A study of the expression and cellular function of the human *FAM111B* gene

by

Cenza Rhoda

A dissertation

Submitted to the Faculty of Health Science

The University of Cape Town

Department of Medicine

Division of Dermatology

In fulfilment of the requirements for the degree of Master of Science in Medicine:

Trichology and Cosmetics degree.

August 2021

Declaration

I, *Cenza Rhoda*, hereby declare that the work on which this dissertation/thesis is based is my original work (except where acknowledgements indicate otherwise) and that neither the whole work nor any part of it has been, is being, or is to be submitted for another degree in this or any other university.

I empower the university to reproduce for the purpose of research either the whole or any portion of the contents in any manner whatsoever.

Signature: Signed by candidate

Date: 19 July 2021

rhdcen001:Cenza_Rhoda_final_body.docx fol. 31/082021

ORIGINALITY REPORT

14%	12%	7%	5%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	hdl.handle.net Internet Source	1%
2	www.nature.com Internet Source	<1%
3	journals.plos.org Internet Source	<1%
4	Diana O. Rios-Szwed, Elisa Garcia-Wilson, Luis Sanchez-Pulido, Vanesa Alvarez et al. "FAM111A regulates replication origin activation and cell fitness", Cold Spring Harbor Laboratory, 2020 Publication	<1%
5	theses.gla.ac.uk Internet Source	<1%
6	Submitted to University of Cape Town Student Paper	<1%
7	fibrogenesis.biomedcentral.com Internet Source	<1%
8	irep.ntu.ac.uk Internet Source	

Table of Contents

Acknowledgements

I am overwhelmed in all humbleness and gratefulness to acknowledge all those who contributed to my thesis directly and indirectly.

Firstly, thank you God for your guidance and providing me with the strength and courage to overcome any challenges.

I would like to thank my esteemed supervisor, Dr. Afolake Arowolo, for her invaluable supervision, support, and motivation during the course of my master's degree. Her vision, ambition, and encouragement are what carried me through the past few years in the lab.

A huge thank you to Prof. Khumalo who has accepted me into the Hair and Skin Research Lab and who has provided intellectual insight and profound knowledge in research and in life which has steered me through my research.

My gratitude extends to the National Research Foundation for funding me the opportunity to undertake my studies at the Department of Medicine, The University of Cape Town.

Thank you to the patients who so kindly and generously donated samples for this research, your willingness to help research for the benefit of future patients is admired.

My appreciation also goes out to my parents, Lorraine and Mark Rhoda for their encouragement and support throughout these very intense academic years. Your prayers, love and sacrifices for educating and preparing me for my future does not go unnoticed. Along with the rest of my family, you have been my greatest source of strength and peace. To my partner, Liam, who has supported and encouraged me throughout my studies. Thank you for listening to my rants about failed experiments and infected cells, for your weekend trips with me to the lab, and reminding me to do my best while living life to the fullest.

Finally, thank you to every single person who has supported me in this research project.

List of tables

List of figures

List of abbreviations

A549: Adenocarcinoma 549

ABC: ammonium bicarbonate

ACN: Acetonitrile

ANOVA: analysis of variance

BAG3: BCL2-Associated Athanogene 3

BCA: Bicinchoninic Acid

BSA: Bovine Serum Albumin

CAFs: carcinoma-associated fibroblasts

CANP: cancer associated nucleoprotein

Cas9: CRISPR associated proteins

cDNA: complementary deoxyribonucleic acid CI: cell index

cm: centimetre

CO2: Carbon dioxide

CRISPR: clustered regularly interspaced short palindromic repeats)

DAPI: 4′,6-diamidino-2-phenylindole

dH_20: distilled water

DMEM: Dulbecco's Modified Eagle Medium

DNA: deoxyribonucleic acid

DPC: DNA-protein crosslink dsDNA: double stranded DNA

ECL: Enhanced chemiluminescence

ECM: extracellular matrix

EDTA: Ethylenediaminetetraacetic acid

EGFP: Enhanced green fluorescent protein

ES: enrichment score

FAM111B: FAM111 Trypsin Like Peptidase B (gene)

FAM111B: FAM111 Trypsin Like Peptidase B (protein)

FASP: Filter Aided Sample Preparation

FBS: Fetal Bovine Serum

FITC: fluorescein isothiocyanate

GAPDH: Glyceraldehyde 3-phosphate dehydrogenase

GO: Gene Ontology

GSEA: gene set enrichment analysis

HCD: High-energy collision dissociation

HCl: Hydrogen chloride

HDR: homology-directed repair

HEK293: Human embryonic kidney 293

HFP: Hereditary Fibrosing Poikiloderma

HGP: Human Genome Project

HPLC: high performance liquid chromatography

HRP: Horseradish Peroxidase

Hrs.: Hours

HSP: Hereditary Sclerosing poikiloderma

HSR: Hair and Skin Research

IAA: iodoacetamide

iBAQ: intensity based absolute quantification

IF: Immunofluorescence

IPF: idiopathic pulmonary fibrosis kDa: Kilodalton

KEGG: Kyoto Encyclopaedia of Genes and Genomes

LC-MS/MS: Liquid chromatography coupled with tandem mass spectrometry

LUAD: Lung Adenocarcinoma

mg: milligram

mm: millimetre

mM: milli molar

mm: millimetre

mRNA: messenger Ribonucleic Acid

MS: Mass spectrometry

NaCl: sodium chloride

NCBI: National Centre for Biotechnology Information NCBI

NGS: Next generation sequencing

NHEJ: non-homologous end joining

nm: nano meter

NT: non-targeting

NX: Normalized expression p-value: probability value

PBS: Phosphate buffer saline

PBST: phosphate buffer saline tween

PCR-RFLP: Polymerase Chain Reaction- Restriction Fragment Length Polymorphism Analysis

PFA: Paraformaldehyde

PI: Propidium iodide

POIKTMP: Poikiloderma with Tendon contractures, Myopathy and Pulmonary fibrosis

PS: Penicillin Streptomycin

PVDF: Polyvinylidene difluoride

Q1-4: Quadrant 1-4

qPCR: quantitative polymerase chain reaction

RISC: RNA-induced silencing complex

RNA: ribonucleic acid

RNAi: ribonucleic acid interference

RT-qPCR: Reverse transcription quantitative polymerase chain reaction

RT: reverse transcription

RT: room temperature

RTS: Rothmund-Thomson syndrome

SD: standard deviation

SDS-PAGE: Sodium Dodecyl Sulphate Polyacrylamide Electrophoresis

SDS: sodium dodecyl sulphate

SEM: standard error of mean

sh-RNA: short hairpin -RNA

siRNA: small interfering ribonucleic acid SNP: single nucleotide polymorphisms

TAFs: tumour-associated fibroblasts

TBS: Tris Buffered Saline

TBST: Tris Buffered Saline

TFA: Trifluoroacetic acid

TME: tumour microenvironment

TMT: tandem mass tag

tracrRNA: trans-activating CRISPR-RNA

Tris-HCl: Tris hydrochloride

U/ml: units per millilitre

UCT: University of Cape town

UHPLC: ultra-high performance liquid chromatography

USA: Unites States of America

UT: Untreated

UV/VIS

V: voltage

α-SMA: smooth muscle alpha-actin

β-Tubulin: beta tubulin

µl: microliter

µM: micromolar

Abstract

POIKTMP, a multi-systemic fibrosing disease, results from mutations in the human *FAM111B* gene. Studies have also suggested high expression of this gene in cancers. Despite rising interest in the pathological effects of *FAM111B* mutations and overexpression of *FAM111B*, knowledge of the physiological role of this gene remains limited. Therefore, this study sought out to provide insights into the cellular function of *FAM111B* and to investigate the pathological effect of the *FAM111B* Y621D mutation.

First, bioinformatics studies coupled with quantitative PCR and Western blots analysis were employed to assess *FAM111B* gene and protein expression in cancerous and non-cancerous cell lines. Subsequently, *FAM111B* gene expression was downregulated and upregulated in the human fibrosarcoma (HT1080) cell line by RNA-interference mediated gene silencing and recombinant gene expression technologies. The effect of these *FAM111B* dysregulations was studied using cell-based functional assays: proliferation, apoptosis, migration, and invasion assays. Furthermore, the functional pathways and interacting proteins of the FAM111B protein was determined using mass spectroscopy proteomics. Finally, preliminary studies in a POIKTMP patient-derived fibroblasts were attempted to recapitulate the results obtained using the HT1080 cell line.

The results from this study indicated that *FAM111B* gene and protein overexpression occurs in cancer cells. Second, the depletion of *FAM111B* suggests a decelerated rate of cell proliferation and migration (14%), and increased apoptosis (1.4-fold). Conversely, overexpression of *FAM111B* resulted in a marked reduction in apoptosis (3-fold) and increased cell migration by 27 %, howbeit, no evidence of increased proliferation. Furthermore, Y621D *FAM111B* mutant cells showed reduced expression of *FAM111B*, decreased apoptosis (1.1-fold), cellular invasion (24%), and indicates an increase in cell proliferation and migration (18 %). The proteomics data suggested wild-type *FAM111B* interacts with HSP7C, a molecular chaperone, which alongside BAG3 and BCL2 to minimise apoptosis. Similarly, Y621D's interaction with G3V3W4, a component of the 20S proteasome complex involved in the proteolytic degradation of damaged proteins, may suggest the rapid clearance of this mutant protein.

Altogether, our data suggest that *FAM111B* promotes the viability of cells and the mutation of the Y621D mutant contributes to the rapid proteolytic clearance of *FAM111B*, thereby leading to reduced cellular fitness.

Chapter 1: Literature review

1.1 Fibrosis

Fibrosis can be defined as the accumulation of extracellular matrix (ECM) proteins, consisting of collagens, elastin, and fibronectin, resulting in scarring, stiffening, and thickening of the affected tissue (Tracy, Minasian, & Caterson, 2016). Fibroblasts are mesenchymal cells that aid in the synthesis of collagen, elastin, and fibronectins. They are essential in supporting and maintaining the structural framework of tissues. Fibroblasts are activated in response to tissue injury by profibrotic cytokines. During this activation, fibroblasts cells change their phenotype from quiescent cells involved in the slow turnover of ECM to highly proliferative and contractile myofibroblasts that secretes more collagen (Neary, Watson, & Baugh, 2015). Myofibroblasts are cells that express protein alpha-smooth muscle actin (αSMA). Connective tissue remodelling is controlled by merging the ECM synthesizing characteristics of fibroblasts with cytoskeletal features of contractile smooth muscle cells (Hinz et al., 2007). It is generally accepted that fibroblasts to myofibroblasts differentiation represents an important event during the physiological reparation of connective tissue after injury. The fibroblasts-myofibroblasts differentiation phenomenon plays a vital role in morphogenesis, organogenesis, inflammation, wound healing and fibrosis in most organs/tissues (Powell et al., 1999). During wound healing, myofibroblasts express various proteins such as collagen, proteoglycans (i.e., fibronectin, tenascin, laminin), and glycoproteins (i.e., antibodies, hormones, and immunoglobulins) (Hinz et al., 2007), which aids in the tissue regeneration. Wound healing involves four meticulously programmed overlapping phases, including the haemostasis, inflammation, proliferative, and remodelling phases (Demidova-Rice, Hamblin, & Herman, 2012; Gosain & DiPietro, 2004). Haemostasis involves forming scar tissue as the blood flow constricts and a fibrin clot forms. The clot releases cytokines and growth factors that control the bleeding. The second phase, the inflammatory phase, involves recruiting neutrophils, macrophages, and lymphocytes, which “decontaminates” the wound by removing foreign debris and apoptotic cells. The proliferative phase occurs after the inflammatory phase when epithelial cells proliferate and migrate to the wounded area. The final phase (i.e., the remodelling phase) is the most prolonged step of wound healing involving the deterioration of capillaries to enable vascular to return to normal functioning in the wound.

In physiological tissue repair, myofibroblasts undergo apoptosis or revert to the quiescent fibroblast phenotype; therefore, the myofibroblast responses diminish (Neary et al., 2015). However, during pathological fibrosis, these myofibroblasts do not undergo apoptosis and, therefore, persist in tissue coagulation, resulting in fibrosis and impairing normal organ functions (Neary et al., 2015).

Idiopathic pulmonary fibrosis (IPF) is characterized by chronic, progressive fibrosis associated with the decline in lung function resulting from scar tissue build up in the lung leading to respiratory failure, and has a high mortality (Barratt, Creamer, Hayton, & Chaudhuri, 2018). A European study interrogating the incidence of IPF showed that about 0.5 in 10000 people were affected annually. These statistics increase over time (Ley, Collard, & King, 2011), while studies in Asia, America, and Africa showed lower rates. However, the global incidence of IPF estimation is more elevated in Asian (0.35 to 1.30 per 10000 persons) and North American (0.75 to 0.93 per 10000 persons) countries compared to Europe (0.09 to 0.49) (Maher et al., 2021). The cause of IPF disease is variable, and the average survival of patients is 2-3 years while some patients live longer. The leading cause of death in patients with IPF is lung failure due to the progression of the disease. Age, gender, and climatic conditions are some of the main factors that affect the mortality rate (Olson, Swigris, Raghu, & Brown, 2009).

1.2 Fibrosis vs cancer

Cancer develops within a tumour microenvironment (TME), supporting tumour survival, growth, and metastasis. The TME comprises immune cells, stromal cells, vasculature, and signalling molecules. It is reported that approximately 20% of cancers are associated or linked to chronic inflammation-related fibrosis (Chandler, Liu, Buckanovich, & Coffman, 2019).

Inflammation is the body's defence mechanism in response to tissue damage and infection. Failure of the body's defence mechanism to facilitate in tissue healing can lead to chronic inflammation, permanent tissue damage and fibrosis. The inflammatory response can be categorized into stages, acute and chronic inflammation, in which respective inflammatory cytokines secreted. The cytokines involved in acute inflammation are namely, Interleukin (IL) -1, Tumour necrosis factor (TNF)-a, IL-6, and IL-11. The cytokines that are secreted during chronic inflammation are grouped into those involved in humoral inflammation, such as IL-3, IL-4, IL-5, IL-6, IL-7, IL-9, IL-10, IL-

13 and transforming growth factor (TGF)-b, while those involved in cellular inflammation are IL-1, IL-2, IL-3, IL-4, IL-7, IL-9, IL-10 IL-12, interferons (IFNs), IFN-g inducing factor (IGIF), TGF-b, and TNF-a and -b. However, pro-inflammatory cytokines such as IL-1 and TNF-a, promote inflammation resulting in tissue destruction and disease conditions, namely fibrosis (Feghali-Bostwick & Wright, 1997).

Fibrosis, as a result of chronic inflammation, may mimic cancerous tissue. There are fundamental hallmarks of cancer, such as genetic mutations, stimulated proliferation, and invasion which are similar to those of fibrotic conditions (Coussens & Werb, 2002). The fibrotic process begins with tissue damage, wound repair occurs, involving haemostasis, inflammation, proliferation, and lastly, tissue remodelling. This process can halt or continue uncontrollably, which leads to a chronic non-healing wound that exacerbates the development of cancer-associated fibrosis. In other words, cancerous tissue imitates this fibrotic process by creating a fibrotic tumour environment that induces altered immune infiltration, chronic pro-inflammatory stimuli, leaking vasculature, and hypoxia (Chandler et al., 2019).

Chronic fibrosis and cancer may result from a persistent healing response; therefore, the similarities that tumours have in common with chronic fibrosis are interesting (Rybinski, Franco-Barraza, & Cukierman, 2014). Studies showed that chronic fibrosis (i.e. liver cirrhosis) caused by viral infections, alcohol abuse or toxic foods could directly predispose one to cancer (El-Serag, 2012). The dense myofibroblastic component with the excessive deposition of ECM is common in fibrosis and cancerous tissue; however, in cancerous tissue, they are referred to as tumour associated fibroblasts (TAFs) or carcinoma-associated fibroblasts (CAFs) (Rybinski et al., 2014). Chronic fibrosis propelled by TAF's and fibrotic myofibroblasts exaggerates ECM synthesis, causing specific paracrine effects on surrounding cells (Rybinski et al., 2014). TAFs express pro-tumorigenic growth factors, proteolytic enzymes, inflammatory cytokines and ECM proteins (Collagen I and II) (Cirri & Chiarugi, 2011). In addition, the role of TAFs is to degrade Collagen IV, associated with the epithelial ECM. There are alterations to the hyperproliferation of stromal compartments and mesenchymal ECM deposition near neoplastic lesions in cancerous tissue. Similarly, the modified ECM typically occurring in chronic fibrosis is also characterized by Collagen I and II depositions. Finally, fibrotic myofibroblasts and TAFs express smooth muscle

actin (α-SMA) and perform a similar contractile activity. Therefore, effective targeting drugs that successfully treat related fibrotic conditions may also stall a cancerous condition (Rybinski et al., 2014).

One profound difference between fibrotic and cancerous tissue is the presence of transformative mutations and epigenetic changes in cancerous epithelial cells, which may be absent in most fibrotic tissue (Coyle, Boudreau, & Marcato, 2017).

1.3 The human *FAM111B* gene and hereditary fibrosing poikiloderma, and cancers.

The *human family with sequence similarity member B* (*FAM111B)* gene located on chromosome 11q12.1 codes for a protein with a minimally characterized function (Hoffmann et al., 2020; Panjawatanan, Ryabets-Lienhard, Bakhach, & Pitukcheewanont, 2019). The expression of the FAM111B protein occurs in many parts of the body, including skin, skeletal muscle, trachea, lung, liver, pancreas, gastrointestinal tract, bone marrow, reproductive system, and brain. Studies have shown that mutations in *FAM111B* causes Hereditary fibrosing poikiloderma (HFP) (Goussot et al., 2017; Mercier et al., 2015; Mercier et al., 2013; Panjawatanan et al., 2019).

HFP is a rare autosomal disease first described in a South African family of European decent (Khumalo et al., 2006; Mercier et al., 2015). The family presented with distinctive phenotypic characteristics involving poikiloderma, tendon contractures and progressive fibrosis after which the syndrome was given the name Hereditary Sclerosing poikiloderma (HSP) with poikiloderma, tendon contractures, myopathy, and pulmonary fibrosis (POIKTMP) (Goussot et al., 2017).

HFP with POIKTMP is characterized by poikiloderma, telangiectasia (enlarged blood vessels visible on cheeks) and hypo- and hyper pigmentation on the face and sun-exposed areas from early childhood as well as muscle dominant fibrotic disorder which was contractures involving the ankles and feet causing a gait disturbance (Mercier et al., 2015). Around the age of puberty, the patients experience pancreatic insufficiency, progressive dyspnoea resulting from lung impairment linked to pulmonary fibrosis. The acronym POIKTMP was used to describe this specific HFP due to the involvement of this disorder with tendon contractures, myopathy, and pulmonary fibrosis. Additional characteristics shown in people with *FAM111B* gene mutations include red, scaly skin patches and mild lymphoedema of the arms and legs; palmoplantar keratoderma, the thickening of

skin on the palms of the hands and soles of the feet; and similarly, the thickening or hardening of the skin of fingers and toes termed sclerosis. Patients with this disorder usually have sparse scalp hair, eyelashes, and eyebrows. Affected individuals have a decreased ability to sweat therefore impairing their ability to tolerate heat. People with POIKTMP have a deficiency of enzymes that digests fat which leads to diarrhoea and poor absorption of fat and fat-soluble vitamins. However, intellectual development is not affected (Chasseuil et al., 2019; Goussot et al., 2017; Kazlouskaya, Feldman, Jakus, Heilman, & Glick, 2018; Küry et al., 2015; Mercier et al., 2015). The prevalence of POIKTMP is low mainly due to health care providers not recognizing the disorder as well as patients showing only one or few features of the disorder. At least 36 affected individuals have been described till date. In 50% of the cases, individuals inherit mutation in an autosomal dominant pattern, which enables only one copy of the mutation in each cell to cause the disorder. In the other half of the cases, new mutations occur spontaneously in individuals with no history of the disorder in the family (Chasseuil et al., 2019; Goussot et al., 2017; Mahajan, Sunkwad, Darkase, & Khopkar, 2020).

The first South African family to be diagnosed with POIKTMP presented with clinical appearances similar to that of the Weary form of poikiloderma, however it appeared to be to be a unique autosomal dominant condition, Werner syndrome(Khumalo et al., 2006). To date, different forms of poikiloderma conditions are considered when differentially diagnosing POIKTMP, namely Rothmund-Thomson syndrome (RTS) and Werner syndrome. RTS is characterized by poikiloderma, alopecia, small stature, skeletal and dental abnormalities, and an increased risk of cancer and cataracts. The identification of variants in *ANAPC1* or *RECQL4* are required to confirm RTS diagnosis. Similarly, Werner syndrome, caused by a mutation in the *WRN* gene, is characterized by alopecia, skin ulcers, cancer predisposition, growth retardation, cataracts, type 2 diabetes, and osteoporosis (Larizza, Roversi, & Volpi, 2010).

The varied phenotypic characteristics of POIKTMP patients are a result of multiple mutations of the *FAM111B* gene amongst patients. To date, at least eleven mutations of the *FAM111B* gene have been identified in patients with POIKTMP (F. Chen et al., 2019; Dokic et al., 2020; Goussot et al., 2017; Khumalo et al., 2006; Mahajan et al., 2020; Mercier et al., 2015; Panjawatanan et al., 2019; Seo et al., 2016; Takeichi et al., 2017; Zhang et al., 2019).

This protein contains a putative peptidase domain in the C-terminus (Figure 1.1). It is suggested that these missense mutations alter the protease activity in the body which explains the varied signs and symptoms of POIKTMP (*FAM111B* gene, 2017) (Kazlouskaya et al., 2018).

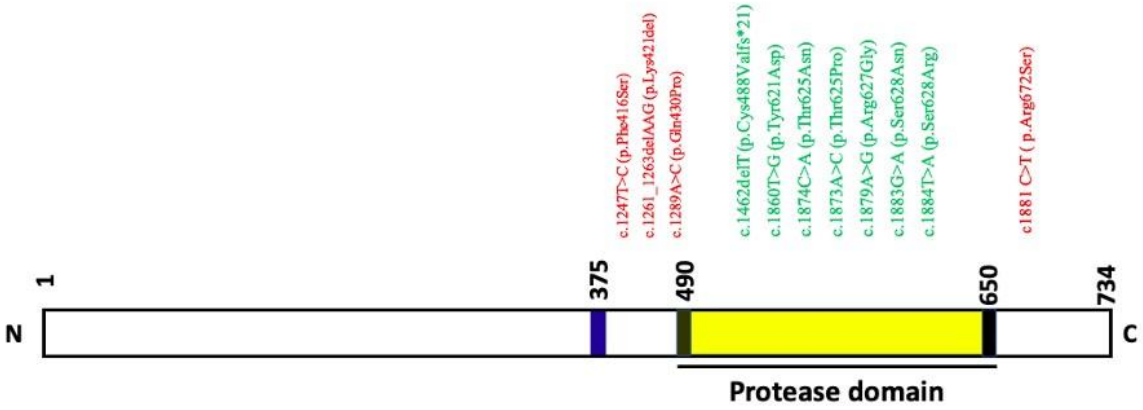


Figure 1. 1 ***A schematic diagram of the FAM111B protease domain illustrating the position of mutations within the catalytic domain (GREEN) and the mutations outside the catalytic domain (RED***

(From manuscript in preparation, 2021).

The FAM111B protein is made up of a trypsin-like peptidase domain located in the C-terminus. The genetic variations in FAM111B gene are associated with Hereditary fibrosing poikiloderma involving tendon contractures, muscle weakness and pulmonary fibrosis.

Although the possible substrates or interacting protein partners of the FAM111B protein is not well established, a recent article suggested that FAM111B interacted with FAM111A (a paralogue protein), with both proteins playing significant roles in DNA replication (Rios-Szwed et al., 2020). However, they also have their independent functions, FAM111B is produced in S-phase of the cell cycle which may indicate the genes association to DNA replication (Aviner, Shenoy, Elroy-Stein, & Geiger, 2015). It was discovered that like FAM111A, FAM111B localized in the nucleus briefly to newly replicated chromatin and when *FAM111B* was downregulated the DNA synthesis was also reduced (Rios-Szwed et al., 2020). In 2019, an article showed that FAM111B plays a role in cell proliferation, migration and apoptosis (Sun et al., 2019). They also performed a Kyoto

Encyclopaedia of Genes and Genomes (KEGG) enrichment analysis which indicated a high correlation of FAM111B with CCNB1 and CDC25C, both of which are imperative in the G2M phase of the cell cycle. Interestingly, an *in silico* analysis by Sun *et al.*, discovered that FAM111B interacts with proteins BAG3 (BCL2-Associated Athanogene 3) and BCL2, therefore, influencing programmed cell death by causing a decline in apoptosis in cancerous tissue resulting in tumours (Sun et al., 2019). Despite all the research done on the *FAM111B* gene, still little is known about its function.

1.4 Functional genomics

1.4.1 What is functional genomics?

On April 14th, 2003, the Human Genome Project (HGP) was announced complete (Hood & Rowen, 2013). The project aimed to identify new genes and map all genes of the human genome. Despite the studies completed, the main challenge of functionally characterizing newly discovered genes remains. Since mutations in the human genome are likely to cause pathological conditions, like HFP with POIKTMP, functional studies are fundamental to understand how these mutations impact human health. Functional genomics is a field of molecular biology that investigates and determines possible functions of genes and non-genetic material of an organism with the support of studies in transcriptomics, proteomics and metabolomics (Hernández-Domínguez et al., 2019). This field involves high-throughput techniques that focus on the interactions of genes and proteins and the transfer of genetic information through transcription and translation while also investigating the regulation of gene expression and protein-protein interactions (Bunnik & Le Roch, 2013). For instance, in 2019, a functional genomics study by Haijun Sun *et al.* suggested a role of FAM111B in cell proliferation, apoptosis, migration and invasion of cancerous cell lines (Sun et al., 2019). However, further studies are required to validate these claims and provide more mechanistic insights into these functions.

