A genomic and histopathological study of the mutated human *FAM111B* gene related to POIKTMP disease

by

Nadine Tambwe

Submitted for the degree of

Master of Science

Division of Medical Biochemistry

Department of Integrative Biomedical Sciences

Faculty of Health Sciences

UNIVERSITY OF CAPE TOWN

Supervisor: Dr Afolake Arowolo

Co-supervisior: Prof. N.P Khumalo

DECLARATION

1

I, *Nadine Tambwe*, 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.

Signed by candidate

Date: 03/02/2023

Table of Contents

Acknowledgements

The projected was funded and completed at the Hair and Research Lab, Groote Schuur Hospital.

To God be the Glory, All the glory, honour and praise belong to the Lord. (Proverbs 3:5-6)

I would like to extend my gratitude and gratefulness to my supervisor Dr Afolake Arowolo, for her guidance, her assistance and mentorship throughout my Honours and master' journey. I have learned a lot under her tutelage. Thank you for helping me lay the foundation Doc.

To Prof. Khumalo who has taken me into the fold of the Hair and Skin Research Lab, I truly appreciate your wonderful insights and contribution towards this project that is so near and dear to her.

The project would not have been possible without Prof. Komala Pillay, who collaborated gratefully by providing us with samples from the patient.

My acknowledgement extends to the National Research Foundation for funding my Masters's degree at The University of Cape Town.

I also extend my gratitude towards the Department of Medicine Masters Fellowship Award for funding me with their prestigious award.

To my HSR family, the ladies who have made the journey fun, thank you for your insights and friendship during the stage.

Lastly, a big thank you to my family, my TSG family and my extended family that developed along the way. Thank you for your support, and assistance through these years. Mostly thank you for understanding and believing in me.

To my father, to whom I dedicate this thesis, thank you for being the best dad a girl could ever ask for.

List of Tables

List of Figures

List of Abbreviations

bFGF: basic fibroblast growth factor

BSA: Bovine Serum Albumin

CANP: cancer-associated nucleoprotein

CCN2: Connective tissue growth factor

cDNA: complementary deoxyribonucleic acid

COL3A1: Collagen alpha-type 111 alpha 1

CT: Cycle threshold

DAB: diaminobenzidine (DAB)

DAPI: 4′,6-diamidino-2-phenylindole

dH_20: distilled water

DNA: deoxyribonucleic acid

ECM: extracellular matrix

FAM111B: FAM111 Trypsin Like Peptidase B (gene)

FFPE: Formalin-fixed paraffin-embedded

GAPDH: Glyceraldehyde 3-phosphate dehydrogenase

GO: Gene Ontology

GSEA: Gene set enrichmentanalysis

GWAS: Genome-wide associated study

HFP: Hereditary Fibrosing Poikiloderma

HGP: Human Genome Project

HIER: Heat-Induced Epitope Retrieval

HRP: Horseradish Peroxidase

HSP: Hereditary Sclerosing poikiloderma

HSR: Hair and Skin Research

IL-1: Interleukin-1

IPF: idiopathic pulmonary fibrosis

ITGB-1: Integrin beta-1

kDa: Kilodalton

KEGG: Kyoto Encyclopaedia of Genes and Genomes

LUAD: Lung Adenocarcinoma

mM: milli molar

MMP3/13: Matrix metalloproteinase

MOPPD: Mutations outside the putative protease domain

mRNA: messenger Ribonucleic Acid

MWPPD: Mutations within the putative protease domain

NGS: Next-generation sequencing

OMIM: Online mendeline Inheritance in man

PBS: Phosphate buffer saline

PBST: phosphate buffer saline tween

PCR: Polymerase Chain Reaction Polymorphism

PDGF: platelet-derived growth factor

POIKTMP: Poikiloderma with Tendon contractures, Myopathy and Pulmonary fibrosis

qPCR: quantitative polymerase chain reaction

RFLP: Restriction fragment length polymorphism(RFLP)

RNA: ribonucleic acid

RNA-seq: RNA Sequencing

RT-qPCR: Reverse transcription quantitative polymerase chain reaction

RT: reverse transcription

RT: room temperature

RTS: Rothmund-Thomson syndrome

SNP: single nucleotide polymorphism

TGF-β:Transforming growth factor beta

TFN-α: tumour necrosis factor-α

THBS-2: Thrombospondin 2

UCT: University of Cape town

USA: United States of America

WES: Whole-exome sequencing

WGS: Whole genome sequencing

α-SMA: smooth muscle alpha-actin

µl: microliter

µM: micromolar

Abstract

Fibrosis is a pathological feature of many chronic inflammatory diseases, eventually leading to organ failure and death. POIKTMP is a rare, multi-organ fibrosing disease which is associated with mutations of the human *FAM111B* gene. *FAM111B* gene codes for a protein whose function is not well characterized. Therefore, elucidating the mechanism of *FAM111B* or its mutations in POIKTMP is beneficial to understanding the complexities surrounding this multisystemic fibrosing disease. The study sought to understand the pathogenesis of fibrosis, its role in POIKTMP and its causative gene mutation: *FAM111B* Y621D.

First, Sanger sequencing was used to confirm the presence of the FAM111B Y621D mutation using DNA isolated and amplified from post-mortem FFPE tissues of a POIKTMP patient first described with the disease in South African Following that, qRT-PCR was employed to assess gene expression changes between the patient and the familial control. The RT^2 Profiler Human fibrosis PCR Array was then used to associate POIKTMP and 84 known fibrotic markers to propose a possible fibrotic pathway associated with POIKTMP disease using mRNA from the lung and skin POIKTMP patient tissues. Gene-set enrichment analysis (GSEA) using Enrichr, a computational GSEA tool, was used to predict enrichment analysis between the identified upregulated fibrosis markers and the *FAM111B* gene. Finally, Immunohistochemistry was used to identify cellular and sub-cellular protein distribution of *FAM111B* and other fibroproliferative markers of interest to annotate pathological changes.

The results from this study validated the *FAM111B Y621D* mutation in the affected tissues. Next, FAM111B mRNA was shown to be downregulated in the lungs and skeletal muscle tissues of the POIKTMP patient. The human fibrosis PCR array experiments identified eight upregulated fibrotic markers: MMP3, MMP13, PDGFA, ITGB-1, THBS-2, COL3A1, TGFβ-3, and CCN2 in the patient lungs and skin tissues, which were validated by qRT-PCR. Furthermore, these genes with *FAM111B* form a gene-list that was used in interrogating various gene-set libraries in the gene-set enrichment analysis. *FAM111B* was enriched in some gene-set libraries within the Diseases/Drugs and Cell type categories. The GSEA terms enriched within these libraries are the pathways associated with SARS-COVID-19 perturbations and cell/tissue types related to the small intestine, breast, oesophagus, thyroid, smooth muscle and stromal cells of some of these organs. Lastly, immunohistochemistry results corroborated this study's mRNA expression analysis by showing that *FAM111B* was more highly expressed in the skin than in the lung patient. TGF-β1 and Ki-67 markers were

assessed from protein expression, which resulted in higher expression in the POIKTMP patient skin tissue than in the lungs.

Altogether, our data suggest that *FAM111B* and mutations in this gene play a pivotal role in POIKTMP and other fibrosing organ diseases, representing a potential disease biomarker and possible therapeutic target in POIKTMP and other fibrotic disorders.

CHAPTER 1

1. Introduction and literature review

1.1 Introduction

Fibrosis is the excessive formation of fibrous connective extracellular matrices in response to tissue injury or damage. It is a pathological feature of many chronic inflammatory diseases, eventually leading to organ failure and death.[1] Hereditary fibrosing Poikiloderma with tendon contractures, myopathy, and pulmonary fibrosis (POIKTMP, Online Mendelian Inheritance in Man [OMIM] #615704) is a rare, autosomal dominant multi-organ fibrosing disease.[2] Other clinical features of this disorder include hypotrichosis and hypohidrosis, which typically appear in early childhood.[3-5] Following that severe, phenotypes become much more prevalent in the second decade of life.[6] In older patients, pulmonary irregularities, such as progressive dyspnea with obstructive lung fibrosis, are more common.

Whole exosome analyses identified the *Family with Sequence Similarity 111-member B* (*FAM111B*) gene as the causal gene of POIKTMP.[3,5] There are thirty-five cases of POIKTMP documented globally and nineteen *FAM111B* mutations reported, adding to the growing evidence connecting *FAM111B* to the disease.[6-8] However, the *FAM111B* gene codes for a protein whose function is not well characterized.

The ability to describe the roles and interactions of genes and proteins using genome-wide techniques has been made possible by more recent applications of functional genomics and molecular biology tools. These techniques allow us to track changes in gene expression at the genome, transcriptome, and proteome levels. Hence, understanding the mechanism of *FAM111B* and its role in POIKTMP is beneficial to elucidate the complexities surrounding the multisystemic fibrosing disease.

This chapter focuses on the pathogenesis of fibrosis and its role in POIKTMP. It discusses the POIKTMP disease phenotype in the context of mutations of the uncharacterized human *FAM111B* gene and the approach toward understanding its functional mechanism. Furthermore, this chapter describes the genomic approaches that can be harnessed to elucidate the function of *FAM111B.* Finally, the chapter highlights the aims and objectives of the study.

1.2 Literature review

1.2.1 Fibrosis

Fibrosis refers to the increased deposition and overproduction of the extracellular matrix (ECM) components, such as collagen, elastin, and fibronectin, which arise due to chronic injury and irregular tissue repair.[7,9,10,11] Although the mechanism of fibrotic aetiology varies and is unknown in some cases,[12] fibrosis is the common endpoint for many tissues and organ dysfunction.[13],[14] So, it is essential to accurately diagnose the fibrosis stage to monitor disease progression and prognosis.[15] Additionally, few treatment strategies are available to address fibrosis.[12] Since chronic fibroproliferative diseases affect many organs and it creates a burden on healthcare resources and the economy, specifically in developed countries, where fibrotic-associated diseases account for approximately 45% of deaths.[1,12,16]

1.2.2 Pathogenesis of Fibrosis

Studying the pathogenesis of fibrotic diseases has advanced significantly, but it remains complex due to disease variability and, most importantly, the profibrotic mediators involved. Chronic progressive fibrosis is a common pathway in diseases like idiopathic pulmonary fibrosis (IPF), where reduced lung function is often associated with scar tissue build-up that causes respiratory failure.[11,17] Inflammatory cells, such as macrophages, present in the injured lung, are the most abundant effector cells as they produce cytokines that regulate fibroblast growth.[17] These cytokines include interleukin 1 (IL-1) and tumour necrosis factor-α (TFN-α).[17] Among these cytokines, some regulate fibroblast function and collagen synthesis, while some recruit and induce cytokine production in other inflammatory cells. Producing pro-inflammatory and pro-fibrotic factors or cytokines, along with the release of growth factors and enzymes, indirectly stimulate myofibroblast activation, ultimately leading to the overproduction of fibroblast. [13, 18]

Transforming growth factor beta (TGF-β) isoforms are an important profibrotic stimulus for fibroblasts; the TGF-β superfamily includes TGF-β1-3, with a shared structure and similar signalling pathways that can overlap in biological effects.[19] The TGF-β superfamily comprises various polypeptides involved in cellular activities such as proliferation, differentiation, migration, adhesion, ECM synthesis, and cell death.[20] There are currently several approaches to combating tissue progressive fibrotic diseases, one of which is by regulating the myofibroblast and TGF-β pathway. TGF-β signalling pathway controls cellular processes and the maintenance of homeostasis.[20] This pathway is a tissue repair regulator

turned on by activated macrophages, lymphocytes, and fibroblasts and is an essential source at injury sites.

Activated myofibroblasts are present in fibrotic disease cells, making differentiation events important in progressive fibrosis and ECM accumulation.[1,21,22] TGF-β signalling occurs primarily through the transmembrane receptor serine/threonine kinases, which activate cytoplasmic SMAD protein complexes.[21,23,24] These complexes function downstream of the TGF-β signal transduction mediators and then translocate into the nucleus to regulate the expression of targeted genes associated with fibrosis (**Figure 1-1**). This complex process has resulted in the discovery of some molecular mechanisms of TGF-β gene activation.[21] Moreover, transgenic mouse studies have previously found a link between fibrotic disease and overexpression of TGF-β1, which frequently promotes multi-organ fibrosis, implying a connection between TGF-β expression and fibrotic disease.[2]

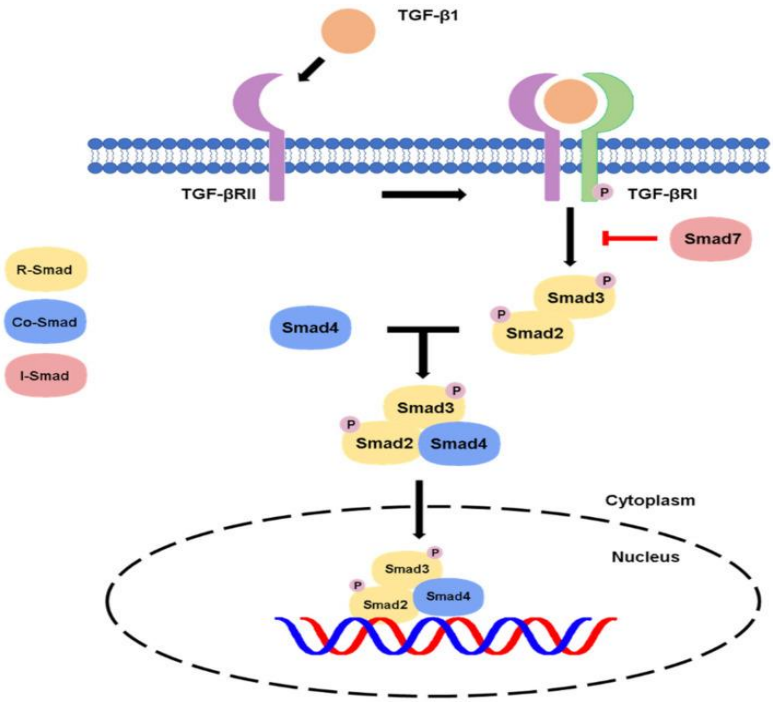


Figure 1-1:Transforming growth factor-β (TGF-β) canonical signalling pathway. TGF-β activation complex activates SMAD proteins via phosphorylation and translocate into the nucleus. This translocation induces gene transcription and promotes cell growth and survival.[24]

1.2.3 Hereditary Poikiloderma Disorders

Inflammatory or neoplastic-acquired illnesses can also bring about poikiloderma, a significant severe genetic disorder caused primarily by errors in DNA repair and characterized by skin photosensitivity.[25] Poikiloderma and other severe multi-organ

manifestations are also indicated in similar inherited disorders.[25] Genetic innovations and advancements have made differentiating between these rare autosomal disorders easier. Some syndromes with inherited poikilodermas conditions have been classified as follows: Rothmund–Thomson syndrome (RTS, OMIM) #268400), Bloom syndrome (OMIM #210900), congenital dyskeratosis (OMIM #305000), Baller–Gerold syndrome (OMIM #218600), poikiloderma with neutropenia (OMIM #604173), hereditary sclerosing poikiloderma of Weary (OMIM #173700) and Kindler syndrome (OMIM#173650).[26,3,5,27]

