; histone deubiquitinase *BAP1*; histone methyltransferase *SETD2*; histone demethylase JARID1C and UTX were also discovered in ccRCC, suggesting that epigenetic regulation of additional functional pathways involved in the disease's pathogenesis and progression

(Dalgliesh et al., 2010; Guo et al., 2012; Ricketts et al., 2018; Varela et al., 2011). Thus, oncogenic metabolism and epigenetic reprogramming have emerged as key characteristics of ccRCC.

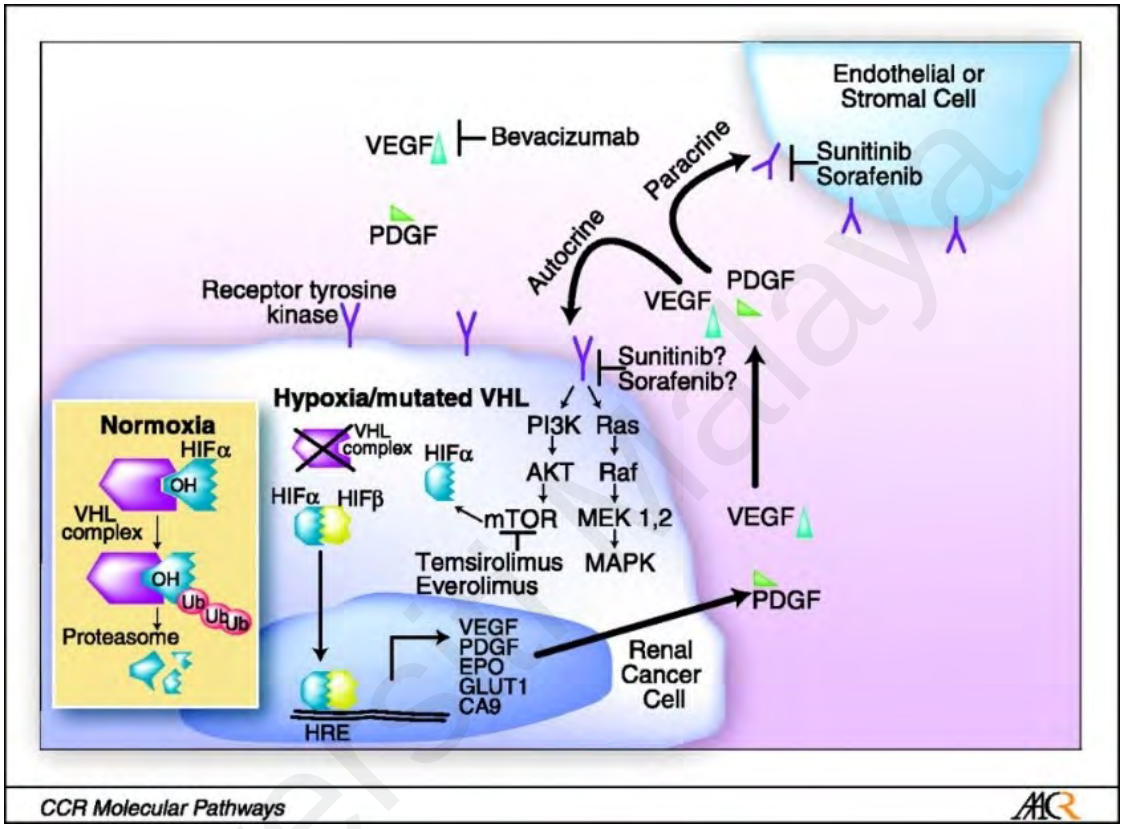


Figure 2.4: Putative paracrine and autocrine effects of VEGF and PDGF within RCC and targets of clinically effective pharmacologic agents. The signal pathways in RCC can be inhibited at several steps: inhibition of VEGF (by bevacizumab), inhibition of tyrosine kinase activity of RTK (by sunitinib and sorafenib), and inhibition of mTOR (by temsirolimus and everolimus). (Photo sourced from Patel et al., 2006)

2.2 Heat Shock Factor 4 (HSF4)

Initially, the heat shock factors (HSFs) were thought to be a biological programme that was initiated in response to a rise in temperature. HSFs are also responsive to other stimuli, such as oxidants, and heavy metals, according to later discoveries (Morimoto, 1993). They play a wide range of roles in stress resistance, including protecting against protein misfolding, inflammation, and environmental threats (Gomez-Pastor et al., 2017). HSF1, HSF2, HSF3, HSF4, HSF5, HSFX, and HSFY are the seven members of the HSF family that have been identified thus far. Except for HSF3, all of these HSFs are encoded in the human DNA. HSFs play a role in a variety of physiological processes and diseases by direct regulation of their target genes as shown in Figure 2.5.

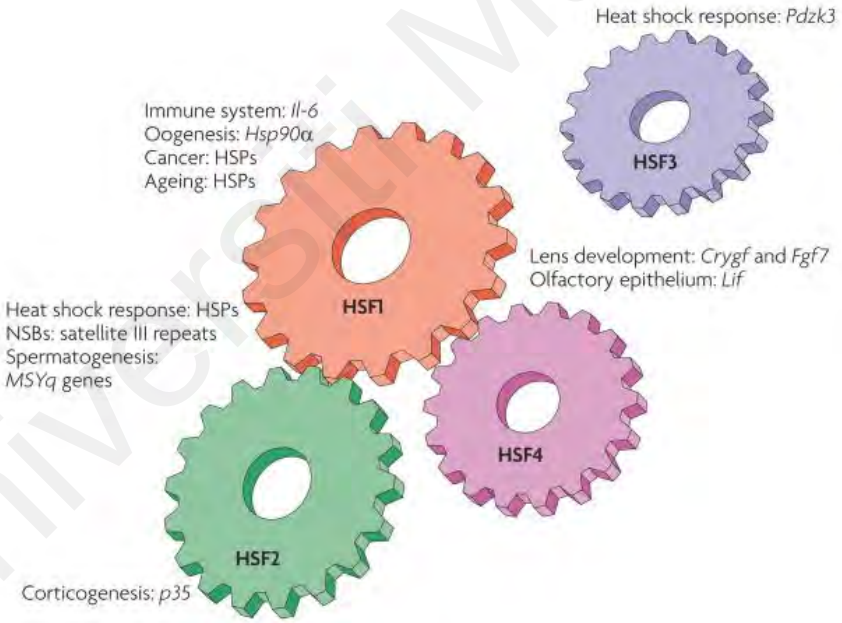


Figure 2.5: The mammalian HSF machinery. An overview of HSFs family member and their contribution to various normal physiological processes and pathologies through direct regulation of their target genes. HSF1 is the most intensively studied in the HSF family that plays role in heat shock response, oogenesis, immune system, cancer and ageing; HSF2 works closely with HSF1 in heat shock response and developmental process such as spermatogenesis, oogenesis and corticogenesis; HSF4 also cooperate with HSF1 in development of several sensory organs but has no role in heat shock response; HSF3 is recent discovered HSF family member which is not known to has related functional role with other HSF family members. (Photo sourced from Åkerfelt et al., 2010)

In the absence of stress stimuli, HSF1 is expressed constitutively in most tissues and cell types, but it remains inactive. Increased temperature and other cellular stressors can cause proteins to misfold. Heat shock proteins bind to misfolded proteins, causing HSF-1 to dissociate. HSF1 is able to form trimers, translocate to the cell nucleus, and trigger transcription as a result of this (Prahlad & Morimoto, 2009). HSF2 is known to influence HSF1-mediated heat shock protein (HSP) gene expression by forming heterocomplexes. HSF1 and HSF2 have been linked to a variety of developmental processes, including oogenesis, spermatogenesis, and corticogenesis. In collaboration with HSF1, HSF4 has a function in the development of several sensory organs especially but it has no effect on the heat shock response (Åkerfelt et al., 2010).

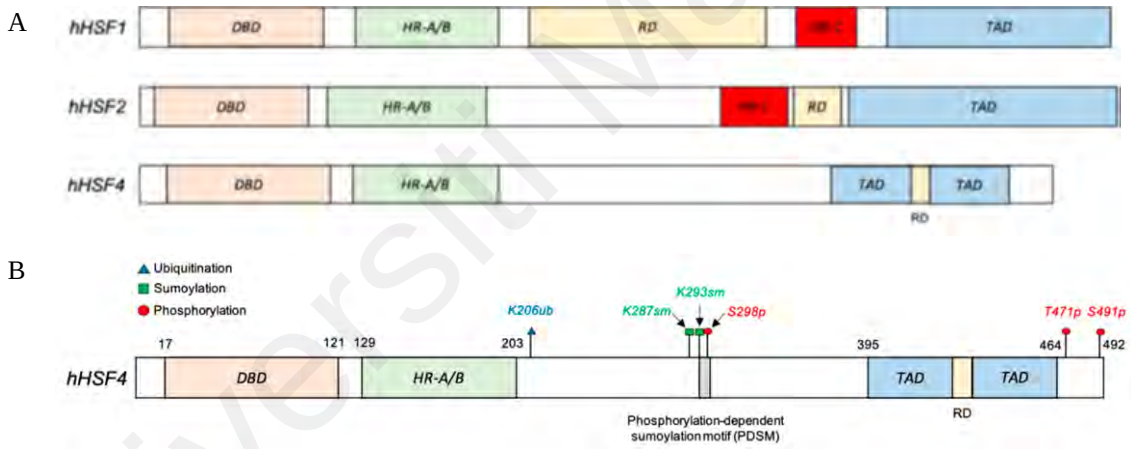


Figure 2.6: A) Comparison of human HSF1, HSF2, and HSF4 protein structure. A highly conserved DNA binding domain (DBD), leucine zipper oligomerization domain (HR-A/B), regulatory domain (RD), and transactivation domain at C-terminus (TAD) are found in all human HSF members. HSF4 lacks the oligomerization repressor domain (HR-C), unlike HSF1 and HSF2, which make HSF4 constitutively active. B) HSF4 protein post-translational modifications (PTMs). HSF4 contains two TADs at the C-terminus, which lead to diverse functions beyond the regulation of heat or stress response transcriptional programs. The post-translational modifications (PTMs) found in HSF4 included three phosphorylation sites, two sumoylation sites, one ubiquitylation site. HSF4 also possess a phosphorylation-dependent sumoylation motif (PDSM) which reduces the expression of the crystallin gene that is participating in the normal lens development and homeostasis. (Photo sourced from Saiful Effendi Syafruddin et al., 2021)

The protein structure difference between HSF4 and other HSFs allow it to stay constitutively active and be able to interact with various protein partners as well as thus has more regulatory functions in transcriptional activities compared to other HSFs (Saiful Effendi Syafruddin et al., 2021). The roles of HSF4 gene in lens development were extensively reported while its functional roles in regulating neuronal genesis in nasal cavity and central nervous system (CNS) development are yet to be elucidated. HSF4 is highly expressed in skeletal muscle, brain, eye and pancreas tissues, according to UniProt Database (https://www.uniprot.org/uniprot/Q9ULV5). Hence, the abnormal expression or mutations in HSF4 could lead to the disease pathogenesis especially cataracts. HSF4 also interacts with several factors of cell growth and maintenance such as p53, proteins that participate in MAPK and mTOR signalling pathways, and regulates HIFα expression which is a key factor of kidney tumorigenesis (R. Chen et al., 2011; Jin et al., 2012). Referring to the database of The Human Protein Atlas (https://www.proteinatlas.org/ENSG00000102878-HSF4/pathology), HSF4 is remarked as prognostic maker and its high expression is unfavourable in renal cell cancer (RCC) and colorectal cancer; the RNA expression from HSF4 gene is also reported most high expression in RCC as compared to other cancers. However, the HSF4 protein profile analysis in RCC is still pending. These database analyses have shown the correlation between the HSF4 gene and RCC prognosis. These connections are yet to be explored and could have high research value about kidney cancer pathogenesis.

2.3 Clustered regularly interspaced short palindromic repeats (CRISPR-Cas9 System)

CRISPR (Clustered Regularly Interspaced Short Palindromic Repeat) sequences were first discovered in the *E. coli* genome in 1987, but their role as a bacteriophage defence mechanism was not fully understood until 2007. Scientists theorised that prokaryotes used CRISPR as part of an adaptive immune system, using multiple CRISPR-associated (Cas) genes to not only capture invading phages but also to eliminate them upon re-exposure.

In endogenous Type II bacterial CRISPR system, a CRISPR array found in bacteria genome codes for many distinct protospacer sequences which are homologous to foreign DNA. Short palindromic repeat sequences separate these protospacers. At the adaption stage of system, specific Cas proteins cut foreign DNA into little fragments of around 20 bp in length (spacer-sized) and paste them into CRISPR arrays, which are contiguous sequences of DNA. During CRISPR RNA biogenesis, the targeted CRISPR loci is expressed into the form of pre-CRISPR RNA (pre-crRNA), which is then processed into individual CRISPR RNAs (crRNAs) by a special trans-activating crRNA (tracrRNA) with homology to the short palindromic repeat (PAM) and RNAse III and Cas9 enzymes. The tracrRNA and Cas9 nuclease form a complex with each individual, unique crRNA, becoming different crRNA:tracrRNA:Cas9 complex (gRNA:Cas9). Finally, at the interference stage, the gRNA:Cas9 complex is able to bind the DNA sequence complementary to the crRNA, Cas9 unwinds the double strand DNA and cleave the strands near to PAM sequence. The invader DNA sequence is destroyed and the gRNA complex is detached. The three stages' process of CRISPR-Cas9 system workflow is shown in Figure 2.7.