1.4.2 Techniques required to study functional genomics

1.4.2.1 Mutagenesis

Mutagenesis is the process of inducing a genetic mutation. These mutations may occur spontaneously or generate by mutagens. A technique known as transposon mutagenesis creates mutations, resulting in random gene knockouts. Site-directed mutagenesis utilizes PCR to introduce desired mutations. The ability to quickly and precisely mutate a protein is vital for mechanical and functional studies (Foster, Oster, Mayer, Avery, & Audus, 1998).

1.4.2.2 RNAi/ gene knockdown approach

Studying the effects of the reduced gene expression can show the physiological role of the gene product (Voorhoeve & Agami, 2003). The RNA interference (RNAi) - mediated gene knockdown approach is widely utilized to analyse the functional characteristics of genes *in vitro* and *in vivo* models.

There are two types of RNAi molecules, small interfering RNA (siRNA) and short hairpin RNA (shRNA). siRNAs are double-stranded RNA molecules, 20-25 nucleotides in length.

The shRNA is approximately 80 base pairs in length consisting of an internal hybridized region that creates a hairpin structure. They are processed in the cell to form siRNA that down-regulated gene expression (Figure 1.2) (Schütze, 2004). Figure 1.2 illustrates how double-stranded DNA (dsDNA), and shRNA are introduced into the cells where it is both processed into shorter siRNA and sometimes even small micro-RNA (miRNA) by the Dicer molecule. While siRNA forms a complex with RISC and cleaves the targeted complementary RNA, miRNA forms protein complexes (miRNP) that targets complementary RNA and inhibits the translation of proteins.

The carefully designed target specific siRNA targets all the variants of a particular gene sequence with the aid of the Dicer molecule (encoded by the DICER1 gene), an enzyme crucial for microRNA biogenesis (Lai, Murgia, Parkkinen, Domanskyi, & Vinnikov, 2019). The Dicer molecule and R2D2, a protein that transfers siRNA onto Ago2 (the central component of RISC), forms an RNA-induced silencing complex (RISC). The resulting siRNA with RISC induces the unwinding of the ds-RNA (Dominska & Dykxhoorn, 2010). Once the ds-RNA unwound, the

antisense RNA strand directs RISC to the targeted mRNA's complementary site, where it activates the endo-nucleolytic activity of Ago2 results in mRNA cleavage (Figure 1.3).

RNAi, specifically siRNA used in this study, has been used to study the function of genes. They are easily introduced into cells, has a high efficiency rate, are easily and rapidly produced, and has reduced off-target effects compared to alternative gene-altering techniques. These are important characteristics that enables the study of genetic functions in vitro (Campeau & Gobeil, 2011).

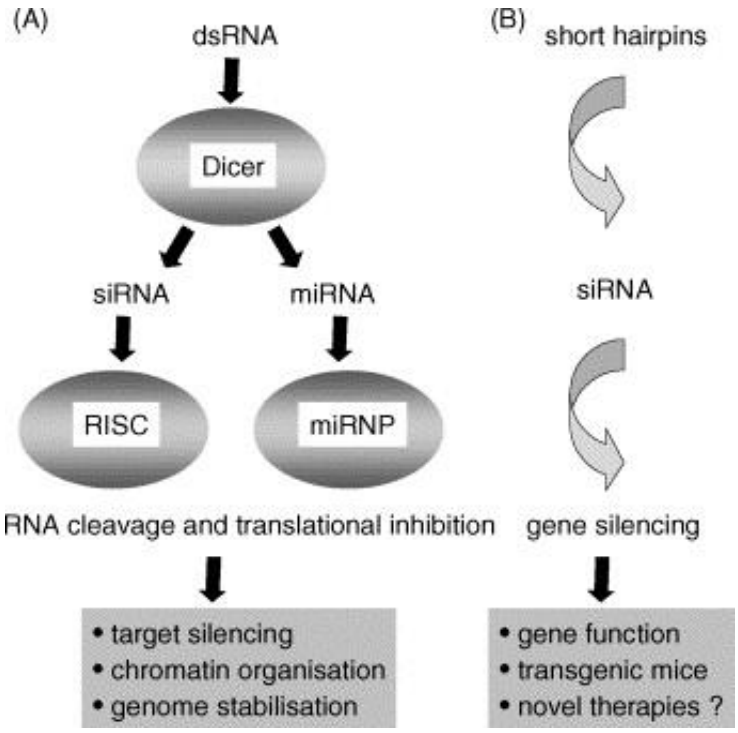


Figure 1. 2 ***Figure 1.2 RNAi as an editing tool for gene function analysis.***

(A) dsDNA is cleaved by the Dicer multiprotein complex to form siRNAs or miRNAs. siRNAs form a complex with RISC that mediates the targeted gene silencing by RNA cleavage. miRNAs form a miRNA-protein complex (miRNP) that is transported to the complementary RNA where it inhibits the translation of proteins. (B) shRNA that are transported into cells via a vector, are converted to siRNA resulting in gene silencing (Schütze, 2004).

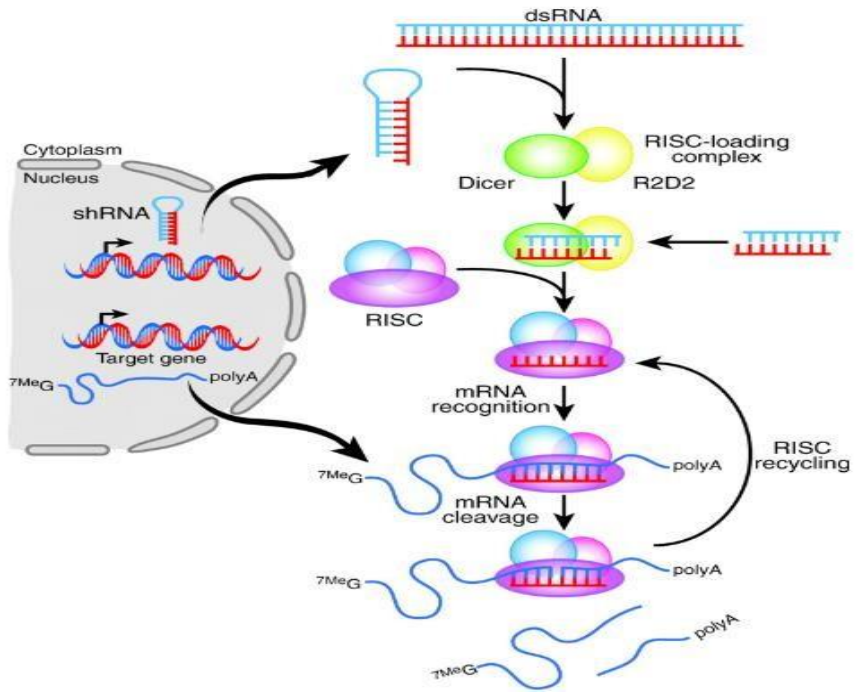


Figure 1. 3 ***<u>The structure and mechanism of siRNA-mediated gene knock-down</u>.***

The siRNA are molecules that facilitates the gene silencing. siRNA molecules are processed from dsRNA and shRNA by a DICER molecule along with the R2D2. The siRNA is taken up by the RISC where the two strands are dissociated into single anti-sense strands. These antisense strands guide the RISC to the targeted mRNA where the mRNA is cleaved (Dominska & Dykxhoorn, 2010).

1.4.2.3 CRISPR/Cas9 screens

CRISPR (clustered regularly interspaced short palindromic repeats) is a natural bacterial immune response utilized by bacteria as a defence mechanism against viruses; when a virus/phage infects a bacterium, the bacterium stores a piece of the viral DNA. The bacterium produces a guide RNA made up of two strands of DNA; one is complementary to the virus, called CRISPR RNA (crRNA), and the other is known as tracrRNA (trans-activating crRNA) that interacts with the cas9 proteins (CRISPR associated proteins). When the virus reinfects the bacterium, the two-RNA structure directs the Cas9 proteins to introduce double-strand breaks in target DNA (Karvelis et al., 2013). The Cas9 proteins attach to the targeted sequence of the viral DNA, unzips the DNA, and matches it to the guide RNA strand. When the match is complete, the Cas9 HNH nuclease domain and Cas9 RuvC-like domain cut the viral DNA by cleaving the complementary and non-complementary strand, respectively. When the cut is repairing, it is prone to error, resulting in a random mutation.

Researchers have used this technique to modify a targeted gene. CRISPR is often used to knock out genes, allowing scientists to observe the phenotypic consequences of loss of gene function.

Additionally, CRISPR can also insert a new gene or segments of DNA into the target locus. For example, a single nucleotide can be changed to model a specific mutation of interest. These knockout and knock-in capabilities make CRISPR a powerful tool to elucidate gene functions. As described previously, knocking out a gene with this technology involves the insertion of the CRISPR-Cas9 into a cell using a guide RNA as a targeting tool to identify the gene of interest. Once the guide RNA locates the targeted gene, Cas9 cuts this gene. The cell's DNA repair mechanism repairs the cut using non-homologous end joining (NHEJ). This repair mechanism is efficient; however, it is inaccurate and introduces errors in the DNA as insertions or deletions that usually knock out the gene (Albadri, Del Bene, & Revenu, 2017).

On the other hand, knocking in a gene requires a mechanism called homology-directed repair (HDR), in which the repair of double-stranded DNA breaks utilizes a DNA sequence known as a repair template to insert a matching DNA sequence into the break (Figure 1.4). This mechanism has been used in research which involved modifying a gene or promotor sequence, inserting fluorescent proteins and creating a SNP for a disease model (Banan, 2020).

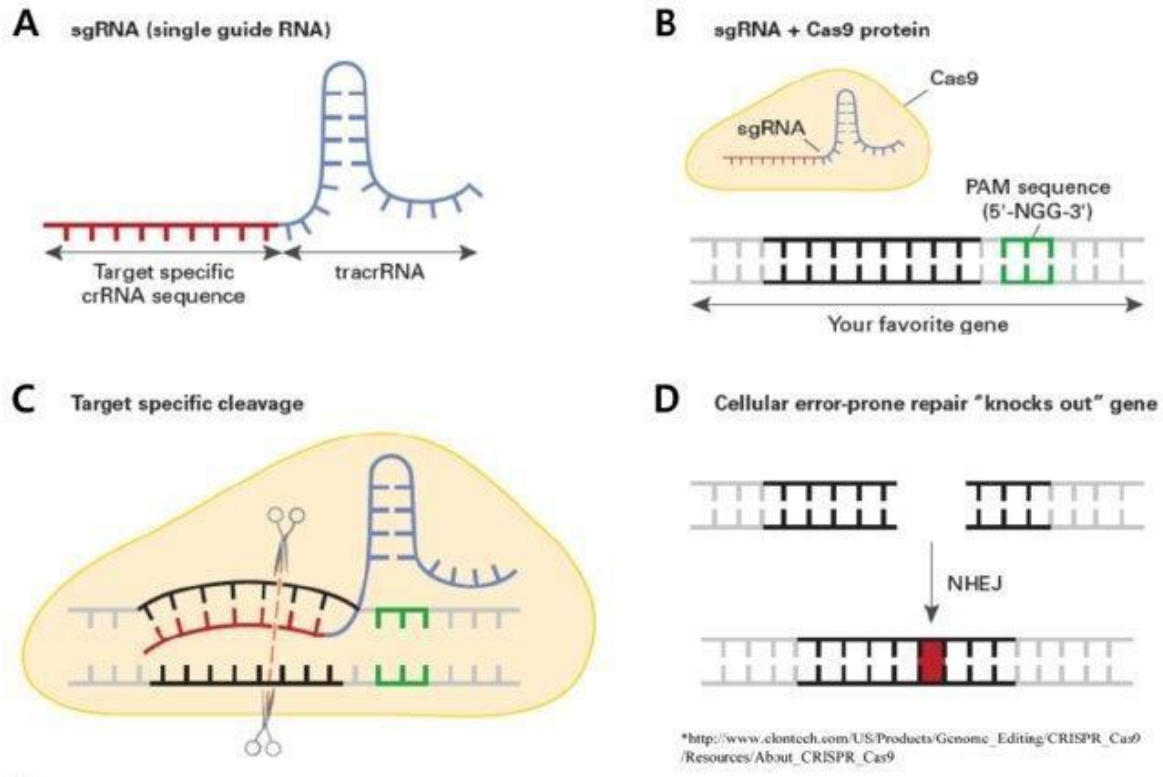


Figure 1. 4 ***<u>The mechanism of CRISPR/Cas9- mediated gene interruption</u>.***

A single guide RNA (sgRNA), composed of a crRNA sequence (complementary to target DNA) and a tracrRNA (interacts with Cas9 protein) (A), binds to a Cas9 protein and is directed to the targeted dsDNA (B). At the site of the dsDNA, the guide RNA binds to the complementary dsDNA sequence. The Cas9 has DNA endonuclease activity which cleaved the dsDNA sequence (C). A nonhomologous end-joining (NHEJ) process repairs this cleaved site, resulting in either insertions or deletions to disrupt the gene's function (D) (Song et al., 2017).

1.4.2.4 RNA sequencing

RNA-sequencing enables the examination of the quantity and sequences of RNA in a sample using next-generation sequencing (NGS) that analyse the continuously changing transcriptome (Chu & Corey, 2012). It allows us to investigate the total RNA content, including messenger RNA (mRNA), ribosomal RNA (rRNA) and transfer RNA (tRNA). RNA-sequencing provides insight into which genes are activated, at what times they are activated, and their expression level (Ozsolak

& Milos, 2011) during the treatment of a specific condition and changing developmental stages. This information is vital for scientists to analyse the changes that may result in a disease or disorder and the function of genes. Scientific studies that use RNA-sequencing are differential gene expression analysis, SNP identification, transcriptional profiling, and RNA altering (Han, Gao, Muegge, Zhang, & Zhou, 2015).

The experimental method of RNA-sequencing involves generating cDNA sequences by reverse transcribing whole RNA molecules, followed by library quantification and sequencing with NGS technology (Kolker et al., 2004) (Figure 1.5).

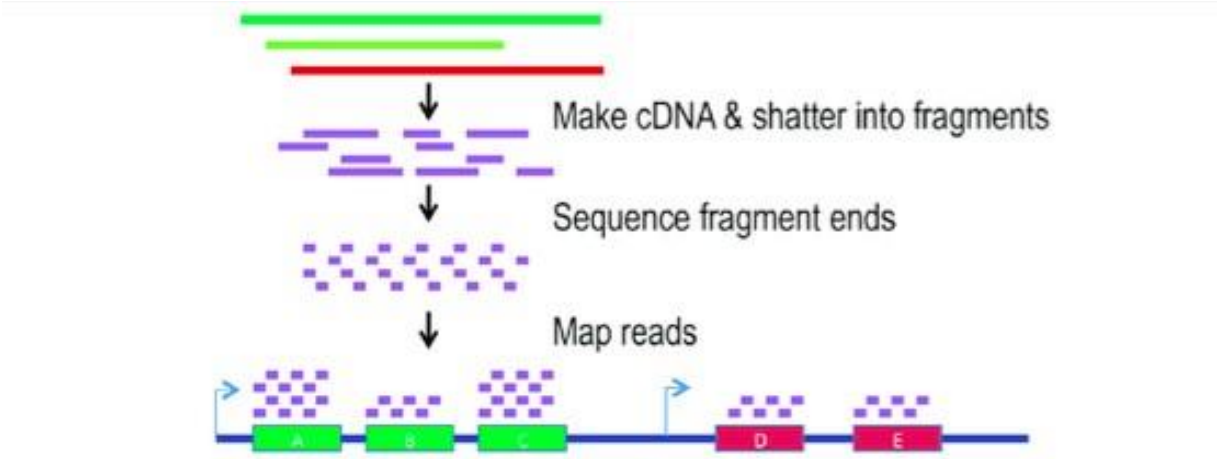


Figure 1. 5 ***A graphic demonstration of the RNA sequencing experiment.***

Extracted mRNA is converted to cDNA by reverse transcription and then fragmented. Fragments that fall within a specified length are ligated with sequence adapters followed by amplification of DNA fragment ends to create single or paired-end reads (Mangul et al., 2013).

1.4.2.5 Proteomics mass spectrometry

Proteomics is the characterization of an organism's proteome consisting of all proteins present in its system. Many conditions can alter the protein content, enabling researchers to investigate protein expression levels between two states. Mass spectrometry (MS) plays an essential role in analysing proteomes and identifying all proteins present in a biological system. However, before

identifying proteins by MS, high-performance liquid chromatography-based (HPLC) separations must be used to categorize proteins (Iadarola, 2019). An article published in 2020 (Rios-Szwed et al., 2020) uses tandem mass tag (TMT) quantitative MS to identify FAM111A putative substrates in whole-cell extracts under three conditions: untreated, si-FAM111A and FAM111A overexpression. FAM111A substrate abundance is increased upon FAM111A overexpression and reduced in the si-FAM111A condition. They found 6928 up-regulated factors in both FAM111A overexpression and si-FAM111A compared to the control. The two factors that qualified as FAM111A substrates, HMCES and RPL26L, were proteins involved in DNA replication and apoptosis. However, FAM111B protein abundance was unaffected, supporting the hypothesis that FAM111B might be a cofactor of FAM111A rather than a substrate (Rios-Szwed et al., 2020).

Thus, protein changes detected in conditions may identify clinically relevant biomarkers. Hence, MS can be used in proteomics to identify diagnostic biomarkers present in biological fluids. Therefore, liquid chromatography coupled with tandem mass spectrometry (LC-MS/MS) is a fundamental investigative tool in proteomics (Tsai et al., 2015).

1.4.3 Techniques to validate or quantify gene expression

Gene expression is a process that links the information encoded within a gene to the final function and structure of all living cells (Volgin, 2014). There are several techniques available for studying and quantifying gene expression regulation. Gene expression is measured by quantifying levels of the gene products, proteins and RNA. At transcriptional levels, mRNA (Messenger RNA) expression is usually measured and quantified by RT-qPCR (Reverse transcription-quantitative polymerase chain reaction). In contrast, protein expression may be measured and quantified with Western blot and Immunofluorescence confocal imaging at translational levels.

1.4.3.1. Western blot analysis

This assay method permits immunodetection, characterisation and quantification of proteins in cell homogenates. It also allows detection of protein modification or degradation even at low abundance (Martins-Gomes & Silva, 2018). This technique has three key steps: separation by size, transfer to a membrane and identifying the target protein using antibodies. Following the extraction of protein from cells, the western blotting method identifies target proteins with antibodies specific

to these proteins (Mahmood & Yang, 2012). There are two methods of antigen detection in Western blotting, direct and indirect method. The direct method is a one-step method consisting of a single primary antibody conjugated to a fluorescent dye. This conjugated primary antibody is able to detect the protein antigen in a rapid process due to minimal incubation times. This method is not widely used due to high cost of conjugated primary antibodies, a weaker signal in detected, and the conjugate may interfere with the primary antibody's attachment to the antigen. The second method is antigen detection, indirect method, includes a primary antibody which directly attaches to the antigen. The membrane is washed and the secondary antibody, which s conjugated with a fluorophore, is incubated. Although a longer process, there are many advantages of the indirect method of antigen detection due to its amplified signal, reduced cost as one secondary antibody can detect multiple primary antibodies, and improves specificity (Lewis, 2016). A chemiluminescent substrate is then used to visualise the protein bands.

1.4.3.2. Immunofluorescence confocal imaging

This immunofluorescence (IF) imaging method makes it easier to resolve and quantify small, faint structures in a cell by eliminating light that is out of focus, thereby improving the image's contrast by reducing the background noise (Peterson, 2010). The determination of colocalization and protein-protein interaction may also be by confocal microscopy using fluorescent makers (Moser, Hochreiter, Herbst, & Schmid, 2017). Confocal microscopy is a form of fluorescence microscopy that generates high-resolution images of cells or tissue stained with fluorescent-tagged probes (Peterson, 2010). This technique involves two methods, indirect and direct. The indirect method labels cellular proteins with targeted primary antibodies and fluorochrome-conjugated secondary antibodies, while direct labels with fluorochrome-conjugated primary antibodies. The staining technique allows for a multiplexed observation of subcellular localization, expression levels and the presence/absence of two or three proteins in one sample. This multiplexed analysis saves both time and reagents as it can study a variation of cells while targeting multiple markers in tissue. In understanding the mechanism of IF, the high energy light is absorbed, which activates the fluorochrome to emit light at its wavelength, therefore allowing the detection of antigen-antibody complexes (Donaldson, 1998).

1.4.3.3 Reverse transcription-quantitative polymerase chain reaction (RT-qPCR)

RT-qPCR is a technique used to quantify mRNA. Reverse transcription makes a DNA template from mRNA. The DNA template (also known as cDNA (complimentary DNA)) undergoes amplification, during which hybridization probes emit varying degrees of fluorescence. The fluorescence detected is used to measure the original number of mRNA copies. The quality and purity of the RNA template are essential for a successful RT-qPCR. There are two main types of qPCR methods, probe-based and dye-based qPCR. Probe-based qPCR requires designed and optimised target-specific probes in addition to primers. Hydrolysis probes containing a 5' reporter fluorophore and a 3' quencher complement the specific target sequence are the most widely used (BioLabs). The fluorescence signal during the PCR cycle resulting from the quencher being cleaved from the reporter fluorophore is proportional to the probe-target sequence present in the sample. This probe-based technique allows for the quantification of multiple targets in a reaction by using different fluorescent dyes for each amplicon-specific probe. The second qPCR technique, dye-based, measures DNA amplification in real-time as it occurs during PCR. This method is popular due to its cost-effective and fast manner of studying large numbers of targets using two sequence-specific primers. SYBR Green is a widely used fluorescent dye in qPCR, which displays fluorescence when it binds to double-stranded DNA (dsDNA). Like probe-based qPCR, the increase in fluorescence is directly proportional to the amount of dsDNA present (BioLabs; Jalali, Zaborowska, & Jalali, 2017).

1.4.4 Cell-based functional assays

1.4.4.1 Real-time cell analysis (RTCA)

Several cell-based functional assays are valuable for the functional characterization of a gene. The real-time cell analysis (RTCA) is one such assay that monitors the change of migration, proliferation, and invasion of cells in real-time with the xCELLigence RTCA System. The impedance, resistance of an electric current of gold microelectrodes in RTCA systems when cells are not present or not adhered onto the electrodes is determined with ionic cell culture medium solution. When cells attach to the surface of the electrode, it alters the ionic medium of the electrode solution, resulting in resistance of the electric surface. This resistance is known as

impedance. The higher the proliferation and adhesion rate of the cells, the higher the impedance. This data is reflected as the cell index (CI), representing the cell number and ratio of cells at different time intervals. Before cell adhesion, the CI would be set at zero and gradually increase with cell adherence. Therefore, the CI in an RTCA system results from the impedance induced by adherent cells to the electron surface (Türker Şener, Albeniz, Dinç, & Albeniz, 2017).

The calculation of CI is as follows:

$$CI = \frac{impedance\ at\ time\ point - imedance\ in\ the\ absence\ of\ cells}{nominal\ impedance\ value}$$

(Türker Şener et al., 2017)

1.4.4.2 Migration/scratch assay

The *in vitro* migration assay, also known as Scratch or Wound healing assay, is an efficient, affordable technique to measure cellular migration, mimicking the migration of fibroblasts after injury. The method involves creating a scratch representing the wound in a cell monolayer capturing the images at 6-12 hours intervals. The cells are then stained and imaged to compare the rate of migration of cells (Liang, Park, & Guan, 2007).

1.4.4.3 Invasion assay

The basement membranes are composed of extracellular matrices comprised of laminins, proteoglycans, and collagens in the human body. This basement membrane forms a sheet that divides the epithelial tissue from the adjacent connective tissue. Cancerous cells can cross this membrane and disseminate and penetrate tissue to cause secondary tumours (Hall & Brooks, 2014). Consequently, this *in vitro* study was developed to imitate the *in vivo* process of tumour invasion. A basement membrane matrix preparation, Matrigel, act as a matrix barrier and DMEM (Dulbecco's Modified Eagles Media) with 10% FBS as a chemoattractant. The results obtained from this assay portrays the *in vivo* ability of tumour cells to invade tissue (Shaw, 2005).

1.4.4.4 Cell apoptosis assay

Apoptosis is a process of programmed cell death for eliminating unwanted cells. There are various techniques which could be used to detect the apoptosis: fluorescent correlation spectroscopy, electron microscopy, western blotting, Annexin V staining, TUNEL assay, and flow cytometry (Martinez, Reif, & Pappas, 2010) . In cases where measuring a large number of cells are required, flow cytometry is the best method of detection. This method is capable of analysing multiple parameters simultaneously and sorting cell populations. Annexin V and Propidium iodide (PI) staining are routinely used to identify apoptotic cells. The first step of apoptosis involves the movement of the membrane phosphatidylserine (PS) from the inside to the outside of the plasma membrane surface. Annexin V, a Ca^{2+} dependent phospholipid-binding protein, has a high affinity for PS, therefore fluorochrome-labelled Annexin V (i.e., FITC) can be used for the detection of visible PS on the surface of cells using flow cytometry. PI is a stain that penetrates only dead or damaged cells. Interestingly, this method enables the differentiation of cells undergoing early apoptosis, late apoptosis, or necrotic processes. When staining with Annexin V in conjunction with a vital PI, you are able to detect viable cells (Annexin V-/PI-), early apoptotic cells (Annexin V+/ PI-), and late apoptotic or necrotic cells (Annexin V+/ PI+) (Biosciences, 2011).