Poikiloderma develops early in all these conditions,[25,28], and affected individuals suffer from photosensitivity. The Rothmund-Thompson syndrome (RTS) interlaces with other poikiloderma syndromes, particularly POIKTMP resulting in initial misdiagnosis with RTS in early childhood due to similar lesions. In fact, in some cases, former RTS cases are now being reported as POIKTMP, with over 300 cases reported in 2010.[29] The RTS disease primarily emerges as a differential diagnosis of hereditary poikiloderma syndrome and has frequently been connected to other clinical characteristics.[3] Besides, the gene mutation responsible for RTS is *RECQL4* on chromosome 8q24.3.[3,5,30,31]

1.2.4 Clinical manifestations of POIKTMP

In the first six months of life, poikiloderma in sun-exposed areas usually manifests in individuals with POIKTMP. In the second decade of life, tendon contractures, myopathy, and pulmonary fibrosis develop and manifests as restrictive lung function impairment with progressive dyspnoeal linked to primary fibrosis.[3,32,33] These clinical manifestations initially occurred in two generations of a South African family affected by POIKTMP, with the proband presenting with poikiloderma, alopecia, tendon contractures, myopathy, and sclerosis of the digits.[2] As a result of complications related to interstitial pulmonary fibrosis, the proband eventually died. The surviving member of the proband also displayed pulmonary complications with a high probability of lung transplant in the future.[2] Other extreme clinical presentations reported in individuals with POIKTMP include exocrine pancreatic insufficiency, mild lymphedema of the extremities, and liver impairment.[32] In fact, recent papers have recommended adding end-stage liver disease to the list of clinical phenotypes of POIKTMP and that the abbreviation be expanded to include POIKTMPL to account for liver involvement.[34] Severe liver fibrosis has presented in some patients with POIKTMP and has resulted in the deaths of these individuals.[29,34]

The consensus regarding these case reports establishes that the clinical features of POIKTMP vary from patient to patient, except for the early onset poikiloderma.[3,5,32,33] The initial differences identified as fibrosis in POIKTMP were thought to be associated with fibrogenesis regulation involving collagen and elastin abnormalities.[1,2] POIKTMP has been linked to dominant connective tissue disorder, which causes vasculopathy with secondary fibrosis, skin elastic tissue degeneration, and muscle fatty infiltration.[1,2,34] Hence, these cases support POIKTMP's classification as a multisystemic disorder associated with fibrosis

1.3 The human *FAM111B* gene

The *human family with sequence similarity member B (FAM111B)* gene on chromosome 11q12.1 encodes for a protein with a poorly understood function.[2,3,5] The *FAM111B* protein is expressed in various organs and tissues throughout the body, including the lungs, pancreas, liver, skin, skeletal muscle, and kidneys.[3,5] This protein contains a putative trypsin-like peptidase domain identified in the C-terminus. Missense mutations alter protease activity in the body, which may explain the various signs and symptoms of POIKTMP.[3,5,35,36] Most FAM111A and FAM111B mutations found in patients are located outside this protease domain. However, the molecular basis of how these mutations cause disease and whether they are caused by loss or gain of function mechanisms is unknown.[37] Other studies have suggested that *FAM111B* interacted with its paralog-*FAM111A*, as both genes encode proteins required for DNA replication.[36,37]

1.3.1 FAM111B mutations and POIKTMP

The *FAM111B* gene mutations surrounding this protease domain span the amino acid sequence of the *FAM111B* protein from position 475 to 665.[6] Previously described mutations in *FAM111B* codons 627 and 628 are linked to early-onset disease in affected individuals. In contrast, the symptoms caused by the heterozygous *de novo* mutation in codon 621 manifested later in the South African family, allowing the disorder to be passed down through familial transmission and explaining genetic heterogeneity.[3,5,6] About eleven different *FAM111B* gene mutations have been reported in five familial cases, which form part of the thirty-six individual cases of POIKTMP worldwide.[6,34,38], and the location of these mutations with regards to the protease domain has been examined as a prediction of disease severity (**Table**).[6]

Table 1-1: FAM111B mutations in POIKTMP cases with the outside (MOPPD) or within (MWPPD) of their putative protease domain

No	Mutation	Origin	Mutation type/ Mode of Inheritance	Mutational location (MWPPD/MOPPD)	Reference
1	c.1861T>G (p.Tyr621Asp)	South Africa	Missense /Autosomal dominant (Paternal)	MWPPD	[3,5]
2	c.1874C > A (p.Thr625Asn)	France	Missense/ *De novo*	MWPPD	[3,5]
3	c.1879A > G (p.Arg627Gly)	Alegria France Italy	Missense/ *De novo* (paternal inheritance)	MWPPD	[3,5]
4	c.1247T>C (p.Phe416Ser)	Chinese	Missense *De novo* (Maternal)	MOPPD	[30]
5	c.1881C>T (p. Arg672Ser)	Mexican	Missense/ *De novo*	MOPPD	[29]
6	c.1873A>C (p.Thr625Pro)	Chinese	Missense/ *De novo*	MWPPD	[39]
7	c.1883G > A (p.Ser628Asn)	France; Morocco; Ireland; Dominican Republic	Missense/ *De novo*	MWPPD	[3,5]
8	c.1289A > C (p.Gln430Pro)	Kuwait	Missense/ *De novo*	MOPPD	

9	c.1261_1263del AAG (p.Lys421del)	United States of America	In-frame deletion/ Autosomal dominant Paternal	MOPPD	40
10	c.1884T > A (p.Ser628Arg)	France	Missense/ Autosomal dominant	MWPPD	32
11	c.1886T > C (p.Phe629Ser)	Germany	Missense/*De novo*	MWPPD	8

Note: Adapted from Mutations within the putative protease domain of the human FAM111B gene may predict disease severity and poor prognosis: A review of POIKTMP cases 6

The varied phenotypic characteristics of POIKTMP patients have been linked to the position of *FAM111B* mutations within or outside the putative peptidase domain in the C-terminus location of the protein (i.e., MWPPD and MOPPD).[3,5,6] This study suggests that clinical manifestations are more severe in patients within the MWPPD FAM111B mutations than in the MOPPD group, with higher rates of life-threatening diseases like liver abnormalities, cirrhosis, and pulmonary complications within the MOPPD group. Furthermore, POIKTMP-associated fatal pancreatic cancer cases occurred within the MWPPD group; hence these findings have shown the progression of elucidating the disease.[8]

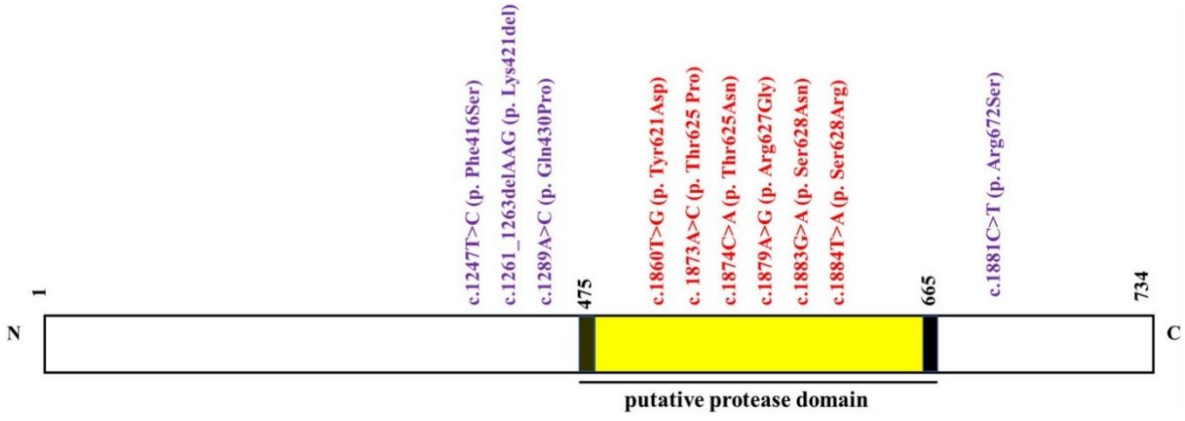


Figure 1-2: FAM111B encoded the putative protease domain regions. The cluster of POIKTMP-associated FAM111B gene mutations outside(MOPPD) and within(MWPPD) the putative domain.[6]

1.3.2 FAM111B in cancers

FAM111B has an alternative nomenclature: Cancer-Associated Nuclear Protein (CANP).[30,34] Genome-wide association study (GWAS) initially identified *FAM111B* as a cancer susceptibility gene in prostate cancer. Furthermore, *FAM111B* gene mutations have also been linked to some cancers, thus, highlighting the cancer-predisposing effects of *FAM111B* deleterious mutations.[7,32] Pancreatic and liver cancers have also been reported in some POIKTMP cases, specifically in older patients.[2,3,7,29] Functional analysis and cross-sectional patient examinations are required to support this hypothesis to confirm the cancer-supporting effects of *FAM111B* mutations.[29,34] There are still unanswered questions about whether POIKTMP is an incidental association or a cancer genetic predisposition.

There are clinical and molecular studies pieces of evidence that suggest the *FAM111B* gene mutations are oncogenic.[41,42] Functional mechanism analysis has tried to elucidate the gene function of *FAM111B,* mostly related to cancer cell-based assay studies. Hallmarks such as cell cycle deregulations,[42] cell cycle developments, progression, and proliferation have been reported in some patients' samples, enhancing the damaging effects of the pathogenic mutation on human tumors.[42] Recent research suggests that the mutation causes cytotoxicity by upregulating a putative serine-like domain in the FAM111B protein.[37] This gene has also increased catalytic activity, DNA replication, transcription repression, and eventual cell death.[6,8,37] The FAM111 gene's fatality has been attributed to progressive tissue damage caused by faulty DNA repair.[43] In POIKTMP patients, chromosomal instability has been demonstrated; these factors, which contribute to impaired DNA repair, may also cause cancer predisposition in the overexpression of the *FAM111B* gene product, which has also been found in some cancers.[37,41,43]

1.4 The use of genomics and histology as tools in elucidating the molecular mechanisms of POIKTMP

The causal heterozygous *FAM111B* missense mutations in POIKTMP were discovered through molecular genetic testing.[44] Advances in molecular genetic testing include gene-targeted and comprehensive genomic testing depending on the phenotype. A clinician must assess the presence of the gene(s) involved, which may be either single-gene or multi-gene panel testing.[44] Alternatively, whole exome and genome sequencing could be performed.[44] Because of the poor prognosis and early manifestation of POIKTMP, genetic testing is the more likely approach to diagnosing POIKTMP.[26]

1.4.1 Genomics Study Tools

Functional genomics is an effective tool that attempts to describe the gene functions and interactions of genes and proteins via genome-wide approaches to establish a correlation between the genome and the phenotype of an organism.[45] This method uses DNA sequences, gene expression, and protein function data to create interactive networks that regulate gene expression, cell differentiation, and cell cycle progression.[45] Functional genomics tools are helpful in the case of pathogenic variants and in understanding how specific mutations affect human health. It is now evident that pathogenic variants like *FAM111B* cause diseases, so understanding pathogenic mechanisms can help with developing tools and designing effective treatments.[46] To achieve this goal, genome analysis involving different fields of study like genomics, epigenomics, transcriptomics, proteomics, and interactomics are necessary to describe the functions of genes and further enable expression profiling.[46] This approach can be effective in the disease management of POIKTMP by studying the part of the *FAM111B* gene relating to molecular pathways affected by the mutation.

1.4.2 Genomics tools for detecting gene mutation

1.4.2.1 Restriction fragment length polymorphism (RFLP)

Restriction fragment length polymorphism (RFLP) is a primary method of analyzing nucleic acids to detect the presence of known sequence variants among different individuals.[47] The molecular basis of RFLP is that nucleotide base substitutions, insertions, deletions, duplications, and inversions can remove or create new restriction sites throughout the genome.[48] The technique involves, examining a specific site of the mutation and amplifying the genomic DNA, this is followed by digesting the genomic DNA into fragments using restriction enzymes and analyzing these fragments.[47-49] This method is often combined with polymerase chain reaction (PCR)-PCR-RFLP to differentiate between the wild-type and mutant genotype. It amplifies the target genomic DNA with short, random primers typically around 10 bp in a PCR reaction and can be used to generate relatively complicated DNA to detect amplified fragment length polymorphisms between organisms.[49] Because the random primers complement different parts of the genomic DNA, PCR products differ in number and size (polymorphism).[48] However, this technique is a bit outdated and has several limitations: RFLPs can only observe specific mutations at enzyme cut sites, which limits the identification of whole genome variation.[48] Additionally, the polymorphism of RFLP markers

is relatively low and needs further detection by radioisotopes, thus limiting the application. In some cases, the application may require different probe/enzyme combinations.

1.4.2.2 Sanger sequencing

DNA sequencing is the "gold standard" for detecting gene mutations. DNA sequencing methods are critical in molecular biology and genomics and have evolved into viable research tools.[45] This method is also helpful in identifying significant evolutionary changes in the genomes of known pathogens and verifying suggestive findings of high diagnostic specificity.[50] Sanger sequencing, often known as the "chain termination method," is a technique for determining the sequence of nucleotide bases in a segment of DNA that is typically less than 1000bp in length. Sequencing is critical for detecting emerging pathogens or novel genotypes of existing pathogens, discovering key evolutionary changes in the genomes of established pathogens, and, most crucially, verifying the discovery of a specific pathogen from a new species or geographic location.[50]

In addition to serving as a confirmatory test with high diagnostic specificity, Sanger sequencing analysis are important in the initial development of molecular assays, monitoring the effectiveness of molecular-based assays, investigating the source of a disease outbreak or agent, and determining the genotype or differentiation of field and vaccine strains.[50] Considerable advancements and modifications of methodologies in sequencing analysis, such as *de novo* sequencing and Next generation sequencing (NGS), including large-scale parallel sequencing have transformed whole-genome sequencing.[50] Thus, contributing to a relatively low error rate and generating millions of short DNA reads around 100-200 bp in length aligns with the human reference sequence.[45,50]

Limitations to this technique include its low-quality sequences within the first few base pairs due to primer binding, also its inability to distinguish single base pair differences in longer segments. Commercial and non-commercial sequence analysis software platforms constantly attempt address these limitations.[51] A small-scale project, such as in this study, would benefits from conventional Sanger sequencing as it provides some degree of accuracy. On the other hand, large-scale projects benefit more from advanced DNA sequencing methods discussed later in this chapter. However, the high operational costs, low throughput, and longer time to the output of Sanger sequencing, have limited its use for whole genome sequencing (WGS) in many species, particularly those with large genomes.