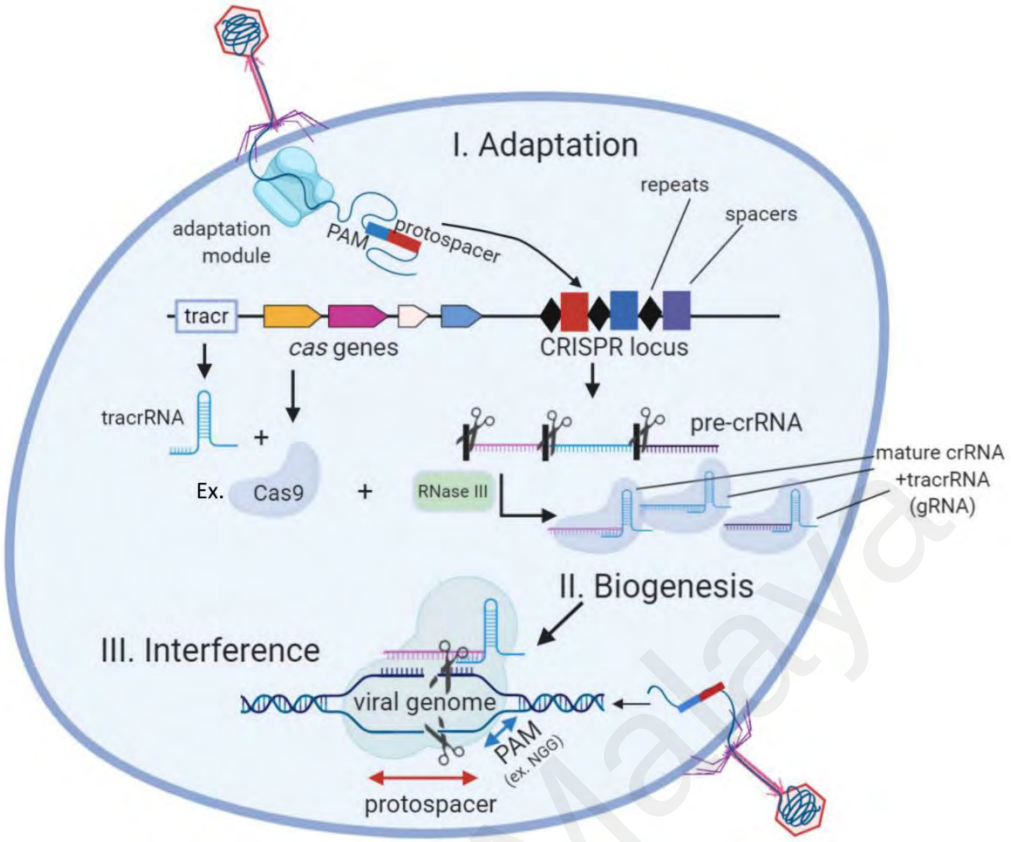


Figure 2.7: An overview of the endogenous Type II bacterial CRISPR system. There are 3 stages in this system: (I) Adaptation, activation of specific Cas protein to convert foreign DNA into spacer-sized fragments; (II) Biogenesis, generation of crRNA and gRNA; (III) Interference, binding of the Cas9:gRNA to the complementary viral genome and cleaving of viral genome. (Photo sourced from Arroyo-Olarte et al., 2021)

Jinek et al. were the first to show that CRISPR could be programmed for targeted DNA breakage in vitro in 2012. CRISPR-based genome editing in mammalian cell culture was described by Cong et al. and Mali et al. in 2013. Since then, more researches are published and contributed to the improvement and refinement of the tool's specificity, orthogonality, multiplex ability in various species, and the development of new applications. Genome engineering methods such as zinc finger nucleases (ZFNs) or transcription-activator-like effector nucleases (TALENs) required scientists to develop and generate a new nuclease pair for each genomic target prior to CRISPR. CRISPR has quickly become the most popular genome engineering method due of its relative ease and versatility.

A guide RNA (gRNA or single guide RNA (sgRNA)) and a CRISPR-associated endonuclease are two components of engineered CRISPR systems (Cas protein). The gRNA is a short synthetic RNA that contains a Cas-binding scaffold sequence as well as a user-defined 20-nucleotide spacer that specifies the genomic target to be changed. It forms a ribonucleoprotein complex with Cas9 protein through interactions between the gRNA scaffold and surface-exposed positively-charged grooves on Cas9. Upon the formation of this complex, Cas9 undergoes a conformational change that transforms it from an inactive, non-DNA binding conformation to an active DNA-binding shape. At the same time, the gRNA's spacer region is still free to interact with target DNA. When Cas9 binds to a target that is specific homologous with gRNA sequence, Cas9 changes conformation again which allowing the nuclease domains RuvC and HNH to cleave opposing strands of the target DNA. The resultant double-strand break (DSB) is subsequently repaired using one of two broad repair pathways: i) efficient but error-prone non-homologous end joining (NHEJ) or ii) less efficient but high-fidelity homology directed repair (HDR) (Rasul et al., 2022), as shown in Figure 2.8.

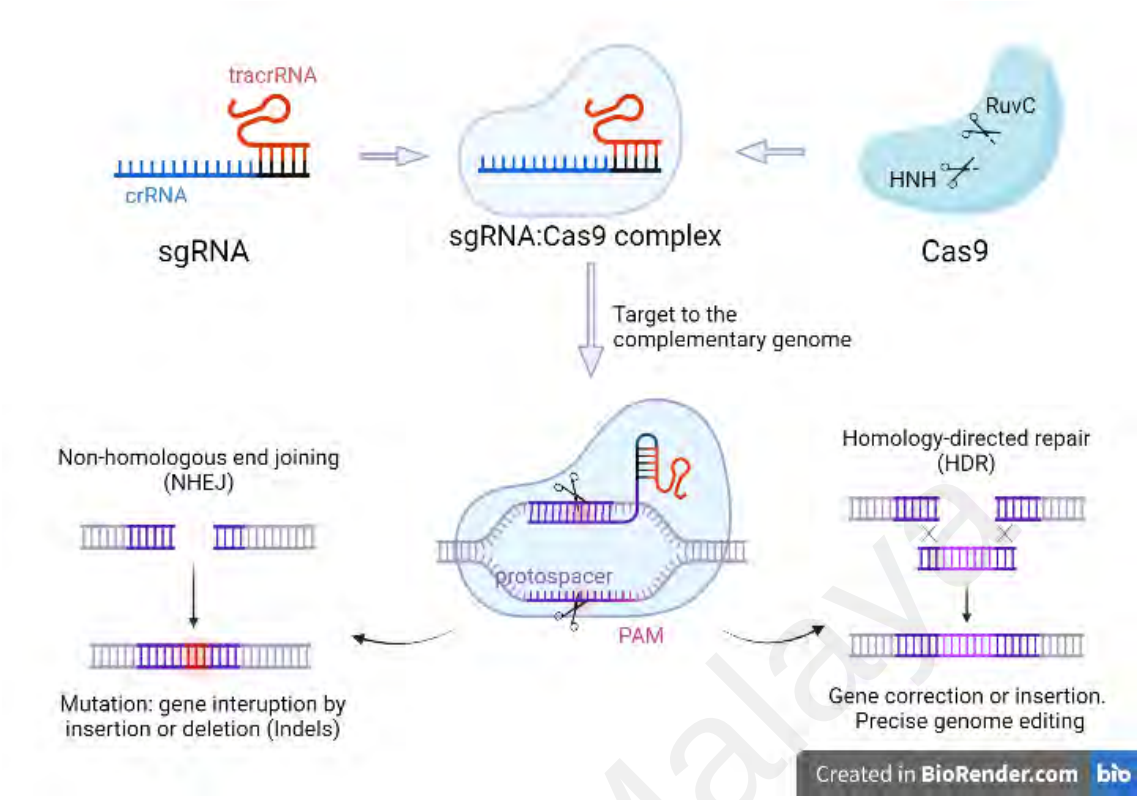


Figure 2.8: Gene editing using CRISPR. The CRISPR/Cas9 system consists of two main drivers: a sgRNA and Cas9 containing two nuclease domains, RuvC and HNH (indicated by the scissors symbol). The sgRNA consists of crRNA or spacer region and a tracrRNA. The spacer region in the crRNA is complementary to a protospacer (a 20 bp-long target DNA indicated by the purple line) with the protospacer adjacent motif (PAM) (indicated by the pink line). The sgRNA and Cas9 combine and form sgRNA:Cas9 complex. Then, the sgRNA targets the complementary protospacer, and Cas9 nucleases generate a double stranded break at the target site, which results in NHEJ or HDR. (Rasul et al., 2022). (Figure created in https://biorender.com/)

A loss-of-function mutation in the targeted gene is the optimal outcome of repair pathways. The genomic target of the Cas protein can be modifying by editing the target sequence in the gRNA however the intensity of the produced knockout phenotype, on the other hand, must be tested experimentally. CRISPR was first developed to selectively knockout the target genes, but thanks to the flexible of gRNA sequence edition, it become a handy tool to activate (CRISPRa)/ repress (CRISPRi) target genes, purify specific areas of DNA, image DNA in living cells, precisely alter DNA and RNA, and ideal for genome-wide screening. In this research, we will target the HSF4 by CRISPR-Cas9 (CRISPRko) in the purpose of study the phenotype effects in ccRCC.

CHAPTER 3: METHODOLOGY

3.1 Cell line and plasmid

The human ccRCC cell line, 786-M1A, a gift from Dr. Sakari Vanharanta at the University of Cambridge is the metastatic derivative of 786-O cells (ATCC® CRL-1932™). The identity of 786-M1A cells was previously confirmed by STR analysis and validated by Sanger sequencing of known VHL mutations in this cell line (Saiful E. Syafruddin et al., 2019). The 786-M1A cell line was cultured in RPMI-1640 medium supplemented with 10 % FBS, penicillin (100 U/mL) and streptomycin (µg/mL). The HEK293T cells was used for lentivirus production and cultured in DMEM, supplemented with 10% FBS, penicillin (100 U/mL) and streptomycin (µg/mL).

The sgRNA expression, pKLV2-U6gRNA(*BbsI*)-PGKpuro2ABFP was a gift from Kosuke Yusa (Addgene plasmid #50946) (Koike-Yusa et al., 2014). The Cas9-expressing plasmid, lentiCas9-Blast was a gift from Feng Zhang (Addgene plasmid # 52962) (NE et al., 2014). The lentiviral packaging plasmids, psPAX (Addgene #12260) and pMD2.G (Addgene #12259), were gifts from Didier Trono.

3.2 Design sgRNA targeting HSF4 gene

The high confidence HSF4-targeting sgRNAs were predicted using from GPP sgRNA Designer webtool (https://portals.broadinstitute.org/gpp/public/analysis-tools/sgrna-design) provided by Zhang Lab at MIT website (https://zlab.bio/guide-design-resources). In total 5 sgRNAs targeting HSF4 were chosen, which hereinafter are referred to as sgHSF4_1 – sgHSF4_5.

Table 3.1 : sgNTC and sgHSF4 sequences

Input	sgRNA sequence (5' to 3')	Exon Number
sgNTC	GGAAACTTACAACCAGATTG	
sgHSF4_1	GCAGAAGTCACACTCGTTTG	9
sgHSF4_2	CAGCACCCGAGCTTCGTGCG	3
sgHSF4_3	CCTCGCCCAGTAGTCGACCC	4
sgHSF4_4	CACCTTCCGAAAACCGTCTG	3
sgHSF4_5	TCAGGCAGAACGAGATCTTG	5

The human HSF4 genomic region was extracted from Ensembl Web Tool (https://www.ensembl.org) and HSF4 genomic region map was generated and visualized by using SnapGene® Viewer. Using this software, we labelled the human HSF4 exon regions, start codon, stop codon and the sgHSF4_1 to sgHSF4_5 target sites (Figure 3.1). To construct the sgRNAs targeting HSF4, the top and bottom strand of each of these sgRNAs were purchased separately from IDT. The *BbsI* restriction site overhang (red highlighted sequence) was incorporated into the 5' end of these top and bottom strands as depicted in Table 3.2.