1.4.5 Bioinformatics methods for functional genomics

The National Centre for Biotechnology Information (NCBI) defines bioinformatics as applying multidisciplinary fields such as mathematics, computer, and biology to facilitate discovering new biological ideas and analyse and interpret the data (Bayat, 2002) (Hernández-Domínguez et al., 2019). Bioinformatics plays a pivotal role in all aspects of protein analysis, including sequence analysis and structure analysis. It involves analysing DNA, RNA, and protein sequences. The analysis of sequences aids in determining potential functions for newly discovered proteins by comparing new sequences with known functional sequences or pathways. In studying the structure of proteins, structural bioinformatics analysis can help predict the interacting factors regulating the folding and stability of proteins. Bioinformatics studies have thus driven major research milestones, including sequence alignment, gene identification, gene expression and interacting proteins, prediction of protein structure and protein structure alignment (Nisbet, Elder, & Miner, 2009).

Furthermore, the functional interpretation of large-scale data from RNA, DNA and protein sequencing is a critically essential component of research to identify affected genes in particular diseases. Functional information is stored on databases like KEGG, Gene Ontology (GO) and Reactome which are accessible for researchers to discover information related to a particular biological phenotype (Subhash & Kanduri, 2016).

1.4.5.1 Functional enrichment analysis

Traditional biological functional analysis research approaches described in 1.4.2 above are generally time-consuming and inefficient in analysing large gene sets. Over time, high-thorough technologies are developed and improved to investigate many genes under specific biological conditions. The retrieval of lengthy gene lists from these technologies is analysed and delineated with public databases (e.g. Gene Ontology) to identify the most suitable and enriched biological phenotype (D. W. Huang, Sherman, & Lempicki, 2009).

Functional enrichment analysis, also known as gene set enrichment analysis (GSEA), involves generating a list of genes and is usually the final stage of genetic, proteomic and metabolic data analysis. This analysis determines whether genes that share particular pathways, or gene sets, are significantly associated with a specific biological phenotype or not (Tipney & Hunter, 2010). For instance, this analysis statistically categorises genes and proteins overexpressed in a large group of genes or proteins that can ultimately associate it to a disease phenotype (Subramanian et al., 2005).

There are three steps involved in the GSEA method (Subramanian et al., 2005). The summarized steps include:

1. Calculating the enrichment score (ES): This score represents the amount to which the gene in the set is over-expressed or under-expressed.
2. Determination of statistical significance.
3. Normalizing the enrichment scores for each set.

1.5 Aim

The aim of this study is to use cell-based functional genomics tools to provide insight into the function of *FAM111B* in a fibrotic related cancer model and to determine the role of *FAM111B* mutations in POIKTMP patients.

1.6 Objectives

1. To assess the gene and protein expression of *FAM111B* in mammalian cell lines and POIKTMP patient fibroblasts by Western blot and qPCR.
2. To knockdown and overexpress *FAM111B* by siRNA and recombinant DNA technology gene editing tools, respectively.
3. To propose a cellular function of *FAM111B* in a fibrotic-disease model and in the South African POIKTMP patient using cell-based functional genomic tools.

Chapter 2: Methodology

2.1 Gene and protein expression studies

2.1.1 Bioinformatics

The Human Protein Atlas was analysed to identify cell lines that express the *FAM111B* gene (https://www.proteinatlas.org/ENSG00000189057-*FAM111B*). The cell lines in the Human Protein Atlas have been studied by RNA sequencing and Immunohistochemistry to determine their *FAM111B* gene expression.

2.1.2 Cell culture

2.1.2.1 Cells lines

The cell lines HEK293, A549, HT1080, HeLa (Table 1) were cultured and maintained in Dulbecco Modified Eagles Media (DMEM, Thermofisher, USA) supplemented with 10% Fetal Bovine Serum (FBS, Thermofisher, USA) at 37°C with 5% CO2.

Table 1: List of cells used in assessing the expression of the human FAM111B gene.

Cell	
Tissue origin	
HEK-293 cell line (ATTC)	Embryonic kidney
HeLa cell line (ATTC)	Cervix cancer
A549 cell line (ATTC)	Adenocarcinoma human alveolar basal epithelium.
HT1080 cell line (ATTC)	Fibrosarcoma connective tissue
Primary Fibroblast cells (Groote Schuur Hospital)	The skin of POIKTMP patient (daughter) and healthy control (mother)

2.1.2.2 Primary cells

Primary fibroblasts from a 4mm skin biopsy were retrieved from a South African patient diagnosed with HFP with POIKTMP at the Groote Schuur hospital, Cape Town, South Africa (Table 1). Ethics approval for the study was granted by the University of Cape Town (UCT) Faculty of Health Sciences Human Research Ethics Committee (R023/2019). The unaffected mother also provided us with a 4mm skin biopsy used as healthy control. The biopsies were maintained at 4°C in DMEM supplemented with 10% FBS and was immediately processed using a standard dissociation protocol established at Fraunhofer IGB with modifications (Heymer, 2009). The biopsies were placed in a 10 cm Petri dish with a few drops of phosphate buffer saline (PBS) and Penicillin Streptomycin (PS). A blade and forceps were used to remove the epidermis, and the dermis was chopped into fine pieces and placed into a collagenase cocktail (1000 U/ml collagenase, 20x dispase (0.1mg/ml), 0.25% Trypsin, and 10% DMEM + 1% PS). The cocktail was placed into a 37°C shaker for 30 minutes or until the media changed colour, indicating tissue dissociation, and passed through a 16G needle (Lasec, RSA) to dissociate the tissue into cells further. These cells were placed into a 37°C incubator for 30 minutes. The cocktail containing cells were filtered through a 100 μm pore sized filter (Corning), with the filtrate centrifuged for 10 minutes at 1000 rpm, and the supernatant was discarded. The resultant pellet was suspended into a T25 flask with DMEM supplemented with 10% FBS and 1% PS and incubated at 37°C with 5% CO_2. The fibroblast cells were sub-cultured into a T75 flask once they reached 80-90% confluency and harvested and stored at -80°C for 24 hours and then at -140°C until use. All primary cells used in this study was not cultured beyond passage number two.

2.1.2.3 Validation of Y621D patient mutation by PCR-RFLP

A fast and easy genetic approach, using polymerase chain reaction-restriction fragment length polymorphism (PCR-RFLP), was developed to identify the Y621D *FAM111B* mutation in a South African patient diagnosed with POIKTMP (Cenza et al. 2021, submitted).

DNA was extracted from patient fibroblasts, and normal fibroblasts cultured in DMEM supplemented with 10% FBS following the Qiagen AllPrep DNA/RNA/protein kit's protocol. The DNA was quantified using a spectrophotometer (BioDrop, WhiteSci). Using the Platinum™ Direct PCR Universal Master Mix kit (ThermoFisher Scientific, USA), according to the manufacturer's

protocol, the *FAM111B* region consisting of 1085bp was amplified. The *FAM111B* targeting primers (Inqaba Biotech) used are:

ATGCTATTAATCTGGATGTCCAAAAGGAGG (Forward primer)
CTAACATTCCATGGGTTCAATCTGATGATC (Reverse primer)

The amplified PCR product was cleaned with a PCR clean-up kit (ThermoFisher Scientific) and digested using the High-Fidelity restriction enzyme, BstZ171-HF (NEB) for 15 minutes at 37 ℃. The digested PCR product is then run on a 0.8% agarose gel at 100V for 85 minutes. The gel was visualized using an Azure Biosystems c400 (BioRad) chemiluminescence imaging system.

2.2 *FAM111B* gene and protein expression studies

2.2.1 Gene expression (qPCR)

2.2.1.1 Total RNA extraction and quantification

Cells were cultured as described above to 80-90% confluency and prepared for RNA extraction. The cell culture media was aspirated, and cells were washed twice with ice-cold PBS. Cell lysis was achieved by adding TriZol reagent (Life Technologies) (600µl) directly into the flask with scrapping using a cell scraper (NEST, WhiteSci, SA), and cell lysate placed into a 1.5 ml microcentrifuge tube. Phase separation of RNA from protein in the lysate was done by adding chloroform (400µl) to the lysate with centrifugation at 10000rpm for 10 minutes at 4°C. After centrifugation, the supernatant containing the RNA was removed and purified using the GeneJet RNA Purification Kit (Thermo Fisher Scientific). The RNA concentration was determined using a nucleic acid and protein quant (BioDrop, WhiteSci) and stored at -80°C before reverse transcribed by cDNA synthesis.

2.2.1.2 cDNA synthesis

LunaScript RT SuperMix Kit (New England, BioLabs) was used to reverse transcribe the total RNA extracted following the manufacturer's guidelines in a total reaction volume of 20 µl (1X

LunaScript RT SuperMix + <1 µg RNA + up to 20 µl nuclease-free water). No-RT control reaction volumes of 20 µl (1X No-RT Control Mix (5X) + <1 µg RNA + up to 20 µl nuclease-free water) and non-template control reactions up to 20 µl (1X µl LunaScript RT SuperMix (5X) + up to 20 µl nuclease-free water) were prepared as negative controls. The reaction mixtures were placed in a Thermal Cycler (BioRad T100,™, Thermofisher) for 20 minutes to synthesize cDNA templates.

2.2.1.3 qPCR reaction

The Luna Universal qPCR Master Mix (New England, BioLabs) was prepared according to the manufacturer's protocol in a reaction volume of 10 µl (1X Luna Universal qPCR Master Mix + 0.25 µg forward primer + 0.25 µg reverse primer + <1 ng cDNA + up to 10 µl nuclease-free water) (BioLabs). *FAM111B* and GAPDH (Housekeeping primer) primers used are shown in Table 2. The qPCR reaction was done using the QuantStudio 6 Flex (Life Technologies) under the following conditions: DNA denaturation at 95°C for 60 seconds for DNA strand separation, followed by a 40x cycle of DNA denaturation and extension (15 sec at 95°C, 30 seconds at 60°C). The temperature of the melting curve was set at 60 – 95°C to ensure amplification specificity (BioLabs).

*Table 2**: Primer sequences***

Primer name	**Sequence**
FAM111B (forward)	GCTAGCATGAATAGCATGAAGACA
FAM111B (reverse)	GGATCCGCACTCCATAGG
GAPDH (forward)	GACCTCAACTACATGGTTTACATG
GAPDH (reverse)	GATCTCGCTCCTGGAAGATG

2.2.2 Protein expression (Western blot analysis)

2.2.2.1 Total protein extraction and quantification

The AllPrep DNA/RNA/protein Kit (QIAGEN) kit was used to precipitate the protein obtained from the phase separation (2.2.1.1). The protein was placed into a microcentrifuge tube, and an

equal volume of buffer APP was added. The mixture is mixed and incubated at RT for 10 minutes to precipitate the protein and then centrifuged for 10 minutes at full speed. The supernatant was removed by careful pipetting, and the protein pellet was washed with 70% ethanol by centrifugation for 1 minute at 13 000 rpm. The protein pellet was dried at RT for 10-15 minutes and solubilized with protein solubilizing buffer (50mM Tris-HCl pH.8; 150mm NaCl; 4% Urea; 2% SDS and dH_20). The solubilised protein was then quantified by a Pierce^{TM} Bicinchoninic Acid (BCA) protein assay kit (Thermofisher, RSA). Protein standards (Bovine Gamma Globuli, BioRad, USA) standards and the samples were prepared by adding 5μ to a 96-well Greiner flat bottom plate (CellStar^{R}). The BCA reagent (200 µl, Thermo Scientific, USA) prepared according to the manufacturer's guidelines was added to each well containing the protein standards and protein sample. The plate was incubated at 37°C for 15 minutes, and absorbance was read at 562 nm on a UV/Vis plate reader (Varioscan Lux, Thermo Scientific, USA). The sample protein concentrations were extrapolated from the standard curve created with the protein standards and were normalized. Sample buffer was added to the diluted protein samples and then boiled at 95°C for 5 minutes to denature the proteins.

2.2.2.2 Western blot

The normalised proteins were denatured in the presence of sample buffer (50mM Tris-HCl pH 8, 150 mM NaCl, 4% Urea, 2% SDS, H_20) at 95°C for 5 min, and resolved on a 12% Sodium Dodecyl Sulphate Polyacrylamide Electrophoresis (SDS-PAGE) in a Mini-PROTEAN Tetra protein electrophoresis system (Bio-Rad, USA) at 200V for 45 mins. The resolved proteins were then transblotted onto a polyvinylidene difluoride (PVDF) membrane using the Trans-Blot^{R} Turbo RTA transfer kit and System (Bio-Rad, USA) for 10 mins.

After protein transfer onto the membrane, the membrane was washed briefly with Tris Buffered Saline (1 x TBS, Bio-Rad, USA) blocking solution (5 % Blotting-Grade Blocker, BioRad, in 1X TBS) for an hour at 4°C on a rotary shaker. After this blocking step, membranes were probed with *FAM111B* (Sigma, USA, 1:5000) and GAPDH (Invitrogen, USA, 1:2000) specific antibody in 10ml Casein Blocker (BioRad) overnight at 4°C with shaking. The following day the membrane was washed twice for 10 minutes with 1 x Tris Buffered Saline-Tween20 before probing with an HRP-conjugated secondary antibody and incubated for at least 1hr with shaking at 4˚C. This incubation was then followed by two wash steps with TBST (15mins per step) and the presence of

protein bands visualized using the Clarity™ Western ECL substrate (BioRad) kit on an Azure Biosystems c400 (BioRad) chemiluminescence imaging system. The band intensity was detected using GelQuant software and normalized to the GAPDH housekeeping gene to determine the relative protein expression of FAM111B protein the cell lines.

2.3 *FAM111B* knock-down and overexpression

The *FAM111B* gene was transiently knocked down by short interference RNA (siRNA) delivered to the targeted *FAM111B* gene by a transfection reagent, Lipofectamine RNAiMAX (Thermofisher, RSA). In addition to knocking down, the *FAM111B* gene was overexpressed by recombinant DNA gene expression technology, and the overexpressed plasmid was facilitated to the targeted *FAM111B* gene by a transfection reagent, Lipofectamine 3000 (Thermofisher, RSA).

2.3.1 *FAM111B* knockdown

Pre-designed, pooled, and validated *FAM111B* targeting siRNA's (siRNA-*FAM111B*, USA) was used to silence the expression of the *FAM111B* gene. A non-targeting (siRNA-NT) was also designed as a control.

HT1080 cells were seeded at a density range of 1×10^4- 2.5×10^5 cells per well subject to the plate size (96,24,12 and 6-well) 24hrs before transfection facilitated by Lipofectamine RNAiMAX Reagent. The HT1080 cell line was optimized for its optimal siRNA transfection efficiency.

According to the manufacture's guidelines, the transfection mixture was prepared in a ratio of 3 µl of Lipofectamine RNAiMAX for every 1 µl of total siRNA transfected (final concentration= 0.40µM). The siRNA, lipofectamine RNAiMAX and Opti-MEM mixture were incubated at room temperature (RT) for 5 minutes. The complex was added to the cells and incubated at 37°C with 5% CO_2 for 24 hours before extraction. Untreated two negative control wells (DMEM supplemented with 10% FBS) and non-targeting siRNA (NT-siRNA) were simultaneously seeded and transfected. After 24 hours, the RNA and protein were extracted according to protocol 2.2.1.1 and 2.2.2.1, respectively.

2.3.2 *FAM111B* overexpression

A midiprep (GeneJet Plasmid Midiprep Kit, Thermofisher, USA) of the EGFP-tagged (negative control), *FAM111B*-EGFP overexpressed Y621D plasmid, which was previously created in our lab, was performed according to the manufacturer's protocol, and the resulted plasmid DNA was transfected into HT0180 cells.

HT1080 cells were seeded at a density range of 1 x 10^4- 2.5 x 10^5 cells per well subject to the plate size (96,24,12 and 6-well) 24hrs prior to transfection facilitated by Lipofectamine 3000 Reagent.

According to the manufacturer's guideline, the transfection mixture was prepared in a ratio of 0.75 µl of Lipofectamine 3000 and 2 µl of P3000 for every 1 µl of total DNA transfected (final concentration= 0.5 µg). The DNA, lipofectamine 3000, P3000, and Opti-MEM mixture were incubated at room temperature (RT) for 15 minutes. The complex was added to the cells and incubated at 37°C with 5% CO_2 for 24 hours before extraction. Untreated (DMEM supplemented with 10% FBS) and EGFP-tagged, two negative controls were simultaneously seeded and transfected. After 24 hours, the RNA and protein were extracted according to protocol 2.2.1.1 and 2.2.2.1, respectively.

2.3.3 Validation of gene knock-down and overexpression

The *FAM111B* gene and protein knockdown and overexpression were validated by qPCR (2.2.1.2-2.2.1.3) and Western blot (2.2.2.2), respectively.

2.4. Immunofluorescence (IF) confocal Imaging

2.4.1. Validation of FAM111B expression

HT1080 cells were seeded at a density of 8 x 10^4 cells/well onto sterilized coverslips in a 12 well plate. After 24 hrs of incubation at 37°C and CO_2, the cells were transfected with siRNALipofectamine RNAiMAX complexes (2.3.1) and DNA-Lipofectamine 3000 complexes (2.3.2).

Following a 24hr transfection, the media was aspirated, and the cells were fixed with 1ml 4% Paraformaldehyde (PFA) at room temperature for 10 minutes. The fixative was aspirated and

washed twice with Phosphate buffer saline (PBS) for 5 minutes each. After aspirating, the cells were incubated for 25 minutes at room temperature in 1.5M HCl. The washing step was repeated twice with PBS for 5 minutes and then aspirated. 1ml of 0.5% of Triton-X was added to each well for 10 minutes at RT followed by 1ml of 1% Bovine Serum Albumin (BSA, Sigma, USA) in PBST (Phosphate buffer saline- tween20) for 30 minutes to block non-specific binding sites. The primary antibodies *FAM111B* (1:500, Sigma, USA) and β-Tubulin (1:500, Invitrogen, Canada) was added to each well and incubated in a humidifying box at 4°C overnight. The following day the cells were washed thrice with PBS for 5 minutes each. The secondary antibodies, AntiR-CY3 (1:500, UCT Confocal facility) and AntiM-Alexa647 (1:500, UCT Confocal facility), was incubated for 1-2 hours in foil at room temperature. After incubation, the cells are washed with 1% BSA in PBS thrice for 5 minutes each. The cells were stained with DAPI (1:5000) for 20 minutes, washed twice with PBS, and mounted onto slides with Flouromount Aqueous Mounting Media (Sigma, USA). Slides were stored at 4˚C and then analysed by the University of Cape Town Confocal Facility.

2.4.2 Localisation of the FAM111B protein

Cells were prepared for confocal staining according to protocol 2.4.1. The confocal images were analysed using Image J's Co-localization Threshold plugin (https://imagej.net/Colocalization_Threshold). The merged image was analysed using the "Colocalization" function, which produced co-localization coefficients exported and saved for data analysis.

2.5 Functional studies of the human *FAM111B* gene

2.5.1 Cell-based functional studies

2.5.1.1 Cell proliferation (RTCA), viability (Flow cytometry), migration and invasion assays.

Cell proliferation, viability, migration, and invasion are essential functions in all cells. These functions were observed in *FAM111B* gene knocked-down and overexpressed HT1080 cells, and patient-derived fibroblasts and normal fibroblasts.

2.5.1.1.4 RTCA

The initial background was read by adding 80μl DMEM supplemented with 10% FBS into an EPlate 96 well (ACEA Biosciences) and incubating at RT for 30 minutes for equilibration. After incubation, the plate was placed into the xCelligence RTCA SP analyser (ACEA Biosciences) at 37°C for 1 minute to obtain the background reading. HT1080 cells were seeded at a density of 1 x 10^4 cells/well. Internal control wells contained DMEM supplemented with 10% FBS only. The plate was incubated in the hood for 30 minutes for equilibration and placed back into the xCelligence RTCA SP analyzer at 37°C and 5 % CO_2 for 24 hrs. After 24 hrs of incubation, the cells were transfected with siRNA-Lipofectamine RNAiMAX (*FAM111B* knockdown, 2.3.1) and DNA-Lipofectamine 3000 complexes (*FAM111B* overexpression, 2.3.2). The cells were incubated in the xCelligence RTCA SP analyser for 48 hrs, and the data were analysed using the RTCA Data Analysis Software1.0. (https://www.agilent.com/en/product/cell-analysis/real-time-cellanalysis/rtca-software/rtca-software-pro-741236).

2.5.1.1.1 Migration/Scratch assay

HT1080 cells were seeded at a density of 8 x 10^4 cells/well in a 12-well plate. After 24 hrs of incubation at 37°C and 5 % CO_2, the cells were transfected with siRNA-Lipofectamine RNAiMAX (2.3.1) and DNA-Lipofectamine 3000 complexes (2.3.2). Once the cells reached 100% confluency, a straight line (wound) was created using a 1ml pipette tip from the top to the bottom of the well. The media was aspirated, and the cells were washed twice with PBS. DMEM supplemented with 10% FBS (1 ml) was added to the wells. Images monitoring the migration of cells into the scratch area was captured every 12 hours for the next 24 hours. After 24 hours, the media was aspirated, washed twice with PBS, and fixed with 4 % Paraformaldehyde (PFA, 400 µl) for 5 minutes. The fixed cells were then washed twice with PBS and stained with 10 µl of 0.5% crystal violet dye solution (0.5g crystal violet powder, 80 ml distilled H_2O, 20 ml methanol) per well for 20 minutes at room temperature. The dye was removed, and the plate was washed twice with PBS. For the *FAM111B* mutant patient fibroblasts and normal control fibroblasts, cells were seeded at a density of 1 x 10^5 cells/well. Due to the slower proliferation of primary fibroblasts, the cells were incubated for 48 hrs at 37°C and 5 % CO_2. Once the cells reached 100% confluency, the scratch assay was repeated on the fibroblasts. The endpoint images were taken 30 hours

postscratching. Image J software (https://imagej.nih.gov/ij/download.html) was used to determine the rate of migration. The image was uploaded, and the colour segmentation tool (http://bigwww.epfl.ch/sage/soft/colorsegmentation/) was used to differentiate the stained cells against the white background of the wounded area.

2.5.1.1.2 Invasion assay

The extracellular matrix (ECM) membrane of a 12-well Cell Invasion Assay Kit (Chemicon, USA) was rehydrated with 300μl serum-free DMEM for 1-2 hours. The media was carefully removed and replaced with an HT1080 cell suspension containing 0.5 $\times$ 10^5 cells/ml prepared in serum-free DMEM. DMEM supplemented with 10% FBS was added to the lower chamber, and the plate was incubated for 24 hrs at 37°C with 5% CO_2. After incubation, the cells were transfected for 24 hrs with siRNA-Lipofectamine RNAiMAX (2.3.1) and DNA-Lipofectamine 3000 complexes (2.3.2) for gene knockdown and overexpression, respectively. Following transfection, the non-invasive cells and ECM were removed from the interior of the inserts using a cotton swab, and the cells were stained with crystal violet solution (500 µl). 0.5g crystal violet powder, 80 ml distilled H_2O (dH_2O), 20 ml methanol) for 20 minutes. The invasive stained cells on the inserts were washed with dH_2O several times and then air-dried and imaged using a microscope at 40x magnification. Image J software (https://imagej.nih.gov/ij/download.html) was used to determine the percentage of invasive cells. The image was uploaded, and the colour segmentation tool (http://bigwww.epfl.ch/sage/soft/colorsegmentation/) was used to differentiate the invasive stained cells against the white background.

2.5.1.1.3 Flow cytometry

HT1080 cells were seeded at a density of 2.5 x 10^5 in a 6-well plate and transfected for 24 hrs with siRNA-Lipofectamine RNAiMAX (2.3.1) and DNA-Lipofectamine 3000 complexes (2.3.2) for gene knockdown and overexpression, respectively. The culture media was aspirated from the culture plates, and the cells were washed twice with 1x PBS. The cells were lifted from the culture wells using 200μl Trypsin-EDTA (Thermofisher). The cells were collected using a four °C centrifuge at 1500rpm and counted. The appropriate amount of 1x binding buffer (BD Pharmingen) was added to the pellet. After thoroughly mixing the cells using a pipette, 200μl

samples were placed into a 5ml glass culture tube and placed on ice. For staining the cells, 5µl of FITC Annexin V and 5µl of Propidium Iodide (PI) (BD Pharmingen, Sigma, USA) was added to the samples in a dark room. FITC Annexin V (FITC +; PI -) and PI (FITC -; PI +) controls contained only 5µl of the FITC Annexin V and PI dye, respectively. A sample with no stain was used as a negative stain control (FITC -; PI -). The samples were vortexed and incubated at RT for 15 minutes in the dark.

The samples were placed into the BD Accuri C6 Plus-BD CSampler to measure cellular apoptosis. FlowJo (http://info.flowjo.com/africa-serial-number-license-request-form) was used to analyse the data.