1.4.2.3 Next Generational sequencing

Next-generation sequencing (NGS) applications are new advanced DNA and RNA sequencing methods. The advent of NGS has opened new possibilities and challenges in genome exploration and clinical diagnosis.[52] NGS differs from traditional Sanger sequencing in that it allows for the simultaneous analysis of millions to billions of sequences from individual strands of DNA.[52] NGS has primarily replaced traditional Sanger sequencing in clinical and non-clinical applications and is applied in genome sequencing, genome diversity investigation, metagenomics, epigenetics, non-coding RNA and protein-binding site discovery, and gene-expression profiling using RNA sequencing.[53] Initially, NGS was used primarily as a targeting strategy, with the desired regions of interest identified using positional mapping sequenced to recognize the changes in the gene of interest.[53]

This tool is widely used to classify sequence variations. It has provided several genetic markers, including standard and uncommon variations, that can be used to investigate diseases like cancer using NGS-based molecular diagnosis. NGS sequencing is categorized into three types: whole genome sequencing (WGS), whole-exome sequencing (WES), and targeted sequencing. WGS supports the sequencing of the entire genome and can be beneficial in identifying a wide range of significant genetic variations like nucleotide substitutions, insertions and deletions, copy number variations (CNVs), chromosomal modifications, and non-coding region examination.[53]

WES was established as an inexpensive method of acquiring the sub-genome directly associated with the coding regions of the genome.[54] In other words, this method searches for rare or common variants in coding regions (exons) of a genome linked with a disorder or phenotypes. Because the exome accounts for 1.5% of the genome, WES can be used to determine the sequence of regulatory regions, such as promoters, and non-protein-coding regions, like microRNAs, describing the functional parts of the genome.[55] WES used for targeted sequencing of a specific disease may be more accurate and cost-effective for clinical applications. It also superspeeds the discovery of genetic causes and contributors to disease in both research and clinical settings.[55] Moreover, WES can achieve more excellent sequence coverage than (WGS).[56]

The standard method for WES includes developing a library and using solution-based target enrichment, which is becoming more popular due to its ease of automation and simplicity. A pool of oligonucleotide probes (DNA/RNA) is synthesized to hybridize target exon regions.[57]

The probes are made up of pull-down tags with beads that allow the hybridized probe-target fragments to be isolated and the unbound regions of genomic DNA to be washed away. The targeted genomic elements are eventually sequenced on the NGS platform.[57]

NGS is advantageous in delivering information on the genome sequences of various diseases and is a tool that future pathologists and clinicians may find invaluable. This platform's data can provide a comprehensive assessment of genomic landscapes associated with the evolution of various disorders.[58] NGS can provide real-time insights into a patient's genomic, transcriptomic, and epigenetic makeup, promising results in personalized healthcare prospects. It is well established to assess gene expression patterns through cDNA sequencing, and it has the potential to provide more insights into disease variability at an individual level.[58]

In diseases like POIKTMP, where there appear to be multiple *FAM111B* variants, genetic/genomic profiles of patients could help inform treatment care and monitor disease progression, diagnosis, and efficacy of treatment.

1.4.3 Tools in gene expression studies

Transcriptomics generally addresses gene expression and regulation, which is essential for cellular development and differentiation. Sequencing of the cDNA rather than genomic DNA analysis essentially targets the transcribed portion of the genome. Transcriptome analysis provides insights into the transcriptional expression of the genome, which expands our understanding of gene structure, gene expression regulation, gene product function, and genome dynamics.[59] This will further ensure the regulation network at work of biological processes and offer instructions in disease diagnosis and treatment strategies.[59]

1.4.3.1 DNA Microarray

DNA microarray technique is based on complementary probe hybridization and is used to concurrently measure the relative abundance of transcripts from two or more samples for thousands of genes.[60] This method primarily monitors expression by measuring gene transcript levels in different physiological/pathological or experimental conditions in tissues and cells to look for regulatory expression patterns. This information is expected to improve understanding of complex networks of cross communications in healthy and diseased individuals. The application also performs polymorphism analysis, which involves scanning polymorphic regions of the genome for disease linkage to identify disease susceptibility

genes and inherited disease genes.[61] Furthermore, this technique uses the technological intersection of biology and computers to screen many genes simultaneously and is automatable. By combining photolithography and oligonucleotide chemistry, gene-specific cDNAs can be spotted on a nylon membrane or glass surface chips, or oligonucleotides can be synthesized *in situ*. The chips' strength lies in their ability to conduct comparative expression studies in diseased versus healthy samples and to document changes at various stages of the disease's natural course or response to treatment. It can help determine the genetic basis of complex diseases.[62]

In DNA microarray, the extracted RNA from the sample of interest must first be labeled directly and converted to a labeled complementary DNA (cDNA) or a T7 RNA promoter-tailed cDNA, which is then converted to cRNA. The microarray is washed after the labeled cRNA or cDNA is hybridized, and the signal is detected by measuring fluorescence at each spot. Signal intensity in each area is used to calculate the level of expression of the corresponding gene. Microarrays can be used with chromatin immunoprecipitation to identify transcription factor (TF) binding sites, where the TFs are cross-linked to DNA. At the core, microarrays measure the relative concentrations of multiple DNA or RNA sequences.

It is typically assumed that the signal measured on a microarray corresponds to the concentration of a single species in solution that can hybridize at that position.[63] So, this result is in the limited range of detection because of background and saturation signals; high background levels result in cross-hybridization and reliance upon existing knowledge of the genome sequence.[63] Furthermore, it is difficult to design arrays in which multiple related DNA/RNA sequences won't bind to the same probe on an array, especially for complex mammalian genomes. Gene families and genes with multiple splice variants can be particularly affected by this. Lastly, DNA arrays can only detect sequences that they were designed to detect. In other words, RNA or DNA from the hybridization solution will not be detected if there are no complementary sequences on the array. This often means that genes that have not yet been annotated in a genome will not be represented on the array for gene expression analysis.[63]

The RT^2 Profiler PCR Array system is a technique that is currently being used to analyse the expression of a specific panel of genes by combining the quantitative performance of RT-PCR with the multi-gene profiling efficiency of microarray.[64] The technology is equipped with specific RT^2 qPCR Primer Assays uniquely designed for analysis. The technology

criteria will ensure that the quantitative PCR (qPCR) reaction generates single, gene-specific amplicons while preventing nonspecific co-amplification. Additionally, it allows for many genes to be assayed simultaneously.[64]

1.4.3.2 RNA-sequencing (RNA-Seq)

As advances toward analyzing gene expression specificity increase, so does RNA-sequencing technology which is an extension of NGS technologies. Microarray analysis and RNA-seq technology are two examples of high-throughput transcriptomics.[65] High-throughput sequencing technologies are increasingly being used in biomedical research because they have tremendously accelerated the characterization and quantification of genomes, epigenomes, and transcriptomes over the last few years since their initial application.[65,66]

RNA-Seq technology produces cDNA sequences derived from the entire RNA molecules, followed by library preparation and eventual massive parallel sequencing. Applying RNA-seq quantification permits each transcript's abundance level or relative changes during defined developmental stages or under specific treatment conditions.[67] Also, RNA-seq is unbiased, allowing transcriptome analysis with single base pair resolution, tremendous detection range, and low background signals.[68] While RNA-seq technology is thought unbiased, it should be noted that RNA preparation and fragmentation, as well as library construction, can all be influenced. Strand-specific sequencing keeps the original RNA transcript's orientation, which may help identify antisense or non-coding RNA.[67] Because data generation is constantly evolving, it is necessary to continuously develop sequencing technology, experimental design, and algorithm development when interpreting RNA-seq datasets. Essential tools for RNA-seq data analysis include quality assessment and read mapping.[66] In contrast to other hybridized-based gene expression tools, RNA-seq is not limited to interrogating selected probes on an array. It can also be used in species where the whole reference genome is not assembled.[69] The appeal of RNA-seq for gene expression profiling is that it can be used for quantitative and exploratory RNA analysis. Because 85% of the human genome can be transcribed, with only 3% protein-coding genes,[65,67] the technology effectively classifies the diversity of novel transcript species such as non-coding RNA, miRNA, siRNA, and other smaller classes involved in RNA stability and protein translation regulation.[65] Additionally, RNA-seq can provide information about transcriptional start sites by revealing promoter usage using mRNA data. It also can explore allele-specific

expression, disease-associated single nucleotide polymorphisms (SNP), and gene fusions, all of which contribute to our understanding of disease causal variants.[65,70]

1.4.4 Proteomics

Proteomics is critical to understanding how proteins function in biological processes.[45] It characterizes the proteome of organisms, thus exploring proteins present in a system. Protein expression analysis between diseased and healthy states of a condition is enabled by disorders or malfunctions that alter the protein. In essence, the proteome can characterize most of the valuable information of genes and is crucial for diagnosis.[71] Mass Spectrometry (MS) platforms have been developed to analyse complex protein mixtures with high sensitivity and require high-throughput techniques like Bioinformatics applications to understand their biological functions. In addition to analysing proteins, the proteomics approach has expanded to include structural and functional proteomics, profiling, and structural analysis.[72] Functional proteomics can also be used to look for proteins that have been post-translationally modified or for protein interactions. With the inclusion of software analysis tools, the proteomics approach has evolved from simply analysing proteins into profiling and characterizing structural and functional proteomics.[72]

1.4.5 Tools for validating or quantifying gene expression

Gene expression is a multistep disciplinary approach that governs the cellular process of gene transcription and translation. In this area, total mRNA and protein abundance are assessed and measured. To achieve a quantitative description of gene expression, combining studies comparing mRNA and protein levels is an excellent method of correlating data. By implementing these validation tools, cellular mRNA and protein expression levels can be quantified in parallel to confirm relative expression.

1.4.5.1 Western Blot Analysis

Western blot analysis helps determine the presence, relative abundance, mass, post-translational modifications (PTM) of proteins, and protein-protein interactions. Because it relies on antibodies to target antigens found in a sample, this is useful in diagnosis. After sample separation on a protein gel and transfer to the membrane, primary and secondary antibodies are used to bind and visualize the target protein.[73] The western blot analysis is ideally suited to visually confirm the presence, absence, size, and modifications of the target protein in a protein sample.[73] It is also a quantitative tool for determining the target protein's

relative expression level.[74] Protein sample preparation and separation, transfer and total protein imaging, antibody incubation, and chemiluminescent detection are all part of the Western blot procedure.[74] It is also possible to accurately analyse and quantify western blots using chemiluminescence technology and sophisticated software tools.[74] Western blots can be used for absolute quantification with a known concentration of recombinant protein in conjunction with a calibration curve or for quantification of samples relative to a control sample. A housekeeping gene or staining of the transferred proteins on the membrane is used to normalize the data for relative quantification. For quantitative analysis, a correlation between the measurable results and the amount of protein in the examined samples must be established. This association is influenced by primary and secondary antibody kinetics, fluorescence detection, image acquisition, processing, and quantification.[75] Another factor that influences the absolute measure of quantity takes into account two factors: Before a more precise comparison can be made, variations in loading and transfer rates between samples in separate lanes on separate blots must be standardized. Second, because the signal produced by detection is not linear across the concentration range of samples, it should not be used to model the concentration.[76]

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

The quantitative real-time polymerase chain reaction (qPCR) is a popular method for measuring target gene expression in various samples because it provides a more comprehensive understanding of gene expression changes.[77] Total RNA is extracted from the sample, and the mRNA is reverse-transcribed into complementary DNA (cDNA), which is then used as the template for the qPCR reaction. [77] The advantage of qPCR over traditional RNA quantification methods, such as Northern blotting and *in situ* hybridization, is that it is quick and straightforward to perform. In addition, compared to other approaches, the method used for qPCR detection is qualitative and specific. As a result, qPCR is currently regarded as one of the most effective methods for quantifying gene expression, as it can detect mRNA with low expression levels.[77] The workflow of gene expression using qPCR includes the acquisition and handling of experimental samples, total RNA extraction, RNA concentration assessment and quality check, cDNA synthesis via reverse transcription, optimizing the conditions of the qPCR assay, running the qPCR reaction under optimized conditions to measure the level of target genes, and finally, data analysis using appropriate normalization methods.[77]

1.4.6 Histology and Immunohistochemistry

Histology and Immunohistochemistry is a powerful diagnostic and research tool that uses novel techniques to detect antigens in tissue sections using specific antibodies that can be visualized through staining. This method is important because it directs researchers by identifying cellular and subcellular protein distribution.[34] This technique has been used in a variety of applications or settings, including cellular differentiation, characterization of primary tumour sites as a detection method, prognostic factor, a predictor of targeted therapy response, and is critical in the identification of structures, microorganisms, and materials secreted by cells of interest.[34] For instance, in POIKTMP-related studies, histology provided insights and suggestive imaging findings in the skin, lungs, and skeletal muscles.[3,5] Muscle histology revealed extensive fatty acid infiltrations, and residual muscle tissue composed of fragmented muscle fascicles has been classified using this tool. Studies in skin histology have discovered a pattern of epidermal atrophy with collagen sclerosis and elastic degeneration in the dermis.[3] Furthermore, post-mortem findings of an affected member of the South African family revealed diffuse fatty infiltration and fibrosis of organs, including the lungs, pancreas, and oesophagus.[2] However, how FAM111B mutations cause these histological changes remains unclear—moreover, studies using immunochemistry to monitor pathological changes of key fibrosis markers are almost non-existent. Therefore, using immunohistochemistry has the potential to offer insights into the molecular changes occurring in the tissues affected by POIKTMP and, more explicitly, relating to FAM111B.

1.5 Aims and Objectives

The aims of this study are: 1) to annotate the pathological changes resulting from the *Y621D FAM111B* gene mutation associated with POIKTMP and understand the dynamic of the processes that leads to the observed changes at the level of the affected tissues. 2) to analyse the expression patterns of FAM111B in different tissue cell types affected by POIKTMP and highlight how the expression profiles differ from control tissues.

1.5.1 The objectives of the study are:

i. Confirm the *FAM111B Y621D* gene mutation in POIKTMP-affected tissues.

ii. Evaluate the gene expression patterns of *FAM111B* and key fibrosis biomarkers, including pathway analysis in a panel of human fibrosis genes.

iii. Assess the protein expression of key fibroproliferative markers (i.e., ki-67; Annexin V; TGFβ-1), including FAM111B in POIKTMP-affected tissues.

Chapter 2

2 Materials and Methods

Formalin-fixed and parafilm-embedded (FFPE) post-mortem tissue sections of POIKTMP-affected patients and three age-matched and gender-matched healthy non-familial controls were donated in collaboration with Prof. Komala Pillay, Head of the Anatomical Pathology department at the University of Cape Town. These samples were part of the original cohorts used in the 2006 study at the Groote Schuur Hospital. The Study obtained Ethical Clearance from the University of Cape Town, with the reference number HREC 057/2021.

2.1 FFPE sample preparations

About two to three 5-10 µm sections from the different FFPE tissue blocks were each placed into a 1.5 ml microcentrifuge tube. For the dewaxing and deparaffinization of these FFPE tissue sections, 1 ml xylene was added into the tubes containing the FFPE sections and vigorously vortexed for 10s, and centrifuged at 12, 000 g for 2 minutes. After that, the samples were rehydrated using a graded system of 1 ml of ethanol at 100, 95, and 70% (Merck,Gremany), with the resultant pellet vortexed and centrifuged at 12, 000 g for 2 minutes to extract residual xylene. The pellet was then incubated at 37℃ for 10 minutes until the residual ethanol had evaporated.