Table 3.2 : sgNTC and sgHSF4 top and bottom strand sequences with *BbsI* restriction site overhang (red highlighted sequences)

Input	sgRNA sequence (5' to 3')
sgNTC	Top **CACCG**GGAAACTTACAACCAGATTG**GT** Bottom **TAAAAC**CAATCTGGTTGTAAGTTTCC**C**
sgHSF4_1	Top **CACCG**GCAGAAGTCACACTCGTTTG**GT** Bottom **TAAAAC**CAAACGAGTGTGACTTCTGC**C**
sgHSF4_2	Top **CACCG**CAGCACCCGAGCTTCGTGCG**GT** Bottom **TAAAAC**CGCACGAAGCTCGGGTGCTG**C**
sgHSF4_3	Top **CACCG**CCTCGCCCAGTAGTCGACCC**GT** Bottom **TAAAAC**GGGTCGACTACTGGGCGAGG**C**
sgHSF4_4	Top **CACCG**CACCTTCCGAAAACCGTCTG**GT** Bottom **TAAAAC**CAGACGGTTTTCGGAAGGTG**C**
sgHSF4_5	Top **CACCG**TCAGGCAGAACGAGATCTTG**GT** Bottom **TAAAAC**CAAGATCTCGTTCTGCCTGA**C**

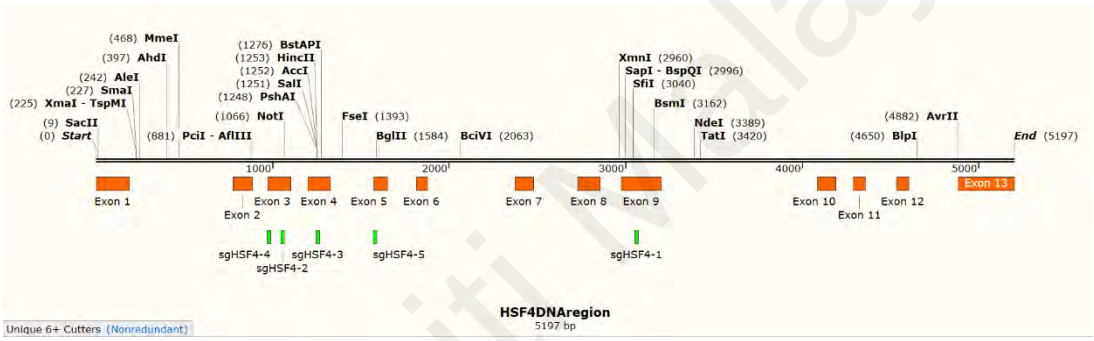


Figure 3.1: Map of HSF4 genomic region using the SanpGene Viewer tool. Orange box, Exon ; Green box. sgRNA binding site. The enlarged map is attached as APPENDIX C1, APPENDIX C2, and APPENDIX C3.

3.3 Cloning sgHSF4 and sgNTC into sgRNA expression vectors

The top and bottom strands of each sgRNA (sgHSF4_1 to sgHSF4_5) were resuspended in nuclease free water to achieve the final concentration of 100uM upon receive. Then, the top and bottom strands of each sgRNAs were annealed and phosphorylated by using T4 Ligase buffer (NEB) and T4 Polynucleotide Kinase (NEB #M0201). The reaction mixtures were incubated at 37 ℃ for 30 min and 95 ℃ for 5 min, followed by ramping down from to 25 ℃ at 5 ℃ /min. The ramping down process is the important step to allow the top and bottom sgRNA strands to properly anneal. The

annealed and phosphorylated sgRNA constructs were diluted 1/200 in DNAse/RNAse-free water for ligation.

To prepare the sgRNA expression vector for subsequent ligation procedure, pKLV-U6-gRNA(*BbsI*)-PGKpuro2ABFP was digested with *BbsI*-HF restriction enzyme (NEB #R3539) in Cutsmart buffer, and then treated with Antarctic phosphatase (AnP) enzyme (NEB #M0289) according to the manufacturer protocol. For ligating the annealed sgRNAs into this AnP treated, *BbsI*-digested vector, the ligation was performed using the T4 Ligase (NEB #M0202) according to manufacturer protocol, followed by incubation at 16 ℃ overnight.

3.4 Bacteria transformation

The ligated plasmids were transformed into the chemically competent *E. coli* DH5α, previously generated *in-house* by the group, for plasmid propagation. Briefly, the ligated plasmids were mixed with bacteria by gentle flicking. Then, the bacteria-plasmid mix were incubated on ice for 30 min, followed by the heat-shock at 42 ℃ for 1 min and incubated on ice for another 5 min. The LB media were added to the transformed bacteria and incubated at 37 ℃ for an hour with 200 to 250 rpm shaking. The transformed bacteria were then plated on LB + ampicillin agar and incubated overnight at 37 ℃.

3.5 Plasmid extraction and Sanger sequencing

The plasmid from positively transformed bacteria were extracted by using Monarch Plasmid miniprep kit (New England Biolabs) following the manufacturer's recommendations. The yield and purity of the extracted plasmids were quantified by using NanoDrop^{TM} 1000 Spectrophotometer (Thermo). The presence of the ligated sgRNAs in the sgRNA expression vector were verified via Sanger sequencing using the U6 Promoter Forward primer.

3.6 Lentiviral production and transduction

The lentiviral particles carrying the sgRNA expression vector were produced by co-transfecting the previously ligated sgRNA expression vector and lentiviral packaging vectors, psPAX and pMD2.G into HEK293T cells using Attractene transfection reagent (Thermo 11668019), according to manufacturer's recommendations. Cell culture media containing the lentiviral particles were harvested 72 h post transfection by centrifugation, followed by filtration using the 0.45 μM syringe filter (Sartorius). The lentivirus containing supernatant were either be used for transduction directly or stored at -80 ℃ until further use. These lentiviruses were transduced into the 786-M1A cell line, which stably express the Cas9 protein. These cells were previously transduced with the Lenticas9-Blast plasmid and treated with blasticidin (Invivogen ant-bl) to select for positively transduced cells.

These Cas9-expressing 786-M1A were transduced with either the sgNTC or the sgHSF4 construct individually. In brief, the Cas9-expressing 786-M1A parental cells, were seeded onto the 6-wells plate and cultured overnight to reach 70 to 80 % confluency on the following day. Next day, the harvested sgNTC or each sgHSF4-containing lentivirus were added drop wise onto the cells and mixed by swirling the plate in the presence of 8 μg/mL Polybrene transfection reagent (Merck Millipore #TR-1003-G). After overnight incubation, the cells were washed and replaced with fresh media. The transduced cells were maintained until the cells reach confluency that were suitable for puromycin (Invivogen ant-pr) treatment. These cells were hereinafter labelled as (i) 786-M1A LB_sgNTC, (ii) 786-M1A LB_sgHSF4_1, (iii) 786-M1A LB_sgHSF4_2, (iv) 786-M1A LB_sgHSF4_3, and (v) 786-M1A LB_sgHSF4_4.

3.7 Assessing HSF4 targeting efficiency and *in-vitro* functional assays

3.7.1 Protein extraction and quantification

To extract protein lysates from the positively transduced cells. The cells were trypsinized using trypsin and spun down at 1200 rpm for 5 min. The pelleted cells were washed once with ice-cold PBS and spun down again to pellet the cells. The pelleted cells were resuspended with enough volume of 1X RIPA buffer containing 1:100 protease inhibitor cocktail (Nacalai Tesque). The cells were lysed on ice for 30 min, followed by vortexing for 15 sec and centrifugation at 14000 rpm for 20 min. The supernatant containing the protein lysates were collected and transferred into fresh tube for subsequent experiment or stored in -80ºC.

3.7.2 Western blotting

The concentrations of the extracted protein lysates were quantified using the Pierce BCA protein assay kit (Thermo) according to manufacturer's protocol. Equivalent mass of protein lysates was reduced by boiling at 95 ℃ for 5 min in 1x Laemelli Sample Buffer (Biorad) containing 8 % of Beta-Mercaptoethanol. The boiled protein samples were separated using *in house* casted 10% SDS-PAGE gel in 1x Running buffer with constant voltage between 120-140 v until the proteins reached the bottom of the gel.

The separated proteins were then transferred onto the 0.45 uM Nitrocellulose membrane (GE Healthcare Amersham) in 1x Transfer buffer. The transfer was performed by assembling the standard transfer sandwich cast and run with constant ampere at 300mA for 1 ½ to 2 h. The membrane was blocked in 5 % skimmed milk prepared in 0.1% Tween20-1X Tris-buffer saline (TBS-T) by shaking at room temperature for 1 hour. The membrane was then incubated overnight at 4 ℃ with HSF4 Antibody (A-12 sc-398645a), diluted 1:500 in 5 % skimmed milk prepared in 0.1% TBS-T. On the following

day, the membrane was washed 3 times with 0.1% TBS-T and incubated for 1 hour at room temperature with rabbit-anti mouse secondary antibody (Dako P026002), diluted 1:1000 in 5% skimmed milk prepared in 0.1% TBS-T. The membrane was washed again 3 times with 0.1% TBS-T. Subsequently, the membrane was incubated with chemiluminescence ECL substrate (Thermo) for 1 min at room temperature and imaged using ChemiDoc XRS+Gel Imaging System (Bio-Rad). The image of membrane was exported to ImageJ to quantify the band area and intensity.

3.7.3 Genomic extraction

Genomic DNA were extracted from the HSF4-targeted 786-M1A LB cells using the Genomic DNA Purification Kit (Monarch NEB #T3010) by following the manufacturer's protocol. The concentration and purity of the extracted genomic DNA were determined by using the NanoDrop™ 1000 Spectrophotometer (Thermo).

3.7.4 HSF4 targeted region PCR amplification

The region targeted by the sgHSF4 was amplified from the genomic DNA previously extracted from the HSF4 targeted cells. The forward and reverse primers flanking the sgHSF4 targeted region were designed using Primer3Plus webtool (https://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi) and purchased from IDT. The HSF4 targeted region was amplified using designed primer pairs and AccuPrime™ *PFX* Supermix (Thermo #12344040) according to manufacturer's protocol. The PCR products were run on 1% agarose gel and purified using Monarch gel extraction and purification kit (New England Biolabs). The concentrations and purity of the purified PCR products were determined by using NanoDrop™ 1000 Spectrophotometer (Thermo).

3.7.5 Alamar Blue cell proliferation assay

The HSF4 targeted and control 786-M1A LB cells were subjected to Alamar Blue assay (BIO-D-DAL1025) to assess the cells proliferative capacity upon HSF4 targeting. The cells were seeded at a low density 500 cells/ well on the 96-well plate in multiple replicates. The cells were maintained for 6 days whereby the cells' growth rate was assessed and quantified every 2 days using the Alamar Blue reagent (day 0, day 2, day 4 and day6). Briefly, on the quantification day, the cells were incubated for 2 hours with Alamar Blue reagent that was diluted 1:10 in the cell culture media. The media + Alamar Blue reagent were then transferred onto the black 96-well plate and the fluorescence intensity were determined using the Varioskan plate reader. The fluorescence reading was positively correlated with the number of cells. The remaining media + Alamar Blue reagent was removed, and the cells were washed with 1X PBS and added with fresh cell culture media. This procedure was repeated for the subsequent quantification days.

3.7.6 Colony formation assay

The HSF4 targeted and control 786-M1A LB were seeded at a low density 500 cells/ well on 6-well plate and cultured for 7-10 days. The seeded cells were observed every 2 days to monitor the formation of cell colonies. At the end of the assay, the cell culture media was removed and the cells was washed with 1x PBS. The cells were fixed with 10% methanol and 10% acetic acid fixation solution by shaking at room temperature for 15 minutes. After removing the fixation solution, the cells were stained with Crystal Violet reagent (Sigma-Aldrich) by shaking at room temperature for 30 minutes. After 30 minutes staining, the Crystal Violet reagent was removed and the cells were washed with tap water and allowed to dry at room temperature for at least overnight. The formed cell colonies were imaged using the ChemiDoc XRS+ Gel Imaging system (Bio-Rad),

followed by eluting the Crystal Violet reagent with 10% Acetic Acid for quantification using the Varioskan plate reader (Thermo).

3.7.7 Tumorspheres formation assay

The HSF4 targeted cell were seeded onto ultra-low attachment plates (Corning, Corning, NY, USA) at a density 500 000 cells/ well in serum-free medium (RPMI) supplemented with B27 (Gibco 17504-044), EGF (20 ng ml^{-1}, PeproTech AF-100-15) and FGF (20 ng ml^{-1}, PeproTech 100-18B). The seeded cells were cultures for 5 days and the formation of tumorspheres were monitored daily. At the end of the assay, the formed tumorspheres were imaged under the light microscope and the images were analysed and quantified using ImageJ software. Moreover, the tumorspheres cell number were determined by collecting the tumorspheres and transferred to 15 mL Falcon tube respectively for gravity sedimentation for 10 mins at room temperature. Next. The media containing dead cells or non-tumorspheres was aspirated carefully without touching the spheroid sediment, left about 200 µL in the tube. Then 1 mL of 1×PBS was added into the tube for another 10 mins of gravity sedimentation at room temperature. The PBS was discarded carefully without touching the sediment and left about 200 µL in the tube. The 1 mL of trypsin-EDTA was added onto the tumorspheres for 4 mins and the tumorspheres were dissociated by pipetting up and down every 30 secs. The reaction of trypsin was inactivated by adding 3 mL of complete media (RPMI) and then centrifuged at 1200 rpm for 3 mins. The supernatant was discarded and 1 mL of complete media (RPMI) was added to resuspend the pellet. The cell number was counted by using Countess Automated Cell Counter (Thermo).

3.8 Statistical analysis

The experiments of the cell proliferation and colony formation assays were repeated three times, in the independent experiments under same conditions, to ensure the result's reproductivity. The result data were analysed by paired t-test to identity significant differences between HSF4-targeted and control 786-M1A cells.

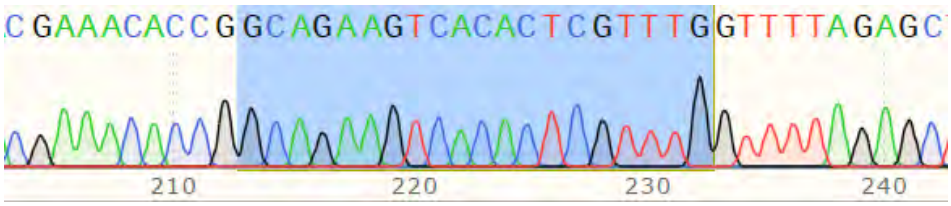
*Please refer to the simplified methodology flow chart at APPENDIX A.