2.5.2 Mass spectroscopy and proteomics analysis

2.5.2.1 Protein extraction

HT1080 cells were seeded at density of 2.5 x 10^5 cells/ml in a 6-well plate and transfected as described in 2.3. The media was aspirated, and the cells washed twice with 1x PBS. The cells were trypsinised and resuspended in DMEM, centrifuged at 1700 rpm and 4°C for 3 minutes. The pelleted cells were washed with PBS and the protein extracted from the cells according to the Qproteome Mammalian Protein Kit (QIAGEN) manufacturers protocol. Protein was quantified by BCA assay (described in 2.5.1.1).

2.5.2.2 Filter Aided Sample Preparation (FASP)

Firstly, the filters were washed by adding 500µl urea to each filter and placed into a centrifuge at 14000g for 10 minutes. After that, 200µl urea and 500µl protein samples (max. 400µl protein) were added to the filters and centrifuged at 14000g for 10 minutes. Another 200µl urea was added and centrifuged at 14000g for 15 minutes. The filter was moved to a fresh Eppendorf tube where 100 µl iodoacetamide (IAA) was added to the filter and agitated in a thermo-mixer at 600rpm for 1 minute. The samples were then incubated in the dark for 20 minutes. After incubation, the samples were spun at 14000g for 10 minutes. A 100µl urea was added to the filter and centrifuged at 14000g for 15 minutes, and then repeated. 100µl of 50mM ammonium bicarbonate (ABC) was added to the filter and centrifuged at 14000g for 10 minutes. This step was repeated until the flow through was visible. The flow-through was discarded. 40µl ABC and Trypsin were added to the filter such

that the ratio of trypsin: protein was 1:100 (i.e., 1μg Trypsin to 100μg of protein sample). The sample was agitated gently, and the pH was checked (pH should be 8 to activate the Trypsin). The samples were then incubated in a wet chamber at 37°C for 18 hours.

After 18 hours, the samples were centrifuged at 14000g for 10 minutes. The flow-through should contain the tryptic peptides. 40μl ABC was added to the filter and centrifuged at 14000rpm for 10 minutes. To acidify the peptides, 50μl of solution B (dH20 + 0.1% Trifluoroacetic acid (TFA)) was added and then centrifuged at 14000g for 10 minutes. This step was repeated until the pH = 2-3For activating the C18 column, 100μl of Solution C (80% ACN + 0.1% TFA) was added and then centrifuged at 4000g at 1min intervals until almost all the liquid was passed through the column filter. This step was repeated thrice until a total volume of 400μl solution C was added. For equilibrating the C18 column, 100μl of Solution E (2% ACN + 0.1% TFA) was added and then centrifuged at 4000g in 1-minute intervals until almost all the liquid was passed through the column. This step was repeated thrice until a total volume of 400μl of solution E was added. The flow-through was discarded. The protein was then loaded onto the stage tip and centrifuged at 4000g in 1-minute intervals until almost all the liquid had passed through. The C18 was washed with 100μl of solution E and then spun at 4000xg in 1-minute intervals until almost all the liquid had passed through and repeated thrice. The stage tip was removed from the collection tube, the flow-through was discarded, and the stage tip was replaced with a glass insert. The peptides were eluded by adding 150μl solution D (60% ACN + 0.1% TFA) and then repeated twice again. The Eppendorf tubes with glass inserts were placed into the MiVac (SP Scientific, USA) at 45 °C. The dried peptides were resuspended in 150μl of solution E.

2.5.2.3 Ultra-high performance liquid chromatography and mass spectrometry.

Samples were analysed by a proteomics analyst at the HSR lab at Groote Schuur hospital.

According to their protocol, "the resuspended peptides were transferred into glass vials and placed into the Dionex UltiMate® 3500 RSnano LC systems (ThermoFisher, USA). Nano ultra-high performance liquid chromatography (UHPLC) was performed, followed by gradient chromatography at 23°C.

Shotgun proteomics analysis was done using a QExactive^{TM} Hybrid Quadrupole-Orbitrap Mass Spectrometer (Thermo Fisher, USA). Flow injection analysis of samples introduced from the

HPLC system was done with the Mass spectra (MS). High-energy collision dissociation (HCD) and normalized collision energy were used for peptide fragmentation. The top 10 most abundant parent ions selected by the quadrupole during the initial scan are further fragmented. The identified proteins are quantified using a label-free method, intensity-based absolute quantification (iBAQ)" (Adeola, 2016).

Downstream computational analysis of shotgun proteomics output data, generated from Maxquant software, was performed by Persues software (version 1.6.15.0). Steps employed in Perseus analysis were as follows: data selected> data filtered > data transformed > categorical row annotation > filtered based on valid values > statistical analysis.

2.6 Statistical analysis

All statistical analyses were performed with GraphPad Prism Version 9.0.0 (GraphPad, San Diego, CA). A p-value of $p \leq 0.05$ was considered statistically significant. The results are presented as a mean ±SEM. T-tests were used to determine significant differences amongst the two sample groups. In groups of three or more, ANOVA tests were done to determine the significant differences.

Chapter 3: Results

3.1 *FAM111B* expression studies

3.1.1 *In-silico* expression studies

In-silico expression analysis of the *FAM111B* gene was investigated on The Human Protein Atlas (https://www.proteinatlas.org/ENSG00000189057-*FAM111B*/cell) database. The results showed that 68 cell lines derived from multiple tissue origins expressed *FAM111B* (Figure 3.1). Based on visual inspection and manual quantifying, about 68% of these cell lines originated from cancerous tissue, the other 32% were from non-cancerous tissue. Based on the consensus of three datasets and normalized expression (NX) of0.1 as detection limit, expression above this cut-off was observed in 16 different tissue types (Figure 3.1). High expression (>17 NX) was observed in the lymphoid, mesenchymal, kidney, and urinary bladder, female reproductive system, and skin tissue. Few tissues showed low expression (i.e., <17 NX), e.g., brain, liver and gallbladder,

gastrointestinal tract, pancreas, male reproductive tract, kidney, and urinary bladder, eye, proximal digestive tract, lung, endothelial and myeloid tissue. Based on these genomics and transcriptomics in which the non-cancerous HEK293 ((human embryonic kidney, NX= 3.3,) and cancerous cell lines A549 (adenocarcinoma human alveolar basal epithelial, NX=6.3), HeLa (cervical cancer, NX=12) cell lines showed a positive expression above NX=0.1, these cell lines were accessed for the gene and protein expression of *FAM111B* in this study by qPCR and western blotting, respectively. In addition, though there is no information on *FAM111B's* gene expression in fibroblasts (considering the role these cell types play in fibrosis), in this database. Hence this study subsequently evaluated two fibroblasts cell lines: HT1080 (fibrosarcoma cell line) and MRC5 (healthy lung cell line) for *FAM111B* expression (results presented in 3.1.2).

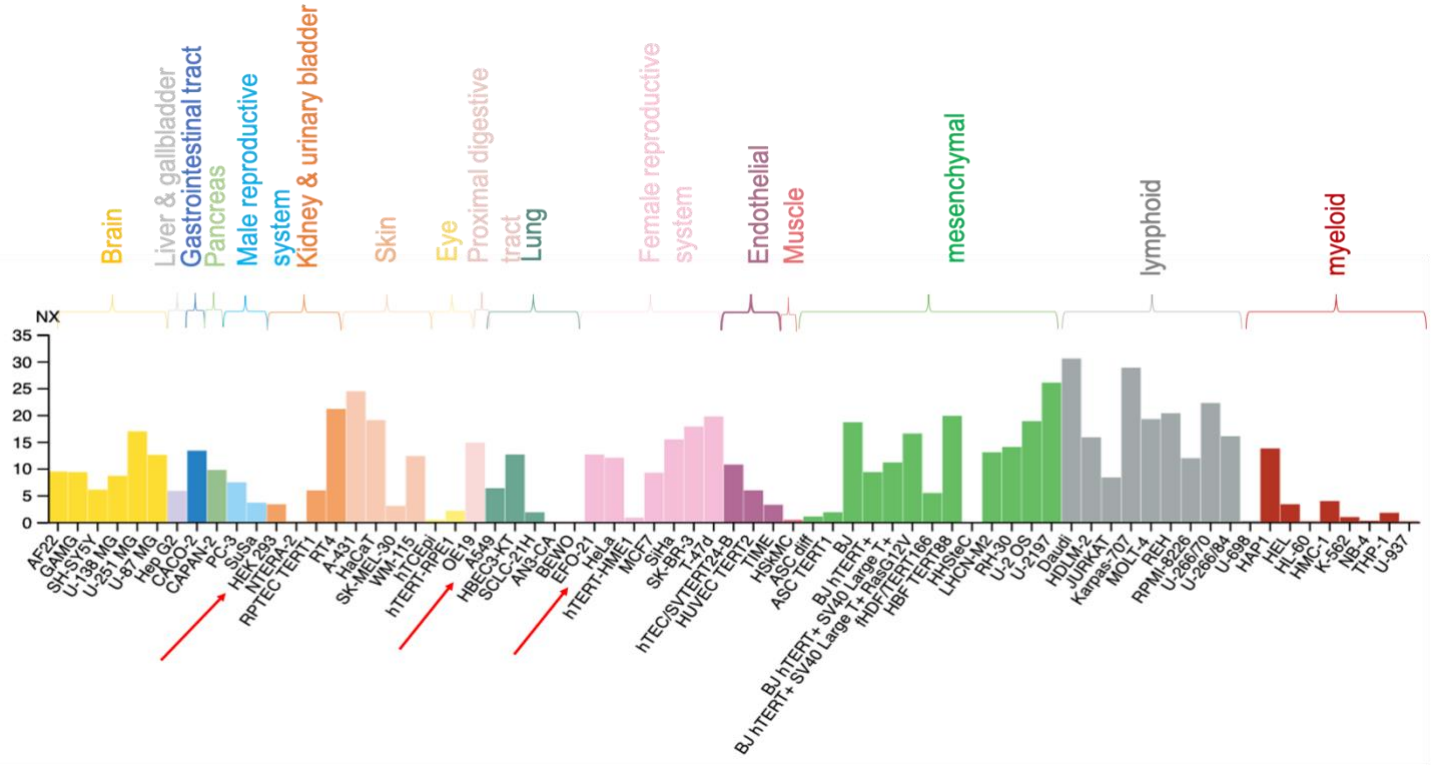


Figure 3. 1 ***<u>An overview of FAM111B gene expression in cell lines derived from various human organ tissues.</u>***

FAM111B gene expression data (RNA sequencing) reported in the Human Protein Atlas are expressed as the number of transcripts per kilobase million (TPM). The histogram shows RNA normalized expression (NX) of the FAM111B gene in 68 cell lines. The cell lines are divided into

16 color-coded groups ordered by tissue or organ source of these cell lines. The red arrows indicate the cell lines that were further investigated in this study.
(https://www.proteinatlas.org/ENSG00000189057-*FAM111B*/cell)
(Adapted from Protein Atlas)

3.1.2 *FAM111B in vitro* gene and protein expression studies

The *FAM111B* mRNA expression of HEK293, the non-cancerous control cell line, showed a lower expression relative to cancerous cell lines, HeLa and A549. We also observed that HEK293 FAM111B protein expression was lower compared to cancerous Hela and HT1080. Although HT1080 *FAM111B* mRNA and A549 FAM111B protein expression did not correspond to the observed trend whereby *FAM111B* is overexpressed in cancerous cells, we must take into consideration that the expressed transcript does not necessarily relate to the protein expression. However, it is observed that the overall expression of *FAM111B* mRNA and protein in HEK293 is relatively reduced compared to the cancerous cell lines (Figure 3.2).

Our study focuses on HT1080 cells due to the insufficient research on *FAM111B*'s role in fibrosis using fibroblast cells. Fibrosis is a process involving replacing connective tissues with parenchymal tissue resulting in tissue remodelling; therefore, the use of HT1080 (derived from connective tissue) is a suitable cell line to use as the fibrotic disease model. Furthermore, the study of the expression of *FAM111B* in HT1080 cells contributes to the information available of cancer cell lines with overexpressed *FAM111B*. The expression of *FAM111B* was studied in HT1080 cells by performing qPCR for gene expression and western blotting for protein expression analysis (Figure 3.2 A-B).

Additionally, we examined the expression of *FAM111B* in HT1080 (cancerous fibroblast cell line) vs MRC5 (normal/noncancerous fibroblast cell line) as a potential control (Figure 3.2 C). However, the western blot illustrates a significantly low or absent expression of FAM111B protein in MRC5 cells in comparison to the expression of FAM111B protein in HT0180 cells. To study gene function, it is important for the cell to express the gene at a protein level, therefore, the absent expression of FAM111B protein in MRC5 cells lead us to the decision to proceed with HT1080 cells as the focus cell line for this study. No further experiment using MRC5 was performed. All

experiment was performed in triplicate and repeated three times with similar results. The bar graphs display mean and standard deviation (SD).

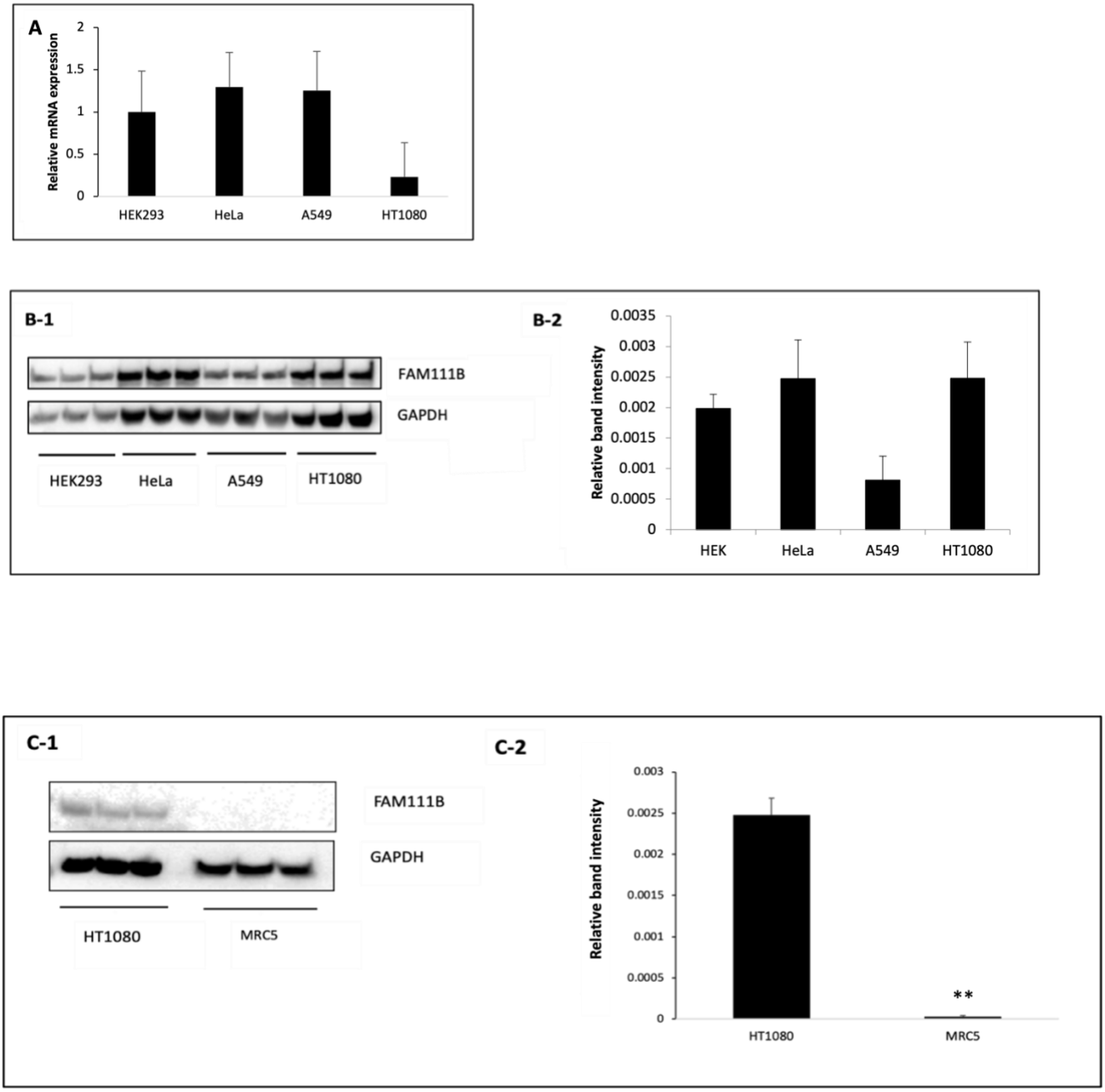


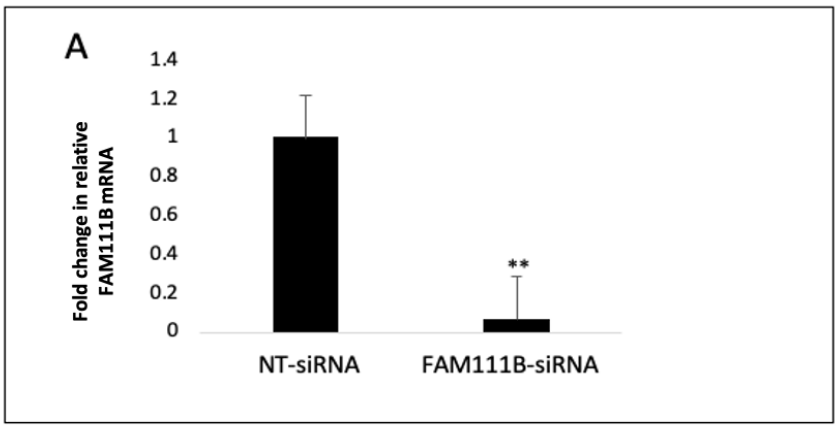
Figure 3. 2 ***FAM111B gene and protein expression in some mammalian cell lines.***

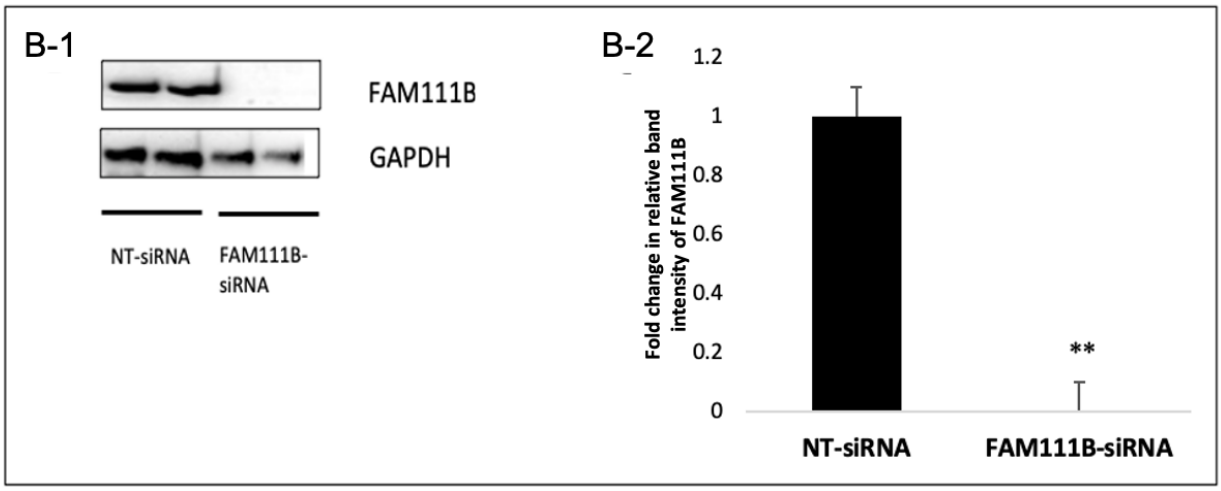
(A) The qPCR graph showing the mRNA expression of FAM111B in four cell lines: HEK293 (human embryonic kidney cell line), HeLa (cervical cancer cell line), A549 (adenocarcinoma

human alveolar basal epithelial cell lines) and HT1080 (Fibrosarcoma cell line). Quantification was relative to HEK293 (as control) mRNA levels; all mRNA levels were normalized to GAPDH. (B) Western blot analysis of FAM111B protein (~85 kDa) expression in HEK, HeLa, A549 and HT1080. (C) Western blot results of FAM111B (~85 kDa) protein expression HT1080 and MRC5. GAPDH (~38 kDa) was used as a loading control. Densitometric measurements were performed on the immunoblots, and values indicate relative band intensities fold change. All protein bands were normalised to their corresponding GAPDH protein expression. Statistical analysis was performed using a t-test, and P-values were provided (=p ≤ 0.05; **=p ≤ 0.005).*

3.2 *FAM111B* siRNA knockdown

The knockdown efficiency of the *FAM111B*-siRNA was determined by qPCR and western blot, as seen in Figure 3.3. A 9.7-fold knockdown of *FAM111B* mRNA expression was observed from the qPCR data (Figure 3.3 A). Western blot results show that the expression of FAM111B protein was utterly knocked down following a 24-hour transfection which was observed by the absence of a protein band present (Figure 3.3 B). Knockdown of FAM111B protein was also confirmed by immunofluorescent confocal imaging, which showed a 5-fold decrease in FAM111B expression (Figure 3.3 C). Each experiment was performed in triplicate and repeated three times with similar results. All results obtained were relative to the non-targeting siRNA (negative control).


A
Fold change in relative FAM111B mRNA
1.4
1.2
1
0.8
0.6
0.4
0.2
0
**
NT-siRNA
FAM111B-siRNA


B-1
FAM111B
GAPDH
NT-siRNA
FAM111B-siRNA
B-2
Fold change in relative band intensity of FAM111B
1.2
1
0.8
0.6
0.4
0.2
0
**
NT-siRNA
FAM111B-siRNA

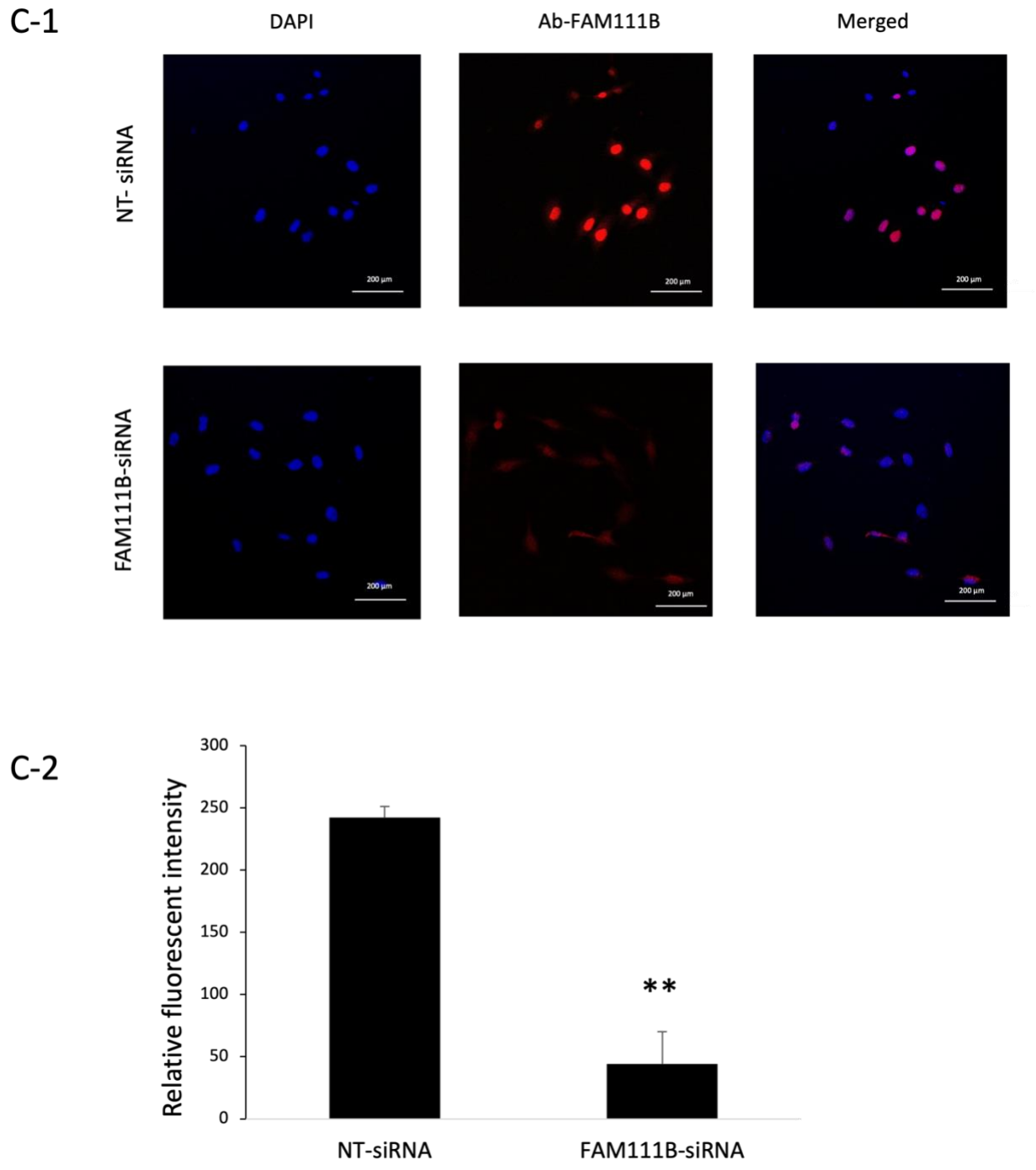


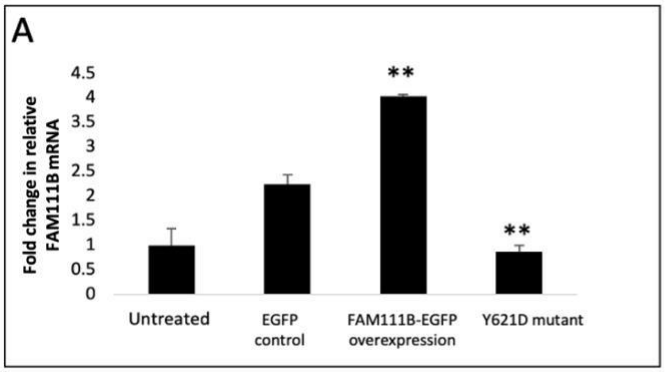
Figure 3. 3 ***siRNA-mediated knockdown of FAM111B mRNA in HT1080 cells.***

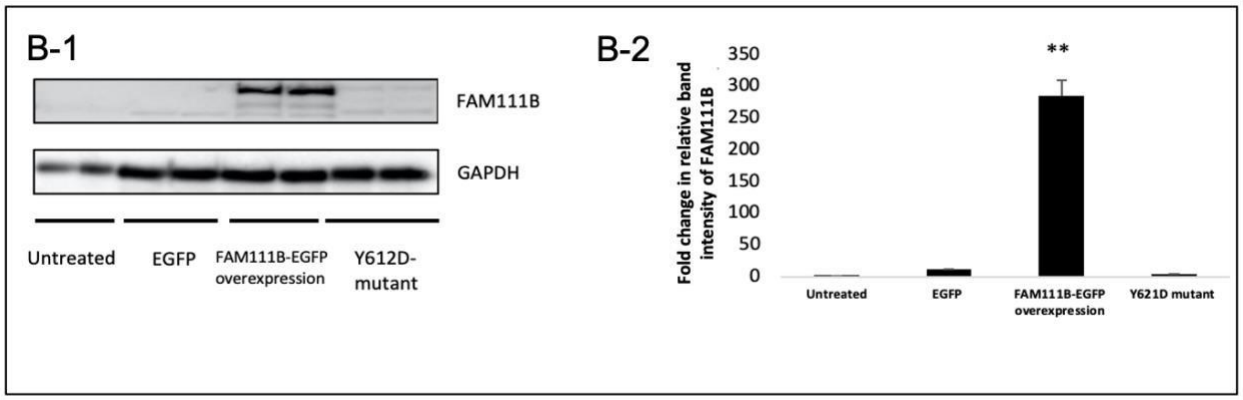
(A) HT1080 cells were transfected with FAM111B specific-siRNA (FAM111B-siRNA), and the mRNA expression levels were quantified by qPCR. The FAM111B-siRNA mRNA quantification was relative to that of the non-targeting siRNA, and measurements were normalized to GAPDH.