2.2 Nucleic Acid and protein extraction

2.2.1 DNA Extraction

Genomic DNA isolation from the POIKTMP post-mortem FFPE patient tissue sections was achieved using the QIAamp DNA FFPE Tissue kit (Cat. No. 56404, Qiagen, Germany) based on the kit's manufacturer's protocol. Following dewaxing and deparaffinization, the kit's tissue lysis buffer was applied to the FFPE tissues to enable the isolation of purified genomic DNA, Proteinase K is added and carried out under denaturing conditions of 56℃ and 90°C for an hour each for further digestion to partially remove formalin crosslinking of the released DNA and to improving DNA yields. The reaction was briefly centrifuged at 6000x g and 2 µl RNase A (100 mg/ml) was added to incubated for 2 min at room temperature. We then carried out DNA binding, the buffer AL and 96% ethanol were mixed and added to the reaction for the DNA to bind to the membrane, the lysate was transferred to the QIAamp MiniElute column and centrifuged at 6000 x g for a minute, this ensures contaminates flow through. Residual contaminants are eliminated using the manufacturers' guidelines to

eliminate cell debris using several washing steps and centrifugation. Lastly, pure DNA is eluted from the membrane by adding 35 µl of DNA elution buffer ATE and centrifuging at 20, 000 x g for 1 min. The DNA was quantified by using the UV/vis spectrophotometer (Biodrop,Biochrom, WhiteSci) and stored at -80°C.

2.2.2 Total RNA Extraction

Total RNA from the FFPE sections was dewaxed and deparaffinization, as described in 2.1, was extracted using the RNeasy FFPE kit (Cat.No.73504, Qiagen, Germany) following the manufacturer's guidelines. Samples are incubated in an optimized lysis buffer which contains proteinase K and briefly centrifuged for 1 minute at 11,000 x g, to release RNA from the sections. A short incubation at a higher temperature of 56°C for 15 minutes, then 80°C for 15 minutes is adhered to partially reverses formalin crosslinking of the released nucleic acids, improving RNA yield and quality, as well as RNA performance in downstream applications such as real-time PCR. After that the lysate is incubated on ice for 3 minutes, then centrifuge for 15 minutes at 20,000 x g. Following this, DNase treatment is optimized to remove all genomic DNA, including very small DNA fragments that are frequently present in FFPE samples after prolonged formalin fixation and/or long storage times.

The lysate is then mixed with Buffer RBC, ethanol is added to provide appropriate RNA binding conditions, and the samples were applied to an RNeasy MinElute spin column, where total RNA binds to the membrane and contaminants are efficiently washed away in the wash buffer provided, and the samples were centrifuged at 8000 x g according to the protocol. The RNA is then eluted in at least 25 µl of RNase-free water and stored at -80°C.

2.2.3 Protein Extraction

The FFPE tissue sections were dewaxed and deparaffinized using xylene and ethanol according to the manufacturer's protocol in the Qproteome FFPE Tissue kit (Cat.No.37623, Qiagen, Germany). Following deparaffinized tissue was incubated at 80°C for 20 minutes and 100°C for two hours on a Thermomixer with agitation at 750 rpm in an optimized lysis and extraction buffer (EXB) supplemented with 6µl of β-mercaptoethanol (Cat.No.1610710,Bio-Rad, United States) to reach a total volume of 100 µl. Following the incubation step, samples were centrifuged for 15 minutes at 14, 000 g, and the supernatant containing the released protein recovered and stored at -80°C.

2.3 Validation of the *FAM111B Y621D* gene mutation

2.3.1 Polymerase chain reaction with restriction Fragment Length Polymorphism (PCR-RFLP)

The *FAM111B* Y621D patient mutation was validated in the South African POIKTMP patient by PCR-RFLP. About 25 ng/µl of DNA was amplified with the *FAM111B* gene targeting primers described below (Inqaba Biotech, South Africa, **Table 2-1**) using the Platinum^{TM} Direct PCR Universal Master Mix (Cat. No.A44647100,Thermo Scientific, United States) according to the manufacturer's protocol. Following amplification and visualization on a 1 % agarose gel, the amplified PCR product was cleaned up using the Thermoscientific Genejet PCR Purification kit (Cat.No.K0702,Thermo Scientific, United States). The purified DNA was digested with the restriction enzyme BstZ171-HF(Cat.No. Lot No.10149198,New England BioLabs, United States) for 15 minutes at 37°C and resolved on a 2 % agarose gel electrophoresis at 100V for 65 minutes. The gel was visualized using the Azure Biosystems c400 (Azure Biosystems, United States) imaging system.

Table 2-1: FAM111B Sequence

FAM111B SEQUENCE

FORWARD PRIMER	ATGCTATTAATCTGGATGTCCAAAAGGAGG
REVERSE PRIMER	CTAACATTCCATGGGTTCAATCTGATGATC

2.3.2 Sanger Sequencing

DNA sequencing of the purified PCR amplicon obtained from 2.3.1 was performed as a service at Stellenbosch University's Central Analytical Facilities (CAF).

2.4 *FAM111B* gene expression and fibrosis pathway analysis

2.4.1 Reverse transcription and quantitative polymerase chain reaction (RT-qPCR)

2.4.1.1 cDNA synthesis

The total RNA extracted in section 2.2.2 was reverse transcribed using the LunaScript RT SuperMix Kit (Cat.No.E3010S,New England BioLabs,United States) following the manufacturer's guidelines in a total of 30 µl reaction containing 10 ng/µl of total RNA, 6 µl LunaScript RT SuperMix (5x) and nuclease-free water. The cDNA synthesis reaction was incubated using the T100 Thermal Cycler (Cat.No.1861096, BioRad, United States) under the conditions: Primer annealing at 25°C for 2 minutes; cDNA synthesis at 55° C for 10

minutes, and finally at 95° C for heat inactivation for a minute. The cDNA was and stored at -20°C until further downstream use.

2.4.1.2 qPCR

To quantify the mRNA expression of all target genes, qPCR was carried out on the synthesised cDNA (2.4.11) using the Luna Universal qPCR Master Mix (Cat.No.M3003,New England BioLabs,United States) in a reaction volume of 30 µl, according to the manufacturer's protocol. The total RNA samples for the healthy control patient was pooled together and optimized as they were from the same population. Primers for the studied genes, *FAM111B* and control house-keeping genes, GAPDH and ACTIN, (**Table 2-2**) were utilised. qPCR amplifications were performed using the QuantStudio 6 Flex (Cat. No.4485699, Applied Biosystems, ThermoFisher Scientific, United States) under the following conditions: DNA denaturation at 95°C for 60 seconds for DNA strand separation, followed by a 40x cycles of DNA denaturation and extension (15 seconds at 95°C, 30 seconds at 60°C). The temperature of the melting curve was set at 60 – 95°C to ensure amplification specificity.

Table 2-2: RT-PCR Primers sequences

Primer	Sequence
FAM111B (forward)	GCTAGCATGAATAGCATGAAGACA
FAM111B(reverse)	GGATCCGCACTCCATAGG
GAPDH (forward)	GACCTCAACTACATGGTTTACATG
GAPDH (reverse)	GATCTCGCTCCTGGAAGATG
ACTIN (forward)	GATTCCTATGGACGAG
ACTIN (reverse)	GTTGGTGACGGCCGTG

2.4.2 Human fibrosis gene markers profiling analysis

This experiment was performed using the catalogued Format A reagents of the RT^2 Profiler PCR for human fibrosis Array kit (Cat. no. PAHS-120Z,Qiagen,Germany) compatible with the model of the real-time cycle for QuantStudio flex 6 (Applied Biosystem, Life Technologies, United States). Following RNA quantification and quality control on the integrity of the FFPE samples and to measure the accurate concentration using the Agilent® 2100 Bioanalyser (Cat. No.G2939BA, Agilent, United States) **(Figure A-0)** with the RNA 6000 Nano Lab Chip, cDNA synthesis was prepared using the RT^2 First Strand synthesis kit

(Cat. no.330401,Qiagen, Germany). The RNA samples from the control samples were pooled and cDNA synthesized using RNA concentration similar to that used for the patient's RNA with the same kit. About 25 ng of RNA from patient or pooled control RNA sample was subjected to genomic DNA elimination using the DNA elimination mix provided in cDNA synthesis kit in a 10 µl reaction volume at 42°C for 5 min. Following this step, cDNA synthesis was initiated by adding 10µl of the reverse transcription mix to the genomic DNA elimination reaction and at 42°C for 15 min and stopped by increasing reaction temperature to 95°C for 5 min. The resultant cDNA solution was diluted with 91µl of RNase-free water and stored on ice for use in PCR.

For the quantitative PCR reaction, a 2700 µl of bulk reaction mix consisting of 102 µl of the cDNA synthesis reaction, 1350 µl of RT² SYBR® Green qPCR Mastermix (Cat. no. 330529, Qiagen, Germany) and RNase free water was prepared 25µl of this bulk mix dispensed in each well of the RT^2 Profiler PCR Array 96-well plate reaction. The array plate was briefly centrifuged at 1000 x g for 1 min target amplification reaction done at the cycling conditions: 1x denaturation at 95°C for 10 min, 40 x cycles of DNA denaturation and extension (15 secs at 95°C, 50 secs at 60°C). The cycle threshold (Ct) values in the RT-PCR were used to measure the target gene amplification and expression. The Ct values were then exported into an Excel file and uploaded to the GeneGlobe® data analysis web portal at https://geneglobe.qiagen.com/. Samples were assigned to controls and test groups. Ct values were normalized based on the reference genes and the fold change of the expressed genes calculated using the delta-delta Ct method of the GeneGlobe® data analysis web portal. These analysed data (fold change) including the data relevant visualizations for the differentially expressed genes were then exported as an Excel or PDF document for further analysis.

2.4.2.1 Validation of differentially expressed genes

From the data analysis reports described in 2.4.2, the eight most highly differentially expressed genes were selected and validated with qPCR using the targeted primers **Table 2-3**.

Table 2-3: RT^2 Profiler PCR Array for human Fibrosis validation Primers

Name	Forward Sequence	Reverse Sequence
Platelet-derived growth factor subunit A (PDGFA)	CAGCGACTCCTGGAGATAGAC	GGACAGCTTCCTCGATGCTT

Transforming growth factor beta 3 (TGFβ-3)	ATCTGGGAAATGGGCTCGG	TTCTGTCCCCTGCTTCTCTCTC
Cellular communication network factor 2 (CCN2)	CACCCGGGTTACCAATGACA	GGATGCACTTTTTGCCCTTCTTA
Matrix metallopeptidase 13 (MMP13)	AAGATGCATCCAGGGGTCCT	ATCTCAGGTAGCGCTCTGCAA
Matrix metallopeptidase 3 (MMP3)	ACAAAGGATACAACAGGGACCA	GGAACCGAGTCAGGTCTGTG
Integrin subunit beta 1 (ITGB1)	TGGTCTCTAAATTGCCGGTGA	AGTGTTGTGGGATTTGCACG
Collagen type III alpha 1 chain (COL3A1)	GAAAGAGGATCTGAGGGCTCC	AAACCGCCAGCTTTTTCACC
Thrombospondin 2 (THBS2)	TCGGTCCGGAACACTGAAAC	CAGATCTGAGGCTTGCGTG

2.4.2.2 Gene-set enrichment analysis (GSEA)

The GSEA of the top eight upregulated profibrotic genes in POIKTMP lung and skin tissues obtained from GeneGlobe®, including the *FAM111B* gene, were further analysed using Enrichr[78-80], a web-based interactive tool for enriching cellular/tissue function or molecular pathways terms, to provide new knowledge on the collective processes of these genes in POIKTMP. A gene-list in the gene symbol format was created with *FAM111B,* and these top-upregulated genes and uploaded onto the online version of this enrichment software (https://maayanlab.cloud/Enrichr/). The gene-set libraries categories selected for our enrichment analysis are Pathways, Ontologies, Diseases/Drugs and Cell Types. In the Pathways category, the Kyoto Encyclopedia of Genes and Genomes (KEGG) and the REACTOME gene-set libraries were used in interrogating the gene-list. In the Ontologies category, the updated Gene ontologies (GO) (i.e., GO 2021) for Biological processes, Cellular components and Molecular function gene-set libraries were used. In the Diseases /Drug category, only the Covid-19-related gene-set library was queried. For the Cell Types category, the gene-set libraries: GTEx Aging signatures and GTEx Tissue expression Up, Human Gene Atlas, Descartes Cell Types and Tissues 2021 were used in the enrichment analysis. The results of the enriched biological and functional terms identified from the inputted gene-list were ranked in descending order of statistical significance based on the combined scoring enrichment computing method **(**a modified form of the Fisher-exact test

method for multiple testing of inputted gene-list in GSEA) described by the authors of the Enrichr software[78]. Furthermore, the top-five of the enriched terms and genes present in the gene-list associated with the enriched biological terms were visualised using the Bar graphs/Tables (enriched terms only) and Clustergrams (enriched terms and the related genes from the gene-list) visualisation tools of the same software[78].