CHAPTER 4: RESULTS

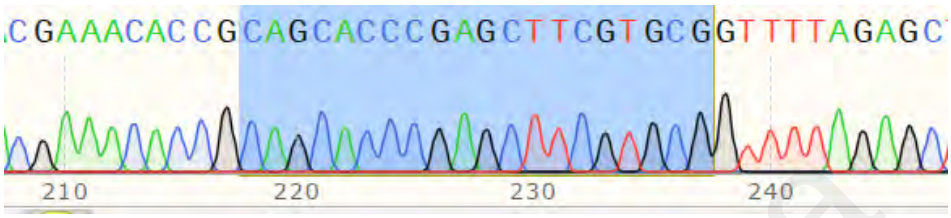
4.1 CRISPR-Cas9 system targeting HSF4 in kidney cancer cells

In this study, we employed dual CRISPR-Cas9 system to target HSF4 in ccRCC, where the sgRNAs targeting HSF4 and Cas-9 enzyme were encoded by separate plasmids. These plasmids were individually delivered into the cells to generate 786-M1A ccRCC cells that stably expressed the Cas9 and sgHSF4. There are five constructs of sgHSF4 designed, namely sgHSF4_1, sgHSF4_2, sgHSF4_3, sgHSF4_3, sgHSF4_4, and sgHSF4_5 which is shown in Table 3.1 and Table 3.2. The sgNTC was the control sgRNA that did not have any complementary binding site at the human genomic region. The sgNTC and five sgHSF4 constructs were annealed and individually ligated into sgRNA expression plasmid and transformed into *in-house* generated chemically competent *E. coli* DH5α. The plasmids were extracted from the positively transformed bacteria and sent for Sanger sequencing to verify the presence of the ligated sgHSF4 constructs.

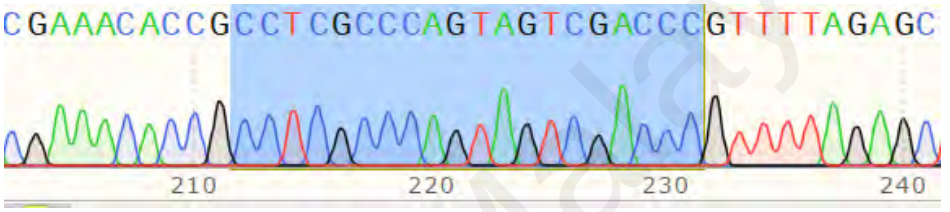
A sgHSF4_1: GCAGAAGTCACACTCGTTTG



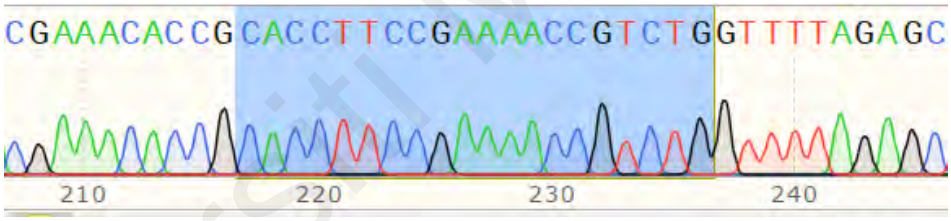
B sgHSF4_2: CAGCACCCGAGCTTCGTGCG



C sgHSF4_3: CCTCGCCCAGTAGTCGACCC



D sgHSF4_4: CACCTTCCGAAAACCGTCTG



E sgHSF4_5: TCAGGCAGAACGAGATCTTG

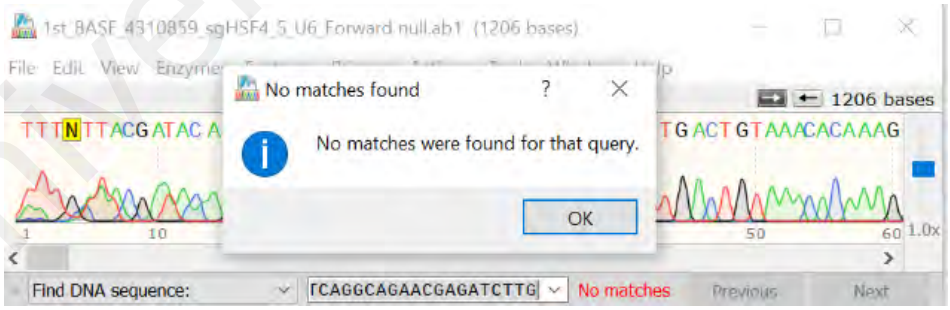


Figure 4.1: Sanger sequencing results of the sgHSF4-ligated sgRNA expression vector. A) positive inserted sgHSF4_1 (designed sequence: GCAGAAGTCACACTCGTTTG), B) positive inserted sgHSF4_2 (designed sequence: CAGCACCCGAGCTTCGTGCG), C) positive inserted sgHSF4_3 (designed sequence: CCTCGCCCAGTAGTCGACCC), D) positive inserted sgHSF4_4 (designed sequence: CACCTTCCGAAAACCGTCTG), E) negative inserted sgHSF4_5 (designed sequence: TCAGGCAGAACGAGATCTTG).

Figure 4.1 shows the Sanger sequencing results of the ligated sgHSF4_1 to sgHSF4_5. All of the sgHSF4 constructs except sgHSF4_5 were successfully ligated and cloned into the sgRNA expression vector. This was indicated by the presence of the sgHSF4_1 to sgHSF4_4 construct sequence in the sgRNA expression plasmids that were sent for Sanger sequencing. The sgNTC was also confirmed to be successfully cloned into the sgRNA expression vector (data not shown). We proceeded to produce lentivirus with the sgNTC, sgHSF4_1, sgHSF4_2, sgHSF4_3, and sgHSF4_4 constructs.

The lentivirus particles carrying the aforementioned sgNTC and sgHSF4 constructs were produced using the HEK293T cells. The HEK 293T cells were observed before and after the co-transfection with the sgHSF4-ligated sgRNA expression vector and lentiviral packaging plasmids (Figure 4.2). The HEK293T cells translated and produced the lentivirus from the plasmids and lentivirus particles would subsequently package the sgNTC or sgHSF4 expression plasmids. The lentivirus particles carrying the sgNTC or sgHSF4 expression plasmids would be released into the media from the HEK293T cells. The lentivirus particles were harvested, filtered and kept at -80 ℃ until the transduction step.

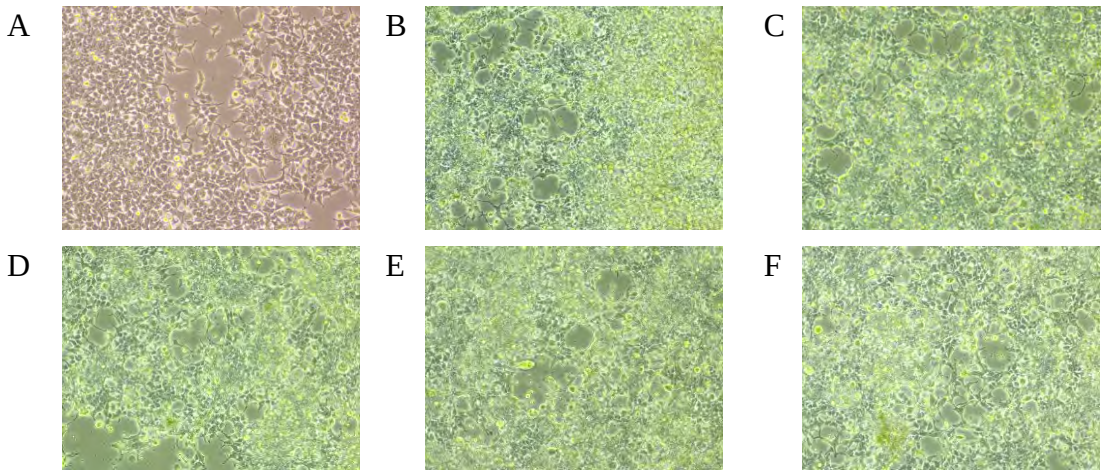


Figure 4.2: Lentiviral production in HEK293T cells (Magnification power at 4×). A) HEK293T cells before co-transfection with the sgRNA expression and lentiviral packaging plasmids. HEK293T cells post co-transfection of the lentiviral packaging plasmids and sgRNA expression plasmid carrying B) sgNTC construct, C) sgHSF4_1 construct, D) sgHSF4_2 construct, E) sgHSF4_3 construct, and F) sgHSF4_4 construct.

The collected lentivirus particles were transduced into the 786-M1A cells that stably expressed the Cas9 protein (786-M1A LB). These cells were previously generated by the member of the group by transducing the 786-M1A parental cells with Lenticas9-Blasticidin plasmid. This was followed by selecting for the positively transduced cells using blasticidin. We next transduced the 786-M1A LB with the lentivirus particle carrying: sgNTC, sgHSF4_1, sgHSF4_2, sgHSF4_3, and sgHSF4_4 expression plasmids. The transduced 786-M1A LB cells were then treated with puromycin to select for the positively transduced cells. Figure 4.3 shows the image of the untransduced vs. transduced cells post-selection with the puromycin. Cells that survived the puromycin selection indicated the cells were positively transduced and carried the sgNTC or sgHSF4 constructs. This was because the sgRNA expression vector, which is the sgNTC or sgHSF4 constructs were ligated with, also encoded for the puromycin resistant gene.

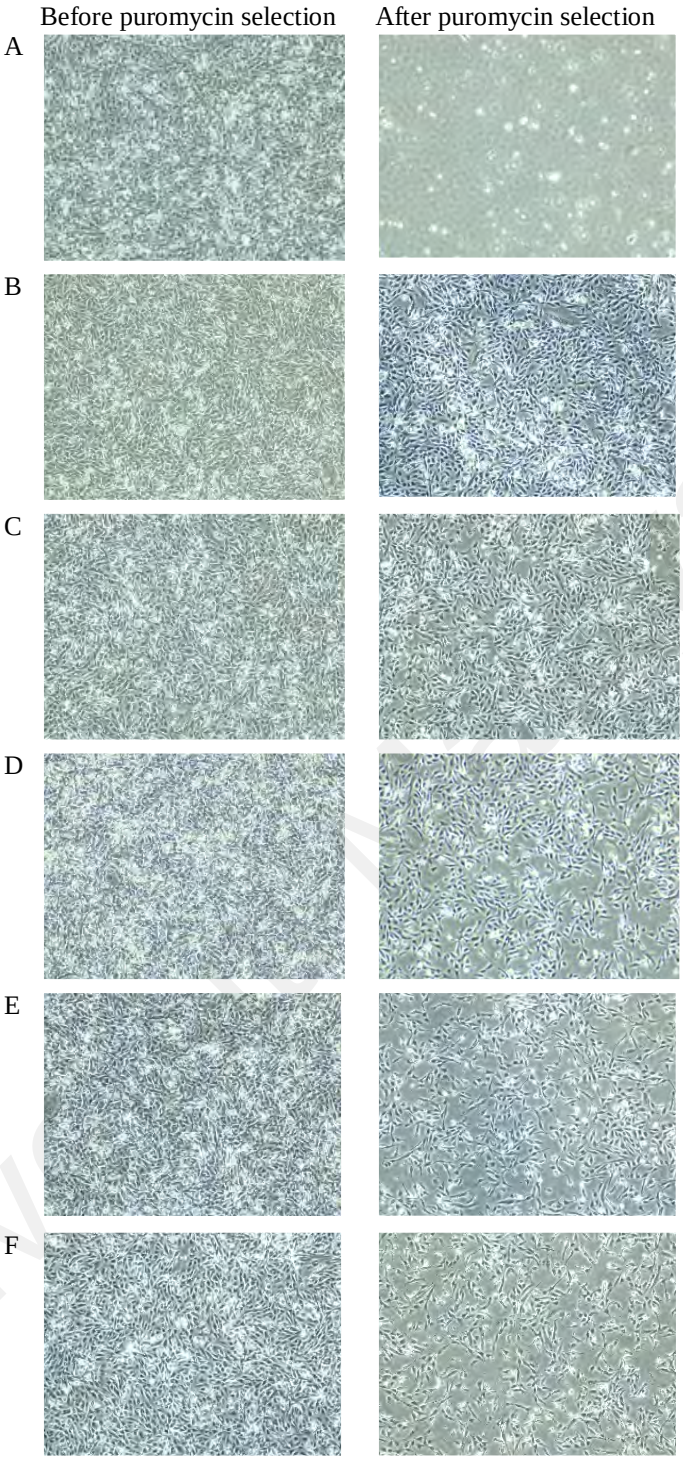


Figure 4.3: Selection of untransduced and transduced 786-MIA LB cells with puromycin (Magnification power at 4×), the selection period was conducted for 3 days. A) untransduced cells before and after selection, the cells were died after selection (after 3 days) because there was absence of the puromycin resistant gene. The cells transduced with lentiviral particles carrying B) sgNTC, C) sgHSF4_1, D) sgHSF4_2, E) sgHSF4_3, and F) sgHSF4_4 expression vector. Cells that were successfully transduced were able to survive puromycin selection due to the presence of puromycin resistant gene in the sgRNA expression vector.