*The qPCR results show a 9.7-fold decrease in FAM111B mRNA expression ($p \leq 0.0005$**). (B-1) Protein expression in FAM111B gene knock-down HT1080 cells. (B-2) Densitometric quantification of the relative FAM111B protein expression to the control (i.e., NT-siRNA). Protein expression levels were normalized to GAPDH. (C-1) Confocal microscopy images FAM111B gene knock-down HT1080 cells stained with FAM111B and Cy3 primary and secondary antibodies, respectively, and counterstained with DAPI to identify the nucleus. The FAM111B and DAPI stained images were merged in the far-right panel. A 5-fold decrease in FAM111B expression was visualized ($p \leq 0.0005$**). The scale bar for all images is 200μm and images were captured at 20X magnification (C-2) A graph showing the relative fluorescent intensity of FAM111B. Quantification was relative to the control non-targeting siRNA (NT-siRNA) fluorescence. The bar graphs display mean and standard deviation (SD). Statistical analysis was performed using a t-test, and P-values were provided (*=$p \leq 0.05$; **=$p \leq 0.005$)*

3.3 *FAM111B* overexpression

FAM111B protein was overexpressed in HT0180 cells by recombinant DNA technology. HT1080 were transfected with an enhanced green fluorescent protein (EGFP) tagged *FAM111B* plasmid DNA (*FAM111B*-EGFP overexpression) and a mutant *FAM111B* (Y621D) also tagged with EGFP. The Y621D plasmid contains the *FAM111B* mutation present in the South African family members with POIKTMP. This experiment validated the overexpression of the *FAM111B* and Y621D mutant gene and protein by qPCR and western blot, respectively (Figure 3.4). The qPCR data shows a 3.9-fold increase in *FAM111B* mRNA expression in cells treated with *FAM111B*EGFP overexpression relative to UT and a 1.7-fold increase relative to EGFP. Y621D-mutant transfected cells showed a 1.25-fold decrease in the expression of *FAM111B* mRNA relative to UT and a 2.8-fold decrease relative to EGFP only (Figure 3.4-A). Western blot results showed a similar trend to the qPCR data results. The FAM111B protein expression was increased by 23-fold in cells treated with the *FAM111B*-EGFP overexpression relative to EGFP only, as shown by the western blot in Figure 3.4-B. Furthermore, a 1.7-fold increase in FAM111B protein expression was observed in the immunofluorescent confocal imaging. There was no significant change in the expression of FAM111B in the Y621D-mutant cells compared to both controls, UT and EGFP.


A
Fold change in relative FAM111B mRNA
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
**
**
Untreated
EGFP control
FAM111B-EGFP overexpression
Y621D mutant


B-1
FAM111B
GAPDH
Untreated
EGFP
FAM111B-EGFP overexpression
Y612D-mutant
B-2
Fold change in relative band intensity of FAM111B
350
300
250
200
150
100
50
0
**
Untreated
EGFP
FAM111B-EGFP overexpression
Y621D mutant

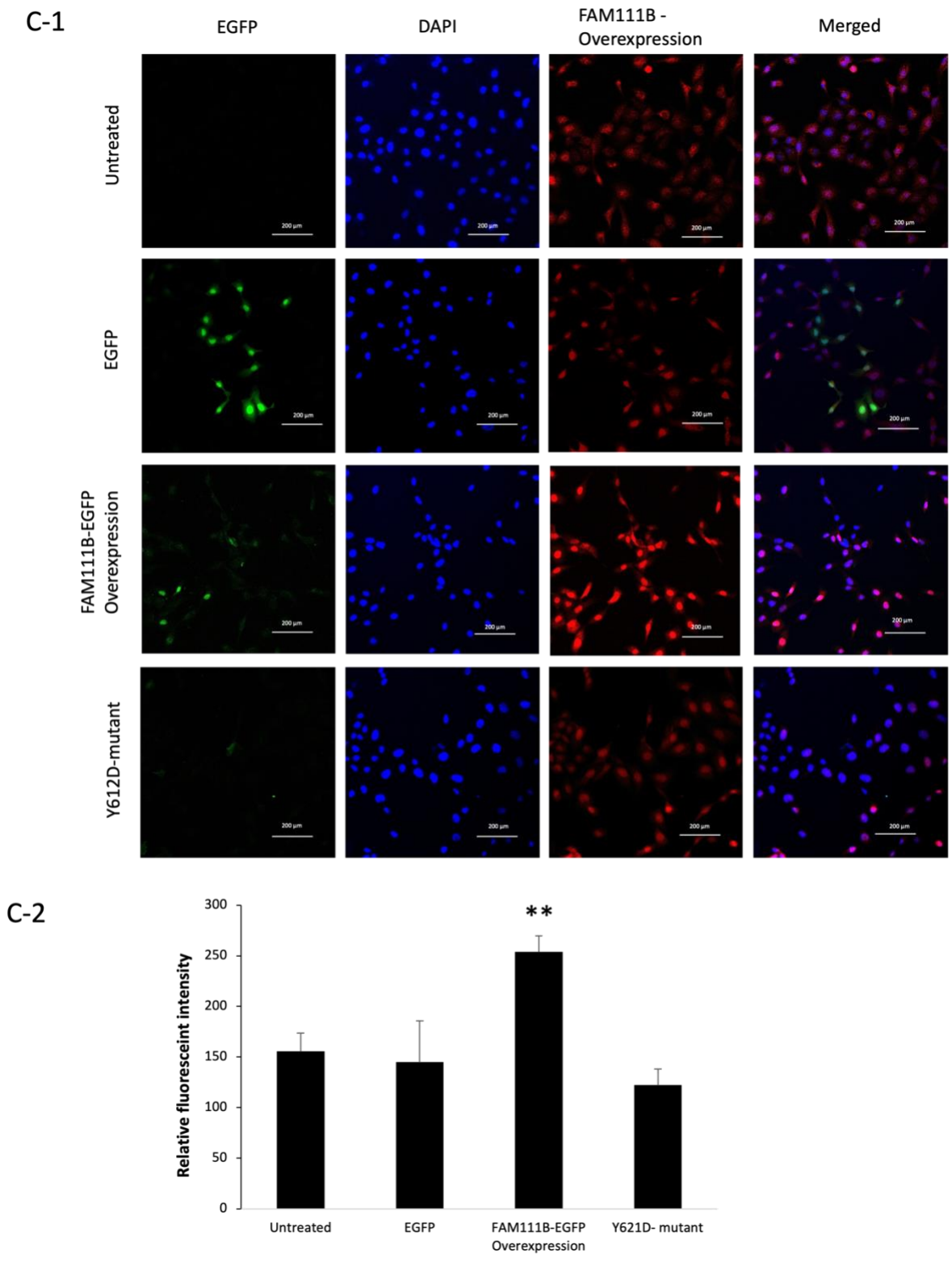


Figure 3. 4 ***FAM111B wild type and Y621D mutant protein overexpression in HT1080 cells***

(A) HT1080 cells were transfected with EGFP-tagged FAM111B plasmid DNA (FAM111BEGFP) and Y621D-mutant plasmid. The mRNA levels (gene expression) in each group were quantified using qPCR (B-1) Western blot showing the protein expression of FAM111B (~ 85 kDa). GAPDH (~38 kDa) served as a loading control. (B-2) Densitometric quantification of the relative FAM111B protein expression in B-1. Quantification was relative to the untreated control (UT) and EGFP, and protein expression levels were normalized to GAPDH. (C-1) HT1080 cells were transfected with an EGFP-tagged overexpressed plasmid and a Y621D-mutant plasmid. The stained cells were analysed by confocal microscopy. Cells were stained with FAM111B antibody and Cy3 secondary antibody (red) and counterstained with DAPI (blue) to identify the nucleus. The images were merged to form the overlapped image in the far-right panel. The scale bar for all images is 200μm and images were captured at 20X magnification. (C-2) A graph showing the relative fluorescent intensity of FAM111B. Quantification was relative to the control EGFP fluorescence. All experiments were performed in triplicate and repeated three times with similar results. The bar graphs represent the mean and standard deviation (SD). Statistical analysis was performed using t-test (= $p \leq 0.05$; ** = $p \leq 0.005$).*

3.4 Subcellular localization of FAM111B

FAM111B protein, also known as cancer-associated nucleoprotein (CANP), is predominantly localized in the nucleus (Kawasaki et al., 2020; Mercier et al., 2019; S. Peng, Yip, Pan, & Hsu, 2008; Sun et al., 2019). Knowledge of the subcellular localization of a protein permits the understanding of the subcellular organization and the function of its proteins (Scott, Calafell, Thomas, & Hallett, 2005).

The colocalization of FAM111B in the cytoplasm was determined by staining cells with FAM111B and Cy3 primary and secondary antibodies, respectively. In contrast, β-Tubulin and Cy5 primary and secondary antibodies respectively were used as a marker for the cytoplasm (Figure 3.5 A-1). The merged image represents the 8% colocalization of FAM111B with cytoplasmic marker β-Tubulin quantified in Figure 3.5 A-2. A 44% colocalization of FAM111B with nuclear marker DAPI (4′,6-diamidino-2-phenylindole) was observed.

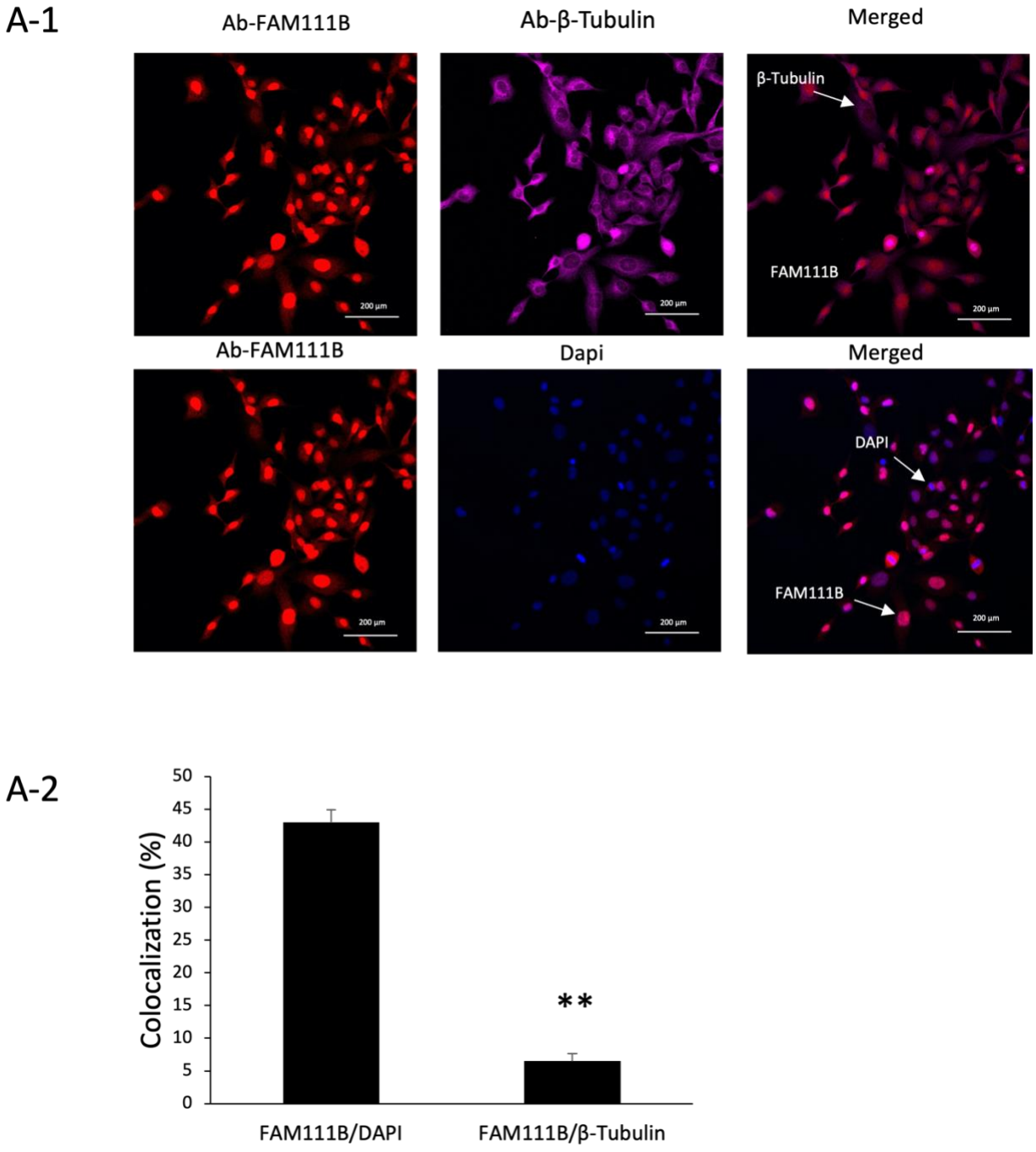


Figure 3. 5 ***<u>Subcellular localization of FAM111B</u>***

(A-1) Representative confocal microscopy images showing the co-localization of FAM111B in the nucleus and cytoplasm. Localization of FAM111B in the cytoplasm was determined by staining cells with FAM111B and Cy3 primary and secondary antibodies (red), respectively, and β-Tubulin

*(cytoplasmic marker) and Cy5 primary and secondary antibodies (magenta), respectively. Colocalization of FAM111B in the nucleus was determined by staining cells with FAM111B and Cy3 primary and secondary antibodies (red) and nuclear marker DAPI (blue). Merged images of FAM111B with β-tubulin and DAPI can be seen in the right panel. The scale bar for all images is 200μm and images were captured at 20X magnification (A-2). The percentage of colocalization of FAM111B with β-tubulin and DAPI. Each experiment was performed three times with similar results. Statistical analysis was performed using a t-test, and p-values were provided (*p ≤ 0.05; **p ≤ 0.005). The scale bar for all images is 100μm.*

3.5 Functional studies of the human *FAM111B* gene

3.5.1 Cell-based functional studies

3.5.1.1 Proliferation assay

Cell proliferation is tightly regulated in normal cells by processes such as programmed cell death and cell cycle. In contrast, cancerous cells show an exaggerated rate of cell proliferation that can develop into a tumour. Altered regulation of gene expression is a crucial determinant for excessive cell proliferation (Y. Peng, Li, & Zhu, 2019). This study investigated the effect of *FAM111B* knockdown and overexpression on cell proliferation (Figure 3.6).

No difference in proliferation between cells treated with NT-siRNA and *FAM111B*-siRNA was observed in the first 20hrs. However, a decline in proliferation was observed from 20-50hrs post-transfection. Within 20-50hrs, the cell index (CI) increased from 0.44 to 3.11 in control (NT), while the CI increased from 0.36 to 2.88 in *FAM111B*-siRNA transfected cells. Figure 3.6-B shows the proliferation rate of cells transfected with FAM111B-EGFP overexpression and the Y621D-mutant. The graph indicates that cells transfected with Y621D-mutant has a significantly reduced rate of proliferation between 30-50hrs post-transfection relative to the control, EGFP. Within 30-50hrs, the negative control (EGFP) showed a CI increase from 1.29 to 3.52, whereas cells transfected with the Y621D mutant revealed a CI increase from 1.05 to 3.14. Meanwhile, the overexpression of the *FAM111B* gene did not show a change in cell proliferation.

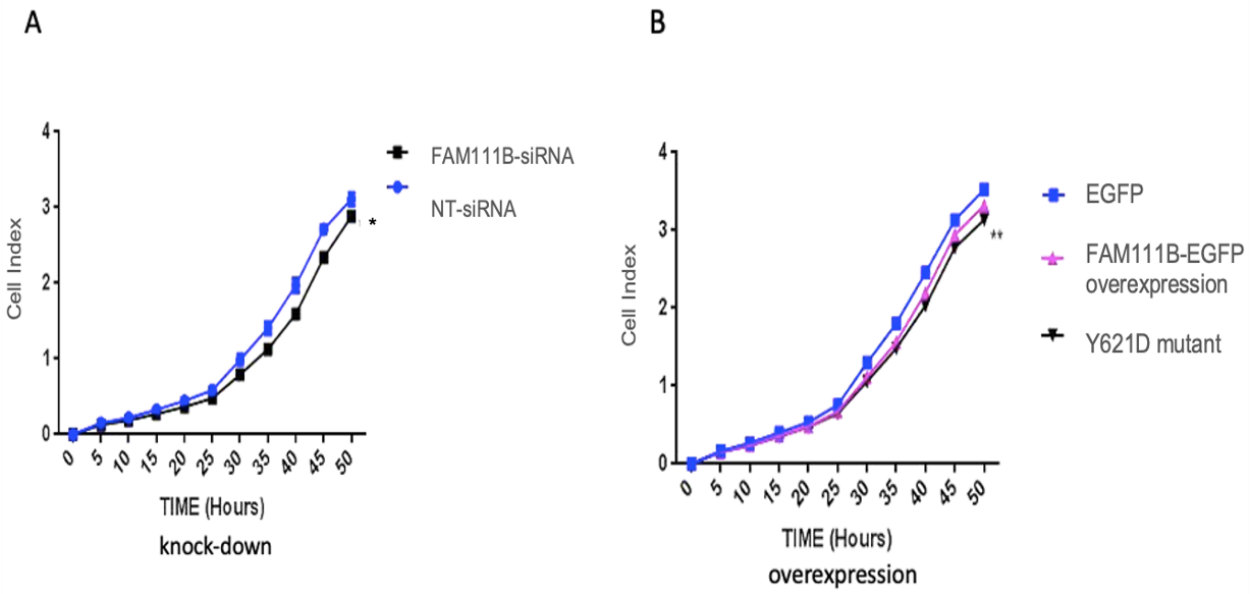


Figure 3. 6 ***The effect of FAM111B knockdown and overexpression on cellular proliferation***

(A) A graph showing the change in the comparative cell index (CI) in FAM111B-siRNA and control, NT-siRNA treated cells 24 hours post-transfection. (B) A graph showing the change in the comparative CI in FAM111B-EGFP overexpressed, Y621D-mutant and negative controls, EGFP, treated cells 24 hours post-transfection. A significant difference in proliferation was observed in Y621D-mutant cells relative to the control EGFP. Each experiment was performed in triplicate and repeated three times. Statistical analysis was performed using repeated measures ANOVA and p-values were provided (= $p \leq 0.05$; **=$p \leq 0.005$).*

3.5.1.2 Cell apoptosis assay

Flow cytometry was used to examine whether *FAM111B* plays a role in apoptosis. Cells were stained with Annexin V-FITC/PI after knocking down and overexpressing the gene. HT1080 cells transfected with *FAM111B*-siRNA showed a 6% increase in apoptotic cells, while the cells transfected with the *FAM111B*-EGFP showed a 24% decrease of apoptotic cells. However, the Y621D-mutant transfected cells also exhibited a slight decreased of 9%.

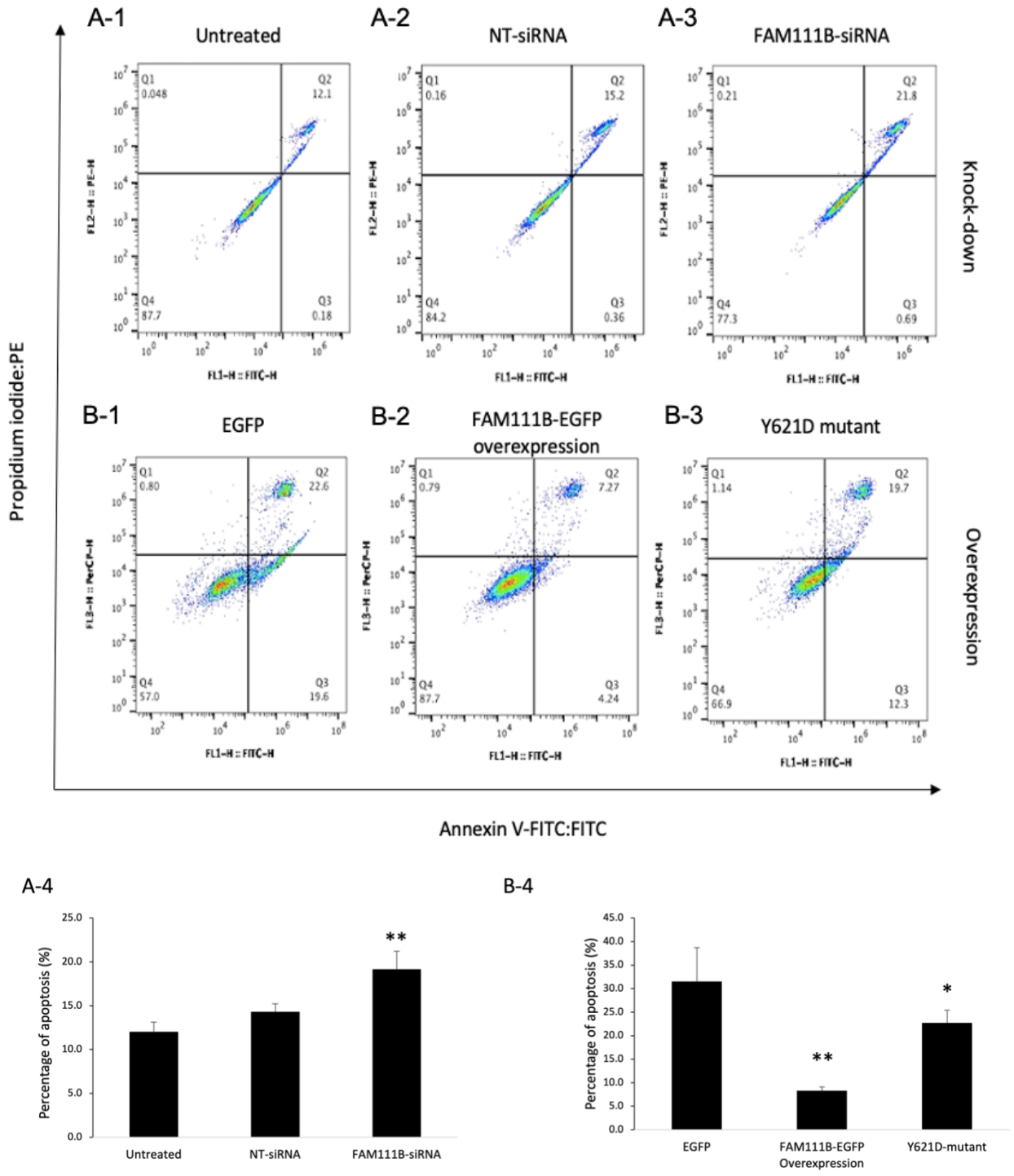


Figure 3. 7 ***The effect of FAM111B knockdown and overexpressing on cell apoptosis.***

Representative flow cytometric scatter plots showing the percentage of cells in viable (annexin VFITC-/PI- cells, Q4), early apoptotic (annexin V-FITC+/ PI-, Q3), late apoptotic (annexin V-

FITC +/ PI +, Q2) and necrotic stages (PI +/ annexin V-FIT-, Q1). (A-4) Quantification of apoptotic FAM111B knocked-down HT1080 cells (Annexin V-FITC positive cells) 24 hours post-transfection. (B-4) Quantification of apoptotic FAM111B overexpressed HT1080 cells (Annexin V-FITC positive cells) 24 hours post-transfection. Each experiment was performed in triplicate and repeated three times. Statistical analysis was performed using a t-test, and p-values were provided (= p ≤ 0. 05; **=p ≤0.005).*

3.5.1.3 Cell migration/scratch assay

We investigated the role of *FAM111B* on cell migration as a potential driving force in the pathogenesis and progression of POIKTMP. Our study showed that when cells were transfected with *FAM1111B*-siRNA, the migration was reduced, as demonstrated by the decrease of the scratch area covered by cells from 38% (NT-siRNA) to 24% (*FAM111B*-siRNA) (Figure 3.8-A). When overexpressing the gene by transfecting with *FAM111B*-EGFP overexpression plasmid, the cells showed a significant increase in migration as the % of the scratch area covered by cells increased from 38% (EGFP) to 65% (*FAM111B*-EGFP overexpression) (Figure 3.8-B). Cells that were transfected with Y621D-mutant plasmid also showed a significant increase in % of the scratch area covered by cells from 38% (EGFP) to 56% (Y621D-mutant), however not as much as the overexpressed plasmid.