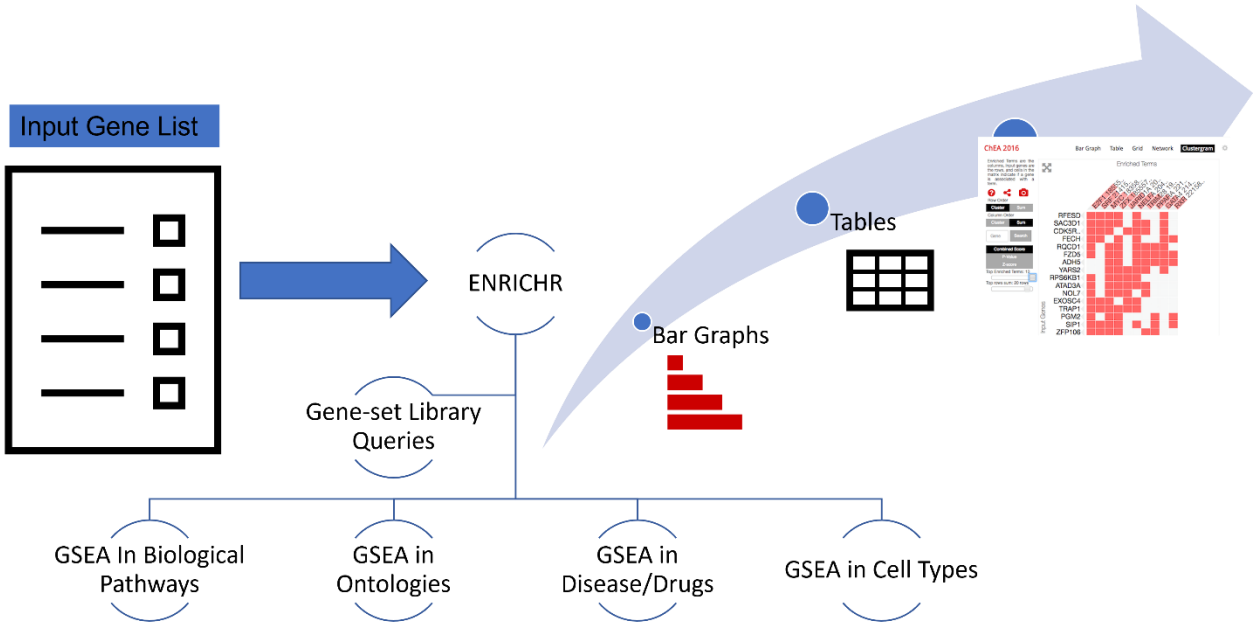


Figure 2-1:Enrichr Workflow for GSEA. Enrichr receives a gene set list as input then submits the gene-set list to the gene-set libraries to compute and query enrichment. The enrichment results are interactively visualized as bar graphs, tables and clustergram grids with the enriched terms highlighted based on the chosen score, and the genes enriched in the terms are also displayed in the clustergram

2.5 Tissue expression of FAM111B and fibroproliferative biomarker proteins

2.5.1 Immunohistochemistry

FFPE tissue sections of about 5-10µm mounted on Histobond® microscope frosted tissue slides were incubated overnight at 37°C, and then for 1 hour further incubated at 60°C. Tissue dewaxing, deparaffinization, and rehydration steps were carried out on the slides using xylene and ethanol (100, 95, and 70 %). The slides were further hydrated for 1 minute in distilled water before being blocked for endogenous peroxidase for 15 minutes in methanol with 3% (v/v) hydrogen peroxidase $(H_2O)_2$. The antigen retrieval method used was Heat Induced Epitope Retrieval (HIER). Briefly, 10mM citrate buffer pH 6.0 was added to the tissue, followed by heating in a scientific microwave at 98°C for 10 minutes and then left to cool for

30 minutes. Non-specific binding was prevented by blocking with 2% bovine serum albumin (BSA) in 1 x phosphate buffer saline (PBS) supplemented with 0.1% Tween® 20 (PBST) for 1 hour. Afterwards, the sections were incubated overnight with the appropriate primary antibodies (**Table 2-4)** diluted in 1 x PBS supplemented with Tween® (Cat.No.9005645,Sigma-Aldrich, United States) detergent at 4°C.

The next morning, the sections were washed three times with 1x PBST, followed by a 30 min incubation with the secondary antibody conjugated to horse-radish peroxidase (DAKO EnVision + System-HRP Labbelled Polymer Anti Rabbit, Cat. No.K4003 and Anti-Mouse Cat. No.K40001, Agilent Technologies, United States). After that, the sections were washed twice with 1 x PBST for 5 minutes at each wash. The slides were developed for 5 minutes using diaminobenzidine (DAB) as the substrate Chromogen System (K346811-2, DAKO, Agilent Technologies, United States), and the reaction was stopped by rinsing in distilled water for 5 minutes. The samples were counterstained in Mayer's hematoxylin for 5 minutes followed by rinsing under running tap water for 5 minutes and then in Scott's TAP water for 5 minutes. The slides were finally dehydrated in a graded ethanol system briefly of 70, 96, and 100%, followed by three times brief incubation in xylene. Subsequently, the slides were mounted onto coverslip using Entellan mounting media (Cat. No. 07960, Merck,United States). The mounted slides were scanned and viewed microscopically using the Olympus VS120 with the Olyvia 2.9 slide viewer at 50 µM. The target tissue expression was quantified using the computer software ImageJ Fiji (Version 1.2, National Institute of Health) as per the published protocol. [81]

Table 2-4: Immunohistochemistry Primary Antibodies

Antibody	**Host Species**	**Catalogue no.**
FAM111B	Rabbit	HPA038637, Atlas Antibodies, Sweden
Annexin V	Mouse	H00000308-M01,Thermoscientific,United States
TGF-B1	Rabbit	Ab215715,ABCAM, United Kingdom
Ki-67	Rabbit	Ab16667,ABCAM, United Kingdom

2.6 Statistical analysis

All statistical analyses were performed with GraphPad Prism Version 9.0.0 (GraphPad, United States). Statistical significance was defined as a p-value ≤ 0.05 when comparing the mean and ±SEM between the control and patient sample groups as determined by t-test.

Chapter 3

3 Results

3.1 Validation of *FAM111B Y621D* mutation in a South African POIKTMP patient tissues

3.1.1 Y621D mutation validation using PCR-RFLP

The post-mortem FFPE tissue samples from a South African POIKTMP patient used in this study formed part of the samples from a previous study that identified mutations in the *FAM111B* gene as the cause of POIKTMP by whole exosome sequencing.

Therefore, the study attempted to use a more cost-effective PCR-RFLP method described previously in our lab to confirm the presence of the *FAM111B* heterozygous [NM_198947.4: c.1861T>G (p. Tyr621Asp or Y621D)] gene mutation. The results showed various inconsistencies in banding patterns that could not confirm the differences in mutation screening (**Figure 3-1**). Lane 1 shows the electrophoretic analysis of *BstZ17I* digested DNA fragments from the South African POIKTMP patient's tissues. The DNA band pattern was supposed to display three bands (1085, 740, and 345 bp), as the heterozygous Y621D mutation led to the loss of the BstZ*17I* site) to confirm the presence of the Y621D mutation. However, this result was not the case with the analysed DNA from POIKTMP patient skin tissue but appeared to be the case with the DNA from POIKTMP-patient lung and skin tissues in **Figure 3-1**(B and B). On the other hand, Lanes 2 to 4 represented BstZ*17I* digested and analysed DNA band patterns from healthy non-familial age and gender-matched controls, which were supposed to show two bands (740 and 345 bp) due to the absence of the mutation did not produce the expected results. Thus, the *FAM111B* Y621D POIKTMP-associated mutation validation with this PCR-RFLP method in our sample was inconclusive.

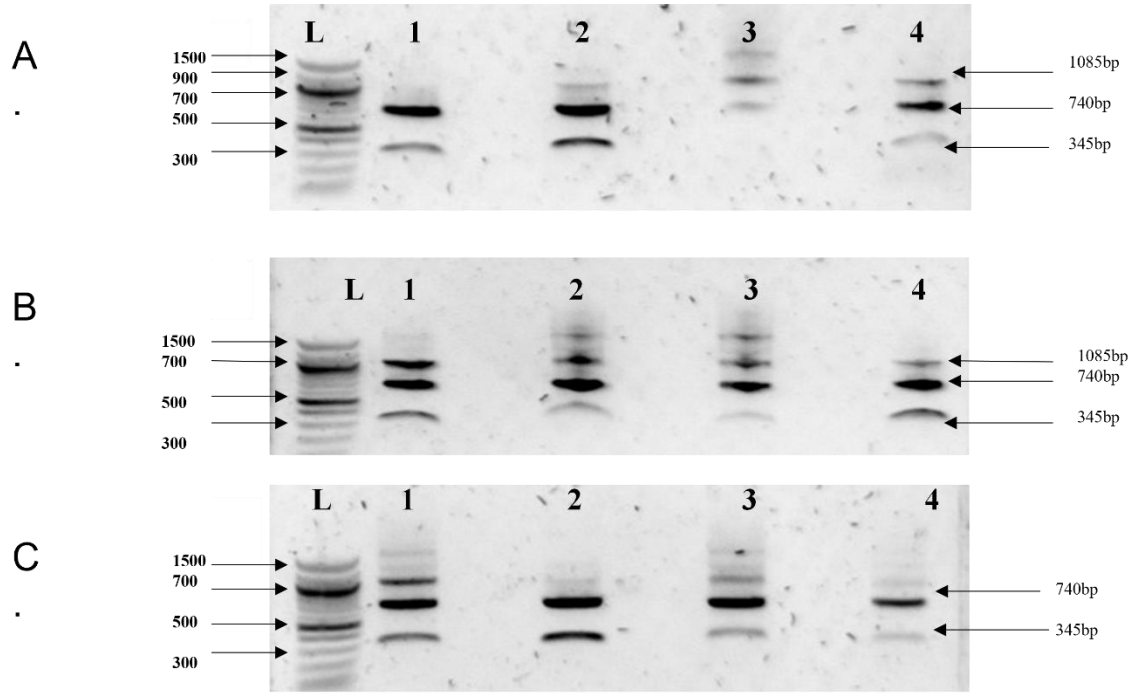


Figure 3-1: PCR-RFLP gene confirmation analysis of the Y621D mutation in the South African-associated POIKTMP patient. A) the skin tissue, **B)** the Lung tissue, and **C)** the skeletal muscle tissue consisting of the POIKTMP patient and familial controls. L = DNA ladder in base pairs (bp); Lane 1: Affected POIKTMP patient; Lane 2-4 Unaffected familial controls

3.1.2 Y621D mutation validation with Sanger Sequencing

Due to the inconsistency observed in the PCR-RFLP method, Sanger sequencing was used to accurately confirm the presence of the FAM111B Y621D gene mutation. An alignment of the POIKTMP patient's electropherogram with that of the controls confirmed the heterozygous T to G mutation of the 1861 nucleotide and the first nucleotide of the codon in the *FAM111B* protein coding sequence (i.e., TAC) that codes for tyrosine 621 amino acid (i.e., Y621) in the patient samples (highlighted in blue and indicated with black arrows). This change resulted in the missense mutation of this tyrosine amino acid to aspartic acid (D) (previously annotated as [NM_198947.4: c.1861T>G (p. Tyr621Asp or Y621D)]). This T to G mutation was present in tissue samples from the POIKTMP used in this study.

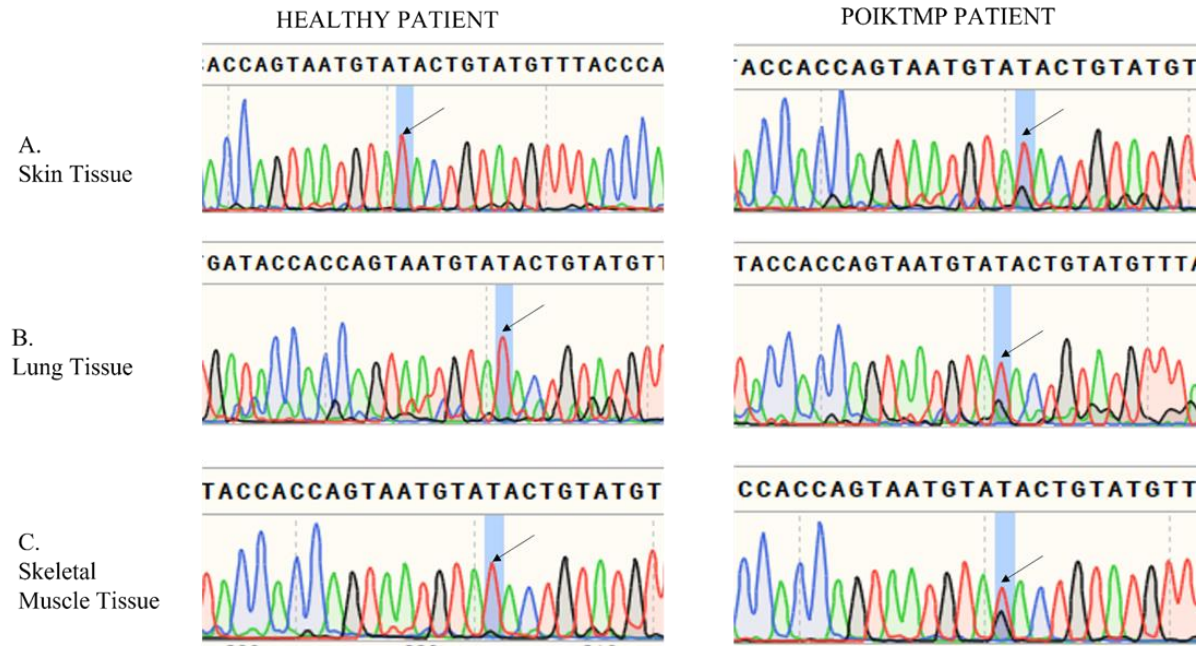


Figure 3-2: Y621D mutation of the FAM111B South African patient using Sanger sequencing. A) Skin Tissue. **B)** Lung Tissue**. C)** Skeletal Muscle Tissue**.** Sanger Sequencing technology allowed us to confirm the presence of the mutation in the POIKTMP patient tissues against the healthy familial controls. Differences in the presence of the *FAM111B* Y621D mutation were shown at the 621-nucleotide position, highlighted, and indicated by the black arrow.

3.2 *FAM111B* gene tissue expression and fibrosis pathway analysis

3.2.1 FAM111B gene expression by RT-qPCR

This study annotated the gene (mRNA) expression of *FAM111B* in the lungs, skin, and skeletal muscle tissue of a POIKTMP with the *FAM111B* Y621D mutation using RT-qPCR.The results show that *FAM111B* gene expression is about two-fold lower in the patient's skeletal muscle tissue than in the controls (**Figure 3-3**). *FAM111B* expression though appearing higher and lower in the POIKTMP skin and lung tissues, respectively, was not statistically significant.

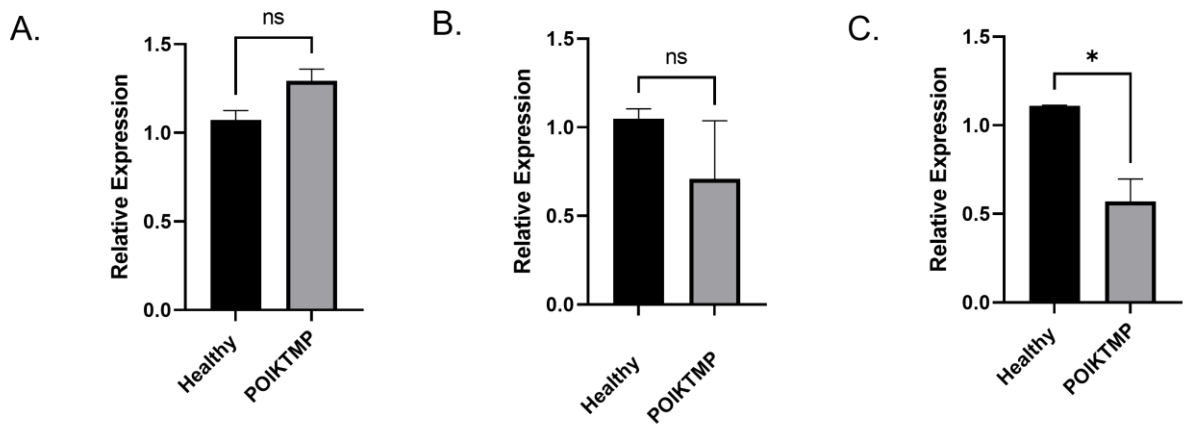


Figure 3-3: Relative FAM111B gene expression studies using RT-qPCR. Quantification was relative to healthy familial mRNA levels; all mRNA levels were normalized to GAPDH and ACTIN**. (A)** FAM11B relative gene expression in the skin tissue. **(B)** FAM111B relative gene expression in the lungs tissue. **(C)** FAM111B relative gene expression in the skeletal muscle tissue. Statistical analysis was performed using a t-test, ns (non-significant), and p-values were provided (*=$p \leq 0.05$)(n=3).