4.2 Accessing HSF4 targeting efficiency in ccRCC

The HSF4 targeting efficiency were tested by Western blot after the Cas9 and sgNTC/sgHSF4 stably expressed 786-M1A LB cells were successfully generated. The proteins were separated on 10% SDS-PAGE gel and then blotted with HSF4 antibody. The HSF4 targeted 786-M1A LB cells were expected to have reduction in the HSF4 protein expression as compared to the 786-M1A LB cells transduced with sgNTC expression vector.

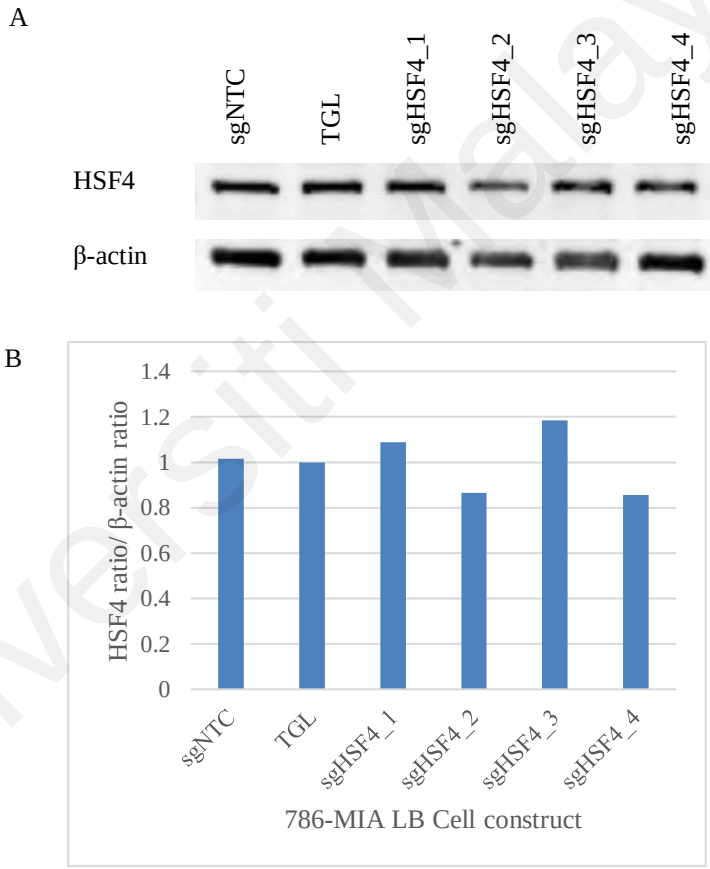


Figure 4.4: The HSF4 Western Blot of the sgNTC and sgHSF4 constructs-transduced 786-M1A LB cells membrane and analysis results. A) HSF4 (55kDa) and β-actin (42kDa) protein bands for all of the transduced cells were observed on the membrane. B) Quantification of the HSF4/ β-actin bands area ratio for all of the transduced cells. The reduction of 15 % of HSF4 protein expression was shown in sgHSF4_2 and sgHSF4_4-transduced cells. The raw data analysis of band measurement by ImageJ is attached as APPENDIX D1.

The expected size of HSF4 protein is about 55kDa, which is similar to the size of HSF4 bands that appeared on the Western blot membrane. β-actin was used as the loading control to ensure consistent loading for all of the protein samples (Figure 4.4A). Moreover, quantification of the HSF4 and β-actin bands area ratio revealed that HSF4 protein was reduced by 15 % in sgHSF4_2 and sgHSF4_4-transduced 786-M1A LB cells as compared to sgNTC-786-M1A LB cells and 786-MIA TGL parental cells (Figure 4.4B). The reduction of HSF4 protein in 786-M1A LB sgHSF4_2 cells was visualised in second batch Western blot membrane, indicated the validity of Western blot results (Figure 4.7) before we proceed to *in-vitro* functional assay.

4.3 Sequencing of sgHSF4 targeted genome region

Both sgHSF4_2 and sgHSF4_4 constructs were designed to target HSF4 exon number 3 (Figure 3.1). We next proceeded to analyse the genomic region targeted by the sgHSF4_2 construct in order to assess the insertion/deletion (InDels) induced by this construct. To achieve this, the sgHSF4_2-targeted region was amplified using the forward and reverse primer pair as shown in Figure 4.5A. For comparison, the same genomic region was also amplified from the sgNTC-transduced 786-MIA LB cells. The amplified products were subjected to Sanger sequencing and the sequence was aligned. As expected, we observed a 100% match of the sequenced amplicon of the sgNTC-transduced 786-M1A LB cells with the queried human Hg38 genome database (Figure 4.5B). This sequencing result also showed the specificity and efficiency of the designed forward and reverse primers to amplify this sgHSF4_2 targeted region.

From the amplicon of the sgHSF4_2-trandsduced cells, we found an insertion of adenine (A) at the expected Cas9 cleaved site; 3 pairs downstream of the PAM sequence. The original sequence of the sgHSF4_2 binding site is 5'-CGC**A**CGAAGCTCGGGTGCTG-3' whereas after the Cas9 cleavage and editing, there

was an additional A as the following: 5'-CGC**AA**CGAAGCTCGGGTGCT-3' (Figure 4.6A & B). Moreover, we also observed several overlapped peaks downstream of the PAM sequence.

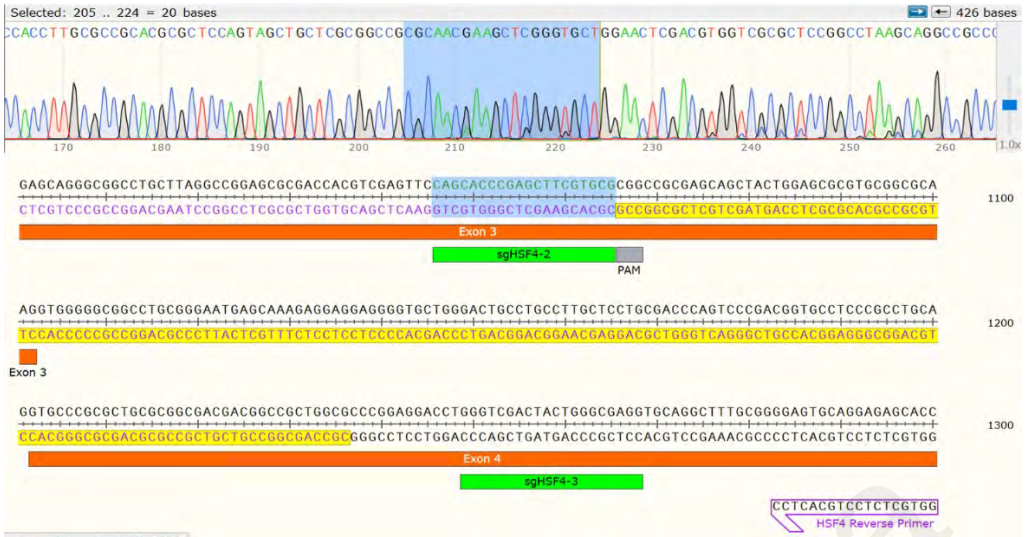
A

B

```
Query: None Query ID: lcl|Query_32605 Length: 310
>
Sequence ID: Query_32607 Length: 414
Range 1: 69 to 378

Score:573 bits(310), Expect:3e-168,
Identities:310/310(100%),  Gaps:0/310(0%), Strand: Plus/Minus

Query  1    TCCTCCGGGCGCCAGCGGCCGTCGTCGCCGCGCAGCGCGGGCACCTGCAGGCGGGAGGCA  60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  378  TCCTCCGGGCGCCAGCGGCCGTCGTCGCCGCGCAGCGCGGGCACCTGCAGGCGGGAGGCA  319

Query  61   CCGTCGGGACTGGGTCGCAGGAGCAAGGCAGGCAGTCCCAGCACCCCTCCTCCTCTTTGC  120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  318  CCGTCGGGACTGGGTCGCAGGAGCAAGGCAGGCAGTCCCAGCACCCCTCCTCCTCTTTGC  259

Query  121  TCATTCCCGCAGGCCGCCCCCACCTTGCGCCGCACGCGCTCCAGTAGCTGCTCGCGGCCG  180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  258  TCATTCCCGCAGGCCGCCCCCACCTTGCGCCGCACGCGCTCCAGTAGCTGCTCGCGGCCG  199

Query  181  CGCACGAAGCTCGGGTGCTGGAACTCGACGTGGTCGCGCTCCGGCCTAAGCAGGCCGCCC  240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  198  CGCACGAAGCTCGGGTGCTGGAACTCGACGTGGTCGCGCTCCGGCCTAAGCAGGCCGCCC  139

Query  241  TGCTCGATGCTCACCACCTTCCGAAAACCGTCTGGGGAGTGGGGCGTGGGGGCGTGGGTC  300
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  138  TGCTCGATGCTCACCACCTTCCGAAAACCGTCTGGGGAGTGGGGCGTGGGGGCGTGGGTC  79

Query  301  GCCCCATCCC  310
            ||||||||||
Sbjct  78   GCCCCATCCC  69
```

Figure 4.5: The sequencing result of the amplified HSF4 genomic region of the sgNTC-transduced 786-MIA LB cells. A) The map of the sgHSF4_2 targeted region and designed primer pairs to amplify this region that is generated using SnapGene Viewer. B) Sequencing alignment result of the sequenced amplicon of sgNTC-transduced-786-M1A LB cells with Human Hg38 genome database. (https://blast.ncbi.nlm.nih.gov/Blast.cgi).

A



B

```
Query: None Query ID: lcl|Query_372381 Length: 310
>
Sequence ID: Query_372383 Length: 315
Range 1: 7 to 315

Score:581 bits(302), Expect:9e-171,
Identities:309/310(99%),  Gaps:1/310(0%), Strand: Plus/Minus

Query  1    CAGGTCCTCCGGGCGCCAGCGGCCGTCGTCGCCGCGCAGCGCGGGCACCTGCAGGCGGGA  60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  315  CAGGTCCTCCGGGCGCCAGCGGCCGTCGTCGCCGCGCAGCGCGGGCACCTGCAGGCGGGA  256

Query  61   GGCACCGTCGGGACTGGGTCGCAGGAGCAAGGCAGGCAGTCCCAGCACCCCTCCTCCTCT  120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  255  GGCACCGTCGGGACTGGGTCGCAGGAGCAAGGCAGGCAGTCCCAGCACCCCTCCTCCTCT  196

Query  121  TTGCTCATTCCCGCAGGCCGCCCCCACCTTGCGCCGCACGCGCTCCAGTAGCTGCTCGCG  180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  195  TTGCTCATTCCCGCAGGCCGCCCCCACCTTGCGCCGCACGCGCTCCAGTAGCTGCTCGCG  136

Query  181  GCCGCGCAACGAAGCTCGGGTGCTGGAACTCGACGTGGTCGCGCTCCGGCCTAAGCAGGC  240
            |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  135  GCCGCGCA-CGAAGCTCGGGTGCTGGAACTCGACGTGGTCGCGCTCCGGCCTAAGCAGGC  77

Query  241  CGCCCTGCTCGATGCTCACCACCTTCCGAAAACCGTCTGGGGAGTGGGGCGTGGGGGCGT  300
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  76   CGCCCTGCTCGATGCTCACCACCTTCCGAAAACCGTCTGGGGAGTGGGGCGTGGGGGCGT  17

Query  301  GGGTCGCCCC  310
            ||||||||||
Sbjct  16   GGGTCGCCCC  7
```

Figure 4.6: The sequencing of amplified HSF4 genome fragment from 786-MIA LB sgHSF4_2. a) The highlighted sequence of genome fragment amplified from sgHSF4_2 cell construct were not founded in HSF4 genone region by checking in the SnapGene Viewer. b) The BLAST results showed an insertion of nucleotide base in 786-MIA LB sgHSF4_2 genome sequence at sgHSF4_2 targeted region (https://blast.ncbi.nlm.nih.gov/Blast.cgi).

4.4 *In-vitro* functional assays

Next, we assessed the phenotypic effects of targeting HSF4 in the 786-M1A LB cells. We focused on the sgHSF4_2 construct that we had verified its HSF4 targeting efficiency via HSF4 Western blotting. In addition, the genomic region targeted by this sgHSF4_2 construct was also characterized using Sanger sequencing. In specific, we performed the Alamar Blue cell proliferation assay, colony formation assay and tumorsphere-forming assay. The proliferative, colony forming and tumorsphere-forming capabilities of the sgHSF4_2 transduced cells were compared with the sgNTC-transduced 786-M1A LB cells. Prior to performing these assays, the downregulation of HSF4 protein expression was confirmed by performing HSF4 Western blotting.