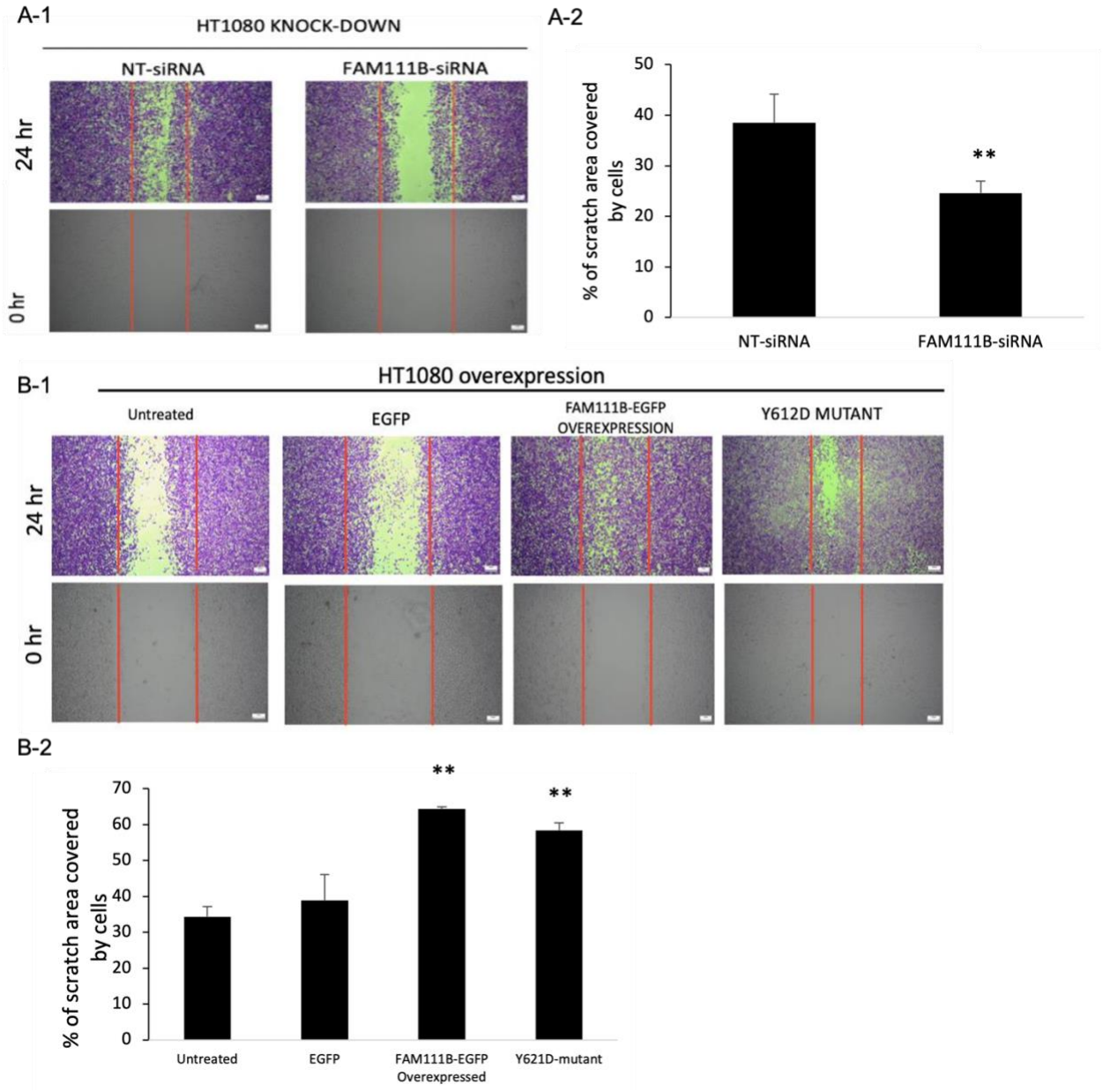


Figure 3. 8 ***Role of FAM111B in cell migration***

The FAM111B-siRNA (A-1) and FAM111B overexpressed (B1) cells were scratched with a 1000 µl pipette tip when the cells reached confluency. The red lines were used to mark the ranges of the scratches. The images were taken at 0 and 24 hours after the scratch was created. All images were captured at 40x magnification. (A-2) Quantification of FAM111B-siRNA was relative to the negative control, NT-siRNA. The graph illustrates the % of the scratch area covered by FAM111B

knocked-down cells. (B-2) Quantification of FAM111B overexpression and Y621D-mutant transfected cells was relative to the negative control, EGFP and Untreated. Data is expressed as the mean ±SEM (= $p \leq 0.05$, **=$p \leq 0.005$) of three independent experiments.*

3.5.1.4 Cell invasion assay

The ability of cancerous cells to invade is one of the hallmarks of the metastatic phenotype. Due to *FAM111B*'s association with cancer, it was crucial to investigate its role in cell invasion. Transwell plates coated with ECM were used to determine cells' invasive potential after knocking down and overexpressing the *FAM111B* gene. The percentage of area covered by cells was quantified by Image J; however, no significant difference in invasive potential was observed when compared to the controls for both knockdown (Figure 3.9-A) and overexpression (Figure 3.9-B). Interestingly, cells transfected with the Y621D-mutant showed a significant decline in invasive potential as the percentage of area covered by cells decreased from 58% (EGFP) to 34% (Y621Dmutant). This result suggests that *FAM111B*'s association with invasion is not related to the lowered expression of *FAM111B* but due to the presence of the Y621D-mutation indicating a potential gain-of-function.

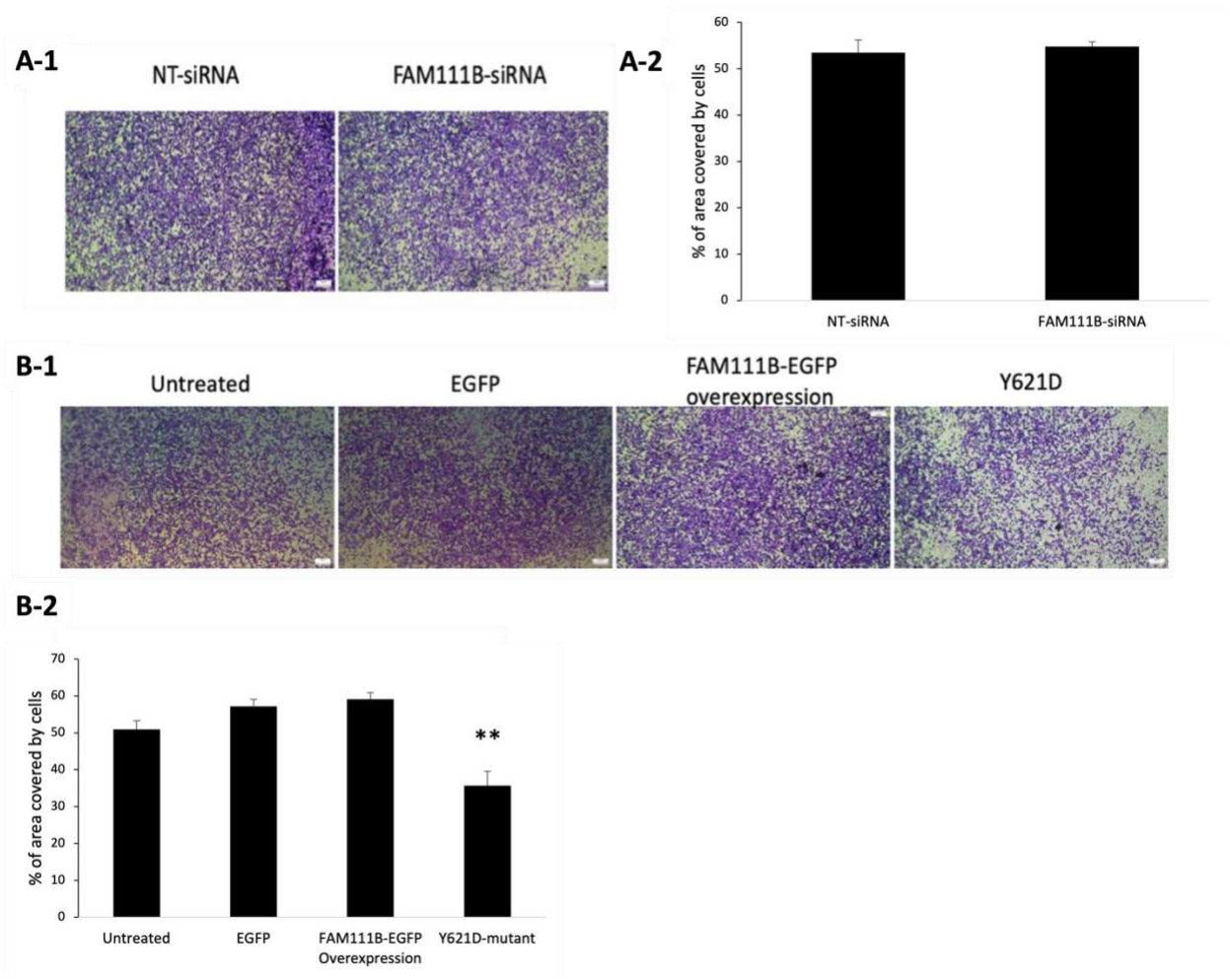


Figure 3. 9 ***<u>The effects of FAM111B knockdown and overexpression on cell invasion.</u>***

*The FAM111B-siRNA (A-1) and FAM111B overexpressed (B1) cells seeded in Transwell plates and images were captured 24 hours post-transfection at 40x magnification. (A-2) Quantification of FAM111B-siRNA was relative to the negative control, NT-siRNA. The graph illustrates the % of the area covered by FAM111B knocked-down cells. Data are expressed as the mean ±SEM (T-test, *p ≤ 0.05, ** p ≤ 0.01) of three independent experiments. (B-2) Quantification of FAM111B overexpression and Y621D-mutant transfected cells was relative to the negative control, EGFP and Untreated. Data is expressed as the mean ±SEM (*=p ≤ 0.05, **=p ≤ 0.005) of three independent experiments.*

3.5.2 Mass spectroscopy and proteomics analysis

3.5.2.1 Differential gene expression analysis for potential biomarkers

Characterizing proteins is essential for determining their functional state, which may have important implications for several biological processes. Protein regulates many cellular processes, such as cell communication, replication, migration, apoptosis, and proliferation. As structural biologists identify a growing number of diseases related to gene mutations resulting in malfunctioning protein and subsequent dysregulation, understanding, and identifying the differences between the two states of the protein (active or inactive) has become a priority. The use of mass spectrometry could help with understanding the pathogenesis of the human FAM111B protein in the progression of POIKTMP, specifically to identify differentially expressed proteins. In this study, we conducted a preliminary label-free shotgun proteomic analysis to identify differentially expressed proteins. Figure 3.10 illustrates proteomics data obtained from cells that were transfected with the *FAM111B*-EGFP overexpression and Y621D-mutant. Figure 3.10-A shows a heatmap that displays the clustering of the groups (Untreated, EGFP, *FAM111B*-EGFP overexpressed, and Y621D) by intensity-based absolute quantification (iBAQ) intensity of proteins expression. The negative controls (EGFP and Untreated) indicating similarity, while *FAM111B*-EGFP overexpression and Y621D clusters separately. Protein enrichment of heat-shock protein, HSP7C, was observed when a t-test was performed to analyse the differentially expressed proteins when comparing groups; control (Untreated/EGFP) vs experimental group (*FAM111B*EGFP overexpression/Y621D-mutant) (±2SD) (Figure 3.10-B-1). A t-test between *FAM111B*EGFP overexpression and Y621D-mutant was performed, and the protein G3V3W4, a proteosome subunit alpha, was shown to be differentially expressed enriched in Y621D-mutant (±2SD) (Figure 3.10-B-2).

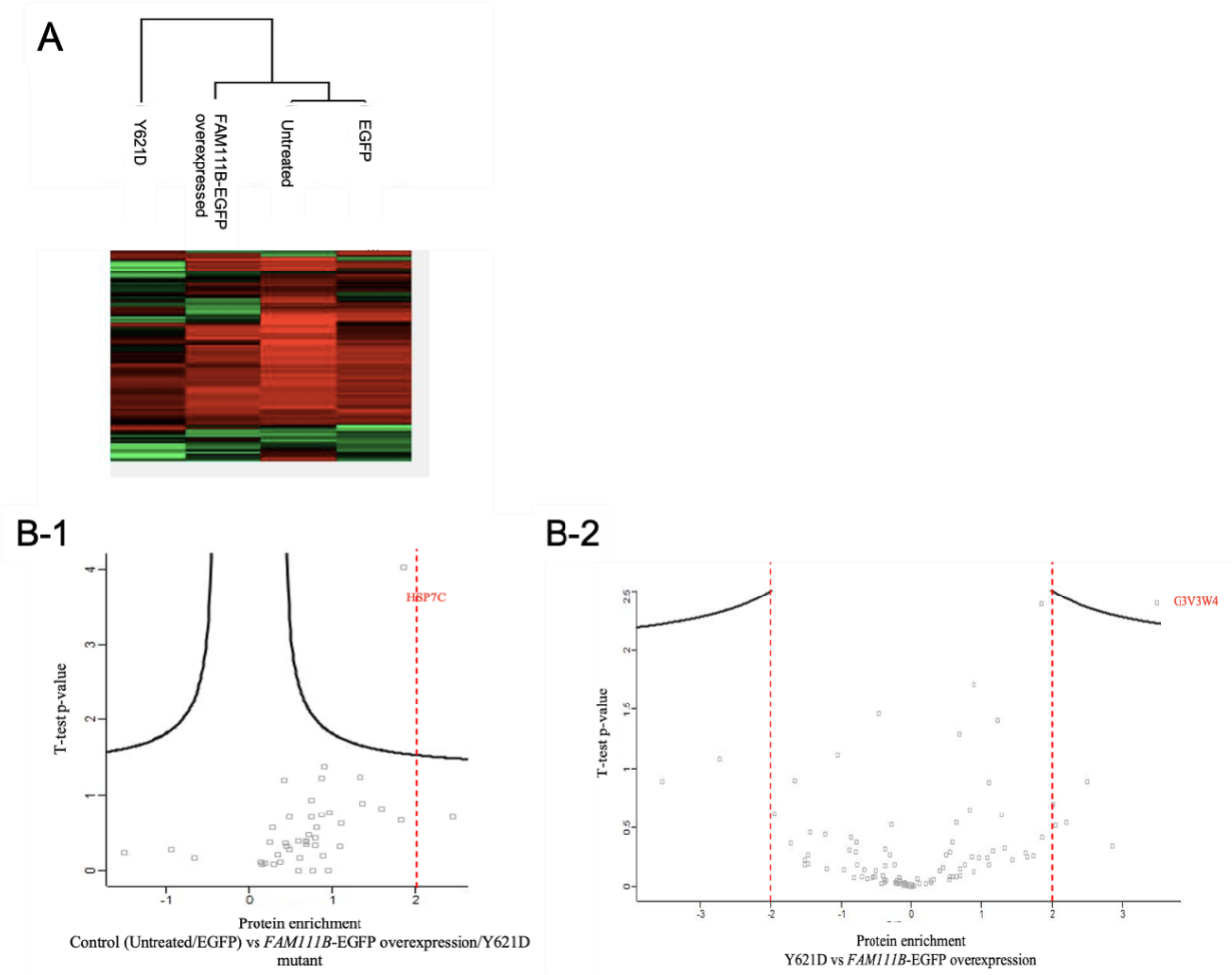


Figure 3. 10 ***Label-free shotgun proteomics analysis findings***

(A) Heatmap of protein abundance pattern in cells transfected with EGFP-tagged (negative control), FAM111B-EGFP overexpression, and Y621D-mutant plasmids. Untreated cells were used as a control. (B-1) Volcano plot with protein enrichment from mass spectrometry (MS) analysis of cells transfected with FAM111B-EGFP overexpression and Y621D-mutant plasmids. The plot shows the differences in protein enrichment between control (Untreated/EGFP) and test group FAM111B-EGFP overexpression/Y621D-mutant with t-test evaluation. In FAM111BEGFP overexpression/Y621D-mutant's proteome, there is one protein, HSP7C, significantly enriched. (B-2) Volcano plot with protein enrichment from MS analysis of cells transfected with FAM111B-EGFP overexpression and Y621D-mutant plasmids. The plot shows the differences in protein enrichment between FAM111B-EGFP overexpression and Y621D-mutant with T-test evaluation.

In the FAM111B-EGFP overexpression proteome, there is one protein significantly enriched, G3V3W4. Statistical analysis was performed using a t-test, and p-values were provided

(= p ≤ 0. 05; **=p ≤ 0.005). The red dotted line indicated ±2Standard deviation (SD).*

3.6 Validation of Y621D patient mutation by PCR-RFLP

Polymerase chain reaction-restriction fragment length polymorphism (PCR-RFLP) was used to confirm the South African *FAM111B* heterozygous [NM_198947.4: c.1861T>G (p. Tyr621Asp)] gene mutation. This method allows for distinguishing mutant-type and wild-type sequences by generating restriction enzymes to digest targeted restriction sites that are only present in mutanttype sequences, producing distinct polymorphic fragments used as markers for mutation identification. We anticipated a loss of the *BstZ17I* restriction site in the Y621D mutated form of this gene, resulting in three DNA fragments with sizes 1085 (non-cleaved DNA), 740 and 345 bp. However, the wild-type gene was expected to display complete cleavage of the DNA to produce fragments with 740 and 345 bp sizes. The data obtained from our PCR-RFLP analysis confirmed these predicted DNA fragments (Figure 3.11). Lane 1 shows the unaffected mother's DNA displaying two bands (740 and 345 bp) because of the absence of the mutation. In comparison, Lane 2 shows the POIKTMP patient DNA band pattern with three bands (1085, 740 and 345 bp) confirming the presence of the Y621D mutation.

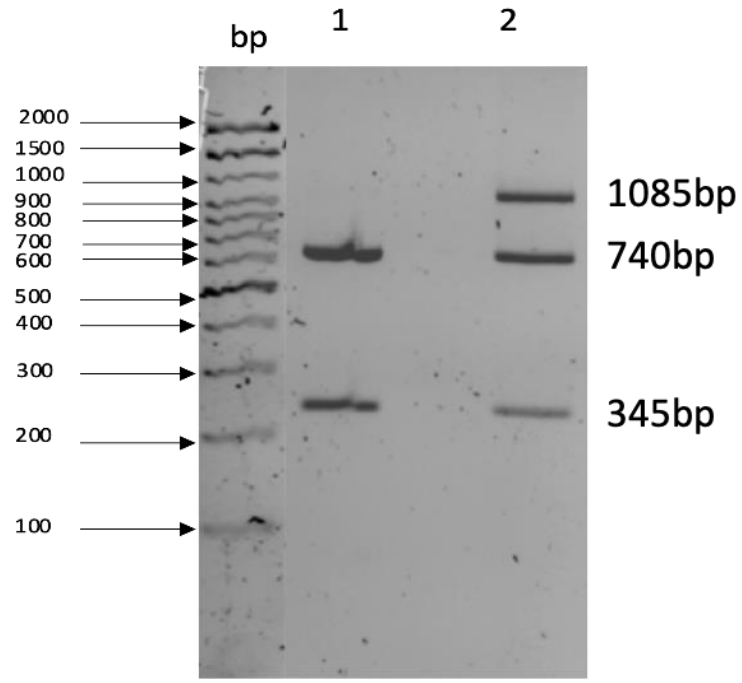


Figure 3. 11 ***PCR-Restriction fragment length polymorphism (PCR-RFLP) genotyping of the South African family with the FAM111B heterozygous Y621D gene mutation.***

Lane 1: Control (unaffected mother), Lane 2: POIKTMP affected daughter.

3.7 *FAM111B*'s expression pattern in POIKTMP patient.

Based on the results obtained from the cell lines, an investigation into the expression of *FAM111B* in POIKTMP patients was of interest. We assessed the gene and protein expression in a POIKTMP patient and a healthy patient (familial control). The *FAM111B* gene expression is down-regulated in POIKTMP patients skin fibroblasts, as shown in Figure 3.12-A (18-fold decrease) relative to the expression of the healthy familial control (mother). The FAM111B protein expression was also significantly downregulated, as observed in the western blot in Figure 3.12-B (2-fold decrease). This data indicates that the patient mutation affects the expression of the *FAM111B* gene and consequently alters its function.

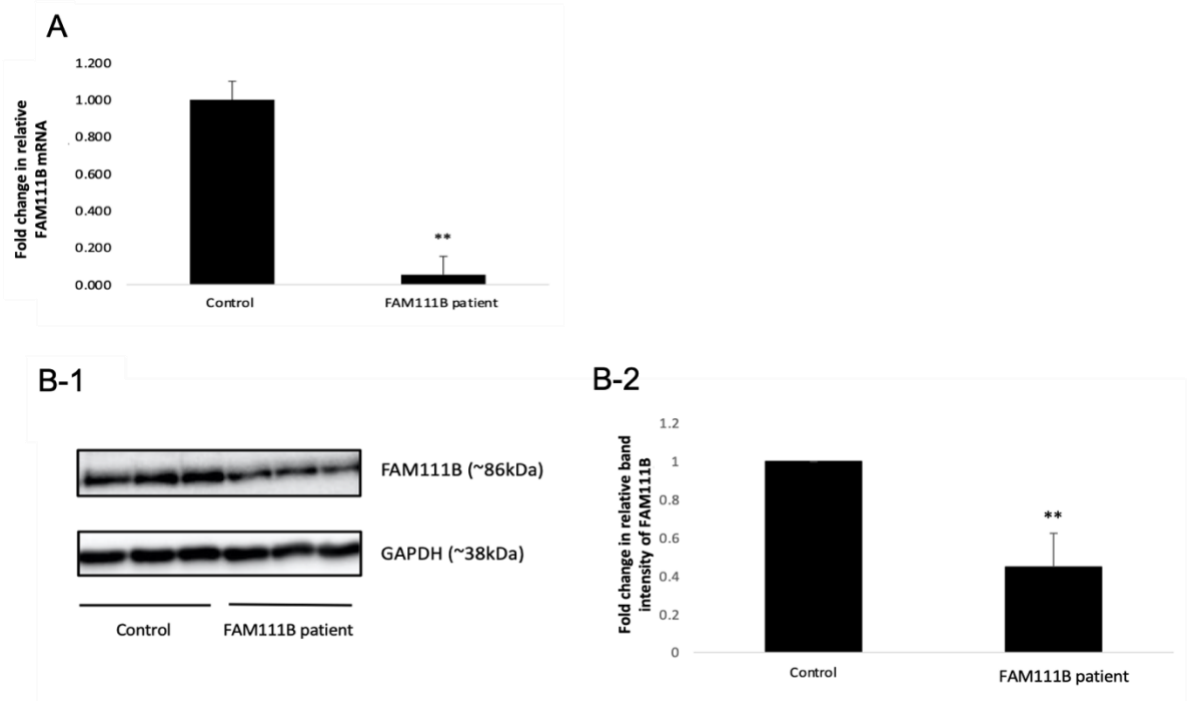


Figure 3. 12 ***<u>The gene and protein expression of FAM111B in patient and familial control by qPCR and Western blot.</u>***

*(A) FAM111B mRNA expression levels in POIKTMP patient and familial control were quantified using qPCR. GADPH primers were used to normalize the results. Results are a representation of three experiments performed in triplicate. (B-1) FAM111B protein expression levels in POIKTMP patient and familial control were quantified by western blotting. GADPH antibody was used as a housekeeping gene to normalize the results. (B-2) A graph showing the relative band intensity of the western blot results. Quantification was relative to the familial control protein expression levels, and results were normalized to GAPDH. Data is expressed as the mean ±SEM (t-test, *=p ≤ 0.05, ** =p ≤ 0.005) of three independent experiments.*

This study's overarching goal was to investigate the cellular function of the human *FAM111B* gene and gene product. The findings in this study suggest that *FAM111B* promotes fibrosis by inhibiting proliferation, migration and invasion and stimulating cell apoptosis in patients with POIKTMP.