3.2.2 RT^2 Profiler PCR Array for Human Fibrosis

Besides accessing the changes in FAM111B mRNA expression in POIKTMP, the changes in the expression of key human fibrosis markers were investigated to identify the molecular pathways perturbed by this disease and its pathogenesis. This objective was achieved using the RT^2 Profiler PCR Array for Human fibrosis, which contains a panel of 84 pathway-focused fibrotic genes in the lung and skin POIKTMP tissue only.

The results showed several upregulated fibrotic genes in the lung and skin tissues (**Table 3-1** and **Table 3-2**, respectively, highlighted in red). For instance, in the POIKTMP lung tissue, fibrotic markers such as platelet-derived growth factor subunit A (PDGFA), Matrix metalloproteinase 3 and 13 (MMP3 & MMP13), Integrin beta-1(ITGB1), transforming growth factor beta-3 (TGFβ-3) and Connective tissue growth factor (CCN2) were upregulated (**Figure 3-4**). In the skin, Collagen alpha-type 111 (COL3A) and Thrombospondin 2 (THBS-2) were upregulated (**Figure 3-5**).

Table 3-1: The RT^2 Human fibrosis array indicating the upregulated fibrosis markers (highlighted in red) in POIKTMP lung tissue

	01	02	03	04	05	06	07	08	09	10	11	12
A	ACTA2 1.99	AGT 2.22	AKT1 -1.26	BCL2 5.37	BMP7 2.22	CAV1 -1.12	CCL11 12.27	CCL2 2.22	CCL3 3.9	CCR2 3.23	CEBPB 21.14	COLA2 -1.5
B	COL3A1 556.41	**CCN2 6667824**	CXCR4 -21476	DCN -40.09	EDN1 185.38	EGF 2.22	ENG -1.04	FASLG 2.22	GRM 1 2.22	HGF -1.32	IFNG 2.22	IL10 2.22
C	IL13 5.64	IL13RA 2 2.22	IL1A 2.22	IL1B 2.22	IL4 2.22	IL5 2.22	ILK 608818	INHBE 2.22	ITGA1 -1.17	ITGA2 -1.17	ITGA3 -7.1	ITGAV 1.03
D	**ITGB1 736247951**	ITGB3 -1.35	ITGB5 2.22	ITGB6 2.22	ITGB8 2.22	JUN 2.22	LOX -107054.28	LTBP1 2.22	MMP1 2.22	**MMP13 500203916.44**	MMP14 -1.99	MMP2 -18.92
E	**MMP3 509692.84**	MMP8 -4491.7	MMP9 2.22	MYC 8.0	NFKB1 -1.39	**PDGFA 14408506**	PDGFB 2.22	PLAT -4913238	PLAU 1.63	PLG 2.22	SERPINA1 2.22	SERPINE1 2363.91
F	SERPINH1 3.91	SMAD2 -3.74	SMAD3 2.22	SMAD4 245432.1	SMAD6 2.22	SMAD7 5876.11	SNAI1 2.22	SP1 2.22	STAT 1 1.11	STAT6 2.22	TGFB1 6.41	TGFB2 2.22
G	**TGFB3 40067362**	TGFBR1 -5.56	TGFBR2 -15.17	TGIF1 -1.14	THBS1 -45.38	THBS2 1.39	TIMP1 -6.23	TIMP2 -30.6	TIMP3 -16.0	TIMP4 2.66	TNF 2.22	VEGFA -24.9

Note: This table is a layout of the genes in reference to the heatmap in Figure 3-4 below. The values represent the log2 fold change in gene expression in POIKTMP lung tissues compared to the healthy control.

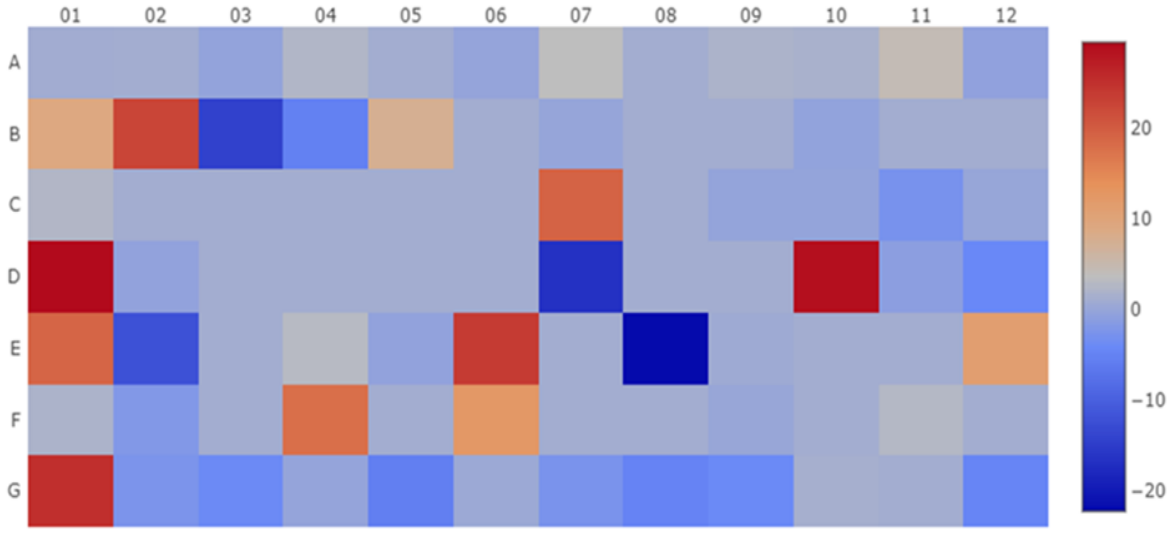


Figure 3-4: A heat map visualisation of the analysed data from the human fibrosis RT^2 Profiler PCR Array analysis report for POIKTMP lung tissues. Heatmap Analysis shows the differentially expressed (red = upregulated and blue = downregulated) fibrosis genes by the normalising the expression of every gene between the patient and healthy (Control group) against one to observe large gene expression changes. The cut-off p-value was 0.05 to assess significance.

Table 3-2: The RT2 Human fibrosis array indicating the upregulated fibrosis markers (highlighted in red) in POIKTMP skin tissue.

	01	02	03	04	05	06	07	08	09	10	11	12
A	ACTA2 -326987.21	AGT -1.24	AKT1 -1.18	BCL2 1.45	BMP7 -1.24	CAV1 -1.24	CCL11 -3.47	CCL2 -1.24	CCL3 -4.36	CCR2 -4.87	CEBPB -1.44	COL1A2 -1.24
B	**COL3A1 49882641.7**	CCN2 -9.66	CXCR4 1662.83	DCN -61.95	EDN1 -19.07	EGF 2156.6	ENG -1.24	FASLG -1.24	GREM1 -1.24	HGF 3269.63	IFNG -6.43	IL10 -1.15
C	IL13 2.31	IL13RA2 -3.85	IL1A -1.24	IL1B -1.24	IL4 -1.24	IL5 -1.24	ILK 789.62	INHBE -1.24	ITGA1 607.13	ITGA2 16897.21	ITGA3 -6.32	ITGAV -4.87
D	ITGB1 -604364.65	ITGB3 -4.88	ITGB5 -1.24	ITGB6 -1.24	ITGB8 2770.49	JUN -10.2	LOX -5.94	LTBP1 -1.24	MMP1 -1.24	MMP13 -9679.85	MMP14 -15.43	MMP2 -23.34
E	MMP3 -1.24	MMP8 -1.24	MMP9 -6.45	MYC 2.07	NFKB1 -1.24	PDGFA -1.24	PDGFB -88.27	PLAT -59744	PLAU -1.24	PLG -22.68	SERPINA1 -1.24	SERPINE1 3095.04
F	SERPINH1 -3.82	SMAD2 -1.24	SMAD3 -1.63	SMAD4 -1.24	SMAD6 -1.24	SMAD7 -1.24	SNAI1 -1.24	SP1 -1.24	STAT1 -3.87	STAT6 -1.24	TGFB1 -1.81	TGFB2 -21.57
G	TGFB3 -1.24	TGFBR1 -1.24	TGFBR2 75499.18	TGIF1 -1.24	THBS1 -52317.28	**THBS2 110666821.11**	TIMP1 -1.32	TIMP2 -1.76	TIMP3 -1.6	TIMP4 -1.24	TNF -1.24	VEGFA -1.14

Note: This table is a layout of the genes in reference to the heatmap in Figure 3-5 below. The values represent the log2 fold change in gene expression in POIKTMP skin tissues compared to the healthy control.

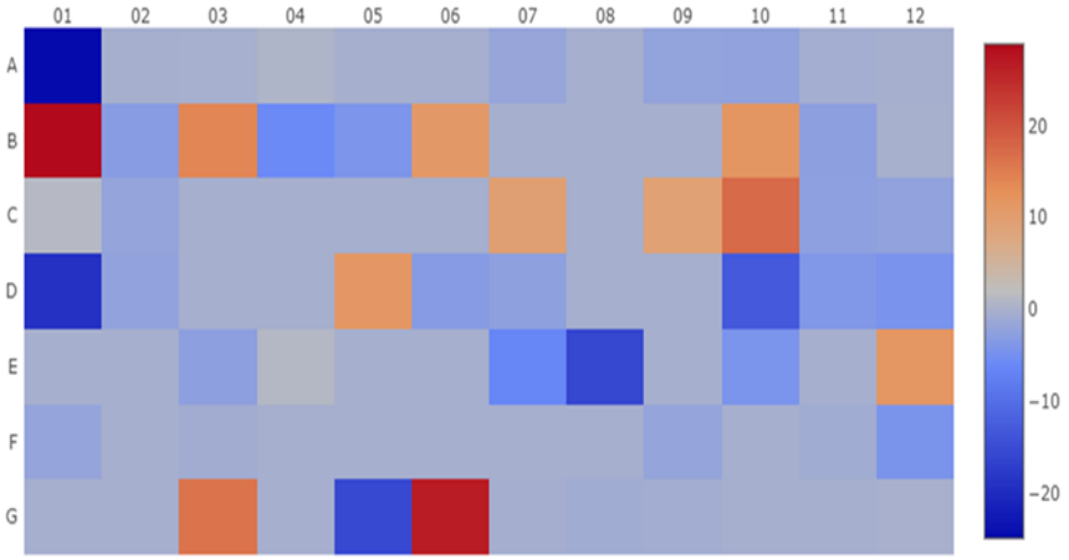


Figure 3-5. Heat map visualisation of the analysed data from the human fibrosis RT^2 Profiler PCR Array analysis report for POIKTMP skin tissues. Heatmap Analysis shows the differentially expressed (red =

upregulated and blue = downregulated) fibrosis genes by normalising the expression of every gene between the patient and healthy (Control group) against one to observe significant gene expression changes. The cut-off p-value was 0.05 to assess significance.

3.2.2.1 Validation of the top-upregulated fibrotic genes obtained from RT^2 Profiler PCR experiments

The expression of the top-regulated human fibrosis-related gene markers identified in POIKTMP, and skin tissues described in 3.2.2 were validated using RT-PCR based. As indicated by the RT^2 Profiler PCR array analysis report for Human fibrosis exported from GeneGlobe®, the genes: TGFβ-3, PDGFA, MMP13, MMP3, CCN2, and ITGB2 were highly expressed in the lung tissue of the POIKTMP patient compared to the healthy familial control (**Figure 3-6**). Likewise, COL3A1 and THBS2 were also significantly higher in the skin tissue of the POIKTMP patient than in the healthy control (**Figure 3-7**). This data indicates that the POIKTMP and possibly the FAM111B Y61D mutation affects the expression of the other fibrotic genes and consequently proposes other functions and pathways in association with FAM111B.

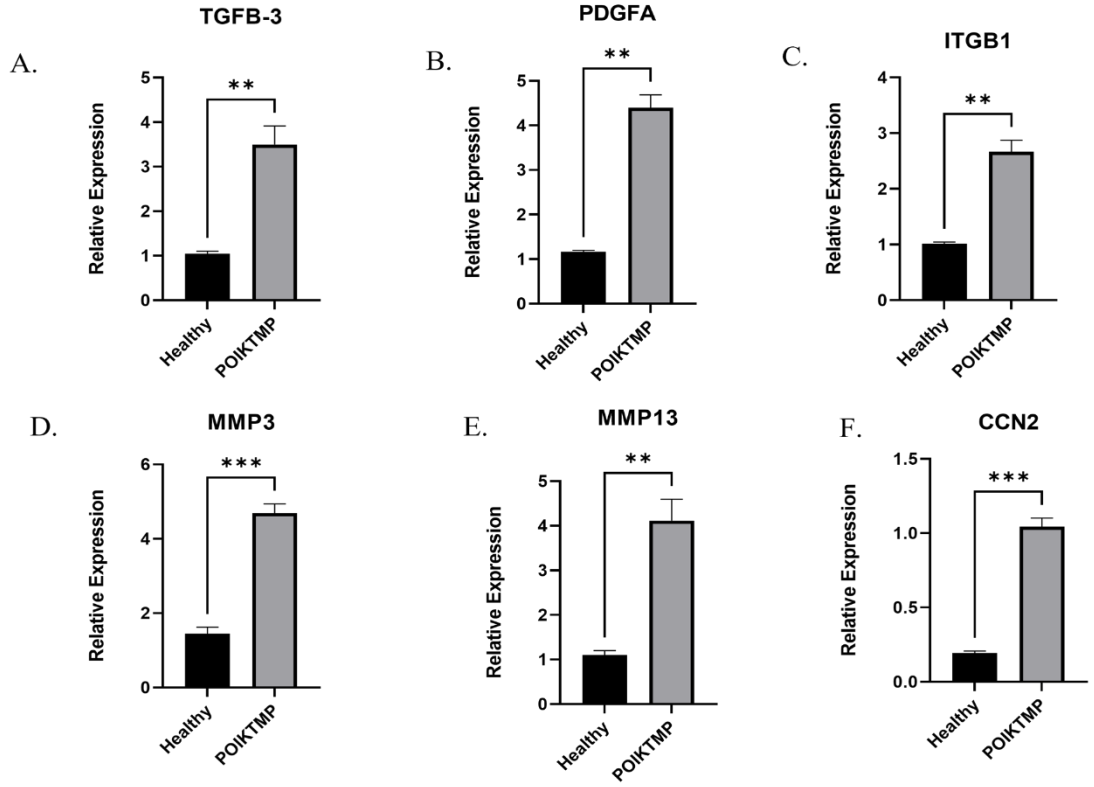


Figure 3-6: Lung tissue validation studies using RT-PCR. A) TGFB-3; **B)** PDGFA; **C)** ITGB1; **D)** MMP3; **E)** MMP13; **F)** CCN2. Relative expression was found to be higher in the upregulated genes of interest in the patient's tissue for all markers. Quantification was relative to the familial healthy control mRNA expression

levels normalized to GAPDH and ACTIN. All experiments were performed in triplicate and repeated three times with similar results. Statistical analysis was performed using t-test (* = $p \leq 0.05$; **= $p \leq 0.005$;***= $p \leq 0.0005$).