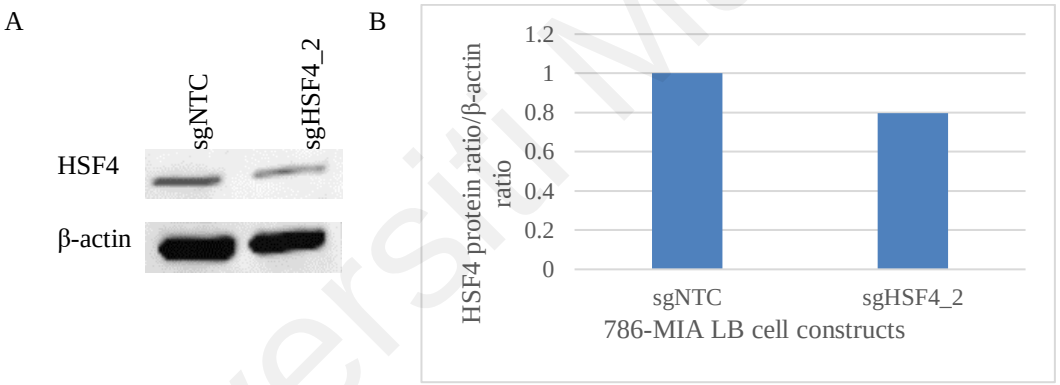


Figure 4.7: The HSF4 Western Blot of the sgNTC and sgHSF4_2-transduced 786-M1A LB cells membrane and analysis results. A) HSF4 and β-actin protein bands for sgNTC and sgHSF4_2-transduced 786-M1A LB cells. B) Quantification of the HSF4 HSF4/ β-actin bands area ratio for sgNTC and sgHSF4_2 cells-transduced 786-M1A LB cells. The reduction of 20% of HSF4 protein expression was shown in sgHSF4_2-transduced cells. The raw data of band measurement by ImageJ is attached as APPENDIX D2.

4.4.1 Cell proliferation assay by Alamar Blue reagent

The cell proliferation rate of 786-MIA LB sgNTC and sgHSF4_2 cells were assayed by Alamar Blue Reagent with five replicates for 6 days. The proliferation rate of 786-MIA LB sgHSF4_2 cells is lower than the 786-MIA LB sgNTC cells (Figure 4.8).

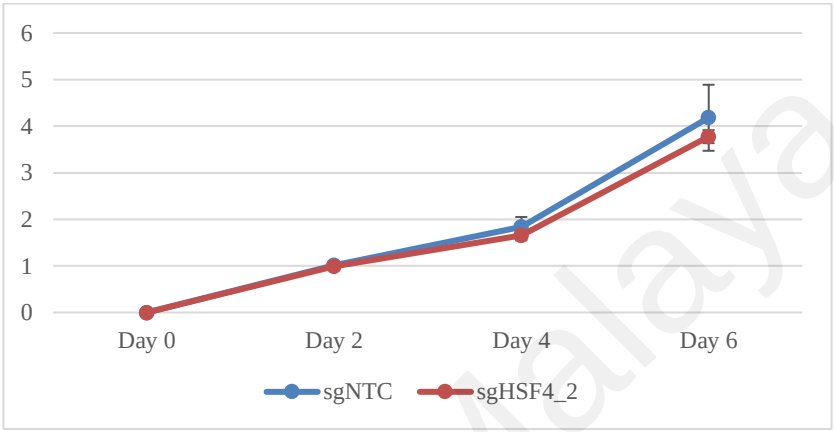


Figure 4.8: Graph of Cell Proliferation Rate of 786-M1A LB sgNTC and sgHSF4_2 (Error bar = standard deviation (SD) value). The proliferation rate of 786-M1A LB sgHSF4_2 cells is lower and less variable (lower SD value) as compared to the 786-M1A LB sgNTC cells. The data of graph is attached as APPENDIX E.

4.4.2 Colony formation assay

The ability of cells to form colony from low cell density was assayed by colony formation assay. This ability is an important indicator of cancer cells to survive and grow eventually from extra low cell number. The results showed difference between 786-MIA LB sgNTC and sgHSF4_2. There are more colonies formed by 786-MIA LB sgNTC as compared with 786-MIA LB sgHSF4_2 (Figure 4.9).

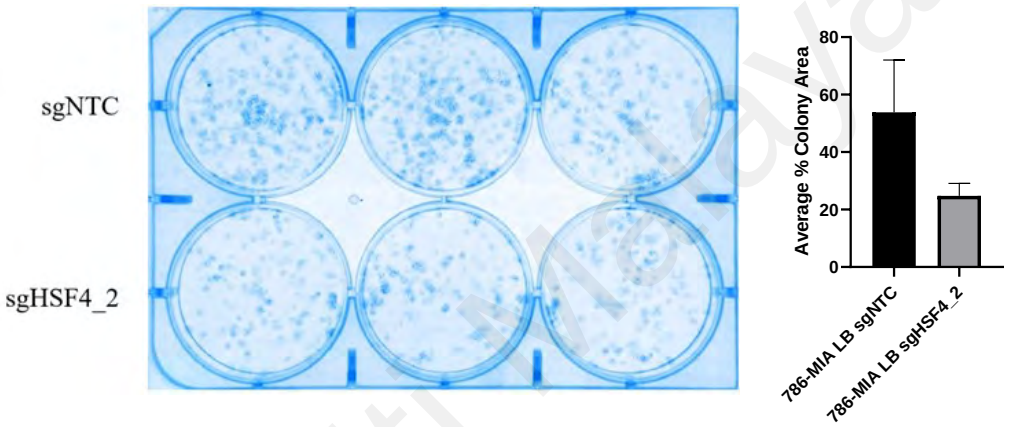


Figure 4.9: Colony formation assay. Less colonies were formed by sgHSF4_2 targeted cell constructs as compared with sgNTC cell construct. The bar chart showed that the average of colony area percentage of sgHSF4_2 targeted cell construct were relatively smaller than sgNTC cell construct (error bar = SD value, P<0.05). The raw data measured by ImageJ and statistical analysis by GraphPad Prism 8 are attached as APPENDIX F1 and APPENDIX F2.

4.4.3 Tumorspheres formation assay

We assayed the ability of sgNTC and sgHSF4_2-transduced 786-M1A LB cells to form tumorspheres. We found that sgNTC-transduced 786-M1A LB cells have higher tumorsphere-forming abilities as compared to the sgHSF4_2-transduced 786-M1A LB cells which are only able to form lesser and smaller tumorspheres as shown in Figure 4.10. The ability of tumorspheres formation between sgNTC and sgHSF4_2-transduced 786-M1A LB cells is significantly different in terms of amount (Figure 4.11 A) and size (Figure 4.11 B) of tumorspheres.

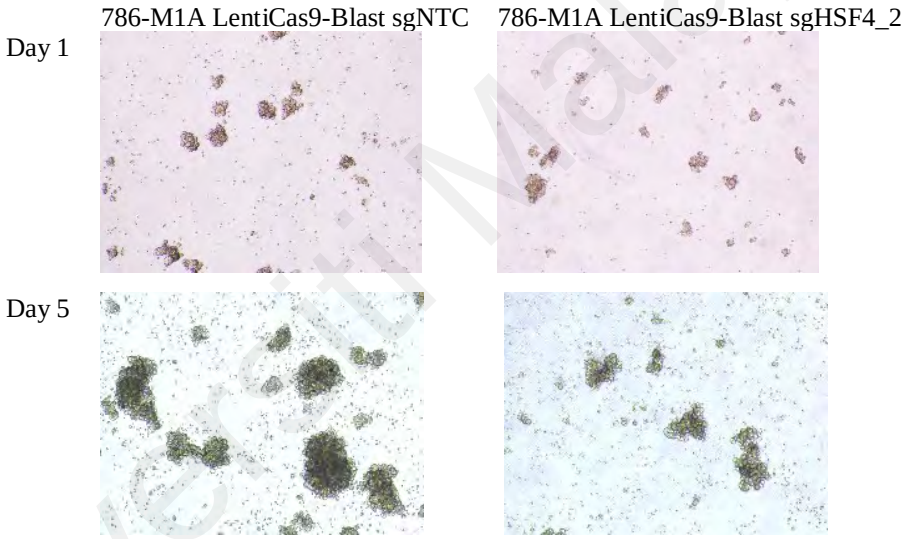


Figure 4.10: Tumorsphere morphology of sgNTC and sgHSF4_2-transduced 786-M1A LB cells in non-adherant conditions. The sgNTC-transduced 786-M1A LB cells is able to form larger tumuorspheres as compared with sgHSF4-transduced 786-M1A LB cells. The other microscopic images were taken and attached as APPENDIX G2.

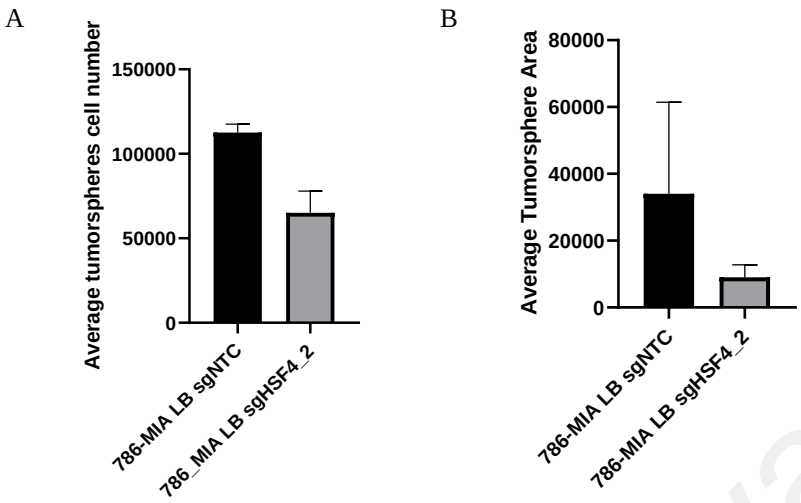


Figure 4.11: Graph of spheroid formation assay analysis (error bar= SD value). A) Graph of average spheroid cell number. 786-MIA LB sgNTC has more tumorsphere cells as compared with 786-MIA LB sgHSF4_2 (P<0.05). B) Graph of average tumorspheres area. 786-MIA LB sgNTC is able to form larger tumorsphres than the 786-MIA LB sgHSF4_2 (P<0.05). The raw data and statistical analysis of graph is attached as APPENDIX G1.

CHAPTER 5: DISCUSSION

The purpose of this research is to interrogate HSF4 functional relevance in supporting ccRCC pathogenesis. The HSF4 gene is one of the highly expressed genes in ccRCC as compared to normal kidney tissues from the prior analysis of ccRCC TCGA RNA-Seq data done by my supervisor, Dr. Saiful Effendi. With the urge to elucidate the underlying HSF4 roles in ccRCC pathogenesis, we targeted the HSF4 gene in ccRCC cells using the CRISPR-Cas9 system, and assessed the HSF4-targeted cells phenotypes *in-vitro*.

5.1 CRISPR-Cas9 targeted HSF4 in ccRCC effectively

The CRISPR-Cas9 gene editing tool was employed to target the HSF4 gene in ccRCC cells in this study. The sgRNA that targeting the HSF4 gene (namely sgHSF4) and Cas9 protein were produced in separate expression plasmid. The success and efficiency of the sgHSF4 cloning steps were determined by selecting the transformed bacteria with ampicillin and subjected the extracted plasmid to Sanger sequencing. This was to confirm the presence of the sgHSF4 into the sgRNA expression plasmid. We successfully delivered the sgHSF4_1, sgHSF4_2, sgHSF4_3, sgHSF4_4, and sgNTC into the ccRCC cells via lentiviral transduction. The 786-MIA LB cell constructs of sgNTC, sgHSF4_1 to 4 were subjected to Western Blot in order to determine the most efficient sgHSF4 constructs.

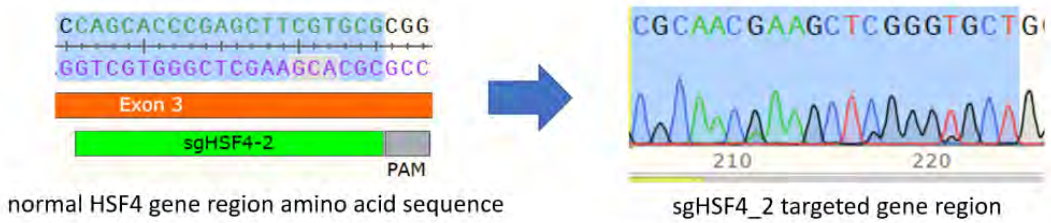
The results of Western Blot showed the reduction in HSF4 protein in 786-MIA LB sgHSF4_2 as compared to the 786-MIA LB sgNTC and 786-MIA TGL control cells (

Figure 4.4). This indicates the sgHSF4_2 has successfully targeted its corresponding HSF4 genomic region. There was no complete knockout observed in Western blot analysis. This was expected because we performed pooled cells CRISPR-Cas9 targeting. The pooled cells CRISPR-Cas9 targeting was chosen instead of single cell targeting

because the pooled cells targeting has several advantages such as, simple and non-laborious, no automation or specialized equipment needed, lesser experimental manipulation, and cost effective (Costa et al., 2004).

To characterize the gene alteration induced by sgHSF4_2, we amplified the genomic region targeted by sgHSF4_2 from sgHSF4_2 and sgNTC-transduced 786-M1A LB cells respectively and then sent for Sanger Sequencing. The genome region's sequence of 786-MIA LB sgNTC has no nucleotide changes as sgNTC is a non-targeting control that does not target any human genomic region and hence cause no alteration (Figure 4.5). In comparison, the HSF4 gene exon number 3 region with the initial sequence of 5'-CGCACGAAGCTCGGGTGCTG-3' which is targeted by sgHSF4_2 has a minor alteration in sequence. The sequence was inserted a nucleotide base of adenine (A) and results the edited sequence of 5'-CGCAACGAAGCTCGGGTGCT-3' (Figure 4.6). This has confirmed the point mutation was occurred at HSF4 gene genomic region.