4. Chapter 4: Discussion and conclusion

4.1 *FAM111B* expression in cell lines

First, an *in-silico* gene expression study of the *FAM111B* gene in cells/tissue on The Human Protein Atlas (https://www.proteinatlas.org/ENSG00000189057-FAM111B/cell) revealed that about 68% of the cell lines expressing *FAM111B* are derived from cancerous tissue and 32% from non-cancerous tissue (Figure 3.1). A 2019 study identified *FAM111B* as a potential therapeutic target for LUAD due to its association with cancer and its oncogenic characteristics (Sun et al., 2019). Later, in 2021, another study showed that cells lines derived from cancerous tissue might have higher expression of the *FAM111B* gene (C. Chen et al., 2021; Kaczkowski et al., 2016; Sun et al., 2019). The study results demonstrated very high expression levels of *FAM111B* in tumour tissues compared to healthy tissues. Initially, cell lines HEK293 (NX=3.3), A549 (NX=6.3) and HeLa (NX=12) were screened based on the genomics and transcriptomics data shown in the Human Protein Atlas database that showed a positive expression above NX=0.1 (Figure 3.1). The low expression of *FAM111B* in HEK293 compared to A549 and HeLa validated the hypothesis that *FAM111B* expression is upregulated in cancer tissue.

Furthermore, we observed that HT1080, a fibrosarcoma connective tissue cell line, was not listed in this database, so we sought to access the expression of the *FAM111B* gene in this cell line to add to the list of cancer-derived cells lines expressing or overexpressing this gene. *FAM111B* gene and protein expression studies were performed on cell lines HEK293, A549, HeLa, and HT1080 cell lines, showing a similar trend whereby *FAM111B* is upregulated cancer-derived cell lines as seen in the literature (Figure 3.2). qPCR data showed that the level of *FAM111B* mRNA is visibly higher in cell lines HeLa and A549, and the protein expression was higher in cell lines HeLa and HT1080. The expression levels of mRNA and protein in a cell line may not always correlate and may be as little as 40% similar (Vogel & Marcotte, 2012). The expression level for protein and mRNA is regulated by intracellular processing involving post-transcription regulations and post-translational modifications. Post-transcriptional regulators include mRNA stability, proteolysis, and translation (Brion, Lutz, & Albert, 2020). Post-translational modifications include proteolytic cleaving and attachment of functional groups to the protein, which modulates the function of the proteins (Uversky, 2013).

Although the expression of *FAM111B* mRNA in HT1080 (Fibrosarcoma cell line) cells were low, the protein expression was significantly high. To find a normal fibroblast cell line as a control to HT1080 cell line, MRC5 was screened for the expression of *FAM111B* (Figure 3.2 C). The western blot data shows that the expression of FAM111B in MRC5 is almost non-existent. The rationale for using the HT1080 cell line for subsequent downstream experiments in this study is the insufficient exploration of the *FAM111B* gene expression and functional characterization in this cell line.

4.2 *FAM111B* knockdown and overexpression

RNA-interference (RNAi) is a powerful tool for the study of gene function. These small interfering RNAs (siRNA) allows for recognition by the enzymatic machinery of RNAi, eventually leading to degradation of targeted mRNA (Mahajan et al., 2020). The efficiency of the siRNA-*FAM111B* knockdown was validated by qPCR for gene expression, western blot, and immunofluorescent confocal microscopy for protein expression (Figure 3.3). Together the results show that *FAM111B* was successfully and significantly knocked down following a 24-hour transfection. Successful knockdown efficiency is often detected between 24-72 hours; however, we observed the optimal transfection at 24 hours in our study. Hence all subsequent experimental transfections were performed following a 24-hour transfection. Sun et al. and Kawasaki et al. also reported the successful siRNA transfection using Lipofectamine RNAiMax between 24-48hrs post-transfection (Kawasaki et al., 2020; Sun et al., 2019). The limitations of this technique include off-target effects, non-specific immune response, and time constraining transient transfection. Thus, to combat these effects, a control non-targeting siRNA (NT-siRNA) is required. GAPDH as a housekeeping gene or normalized control gene in expression studies is common as it is often expressed in many cell lines and remains unchanged across treatment/experimental conditions (similar in our study as well).

Gene silencing can be achieved by various techniques, including RNAi and CRISPR, which are global techniques used to explore the cellular function of genes. In this study we used FAM111B gene-specific, siRNA to downregulate the gene in HT1080 cells. Hence, all expression data (i.e., qPCR and western blot) in the study were normalized GAPDH. The primary difference between RNAi and CRISPR is that RNAi transiently knocks down gene expression at the mRNA level, while CRISPR completely and permanently knockouts the gene at the DNA level (Boettcher &

McManus, 2015). Kawasaki et al. and Hoffmann et al. both performed CRISPR to knockout the *FAM111B* to study its function; however, the knockout of essential genes providing only limited information regarding gene function in studies where the gene of interest is critical to the survival of the organism (Hoffmann et al., 2020; Kawasaki et al., 2020). In such cases, transient gene knockdown is safer. We can better understand the gene function and its effect on the disease phenotype because reducing protein levels to different degrees can be studied.

Furthermore, the reversible nature of knockdowns makes it possible to verify the phenotypic effect by reverting protein expression to its normal level in the same cells (Gilbert et al., 2014). These features provided this study with a rationale to perform *FAM111B*-siRNA knockdown experiments to functionalize the protein.

To understand the cause and pathogenesis of a phenotype, the identification of genetic mutations is typically one of the first steps to explore potential biological pathways. However, overexpression of a wild-type gene product can result in the formation of mutant phenotypes, which allows for the identification of pathway components that are not detected when using traditional gene editing tools and analysis to study loss-of-function (Prelich, 2012). Therefore, we explored the effects of overexpressing the *FAM111B* gene in HT1080 cells. We validated the *FAM111B* gene and protein overexpression by qPCR, western blot, and immunofluorescent confocal microscopy, respectively (Figure 3.4). Together, the results again show that *FAM111B* was successfully and significantly overexpressed following a 24-hour transfection. Hence, all ensuing experiments was performed at 24h post-transfection. Elevated levels of *FAM111B* were detected by qPCR, western blot, and immunofluorescent confocal imaging. In addition to the overexpression of *FAM111B*, the effect of transfecting HT1080 cells with a Y621D-mutant plasmid was of interest. The Y621D plasmid contains the *FAM111B* mutation that is present in a South African POIKTMP patient. The expression of *FAM111B* was downregulated in HT1080 cells transfected with the Y621D-mutant plasmid. Sun et al. reported that *FAM111B* was upregulated in lung adenocarcinoma (LUAD) tumour tissue compared to normal cells. Thus, they performed a transfection with an overexpressed *FAM111B* plasmid in lung tissue. This increase in expression led to an increase in proliferation which suggests that *FAM111B* is involved in the development of LUAD (Sun et al., 2019). This result encouraged an investigation into the effect

of overexpressing *FAM111B* in HT1080 cells to establish the impact of this gene on cellular proliferation, apoptosis, migration, and invasion.

4.3 FAM111B subcellular localization

The subcellular localization of FAM111B and co-localization with cytoplasmic and nuclear proteins were investigated (Figure 3.5). FAM111B showed a 44% co-localization with the nucleus and only 8% with the cytoplasm. Thus, this data suggests that FAM111B is primarily localized in the nucleus, an observation supported by The Human Protein Atlas (https://www.proteinatlas.org/ENSG00000189057-FAM111B/cell). Furthermore, this strong nuclear colocalization may imply that FAM111B plays a crucial cellular function in the nucleus. Protein subcellular localization prediction is an essential aspect of protein function prediction and genome annotation, aiding in discovering therapeutic drug targets.

4.4 The effects of knockdown and overexpression of *FAM111B* on critical cellular functions

4.4.1 Proliferation

A real-time cell analyser assay was performed to determine the role of *FAM111B* in proliferation. Our results showed that when *FAM111B* was knocked down, the rate of proliferation decreases (Figure 3.6 A). Furthermore, overexpressing *FAM111B* showed no effect on cellular proliferation (Figure 3.6 B). This observation could be due to HT1080 cells already overexpressing *FAM111B* being a cancerous cell line. Hence, the overexpression by transfection could result in no significant effect on the expression, therefore, having no impact on the proliferation. The effect of the *FAM111B* mutations on the functional characterization of the gene is not previously explored. We observed that cells transfected with the Y621D-mutant plasmid showed a similar trend to the knocked down cells in which the rate of proliferation decreased. A study that has been published showed that *FAM111B* gene knockdown causes a decrease in proliferation (Sun et al., 2019). A report by Kawasaki *et al.* confirms these results by illustrating that the knockout of the *FAM111B* gene by CRISPR results in a reduced rate of proliferation in A549 cells; however, these results were only obtained when cells were cultured in serum-free media (Kawasaki et al., 2020). The discrepancy might be due to the different methods of gene silencing: siRNA mediated gene

silencing vs CRISPR gene knockout. Knockdown by siRNA transiently downregulates gene expression at the mRNA level, whereas CRISPR/Cas9 knockout system entirely silences the gene at the DNA level (Prabhune, 2021).With recent publications, data suggests that the *FAM111B* mutation is associated with the regulation of proliferation in POIKTMP.

4.4.1 Apoptosis

We evaluated *FAM111B*'s role in cellular apoptosis by performing flow cytometry with Annexin V-FITC/PI staining. These stains can differentiate viable, apoptotic, and necrotic cells. These cells can be detected as apoptotic cells. In apoptotic cells, the phosphatidylserine of the plasma membrane flips from the inside surface to the outside surface allowing Annexin V to bind specifically to phosphatidylserine. In healthy viable cells and cells undergoing apoptosis, the cell membrane integrity excludes Propidium Iodide (PI); however, PI can penetrate necrotic cells due to the loss of its cell membrane integrity (Rieger, Nelson, Konowalchuk, & Barreda, 2011).

FAM111B expression was reduced by siRNA-induced knockdown, and the results showed a significant increase in the percentage of apoptotic cells (Figure 3.7 A). Further, the overexpression of *FAM111B* showed a decline in the rate of apoptotic cells (Figure 3.7 B). These results were consistent with the results reported by Sun et al., who also demonstrated that siRNA mediated knockdown of the *FAM111B* gene causes an increase in cellular apoptosis (Sun et al., 2019). In the same study, they observed that silencing *FAM111B* in cells caused cell cycle arrest in the G2/M phase, leading to more apoptosis. Cell apoptosis is tightly associated with pulmonary fibrosis, as indicated by *Kawasaki et al.,* who observed an increase in apoptosis when *FAM111B* was knocked down in lung adenocarcinoma cells (Kawasaki et al., 2020). Cells transfected with the Y621D mutant plasmid showed a reduction in apoptosis, suggesting that the mutation applies a different mechanism to the knocked down cells, affecting cell survival.

Contrary to *Hoffman et al., FAM111B* disease-associated mutants transfected into cancerous epithelial cell lines showed an elevated apoptotic reaction (Hoffmann et al., 2020). The inconsistency of these results could result from the difference in cell lines used to transfect the

disease-associated mutants. Together these results encourage future work in cell lines derived from different types of tissues.

Studies have reported anti-apoptotic effects resulted in a significant decline in the progression of fibrosis (Johnson & DiPietro, 2013). However, it is essential to note that a wide variety of stimuli can trigger apoptotic death, and, significantly, not all programmed cell death is triggered by the same stimulus. Dysregulation of apoptosis is implicated in various diseases, including fibrosis (Uhal, 2008). A study that investigated the importance of apoptosis in the progression of pulmonary fibrosis showed that epithelial apoptosis was stimulated. In contrast, apoptosis of myofibroblasts was inhibited in the lungs of patients diagnosed with IPF. These findings imply a relationship between dysregulated apoptosis and the development of pulmonary fibrosis (Uhal, 2008). These results coincide with our results that revealed a decrease in apoptosis observed in HT1080 cells transfected with the Y621D mutation present in the South African POIKTMP patient. This data suggests that the presence of the mutation inhibited the cellular apoptosis in the fibroblasts leading to the progression of fibrosis.

4.4.3 Migration

The migration of fibroblasts may be necessary for the pathogenesis of pulmonary fibrosis (Suganuma, Sato, Tamura, & Chida, 1995). We investigated the role of *FAM111B* in migration. Our data showed that when cells were deprived of *FAM111B,* there was a decrease in migration from 38% to 24% (Figure 3.8 A). Overexpressing the *FAM111B* gene showed an increase in migration from 35% to 65% (Figure 3.8 B). The data generated from this assay suggests that FAM111B might play a significant role in developing fibrosis by influencing the rate of migration. Cell migration of fibroblasts is tightly associated with the progression of pulmonary fibrosis, and our results showed that silencing and overexpression of *FAM111B* regulated the migration of HT1080 cells. Suganuma *et al.* indicated that the migration of cells is intensified as the stages of fibrosis progress (Suganuma et al., 1995). The authors of this article concluded that the migration rate of fibroblasts from fibrotic lungs is amplified in comparison to the healthy controls, further demonstrating the association between migration and the pulmonary fibrotic process. Cells transfected with the Y621D mutation also showed an increase in migration from 35% to 56% (Figure 3.8 B). This data implies that the manifestation of the *FAM111B* mutation directly

regulates the migration of the fibroblasts and alludes to the involvement of *FAM111B* in the pathogenesis of POIKTMP.

4.4.4 Invasion

The mechanism of fibroblast invasion through interstitial tissue is one pivotal landmark in the process of tissue remodelling and repair, a process that also plays a vital role in fibrogenesis, carcinogenesis and metastasis (Oehrle et al., 2015). A tumour's environment consists of CAF's that are known to stimulate invasion (Attieh et al., 2017). In this study, we explored whether *FAM111B* induces or inhibits cellular invasion. Our results show that depletion and overexpression of *FAM111B* led to no effect on the rate of invasion of HT1080 cells (Figure 3.9). These results are consistent with Kawasaki *et al.,* who also showed no significant effect of knocking down *FAM111B* on the invasive capabilities of A549 cells (Kawasaki et al., 2020). However, contradicting results were observed in an article by *Sun et al*. that displayed a decline in the invasion of A549 cells in the absence of *FAM111B*. Moreover, we investigated the effects of transfecting HT1080 cells with the Y621D mutation, resulting in decreased invasion as the area covered by cells reduced from 58% (control) to 34% (Figure 3.9 B). The conflicting results obtained encourage further investigation into the role of *FAM111B* in cell invasion.

4. 5 Identification of interacting proteins by Mass Spectrometry

A preliminary study was performed to identify novel FAM111B-interacting proteins by mass spectrometry (MS). Using unsupervised hierarchical clustering of the protein groups untreated, EGFP, EGFP-*FAM111B* in overexpressed wild type *FAM111B* and Y621D groups, a similar molecular signature was found for untreated and EGFP groups (negative controls) (Figure 3.10 A). However, there was a distinct signature for EGFP-*FAM111B* overexpressed, utterly different from untreated and EGFP. There was also a distinct difference between the Y621D group compared to all other groups. Our hierarchical clustering result interestingly differentiated all groups quite clearly due to the mass spectrometry instrument's high mass accuracy and resolution.

We further demonstrated using this method to identify proteins that specifically interact with mutant forms of *FAM111B.* MS is one of the most potent high-thorough tools to identify protein-

protein interactions and novel protein pathways, leading to increased knowledge available regarding the systemic inflammatory response of tissue/cells to therapeutic drugs and therapies (Y. Huang et al., 2012). Furthermore, we identified differentially expressed peptides between the groups (Figure 3.10 B). The dependent variables were normalized iBAQ values for proteins, while the independent variables were untreated, EGFP, EGFP-*FAM111B* overexpressed and Y621D. A pair-wise independent sample t-test was performed with this method, which found that two proteins were differentially expressed. Firstly, the control group (Untreated/EGFP) *vs* test group (*FAM111B*EGFP overexpression/Y621D-mutant) identified protein Heat shock protein cognate 71 (HSP7C) (Figure 3.10 B1). Protein HSP7C (gene name Heat shock protein family A (Hsp70) member 8 (HSPA8)) is a molecular chaperone of the heat shock protein 70 family that functions in ATP binding; ATPase activity; C_3HC4-type RING finger domain binding that plays an important role in the ubiquitination pathway; chaperone binding; enzyme binding; G protein-coupled receptor binding; heat shock protein binding; phosphatidylserine binding; protein binding and bridging; ubiquitin protein ligase binding; unfolded protein binding all mediated by co-chaperones. It plays a crucial role in controlling the quality of proteins by protein homeostasis that degrades misfolded proteins and ensures proper protein folding. These functions are achieved through cycles of ATP binding, ATP hydrolysis and ADP release mediated by co-chaperones (https://www.phosphosite.org/proteinAction.action?id=4119&showAllSites=true).

Interestingly, co-chaperones known as nucleotide exchange factors (NEF) such as BAG1/2/3, facilitates the conversion of HSP7C from the ADP-bound to the ATP-bound state, thereby promoting substrate release. A study by Sun *et al*. found that FAM111B interacts with BAG3. A significant association between proteins BAG3 and BCL2 was observed. This observation encourages further investigation into the biological pathway which found that upregulation of BAG3 and BCL2 caused a decrease in apoptosis, which suggests the essential role of these proteins in the progression of cancerous tissue (Sun et al., 2019). Further studies showed that by downregulating FAM111B, both BAG3 and BCL2 proteins were proportionally decreased. These results, together with ours, suggests that dysregulation of FAM111B affects the expression of BAG3 and BCL2 alongside HSP7C. These effects result in the aggregation of misfolded proteins in the endoplasmic reticulum (ER), impair normal cellular function, and creates a toxic cellular environment, leading to cell death (Haeri & Knox, 2012). As reported in our study, *FAM111B*

plays an imperative role in cell apoptosis, further confirming the potential pathways of interactions resulting in fibrosis.

Further, Hoffmann *et al* report suggested an increase in catalytic activity of FAM111B resulting from reported patient mutations. This increase in catalytic activity affects DNA replication and transcription, leading to increased apoptosis (Hoffmann et al., 2020). These results were further supported by recent studies that have also reported FAM111B's role in DNA replication, metabolism, and damage repair (Roversi et al., 2021). FAM111B interacts with CAPNS1, a protease from the Calpain family, that is downregulated in FAM111B mutant patients. CAPNS1 forms a complex with CAPN1 that favours the DNA damage response and DNA metabolic pathways by interacting with FANCD2, a protein functioning in cell cycle and repair to DNA damage, and PCNA, a novel protein involved in DNA replication. Additionally, they demonstrated the diminishing effects that chromosomal breaks in lymphocytes of POIKTMP patients has on DNA replication and damage repair pathways (Roversi et al., 2021).

Additionally, our study shows that the differentially expressed protein identified between groups *FAM111B-EGFP* and Y621D-EGFP mutant plasmids, G3V3W4 (gene name, Proteasome subunit alpha type-3 (PSMA3)) (Figure 3.10 B2). Its molecular function includes endopeptidase activity, protease activity, and hydrolase activity. It has been reported that *FAM111B*, along with *FAM111A*, protects the replication forks from stalling by removing DNA-protein crosslinks (DPC's), which are complexes that performs as a hindrance preventing optimal DNA replication. It achieves this by instructing polymerase activity causing a decrease in cell apoptosis (Kojima et al., 2020).

It is hypothesized that protein G3V3W4 may rapidly cleave the mutated *FAM111B* protein, therefore, preventing the removing of harmful DPC'S causing an increase in replication stalling and increased apoptosis. Although we are not yet able to prove it with our data, the Y621D mutation could lead to the rapid cleavage of *FAM111B* which might be important in the removal of DPC's and therefore affecting genome instability and cellular fitness.

4.6 Validation of the Y621D mutation by PCR-RFLP

We described for the first time the use of PCR-RFLP in genotyping reported mutations of the human *FAM111B* gene, which has previously been obtained by Sanger sequencing (Cenza 2021, submitted). To date, there are eleven *FAM111B* mutations reported (F. Chen et al., 2019; Goussot et al., 2017; Kazlouskaya et al., 2018; Mercier et al., 2015; Panjawatanan et al., 2019; Sato et al., 2020; Seo et al., 2016; Takeichi et al., 2017). Four of these cases are from South African patients who possess the Y621D mutation. Thus, in this study, we utilized the PCR-RFLP method for screening for the presence of the POIKTMP-associated *FAM111B* mutations and demonstrated this method's use in screening and genotyping the heterozygous NM_198947.4: c.1861T>G (p. Tyr621Asp) *FAM111B* mutation of the South Africa family affected by the POIKTMP disease (Figure 3.11). The inaccessibility of next-generation sequencing (NGS) technologies in developing countries could explain the very few reported cases in these countries. PCR-RFLP is an inexpensive, simple, and fast method to genotype genetic mutations. It will be a valuable tool for the large-scale validation or screening of families or patients suspected of these disease-causing mutations.

4.7 *FAM111B* expression in patient cells

It would be interesting to determine the proliferation rate, apoptosis, migration, and invasion between POIKTMP patients and healthy familial fibroblasts; however, this was not possible due to insufficient patient fibroblasts. Protein and RNA quantification was performed to determine the difference in *FAM111B* expression in HFP-POIKTMP patients versus healthy familial control. The results showed a significant decline in patient *FAM111B* expression (Figure 3.12). This data suggests that the presence of the mutation has a profound effect on the level of expression and consequently the function of the gene.

4.8 Conclusion

Investigating the overexpression of FAM111B in cancerous cells, HT1080, is important to understand the pathogenesis of FAM111B in relation to cancer and the fibrotic diseases, POIKTMP. Our in silico and in vitro findings show that FAM111B is upregulated in cancerous

cell lines compared to normal cell lines. Expression studies in patient-derived fibroblasts showed a decreased expression of *FAM111B.*

To functionalize the *FAM111B* gene and understand its mutant role in the development of POIKTMP, we investigated the effects of *FAM111B* knockdown and overexpression in important cellular functions, namely, proliferation, apoptosis, migration, and invasion. *FAM111B* deprived cells, and Y621D mutant cells showed a reduction in proliferation rate, while cells with overexpressed FAM111B showed no effect on proliferation. Cellular apoptosis increased in cells deprived of FAM111B. However, the apoptotic cells decreased in overexpressed *FAM111B* and Y621D mutant cells. Migration assay results revealed a decline in cellular migration in *FAM111B* knocked down cells and increased overexpressed *FAM111B* and Y621D mutant cells. An investigation into the invasive role of *FAM111B* was inconclusive as no significant change in invasive capabilities were observed.

Shotgun proteomics and mass spectrometry studies identified two proteins HSP7C and G3V3W4. Heat shock protein, HSP7C, may interact with protein BAG3 and BCL2, resulting in the aggregation of misfolded proteins in the ER and consequently impairing normal cellular function, leading to cell death (Haeri & Knox, 2012; Sun et al., 2019). This observation highlights the potential pathways of *FAM111B*'s role in apoptosis, leading to fibrosis. It is hypothesised that protein G3V3W4, identified in Y621D mutant cells, plays a role in the proteolytic cleavage of FAM111B, therefore causing replication stalling as DPC's aggregate. The effects of both protein HSP7C and G3V3W4 could lead to reduced cellular fitness.

In conclusion, this study revealed that the *FAM111B* gene plays a role in cellular proliferation, apoptosis, migration, and proteolytic activity. Future work will focus on further characterising the *FAM111B* gene and protein, specifically in POIKTMP patient-derived cells, using advanced functional genomics methods including mass spectrometry, RNA and DNA sequencing. Furthermore, cell-based functional assay, as those described in this study, will be performed on POIKTMP patient-derived cells to further validate the results. Lastly, a study into the localization of synchronized POIKTMP patient-derived cells will be done to obtain an accurate representation of the expression of FAM111B during the different stages of the cell-cycle. Protein will be

extracted into compartments (cytoplasm, nucleus, membrane, and cytoskeletal proteins) followed by Western blots to confirms the expression of FAM111B in the different cellular compartments.

References

Adeola, H. (2016). *Novel urinary and serological markers of prostate cancer using proteomics techniques: an important tool for early cancer diagnosis and treatment monitoring*
. (PhD). University of Cape Town, University of Cape Town. Retrieved from https://open.uct.ac.za/bitstream/handle/11427/20955/thesis_hsf_2016_adeola_henry_ademola..pdf?sequence=1

Albadri, S., Del Bene, F., & Revenu, C. (2017). Genome editing using CRISPR/Cas9-based knock-in approaches in zebrafish. *Methods (San Diego, Calif.), 121-122*, 77-85. doi:10.1016/j.ymeth.2017.03.005

Attieh, Y., Clark, A. G., Grass, C., Richon, S., Pocard, M., Mariani, P., . . . Vignjevic, D. M. (2017). Cancer-associated fibroblasts lead tumor invasion through integrin-β3-dependent fibronectin assembly. *The Journal of cell biology, 216*(11), 3509-3520. doi:10.1083/jcb.201702033

Aviner, R., Shenoy, A., Elroy-Stein, O., & Geiger, T. (2015). Uncovering Hidden Layers of Cell Cycle Regulation through Integrative Multi-omic Analysis. *PLoS Genet, 11*(10), e1005554. doi:10.1371/journal.pgen.1005554

Banan, M. (2020). Recent advances in CRISPR/Cas9-mediated knock-ins in mammalian cells. *Journal of Biotechnology, 308*, 1-9. doi:https://doi.org/10.1016/j.jbiotec.2019.11.010

Barratt, S. L., Creamer, A., Hayton, C., & Chaudhuri, N. (2018). Idiopathic Pulmonary Fibrosis (IPF): An Overview. *Journal of clinical medicine, 7*(8), 201. doi:10.3390/jcm7080201

Bayat, A. (2002). Science, medicine, and the future: Bioinformatics. *BMJ (Clinical research ed.), 324*(7344), 1018-1022. doi:10.1136/bmj.324.7344.1018

BioLabs. Luna® Universal qPCR Master Mix Protocol Retrieved from https://international.neb.com/protocols/2016/11/08/luna-universal-qpcr-master-mix-protocol-m3003

BioLabs. Probe-based qPCR & RT-qPCR. Retrieved from https://international.neb.com/applications/dna-amplification-pcr-and-qpcr/qpcr-and-rt-qpcr/probe-based-qpcr-and-rt-qpcr

Biosciences, B. (2011). Detection of Apoptosis Using the BD Annexin V FITC Assay on the BD FACSVerse™ System. In.