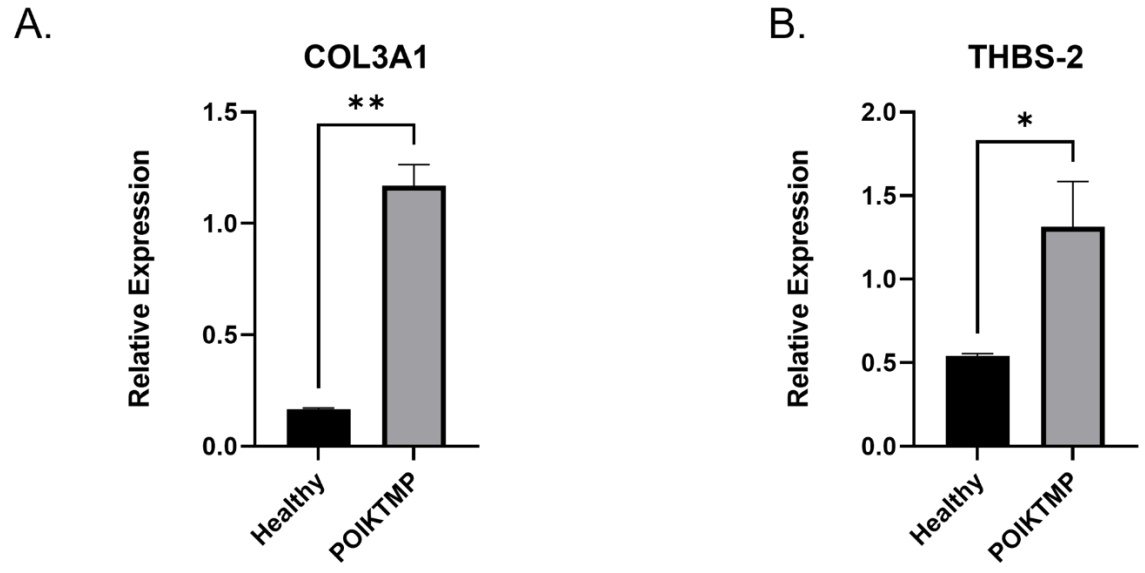


Figure 3-7: Skin tissue validation studies using RT-PCR. A) COL3A1; **B)** THBS-2. A Relative expression was higher in the upregulated genes of interest in the patient's tissue for all markers. Quantification was relative to the familial healthy control mRNA expression levels normalized to GAPDH and ACTIN. All experiments were performed in triplicate and repeated three times. Statistical analysis was performed using a t-test (* = $p \leq 0.05$; **= $p \leq 0.005$).

3.3 Gene-set enrichment analysis (GSEA)

The GSEA results obtained using the Enrichr tool on the gene-list, including the human *FAM111B* gene described in 2.4.2.2, are presented below based on the four categories (i.e., Pathways, Ontologies, Diseases/Drugs and Cell Types) of gene-set libraries interrogated.

3.3.1 GSEA with gene-set libraries in Pathways

The biological pathway query gene-list was used with two get-set libraries for this category of GSEA. The queried library investigated The KEGG 2021 Human and REACTOME 2022 pathways with enriched terms are presented in **Figure 3-8**. In the GSEA with KEGG 2021 Human pathway, the significantly enriched or upregulated terms based on the combined scoring method indicated by the top red horizontal bar graph, (**Figure3-8a** and **Table A-1**) consist of the terms involved in Interleukin-17 (IL-17) signalling pathways and prostate cancer. However, only five (MMP13, COL3A1, PDGFA, ITGB1 and MMP3) of the eight pro-fibrotic genes are in these pathway, and more interesting *FAM111B* was not associated with these terms (**Figure A-3a**). The enrichment analysis involving the REACTOME 2022

gene-set library found the MET interacts with TNS Proteins and Fibronectin Matrix formation terms to be significantly enriched amongst other terms shown in **Figure 3-8b and Table A-2**. Similarly, the previously mentioned five profibrotic genes, with the addition of CCN2 (but not FAM111B) were represented in these the processes (**Figure A-3b**).

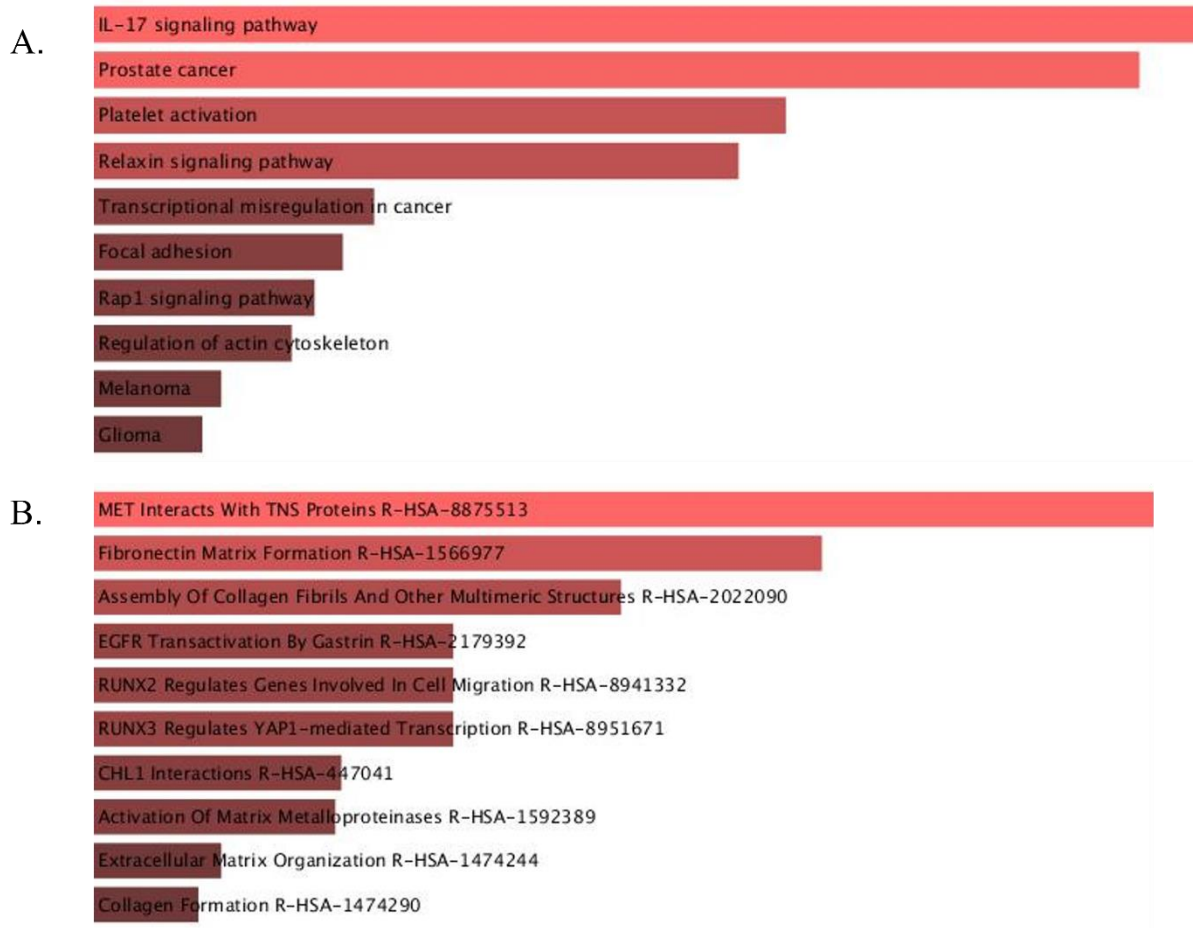


Figure 3-8: GSEA with gene set libraries in Biological Pathways. **A)** KEGG 2021 Human library enrichment analysis observed the Human cell signalling and metabolic pathways in a curated database of the gene-set list in correlation to FAM111B to establish common pathway. **B)** REACTOME 2022 biological pathway enrichment analysis to establish predicative models of the POIKTMP disease variant pathway by correlating the gene-set list in correlation to the FAM111B by associating common Homo sapiens pathways. (the long light red-coloured horizontal bars indicate the terms with the highest enrichment ranking score).

3.3.2 GSEA with gene-set libraries in Ontologies

In this category of GSEA, three get-set libraries were interrogated with our query gene-list, which are the gene ontologies (GO) for 1) Biological process, 2) Molecular Function, and 3) Cellular Component. The enrichment terms using these libraries are presented in **Figure 3-9.**

In the GSEA with the GO Biological process, the top five terms ranked the highest (i.e., significantly enriched or upregulated based on the combined scoring method and indicated by

the long length of the red horizontal bar graph, **Figure 3-9a and Table A-3**) included those involved in cell proliferation, neuroinflammatory response, cell-cell adhesion mediated by integrins, regulation of phosphatidylinositol synthetic process and regulation of inward rectifier potassium channel activity. Other key terms included regulation of hydrogen peroxide metabolic activity, negative regulation of neuron migration, positive regulation of oxidative stress-induced cell death and extracellular structure organization. However, only five (MMP13, COL3A1, PDGFA, ITGB1 and MMP3) of the eight pro-fibrotic genes are in these processes, and more interesting FAM111B was not associated with these terms (**Figure A-4a**).

In the analysis involving the Molecular Function get-set library only a term was significantly enriched which is the function involved in platelet-derived growth factor binding amongst other terms shown **Figure 3-9b and Table A-4**. Similarly, the previously mentioned five profibrotic genes (and again not FAM111B) represented in this process (**Figure A-4b**). Lastly, in the Cellular component gene-set library, only the cellular component for glial cell projection was the most upregulated among other components (**Figure 3-9c, Table A-5**). Only three (PDGFA, COL3A1 and ITGB1) genes are represented in this cellular component. (**Figure A-4c**).

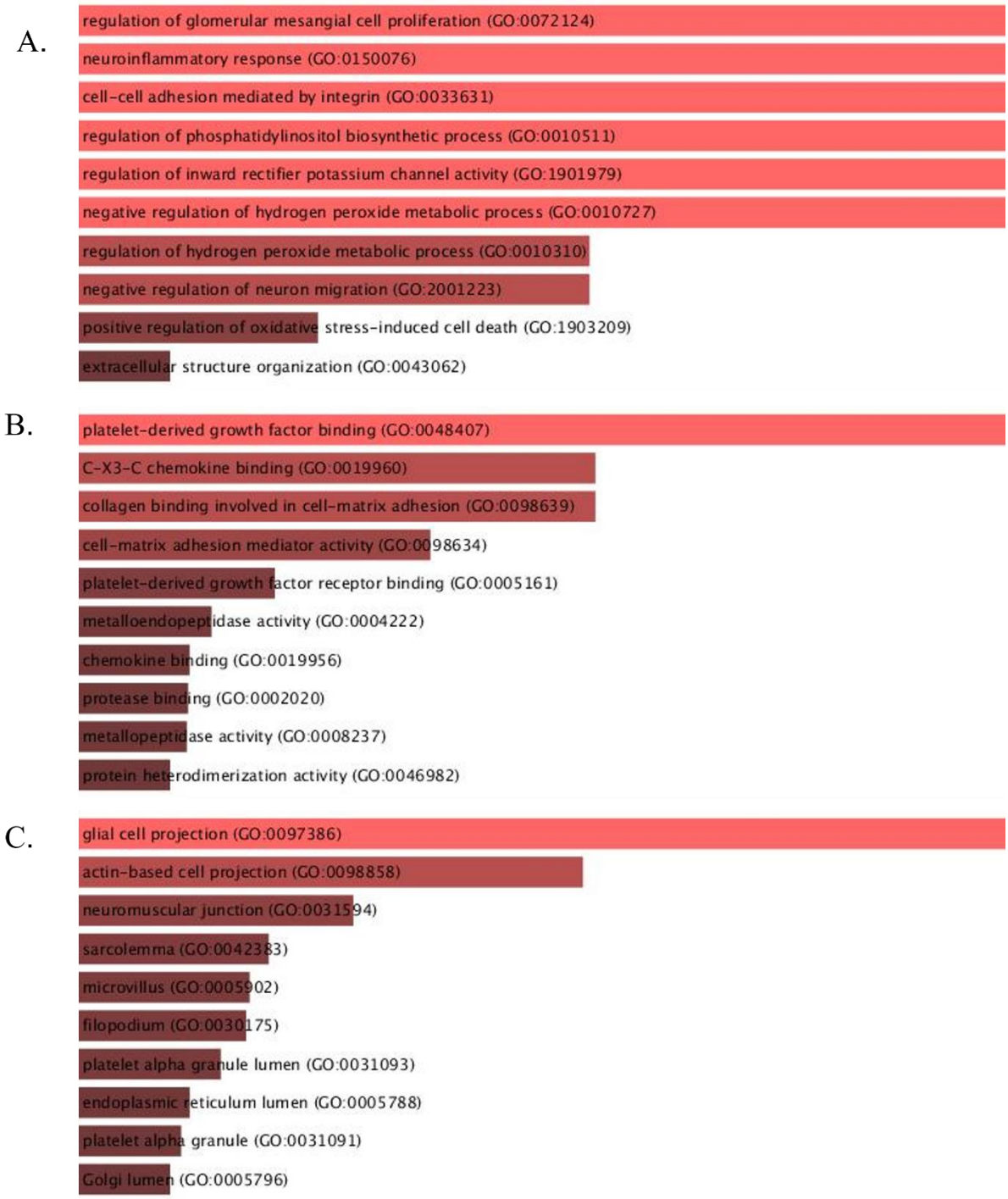


Figure 3-9: GSEA gene-set libraries in Ontologies. A) Biological processes observed. **B)** Molecular Functions observed in a POIKTMP patient. **C)** The cellular components associated in a POIKTMP patient with that of the 8 Human fibrosis markers observed and validated in the RT^2 Profiler PCR Array. (The chosen ontologies were based on the FAM111B gene, and the 8 gene markers were validated in the RT^2 Profiler PCR array for human fibrosis.