The point mutation occurred in 786-M1A LB sgHSF4_2 was insertion of a base that led to frameshift mutation. According to the principle of central dogma, the frameshift mutation on nucleotide sequence can results transcription of mRNA that contains different sequence from normal nucleotide sequence and causing big changes in translation of mRNA to protein (Clark et al., 2020; Elaine, 2022). The effects of a single insertion of nucleotide base in exon region was showed in Figure 5.1.



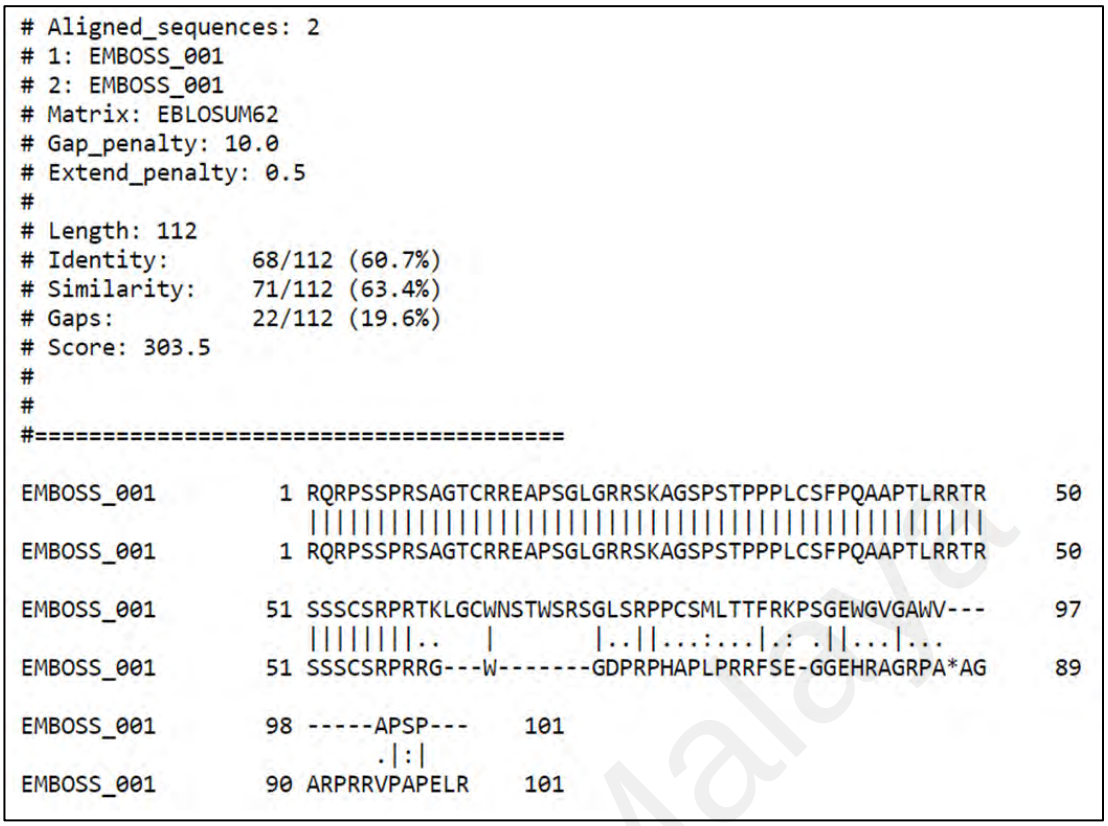
normal HSF4 gene region amino acid sequence

sgHSF4_2 targeted gene region

normal HSF4 gene region	DNA sequence:	5'-CGC ACG AAG CTC GGG TGC TGG-3'
	mRNA sequence:	3'-GCG UGC UUC GAG CCC ACG ACC-5'
	Protein sequence:	5'-ALA CYS PHE GLU PRO THR THR-3'
sgHSF4_2 targeted gene region	DNA sequence:	5'-CGC AAC GAA GCT CGG GTG CTG-3'
	mRNA sequence:	3'-GCG UUG CUU CGA GCC CAC GAC-3'
	Protein sequence:	5'-ALA LEU LEU ARG ALA HIS ASP-3'

Figure 5.1: Frameshift mutation caused by point mutation at sgHSF4_2 targeted gene region (translated by Nucleic acid convertor at https://skaminsky115.github.io/nac/DNA-mRNA-Protein_Converter.html).

Starting from the insertion point, the triplet reading frame was completely changed, hence altering the message for every subsequent amino acid as shown in. The pairwise sequence alignment of protein sequence from sgNTC and sgHSF4_2-transduced cells amplicons have shown the identity (60.7%) and similarity (63.4%) which are contributed by the upstream of sgHSF4_2 targeted site. The sequence alignment after the insertion point where after sgHSF4_2 cleaved site has high gap value of 19.6% (Figure 5.2). Form the results, the insertion of adenine base has shifted the reading frame massage that may lead to interruptions on mRNA transcription activity and level as well as on the production of functional HSF4 protein as the amino acid sequence sets the reading frame message in mRNA and also is the backbone of protein structure (Alberts et al., 2002; Clark et al., 2020).



```
# Aligned_sequences: 2
# 1: EMBOSS_001
# 2: EMBOSS_001
# Matrix: EBLOSUM62
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 112
# Identity:      68/112 (60.7%)
# Similarity:    71/112 (63.4%)
# Gaps:          22/112 (19.6%)
# Score: 303.5
#
#
#=======================================

EMBOSS_001         1 RQRPSSPRSAGTCRREAPSGLGRRSKAGSPSTPPPLCSFPQAAPTLRRTR     50
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001         1 RQRPSSPRSAGTCRREAPSGLGRRSKAGSPSTPPPLCSFPQAAPTLRRTR     50

EMBOSS_001        51 SSSCSRPRTKLGCWNSTWSRSGLSRPPCSMLTTFRKPSGEWGVGAWV---     97
                     ||||||||..   |       |..||...:...|.: .||...|...
EMBOSS_001        51 SSSCSRPRRG---W-------GDPRPHAPLPRRFSE-GGEHRAGRPA*AG     89

EMBOSS_001        98 -----APSP---    101
                          .|:|
EMBOSS_001        90 ARPRRVPAPELR    101
```

Figure 5.2: Results of pairwise sequence alignment of protein sequences of sgNTC and sgHSF4_2-transduced 786-M1A LB cells amplified genomic region by using EMBOSS Needle webtool (https://www.ebi.ac.uk/Tools/psa/emboss_needle/). The information supplement for Figure 5.2 is attached as APPENDIX H1 and APPENDIX H2.

In order to understand the exact mechanisms that lead to reduction of HSF4 protein, further analysis at the RNA level is required. It could be another interesting insight into the regulation of HSF4 gene expression in ccRCC. However, the focus of this research is to investigate the functional role of HSF4 gene in cell growth and fitness. Therefore, the following experiments that were carried out are *in-vitro* functional assays including cell proliferation assay, colony formation assay, and tumorspheres formation assay.

5.2 Phenotypic effects of HSF4 targeting in ccRCC cells

The phenotypic effects of HSF4 targeting in ccRCC cells were assayed *in-vitro* in terms of cell proliferation, colony formation and spheroid formation capabilities. Generally, the 786-M1A LB sgHSF4_2 cells proliferated at lower rate (Figure 4.8), formed lesser colonies (Figure 4.9) and, lesser and smaller tumorspheres in comparison to 786-M1A LB sgNTC cells (Figure 4.10 & Figure 4.11).

2D cultures have a number of drawbacks, including disrupting interactions between the cellular and extracellular environments, as well as alterations in cell shape, polarity, and division mechanism. These drawbacks prompted the development of models that are better suited to simulate *in vivo* situations. Three-dimensional (3D) culture is one such better way (Kapałczyńska et al., 2018). Therefore, in this research we optimised non-adherent culture conditions for 786-M1A cell line in 3D tumorsphere formation assay with the hope to lead to a better knowledge of kidney cancer cells biology. The tumorsphere formation assay shown more remarkable and consistent results as we compared between the sgNTC and sgHSF4_2-transduced 786-M1A LB cells (Figure 4.10 & Figure 4.11).

All in all, both three assays results showed the decreased cell activity in 786-M1A LB sgHSF4_2 as compared with 786-M1A sgNTC. As aforementioned in 5.1, the point mutation could lead to frameshift mutation and it will have consequential impacts to the synthesis of protein coded. In our case, the mutation has impacts in functional HSF4 protein synthesis and also ensuring the lower cell growth activity in 786-M1A cells. This combination of findings provides some support for the conceptual premise that the HSF4 protein has possessed functional roles in 786-M1A cells growth in 2D and more notable in 3D forms. More research needed to be done to draw a clearer picture of the mechanisms how HSF4 protein synthesis affect the cell growth and cellular fitness.

In vitro research has several advantages over *in vivo* research, including strict chemical and physical environment control, higher throughput, lower cost and reduced animal use. *In vitro* investigations, on the other hand, fail to mimic the conditions of cells in the presence of biokinetics in an organism, limiting the value of *in vitro* data to predict *in vivo* behavior (Graudejus et al., 2018). Hence, to observe more obvious phenotype of cells, experiment *in vivo* would be required as the cells might be more dependent to HSF4 *in vivo* while under living stresses and adaption to grow in a living system compare to growing *in-vitro* culture system.

CHAPTER 6: CONCLUSION

6.1 Conclusion

In summary, we successfully targeted HSF4 gene by CRISPS-Cas9 system and determined that HSF4 protein has functional role in maintaining 786-M1A cell line growth and fitness. This research study is an early stage of targeting a specific gene in ccRCC pathogenesis. The downstream molecular pathways in ccRCC that are regulated by HSF4 gene remains to be elucidated. Future research should therefore concentrate on the investigation of functional roles of HSF4 protein in other ccRCC cell lines *in-vitro* and *in vivo*, examination on whether the reintroduction of exogenous HSF4 able to rescue the HSF4-targeted ccRCC cells phenotypes, and identification of putative HSF4 downstream targets. These findings will be worthwhile contributions to strengthen the knowledge on ccRCC pathogenesis and pave the way to development of efficient therapeutic strategy for ccRCC.

6.2 Research Timelines and Milestones

The research was planned to start from May 2021, and the estimated duration is tabulated as Table 6.1 and the milestones are tabulated as Table 4.2. The whole research project laboratory works are estimated to completed in 7 months. However due to the pandemic condition, the progress was dragged heavily. The practical lab work was started from end of September 2021 and tabulated as Table 4.2 while the milestones achieved was tabulated as Table 4.3. The objective and milestones of this research were completely achieved.

Table 6.1 : Estimated Research Duration

Activity	Start	End	Duration
Designing the sgRNA targeting HSF4 gene and make purchase order	1/4/2021	30/4/2021	1 month
Constructing the sgRNA expression vectors	5/5/2021	31/5/2021	1 month
Lentiviral particles production	1/6/2021	30/6/2021	1 month
Transduction of lentiviral particles into kidney cancer cells	1/7/2021	31/7/2021	1 month
Selecting positive HSF4 targeted and control cells, protein extraction	1/8/2021	31/8/2021	1 month
HSF4 Western Blotting and *in-vitro* functional assays	1/9/2021	30/10/2021	2 months

Table 6.2 : Real-time Research Duration

Activity	Start	End	Duration
Designing the sgRNA targeting HSF4 gene and make purchase order	21/4/2021	5/5/2021	2 weeks
MCO and Lab temporally closure	6/5/2021	28/9/2021	5 months
Constructing the sgRNA expression vectors	29/9/2021	29/10/2021	1 month
Lentiviral particles production	29/10/2021	5/11/2021	1 week
Transduction of lentiviral particles into kidney cancer cells	8/11/2021	9/11/2021	2 days
Selecting positive HSF4 targeted and control cells	10/11/2021	15/11/2021	5 days
Protein extraction and HSF4 Western Blotting	16/11/2021	31/12/2021	$1\frac{1}{2}$month
In-vitro functional assays	3/1/2022	31/1/2022	1 month
Genome extraction and sequencing	8/1/2022	21/1/2022	2 weeks

Table 6.3 : Milestones

Date Planned	Date Achieved	Milestone achieved
30/6/2021	5/11/2021	Milestone #1: Construct CRISPR-Cas9 system targeting HSF4 in kidney cancer cells
31/7/2021	15/11/2021	Milestone #2: Positively Transduced kidney cancer cells with lentiviral CRISPR-Cas9 system
31/8/2021	31/12/2021	Milestone #3: Complete assess the HSF4 targeting efficiency in kidney cancer cells
30/10/2021	31/1/2022	Milestone #4: Determined the phenotypic effects of targeting HSF4 in kidney cancer cells

*The Gantt Charts are available in APPENDIX B.

REFERENCES

Adli, M. (2018). The CRISPR tool kit for genome editing and beyond. *Nature Communications*, *9*(1), 1–13.

Åkerfelt, M., Morimoto, R. I., & Sistonen, L. (2010). Heat shock factors: integrators of cell stress, development and lifespan. *Nature Reviews Molecular Cell Biology*, *11*(8), 545–555.