Boettcher, M., & McManus, M. T. (2015). Choosing the Right Tool for the Job: RNAi, TALEN, or CRISPR. *Molecular cell, 58*(4), 575-585. doi:10.1016/j.molcel.2015.04.028

Brion, C., Lutz, S. M., & Albert, F. W. (2020). Simultaneous quantification of mRNA and protein in single cells reveals post-transcriptional effects of genetic variation. *eLife, 9*, e60645. doi:10.7554/eLife.60645

Bunnik, E. M., & Le Roch, K. G. (2013). An Introduction to Functional Genomics and Systems Biology. *Advances in wound care, 2*(9), 490-498. doi:10.1089/wound.2012.0379

Campeau, E., & Gobeil, S. (2011). RNA interference in mammals: Behind the screen. *Briefings in Functional Genomics, 10*, 215-226. doi:10.1093/bfgp/elr018

Chandler, C., Liu, T., Buckanovich, R., & Coffman, L. G. (2019). The double edge sword of fibrosis in cancer. *Translational Research, 209*, 55-67. doi:https://doi.org/10.1016/j.trsl.2019.02.006

Chasseuil, E., McGrath, J. A., Seo, A., Balguerie, X., Bodak, N., Chasseuil, H., . . . Barbarot, S. (2019). Dermatological manifestations of hereditary fibrosing poikiloderma with tendon

contractures, myopathy and pulmonary fibrosis (POIKTMP): a case series of 28 patients. *British Journal of Dermatology, 0*(0). doi:10.1111/bjd.17996

Chen, C., Tang, Y., Qu, W.-D., Han, X., Zuo, J.-B., Cai, Q.-Y., . . . Ke, X.-X. (2021). Evaluation of clinical value and potential mechanism of MTFR2 in lung adenocarcinoma via bioinformatics. *BMC Cancer, 21*(1), 619. doi:10.1186/s12885-021-08378-3

Chen, F., Zheng, L., Li, Y., Li, H., Yao, Z., & Li, M. (2019). Mutation in FAM111B Causes Hereditary Fibrosing Poikiloderma with Tendon Contracture, Myopathy and Pulmonary Fibrosis. *Acta dermato-venereologica, 99*(7-8), 695-696.

Chu, Y., & Corey, D. R. (2012). RNA sequencing: platform selection, experimental design, and data interpretation. *Nucleic acid therapeutics, 22*(4), 271-274. doi:10.1089/nat.2012.0367

Cirri, P., & Chiarugi, P. (2011). Cancer associated fibroblasts: the dark side of the coin. *American journal of cancer research, 1*(4), 482-497. Retrieved from https://pubmed.ncbi.nlm.nih.gov/21984967

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3186047/

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3186047/pdf/ajcr0001-0482.pdf

Coussens, L. M., & Werb, Z. (2002). Inflammation and cancer. *Nature, 420*(6917), 860-867. doi:10.1038/nature01322

Coyle, K. M., Boudreau, J. E., & Marcato, P. (2017). Genetic Mutations and Epigenetic Modifications: Driving Cancer and Informing Precision Medicine. *Biomed Res Int, 2017*, 9620870. doi:10.1155/2017/9620870

Demidova-Rice, T. N., Hamblin, M. R., & Herman, I. M. (2012). Acute and impaired wound healing: pathophysiology and current methods for drug delivery, part 1: normal and chronic wounds: biology, causes, and approaches to care. *Advances in skin & wound care, 25*(7), 304-314. doi:10.1097/01.ASW.0000416006.55218.d0

Dokic, Y., Albahrani, Y., Phung, T., Patel, K., de Guzman, M., Hertel, P., & Hunt, R. (2020). Hereditary fibrosing poikiloderma with tendon contractures, myopathy, and pulmonary fibrosis: Hepatic disease in a child with a novel pathogenic variant of FAM111B. *JAAD case reports, 6*(12), 1217-1220. doi:10.1016/j.jdcr.2020.09.025

Dominska, M., & Dykxhoorn, D. M. (2010). Breaking down the barriers: siRNA delivery and endosome escape. *Journal of Cell Science, 123*(8), 1183. doi:10.1242/jcs.066399

Donaldson, J. G. (1998). Immunofluorescence staining. *Current protocols in cell biology*(1), 4.3. 1-4.3. 6.

El-Serag, H. B. (2012). Epidemiology of Viral Hepatitis and Hepatocellular Carcinoma. *Gastroenterology, 142*(6), 1264-1273.e1261. doi:10.1053/j.gastro.2011.12.061

Feghali-Bostwick, C., & Wright, T. (1997). Cytokines acute and chronic inflammation. *Frontiers in bioscience : a journal and virtual library, 2*, d12-26. doi:10.2741/A171

Foster, K. A., Oster, C. G., Mayer, M. M., Avery, M. L., & Audus, K. L. (1998). Characterization of the A549 cell line as a type II pulmonary epithelial cell model for drug metabolism. *Exp Cell Res, 243*(2), 359-366. doi:10.1006/excr.1998.4172

Gilbert, L. A., Horlbeck, M. A., Adamson, B., Villalta, J. E., Chen, Y., Whitehead, E. H., . . . Weissman, J. S. (2014). Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation. *Cell, 159*(3), 647-661. doi:10.1016/j.cell.2014.09.029

Gosain, A., & DiPietro, L. A. (2004). Aging and wound healing. *World J Surg, 28*(3), 321-326. doi:10.1007/s00268-003-7397-6

Goussot, R., Prasad, M., Stoetzel, C., Lenormand, C., Dollfus, H., & Lipsker, D. (2017). Expanding phenotype of hereditary fibrosing poikiloderma with tendon contractures, myopathy, and pulmonary fibrosis caused by FAM111B mutations: Report of an additional family raising the question of cancer predisposition and a short review of early-onset poikiloderma. *JAAD case reports, 3*(2), 143-150. doi:10.1016/j.jdcr.2017.01.002

Haeri, M., & Knox, B. E. (2012). Endoplasmic Reticulum Stress and Unfolded Protein Response Pathways: Potential for Treating Age-related Retinal Degeneration. *Journal of ophthalmic & vision research, 7*(1), 45-59. Retrieved from https://pubmed.ncbi.nlm.nih.gov/22737387 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3381108/

Hall, D. M. S., & Brooks, S. A. (2014). In Vitro Invasion Assay Using Matrigel™: A Reconstituted Basement Membrane Preparation. In M. Dwek, U. Schumacher, & S. A. Brooks (Eds.), *Metastasis Research Protocols* (pp. 1-11). New York, NY: Springer New York.

Han, Y., Gao, S., Muegge, K., Zhang, W., & Zhou, B. (2015). Advanced Applications of RNA Sequencing and Challenges. *Bioinform Biol Insights, 9*(Suppl 1), 29-46. doi:10.4137/bbi.S28991

Hernández-Domínguez, E., Castillo-Ortega, L., García-Esquivel, Y., Mandujano-González, V., Díaz-Godínez, G., & Alvarez Cervantes, J. (2019). Bioinformatics as a Tool for the Structural and Evolutionary Analysis of Proteins. In.

Hinz, B., Phan, S. H., Thannickal, V. J., Galli, A., Bochaton-Piallat, M.-L., & Gabbiani, G. (2007). The myofibroblast: one function, multiple origins. *The American journal of pathology, 170*(6), 1807-1816. doi:10.2353/ajpath.2007.070112

Hoffmann, S., Pentakota, S., Mund, A., Haahr, P., Coscia, F., Gallo, M., . . . Mailand, N. (2020). FAM111 protease activity undermines cellular fitness and is amplified by gain-of-function mutations in human disease. *EMBO reports, 21*(10), e50662-e50662. doi:10.15252/embr.202050662

Hood, L., & Rowen, L. (2013). The Human Genome Project: big science transforms biology and medicine. *Genome Medicine, 5*(9), 79. doi:10.1186/gm483

Huang, D. W., Sherman, B. T., & Lempicki, R. A. (2009). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic acids research, 37*(1), 1-13. doi:10.1093/nar/gkn923

Huang, Y., Jeong, J. S., Okamura, J., Sook-Kim, M., Zhu, H., Guerrero-Preston, R., & Ratovitski, E. A. (2012). Global tumor protein p53/p63 interactome: making a case for cisplatin chemoresistance. *Cell cycle (Georgetown, Tex.), 11*(12), 2367-2379. doi:10.4161/cc.20863

Iadarola, P. (2019). Special Issue: Mass Spectrometric Proteomics. *Molecules (Basel, Switzerland), 24*(6), 1133. doi:10.3390/molecules24061133

Jalali, M., Zaborowska, J., & Jalali, M. (2017). Chapter 1 - The Polymerase Chain Reaction: PCR, qPCR, and RT-PCR. In M. Jalali, F. Y. L. Saldanha, & M. Jalali (Eds.), *Basic Science Methods for Clinical Researchers* (pp. 1-18). Boston: Academic Press.

Johnson, A., & DiPietro, L. A. (2013). Apoptosis and angiogenesis: an evolving mechanism for fibrosis. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology, 27*(10), 3893-3901. doi:10.1096/fj.12-214189

Kaczkowski, B., Tanaka, Y., Kawaji, H., Sandelin, A., Andersson, R., Itoh, M., . . . Forrest, A. R. (2016). Transcriptome Analysis of Recurrently Deregulated Genes across Multiple Cancers Identifies New Pan-Cancer Biomarkers. *Cancer Res, 76*(2), 216-226. doi:10.1158/0008-5472.Can-15-0484

Karvelis, T., Gasiunas, G., Miksys, A., Barrangou, R., Horvath, P., & Siksnys, V. (2013). crRNA and tracrRNA guide Cas9-mediated DNA interference in Streptococcus thermophilus. *RNA biology, 10*(5), 841-851. doi:10.4161/rna.24203

Kawasaki, K., Nojima, S., Hijiki, S., Tahara, S., Ohshima, K., Matsui, T., . . . Morii, E. (2020). FAM111B enhances proliferation of KRAS-driven lung adenocarcinoma by degrading p16. *Cancer science, 111*(7), 2635-2646. doi:10.1111/cas.14483

Kazlouskaya, V., Feldman, E. J., Jakus, J., Heilman, E., & Glick, S. (2018). A case of hereditary fibrosing poikiloderma with tendon contractures, myopathy and pulmonary fibrosis (POIKTMP) with the emphasis on cutaneous histopathological findings. *Journal of the European Academy of Dermatology and Venereology, 32*(12), e443-e445. doi:10.1111/jdv.14968

Khumalo, N. P., Pillay, K., Beighton, P., Wainwright, H., Walker, B., Saxe, N., . . . Bateman, E. D. (2006). Poikiloderma, tendon contracture and pulmonary fibrosis: a new autosomal dominant syndrome? *Br J Dermatol, 155*(5), 1057-1061. doi:10.1111/j.1365-2133.2006.07473.x

Kojima, Y., Machida, Y., Palani, S., Caulfield, T. R., Radisky, E. S., Kaufmann, S. H., & Machida, Y. J. (2020). FAM111A protects replication forks from protein obstacles via its trypsin-like domain. *Nature Communications, 11*(1), 1318. doi:10.1038/s41467-020-15170-7

Kolker, E., Makarova, K. S., Shabalina, S., Picone, A. F., Purvine, S., Holzman, T., . . . Galperin, M. Y. (2004). Identification and functional analysis of 'hypothetical' genes expressed in Haemophilus influenzae. *Nucleic acids research, 32*(8), 2353-2361. doi:10.1093/nar/gkh555

Küry, S., Mercier, S., Shaboodien, G., Besnard, T., Barbarot, S., Khumalo, N. P., . . . Bézieau, S. (2015). CUGC for hereditary fibrosing poikiloderma with tendon contractures, myopathy, and pulmonary fibrosis (POIKTMP). *European Journal Of Human Genetics, 24*, 779. doi:10.1038/ejhg.2015.205

Lai, Q., Murgia, N., Parkkinen, I., Domanskyi, A., & Vinnikov, I. A. (2019). Chapter 8 - Roles of microRNAs in Parkinson's and other neurodegenerative diseases. In B. Mallick (Ed.), *AGO-Driven Non-Coding RNAs* (pp. 209-232): Academic Press.

Larizza, L., Roversi, G., & Volpi, L. (2010). Rothmund-Thomson syndrome. *Orphanet journal of rare diseases, 5*(1), 2. doi:10.1186/1750-1172-5-2

Lewis, M. (2016). Direct and Indirect Western blotting. Retrieved from https://www.jacksonimmuno.com/secondary-antibody-resource/immuno-techniques/directandindirectwesternblotting/

Ley, B., Collard, H. R., & King, T. E. (2011). Clinical Course and Prediction of Survival in Idiopathic Pulmonary Fibrosis. *American Journal of Respiratory and Critical Care Medicine, 183*(4), 431-440. doi:10.1164/rccm.201006-0894CI

Liang, C.-C., Park, A. Y., & Guan, J.-L. (2007). In vitro scratch assay: a convenient and inexpensive method for analysis of cell migration in vitro. *Nature Protocols, 2*, 329. doi:10.1038/nprot.2007.30

Mahajan, S., Sunkwad, A., Darkase, B., & Khopkar, U. (2020). Poikiloderma with novel gene mutation. *Indian Journal of Paediatric Dermatology, 21*(1), 63-65. doi:10.4103/ijpd.IJPD_98_19

Maher, T. M., Bendstrup, E., Dron, L., Langley, J., Smith, G., Khalid, J. M., . . . Kreuter, M. (2021). Global incidence and prevalence of idiopathic pulmonary fibrosis. *Respiratory Research, 22*(1), 197. doi:10.1186/s12931-021-01791-z

Mahmood, T., & Yang, P.-C. (2012). Western blot: technique, theory, and trouble shooting. *North American journal of medical sciences, 4*(9), 429-434. doi:10.4103/1947-2714.100998

Mangul, S., Al Seesi, S., Mandoiu, I., Caciula, A., Zelikovsky, A., & Brinza, D. (2013). *Transcriptome assembly and quantification from Ion Torrent RNA-Seq data* (Vol. 15 Suppl 5).

Martinez, M. M., Reif, R. D., & Pappas, D. (2010). Detection of apoptosis: A review of conventional and novel techniques. *Analytical Methods, 2*(8), 996-1004. doi:10.1039/C0AY00247J

Martins-Gomes, C., & Silva, A. M. (2018). Western Blot Methodologies for Analysis of In Vitro Protein Expression Induced by Teratogenic Agents. In L. Félix (Ed.), *Teratogenicity Testing: Methods and Protocols* (pp. 191-203). New York, NY: Springer New York.

Mercier, S., Küry, S., Salort-Campana, E., Magot, A., Agbim, U., Besnard, T., . . . Bézieau, S. (2015). Expanding the clinical spectrum of hereditary fibrosing poikiloderma with tendon contractures, myopathy and pulmonary fibrosis due to FAM111B mutations. *Orphanet journal of rare diseases, 10*, 135-135. doi:10.1186/s13023-015-0352-4

Mercier, S., Küry, S., Shaboodien, G., Houniet, D. T., Khumalo, N. P., Bou-Hanna, C., . . . Mayosi, B. M. (2013). Mutations in FAM111B cause hereditary fibrosing poikiloderma with tendon contracture, myopathy, and pulmonary fibrosis. *American journal of human genetics, 93*(6), 1100-1107. doi:10.1016/j.ajhg.2013.10.013

Moser, B., Hochreiter, B., Herbst, R., & Schmid, J. A. (2017). Fluorescence colocalization microscopy analysis can be improved by combining object-recognition with pixel-intensity-correlation. *Biotechnol J, 12*(1). doi:10.1002/biot.201600332

Neary, R., Watson, C. J., & Baugh, J. A. (2015). Epigenetics and the overhealing wound: the role of DNA methylation in fibrosis. *Fibrogenesis & tissue repair, 8*, 18-18. doi:10.1186/s13069-015-0035-8

Nisbet, R., Elder, J., & Miner, G. (2009). Chapter 15 - Bioinformatics. In R. Nisbet, J. Elder, & G. Miner (Eds.), *Handbook of Statistical Analysis and Data Mining Applications* (pp. 321-334). Boston: Academic Press.

Oehrle, B., Burgstaller, G., Irmler, M., Dehmel, S., Grün, J., Hwang, T., . . . Eickelberg, O. (2015). Validated prediction of pro-invasive growth factors using a transcriptome-wide invasion signature derived from a complex 3D invasion assay. *Scientific Reports, 5*(1), 12673. doi:10.1038/srep12673

Olson, A. L., Swigris, J. J., Raghu, G., & Brown, K. K. (2009). Seasonal Variation: Mortality From Pulmonary Fibrosis Is Greatest in the Winter. *Chest, 136*(1), 16-22. doi:https://doi.org/10.1378/chest.08-0703

Ozsolak, F., & Milos, P. M. (2011). RNA sequencing: advances, challenges and opportunities. *Nature Reviews Genetics, 12*(2), 87-98. doi:10.1038/nrg2934

Panjawatanan, P., Ryabets-Lienhard, A., Bakhach, M., & Pitukcheewanont, P. (2019). MON-512 A De Novo Frameshift Mutation of FAM111B Gene Resulting in Progressive Osseous Heteroplasia in an African American Boy: First Case Report. *Journal of the Endocrine Society, 3*(Suppl 1), MON-512. doi:10.1210/js.2019-MON-512

Peterson, D. A. (2010). Confocal Microscopy. In K. Kompoliti & L. V. Metman (Eds.), *Encyclopedia of Movement Disorders* (pp. 250-252). Oxford: Academic Press.

Powell, D. W., Mifflin, R. C., Valentich, J. D., Crowe, S. E., Saada, J. I., & West, A. B. (1999). Myofibroblasts. I. Paracrine cells important in health and disease. *American Journal of Physiology-Cell Physiology, 277*(1), C1-C19. doi:10.1152/ajpcell.1999.277.1.C1

Prabhune, M. (2021). RNAi vs. CRISPR: Guide to Selecting the Best Gene Silencing Method . Retrieved from https://www.synthego.com/blog/rnai-vs-crispr-guide

Prelich, G. (2012). Gene overexpression: uses, mechanisms, and interpretation. *Genetics, 190*(3), 841-854. doi:10.1534/genetics.111.136911

Rieger, A. M., Nelson, K. L., Konowalchuk, J. D., & Barreda, D. R. (2011). Modified annexin V/propidium iodide apoptosis assay for accurate assessment of cell death. *Journal of visualized experiments : JoVE*(50), 2597. doi:10.3791/2597

Rios-Szwed, D., Garcia-Wilson, E., Sanchez-Pulido, L., Alvarez, V., Jiang, H., Bandau, S., . . . Alabert, C. (2020). FAM111A regulates replication origin activation and cell fitness. In: bioRxiv.

Roversi, G., Colombo, E. A., Magnani, I., Gervasini, C., Maggiore, G., Paradisi, M., & Larizza, L. (2021). Spontaneous chromosomal instability in peripheral blood lymphocytes from two molecularly confirmed Italian patients with Hereditary Fibrosis Poikiloderma: insights into cancer predisposition. *Genet Mol Biol, 44*(3), e20200332. doi:10.1590/1678-4685-gmb-2020-0332

Rybinski, B., Franco-Barraza, J., & Cukierman, E. (2014). The wound healing, chronic fibrosis, and cancer progression triad. *Physiological genomics, 46*(7), 223-244. doi:10.1152/physiolgenomics.00158.2013

Sato, M. T., Suzuki, M., Ikari, T., Nakamura, T., Takahashi, K., Sasaki, M., . . . Otsuka, N. (2020). The First Case of Hereditary Fibrosing Poikiloderma with Tendon Contractures, Myopathy, and Pulmonary Fibrosis (POIKTMP) Who Showed Efficacy of Corticosteroids for Pulmonary Involvement. A1463-A1463. doi:10.1164/ajrccm-conference.2020.201.1_MeetingAbstracts.A1463

Schütze, N. (2004). siRNA technology. *Molecular and Cellular Endocrinology, 213*(2), 115-119. doi:https://doi.org/10.1016/j.mce.2003.10.078

Seo, A., Walsh, T., Lee, M. K., Ho, P. A., Hsu, E. K., Sidbury, R., . . . Shimamura, A. (2016). FAM111B Mutation Is Associated With Inherited Exocrine Pancreatic Dysfunction. *Pancreas, 45*(6), 858-862. doi:10.1097/MPA.0000000000000529

Shaw, L. M. (2005). Tumor Cell Invasion Assays. In J.-L. Guan (Ed.), *Cell Migration: Developmental Methods and Protocols* (pp. 97-105). Totowa, NJ: Humana Press.

Song, J.-Y., Niño, M., Nogoy, F. M., Jung, Y.-J., Kang, K.-K., & Cho, Y.-G. (2017). CRISPR/CAS9 as a powerful tool for crop improvement. *Journal of Plant Biotechnology, 44*, 107-114. doi:10.5010/JPB.2017.44.2.107

Subhash, S., & Kanduri, C. (2016). GeneSCF: a real-time based functional enrichment tool with support for multiple organisms. *BMC Bioinformatics, 17*(1), 365. doi:10.1186/s12859-016-1250-z

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., . . . Mesirov, J. P. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America, 102*(43), 15545-15550. doi:10.1073/pnas.0506580102

Suganuma, H., Sato, A., Tamura, R., & Chida, K. (1995). Enhanced migration of fibroblasts derived from lungs with fibrotic lesions. *Thorax, 50*(9), 984-989. doi:10.1136/thx.50.9.984

Sun, H., Liu, K., Huang, J., Sun, Q., Shao, C., Luo, J., . . . Ren, B. (2019). FAM111B, a direct target of p53, promotes the malignant process of lung adenocarcinoma. *OncoTargets and therapy, 12*, 2829-2842. doi:10.2147/OTT.S190934

Takeichi, T., Nanda, A., Yang, H. S., Hsu, C. K., Lee, J. Y., Al-Ajmi, H., . . . McGrath, J. A. (2017). Syndromic inherited poikiloderma due to a de novo mutation in FAM111B. *Br J Dermatol, 176*(2), 534-536. doi:10.1111/bjd.14845

Tipney, H., & Hunter, L. (2010). An introduction to effective use of enrichment analysis software. *Human Genomics, 4*(3), 202. doi:10.1186/1479-7364-4-3-202

Tracy, L. E., Minasian, R. A., & Caterson, E. J. (2016). Extracellular Matrix and Dermal Fibroblast Function in the Healing Wound. *Advances in wound care, 5*(3), 119-136. doi:10.1089/wound.2014.0561

Tsai, T.-H., Song, E., Zhu, R., Di Poto, C., Wang, M., Luo, Y., . . . Ressom, H. W. (2015). LC-MS/MS-based serum proteomics for identification of candidate biomarkers for hepatocellular carcinoma. *Proteomics, 15*(13), 2369-2381. doi:10.1002/pmic.201400364

Türker Şener, L., Albeniz, G., Dinç, B., & Albeniz, I. (2017). iCELLigence real-time cell analysis system for examining the cytotoxicity of drugs to cancer cell lines. *Experimental and therapeutic medicine, 14*(3), 1866-1870. doi:10.3892/etm.2017.4781

Uhal, B. D. (2008). The role of apoptosis in pulmonary fibrosis. *European Respiratory Review, 17*(109), 138. doi:10.1183/09059180.00010906

Uversky, V. N. (2013). Posttranslational Modification. In S. Maloy & K. Hughes (Eds.), *Brenner's Encyclopedia of Genetics (Second Edition)* (pp. 425-430). San Diego: Academic Press.

Vogel, C., & Marcotte, E. M. (2012). Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nat Rev Genet, 13*(4), 227-232. doi:10.1038/nrg3185

Volgin, D. V. (2014). Chapter 17 - Gene Expression: Analysis and Quantitation. In A. S. Verma & A. Singh (Eds.), *Animal Biotechnology* (pp. 307-325). San Diego: Academic Press.

Voorhoeve, P. M., & Agami, R. (2003). Knockdown stands up. *Trends Biotechnol, 21*(1), 2-4.

Zhang, Z., Zhang, J., Chen, F., Zheng, L., Li, H., Liu, M., . . . Yao, Z. (2019). Family of hereditary fibrosing poikiloderma with tendon contractures, myopathy and pulmonary fibrosis caused by a novel FAM111B mutation. *The Journal of Dermatology, 46*(11), 1014-1018. doi:10.1111/1346-8138.15045