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., & Walter, P. (2002). *Molecular Biology of the Cell* (4th ed.). New York: Garland Science.

Arroyo-Olarte, R. D., Bravo Rodríguez, R., & Morales-Ríos, E. (2021). Genome Editing in Bacteria: CRISPR-Cas and Beyond. *Microorganisms*, *9*(4), Article#844.

Athanazio, D. A., Amorim, L. S., Cunha, I. W. da, Leite, K. R. M., Paz, A. R. da, Gomes, R. de P. X., … Bezerra, S. M. (2021). Classification of renal cell tumors – current concepts and use of ancillary tests: recommendations of the Brazilian Society of Pathology. *Surgical and Experimental Pathology*, *4*(1), 1–21.

Barrangou, R., & Marraffini, L. A. (2014). CRISPR-Cas systems: prokaryotes upgrade to adaptive immunity. *Molecular Cell*, *54*(2), 234–244.

Behan, F. M., Iorio, F., Picco, G., Gonçalves, E., Beaver, C. M., Migliardi, G., … Garnett, M. J. (2019). Prioritization of cancer therapeutic targets using CRISPR–Cas9 screens. *Nature*, *568*(7753), 511–516.

Cai, Q., Christie, A., Rajaram, S., Zhou, Q., Araj, E., Chintalapati, S., … Kapur, P. (2020). Ontological analyses reveal clinically-significant clear cell renal cell carcinoma subtypes with convergent evolutionary trajectories into an aggressive type. *EBioMedicine*, *51*, Article#102526.

Chen, R., Liliental, J. E., Kowalski, P. E., Lu, Q., & Cohen, S. N. (2011). Regulation of transcription of hypoxia-inducible factor-1α (HIF-1α) by heat shock factors HSF2 and HSF4. *Oncogene*, *30*(22), 2570–2580.

Chen, W., Hill, H., Christie, A., Kim, M. S., Holloman, E., Pavia-Jimenez, A., … Brugarolas, J. (2016). Targeting renal cell carcinoma with a HIF-2 antagonist. *Nature*, *539*(7627), 112–117.

Cho, H., Du, X., Rizzi, J. P., Liberzon, E., Chakraborty, A. A., Gao, W., … Kaelin, W. G. (2016). On-target efficacy of a HIF-2α antagonist in preclinical kidney cancer models. *Nature*, *539*(7627), 107–111.

Choueiri, T. K., & Motzer, R. J. (2017). Systemic therapy for metastatic renal-cell carcinoma. *The New England Journal of Medicine*, *376*(4), 354–366.

Clark, M. A., Choi, J. H., Douglas, M., & *OpenStax* (2020). *Introduction to Molecular and Cell Biology*. Retrieved from https://rwu.pressbooks.pub/bio103/chapter/the-central-dogma-genes-to-traits/

Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., … Zhang, F. (2013). Multiplex genome engineering using CRISPR/Cas systems. *Science*, *339*(6121), 819–823.

Costa, J. R., Bejcek, B. E., McGee, J. E., Fogel, A. I., Brimacombe, K. R., & Ketteler, R. (2004). Genome editing using engineered nucleases and their use in genomic screening. In *Assay Guidance Manual* (pp. 1–24). Bethesda (MD): Eli Lilly & Company and the National Center for Advancing Translational Sciences.

Creighton, C. J., Morgan, M., Gunaratne, P. H., Wheeler, D. A., Gibbs, R. A., Robertson, G., … Sofia, H. J. (2013). Comprehensive molecular characterization of clear cell renal cell carcinoma. *Nature*, *499*(7456), 43–49.

Dabestani, S., Thorstenson, A., Lindblad, P., Harmenberg, U., Ljungberg, B., & Lundstam, S. (2016). Renal cell carcinoma recurrences and metastases in primary non-metastatic patients: a population-based study. *World Journal of Urology*, *34*(8), 1081–1086.

Dalgliesh, G. L., Furge, K., Greenman, C., Chen, L., Bignell, G., Butler, A., … Futreal, P. A. (2010). Systematic sequencing of renal carcinoma reveals inactivation of histone modifying genes. *Nature*, *463*(7279), 360–363.

Dempster, J. M., Pacini, C., Pantel, S., Behan, F. M., Green, T., Krill-Burger, J., … Iorio, F. (2019). Agreement between two large pan-cancer CRISPR-Cas9 gene dependency data sets. *Nature Communications*, *10*(1), 1–14.

Elaine A. Ostrander. (2022). *Frameshift Mutation*. Retrieved February 2, 2022, from https://www.genome.gov/genetics-glossary/Frameshift-Mutation

Gebhard, R. L., Clayman, R. V., Prigge, W. F., Figenshau, R., Staley, N. A., Reesey, C., & Bear, A. (1987). Abnormal cholesterol metabolism in renal clear cell carcinoma. *Journal of Lipid Research*, *28*(10), 1177–1184.

Gerlinger, M., Horswell, S., Larkin, J., Rowan, A. J., Salm, M. P., Varela, I., … Swanton, C. (2014). Genomic architecture and evolution of clear cell renal cell carcinomas defined by multiregion sequencing. *Nature Genetics*, *46*(3), 225–233.

Gomez-Pastor, R., Burchfiel, E. T., & Thiele, D. J. (2017). Regulation of heat shock transcription factors and their roles in physiology and disease. *Nature Reviews Molecular Cell Biology*, *19*(1), 4–19.

Graudejus, O., Ponce Wong, R., Varghese, N., Wagner, S., & Morrison, B. (2018). Bridging the gap between in vivo and in vitro research: reproducing in vitro the mechanical and electrical environment of cells in vivo. *Frontiers in Cellular Neuroscience*, *12*.

Guo, G., Gui, Y., Gao, S., Tang, A., Hu, X., Huang, Y., … Wang, J. (2012). Frequent mutations of genes encoding ubiquitin-mediated proteolysis pathway components in clear cell renal cell carcinoma. *Nature Genetics*, *44*(1), 17–19.

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

Hart, T., Chandrashekhar, M., Aregger, M., Steinhart, Z., Brown, K. R., MacLeod, G., … Moffat, J. (2015). High-resolution CRISPR screens reveal fitness genes and genotype-specific cancer liabilities. *Cell*, *163*(6), 1515–1526.

Hsieh, J. J., Purdue, M. P., Signoretti, S., Swanton, C., Albiges, L., Schmidinger, M., … Ficarra, V. (2017). Renal cell carcinoma. *Nature Reviews Disease Primers*, *3*.

Jin, X., Eroglu, B., Cho, W., Yamaguchi, Y., Moskophidis, D., & Mivechi, N. F. (2012). Inactivation of heat shock factor Hsf4 induces cellular senescence and suppresses tumorigenesis *in vivo*. *Molecular Cancer Research*, *10*(4), 523–534.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., & Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*, *337*(6096), 816–821.

Kaelin, W. G. (2017). The VHL tumor suppressor gene: insights into oxygen sensing and cancer. *Transactions of the American Clinical and Climatological Association*, *128*, 298–307.

Kapałczyńska, M., Kolenda, T., Przybyła, W., Zajączkowska, M., Teresiak, A., Filas, V., … Lamperska, K. (2018). 2D and 3D cell cultures – a comparison of different types of cancer cell cultures. *Archives of Medical Science : AMS*, *14*(4), Article#910.

Koike-Yusa, H., Li, Y., Tan, E. P., Velasco-Herrera, M. D. C., & Yusa, K. (2014). Genome-wide recessive genetic screening in mammalian cells with a lentiviral CRISPR-guide RNA library. *Nature Biotechnology*, *32*(3), 267–273.

Lander, E. S. (2016). The heroes of CRISPR. *Cell*, *164*(1–2), 18–28.

Lopez-Beltran, A., Carrasco, J. C., Cheng, L., Scarpelli, M., Kirkali, Z., & Montironi, R. (2009). 2009 update on the classification of renal epithelial tumors in adults. *International Journal of Urology*, *16*(5), 432–443.

Luo, J., Solimini, N. L., & Elledge, S. J. (2009). Principles of cancer therapy: oncogene and non-oncogene addiction. *Cell*, *136*(5), 823–837.

Mali, P., Yang, L., Esvelt, K. M., Aach, J., Guell, M., DiCarlo, J. E., … G. M. (2013). RNA-guided human genome engineering via Cas9. *Science*, *339*(6121), 823–826.

Mazza, C., Escudier, B., & Albiges, L. (2017). Nivolumab in renal cell carcinoma: latest evidence and clinical potential. *Therapeutic Advances in Medical Oncology*, *9*(3), 171–181.

Mehdi, A., & Riazalhosseini, Y. (2017). Epigenome aberrations: emerging driving factors of the clear cell renal cell carcinoma. *International Journal of Molecular Sciences*, *18*(8), Article#1774.

Morimoto, R. I. (1993). Cells in stress: transcriptional activation of heat shock genes. *Science*, *259*(5100), 1409–1410.

Patel, P. H., Chadalavada, R. S. V., Chaganti, R. S. K., & Motzer, R. J. (2006). Targeting von Hippel-Lindau pathway in renal cell carcinoma. *Clinical Cancer Research*, *12*(24), 7215–7220.

PDQ Adult Treatment Editorial Board (2002). Renal cell cancer treatment (PDQ®): patient version. *PDQ Cancer Information Summaries*. Bethesda (MD): National Cancer Institute (US)

Prahlad, V., & Morimoto, R. I. (2009). Integrating the stress response: lessons for neurodegenerative diseases from *C. elegans*. *Trends in Cell Biology*, *19*(2), 52–61.

Rasul, M. F., Hussen, B. M., Salihi, A., Ismael, B. S., Jalal, P. J., Zanichelli, A., … Taheri, M. (2022). Strategies to overcome the main challenges of the use of CRISPR/Cas9 as a replacement for cancer therapy. *Molecular Cancer*, *21*(1), 1–30.

Reeja, T. (2013). *Renal Cell Carcinoma*. Retrieved 08, 07, 2021, from https://www.medindia.net/patients/patientinfo/renal-cell-carcinoma.htm

Ricketts, C. J., Cubas, A. A. De, Fan, H., Smith, C. C., Lang, M., Reznik, E., … Linehan, W. M. (2018). The cancer genome atlas comprehensive molecular characterization of renal cell carcinoma. *Cell Reports*, *23*(1), Article#313.

Rini, B. I., & Atkins, M. B. (2009). Resistance to targeted therapy in renal-cell carcinoma. *The Lancet Oncology 10*(10), 992–1000.

Rini, B. I., Campbell, S. C., & Escudier, B. (2009). Renal cell carcinoma. *The Lancet*, *373*(9669), 1119–1132.

Sakai, Y., Kanomata, N., Itami, H., Kajimoto, K., Sakuma, T., & Ohbayashi, C. (2010). Signet-ring cell carcinoma of the stomach metastasizing to renal cell carcinoma: A case report and review of the literature. *The Kobe Journal of Medical Sciences*, *55*(6), E122-31.

Sanjana, N. E., Shalem, O., & Zhang, F. (2014). Improved vectors and genome-wide libraries for CRISPR screening. *Nature Methods*, *11*(8), 783–784.

Sundelin, J. P., Ståhlman, M., Lundqvist, A., Levin, M., Parini, P., Johansson, M. E., & Borén, J. (2012). Increased Expression of the very low-density lipoprotein receptor mediates lipid accumulation in clear-cell renal cell carcinoma. *PLOS ONE*, *7*(11), Article#e48694.

Saiful Effendi Syafruddin, Rodrigues, P., Vojtasova, E., Patel, S. A., Zaini, M. N., Burge, J., Warren, A. Y., Stewart, G. D., Eisen, T., Bihary, D., Samarajiwa, S. A., & Vanharanta, S. (2019). A KLF6-driven transcriptional network links lipid homeostasis and tumour growth in renal carcinoma. *Nature Communications*, *10*(1), 1–13.

Saiful Effendi Syafruddin, Ling, S., Low, T. Y., & Aiman Mohtar, M. (2021). More than meets the eye: revisiting the roles of heat shock factor 4 in health and diseases. *Biomolecules 11*(4), Article#523.

Turajlic, S., Xu, H., Litchfield, K., Rowan, A., Chambers, T., Lopez, J. I., … Swanton, C. (2018). Tracking cancer evolution reveals constrained routes to metastases: TRACERx renal. *Cell, 173*(3), 581-594.

Turajlic, S., Xu, H., Litchfield, K., Rowan, A., Horswell, S., Chambers, T., … Swanton, C. (2018). Deterministic evolutionary trajectories influence primary tumor growth: TRACERx renal. *Cell, 173*(3), 595-610.

Tzelepis, K., Koike-Yusa, H., De Braekeleer, E., Li, Y., Metzakopian, E., Dovey, O. M., … Yusa, K. (2016). A CRISPR dropout screen identifies genetic vulnerabilities and therapeutic targets in acute myeloid leukemia. *Cell Reports*, *17*(4), 1193–1205.

Vanharanta, S., & Massagué, J. (2013). Origins of metastatic traits. *Cancer Cell*, *24*(4), 410–421.

Varela, I., Tarpey, P., Raine, K., Huang, D., Ong, C. K., Stephens, P., … Futreal, P. A. (2011). Exome sequencing identifies frequent mutation of the SWI/SNF complex gene PBRM1 in renal carcinoma. *Nature*, *469*(7331), 539–542.