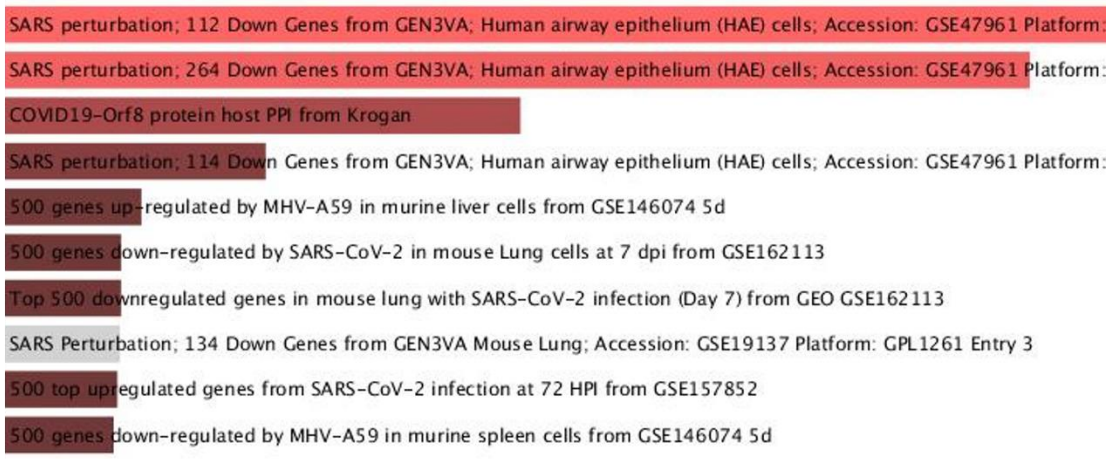
3.3.3 GSEA with gene-set libraries in Disease and Drugs

The disease and drugs category of the GSEA was investigated to query the annotated gene-set libraries of drugs and diseases that was associated with the gene list. Two libraries of interest were observed in **Figure 3-10**. The COVID-19 Related Gene sets 2021, which identified high enrichment in two terms ranked the highest based on the combined scoring method and indicated by the long red horizontal bar graph. The first being SARs-perturbation; 112 Down Genes from GEN3VA; Human airway epithelium cells and SARS perturbation and 264 Down Genes from GEN3VA; Human airway epithelium **(Figure 3-10a and Table A-6).** Subsequently, four (MMP3,MMP13, ITGB1 and COL3A1) pro-fibrotic genes were expressed and interestingly *FAM111B* was found to be associated with these terms **(Figure A-5a).** The other gene library set of interest was the GTEx Aging Signatures 2021, here aging signatures from various tissue were created from the Genotype-Tissue Expression database. It highlighted the significantly enriched terms involved in GTEx Small intestines 20-29 vs 40-49 Down, GTEx Breast 20-29 vs 50-59 Down and GTEx Esophagus 20-29 vs 40-49 Down presented in **Figure 3-10b and Table A-7**. Similarly, the profibrotic genes (MMP13, MMP3and CCN2) and *FAM111B* were enriched in the terms (**Figure A-5b**).

3.3.4 GSEA with gene-set libraries in Cell Types

Lastly, in **Figure 3-11**, the study hoped to look at GSEA Cell types categories of the gene-set libraries in human tissues. Interestingly, we observed high enrichment in three libraries. The GTEx Tissue Expression Down project highlighted two enriched terms, the GTEX-WYVS-0326-SM-SNM9V thyroid female 40-49 years and the GTEX-QLQ7-0726-SM-215G2 thyroid male 60-69 years **(Figure 3-11a and Table A-8).** These included four pro-fibrotic genes (MMP3, PDGFA, ITGB1 and COL3A1) and *FAM111B* to be enriched in the processes **(Figure A-6a).** Next, the Human Gene Atlas library investigated the Smooth muscle term to be highly enriched as indicated in **(Figure 3-11b and Table A-9)**. Similarly, the previous mentioned profibrotic genes and *FAM111B* was associated in these terms **(Figure A-6b)**. Lastly, the Descartes Human Cell types and Tissue analysis exhibited upregulation in the stromal cells in the muscle term of the gene-set library was found to be enriched (**Figure 3-11c and Table A-10**). Also, only the COL3A1 and *FAM111B* genes are represented in this enriched component. (**Figure A-6c**).

A.



B.

Figure 3-10: GSEA with gene-set libraries in Disease and Drug. A) COVID-19 Related Gene set 2021. **B)** GTEx Aging Signatures 2021. Drugs gene set enrichment analysis which identified *FAM111B* to be associated in drug related libraries and associated with other common diseases like Covid-19 and the Aging signatures from the GTEx project. (Enrichment Analysis was based on a combined score ranking).

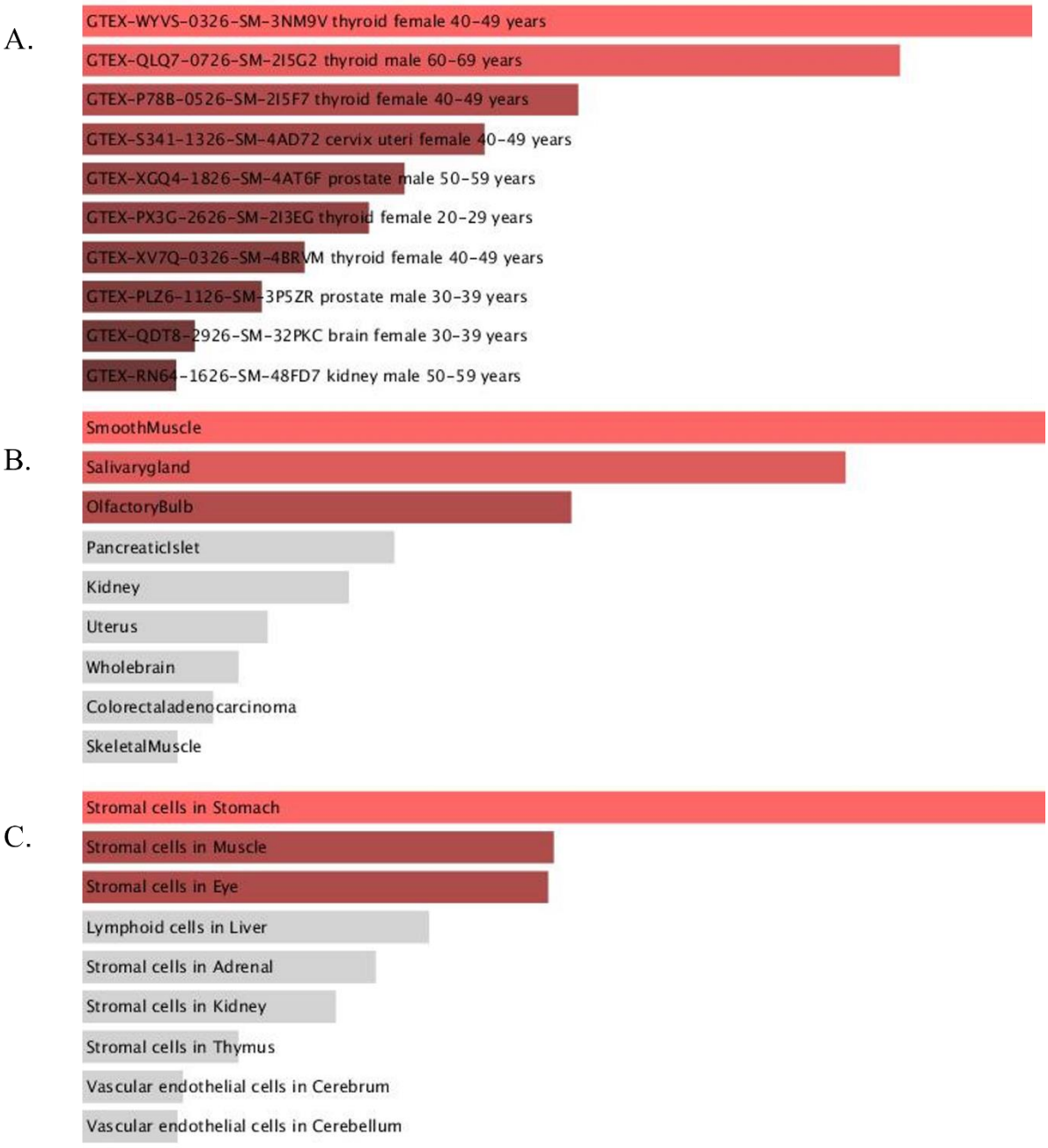


Figure 3-11: Cell types of gene set libraries associated with observed with FAM111B. A) upregulated GTEx Tissue expression signatures from the GTEx project 2021. **B)** Human Gene Atlas library of highly expressed genes in tissues. **C)**The Descartes Human Cell types and Tissues library which observes gene expression profiles used as a cell atlas in foetal gene expression. (Enrichment analysis was based on a combined score ranking.)

3.4 Tissue expression of FAM111B and fibroproliferative biomarker proteins

The changes in the tissue expression of FAM111B, TGFβ-1 (a fibrosis marker), ki-67 (nuclear proliferation marker) and Annexin V (a biomarker of early apoptosis) were accessed in POIKTMP-affected tissue by immunohistochemistry.

The immunohistochemistry analysis on the skin tissue in **Figure 3-12** results showed overall significant protein expression occurred in the patient's tissue. The FAM111B staining efficiency of the patient was 13% higher than that of the familial control at 8% **(Figure 3-12b**). The early apoptosis biomarker in the patient was 25% higher than the 9,5% of the familial control. Likewise, TGFβ-1 and ki-67 also resulted in a much higher expression in the patient, with 13% to 5% difference of the ki-67 and 4% in contrast to 1,7% for the TGF-β1respectively (**Figure 3-12b**).

The study then analysed the Lung tissue of the patient **Figure 3-13**, the Lung tissue displayed differences in morphology compared to the familial control in **Figure 3-13a**, based on the presence of the many air-filled alveoli in familial healthy control and the absence of the alveoli, displaced by connective tissue growths in the patient, indicated by the black arrow. The study found that *FAM111B* expression was much higher in the lung familial control tissues compared to the POIKTMP patient with a 7% difference, the fibrotic marker TGF-β1 also showed similar expression with a 4% difference. The patient tissue however, exhibited a higher protein expression with Annexin V and ki-67, compared to the healthy control with a difference of 17% for Annexin V and mere 2% for the ki-67 marker **(Figure 3-13b).**

Lastly, the study analysed the skeletal muscle tissue for differences in expression (**Figure 3-14**). Visual scans of the tissue appeared to indicate that the skeletal muscle tissue had a higher affinity for the *FAM111B* and Annexin V protein for POIKTMP patient tissue than for the familial healthy control (**Figure 3-14a**). To support this analysis, the staining efficiency in **Figure 3-14b** for *FAM111B* and Annexin V was much higher, with FAM111B showing a significant difference of 10% and Annexin V 24,5% difference in significance. The with TGF-1 and ki-67 appear to have minimal reductions in expression with little significant difference expressed (**Figure 3-14b).**

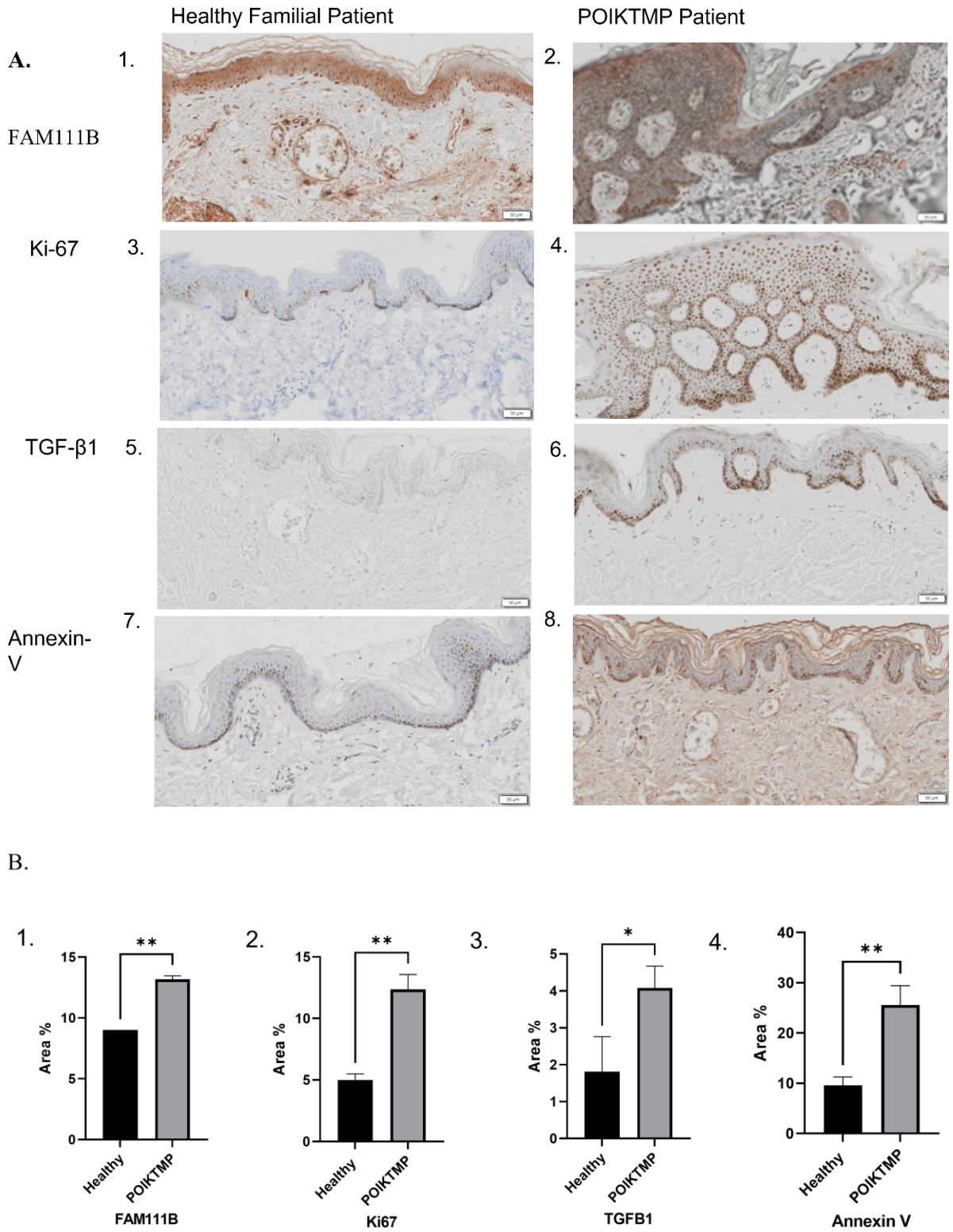


Figure 3-12: Skin Tissue protein expression Analysis. A) DAB staining in Skin tissue (Scale bar-50 µM). 1A-Healthy patient-FAM111B marker; 2A-POIKTMP patient-FAM111B marker; 3A-Healthy Familial patient-ki-67 Marker; 4A-POIKTMP patient-ki-67 marker; 5A-Healthy familial patient TGF-β1; 6A-POIKTMP patient TGFβ-1 marker; 7A-Healthy familial patient-Annexin V marker; 8A-POIKTMP-Annexin V marker. **B)** Target Tissue expression; 1B-FAM111B marker; 2B- ki-67 marker; 3B-TGFβ-1 marker; 4B-Annexin V marker. *Statistical analysis (*=$p \leq 0.05$, **=$p \leq 0.005$) of three independent experiments.*

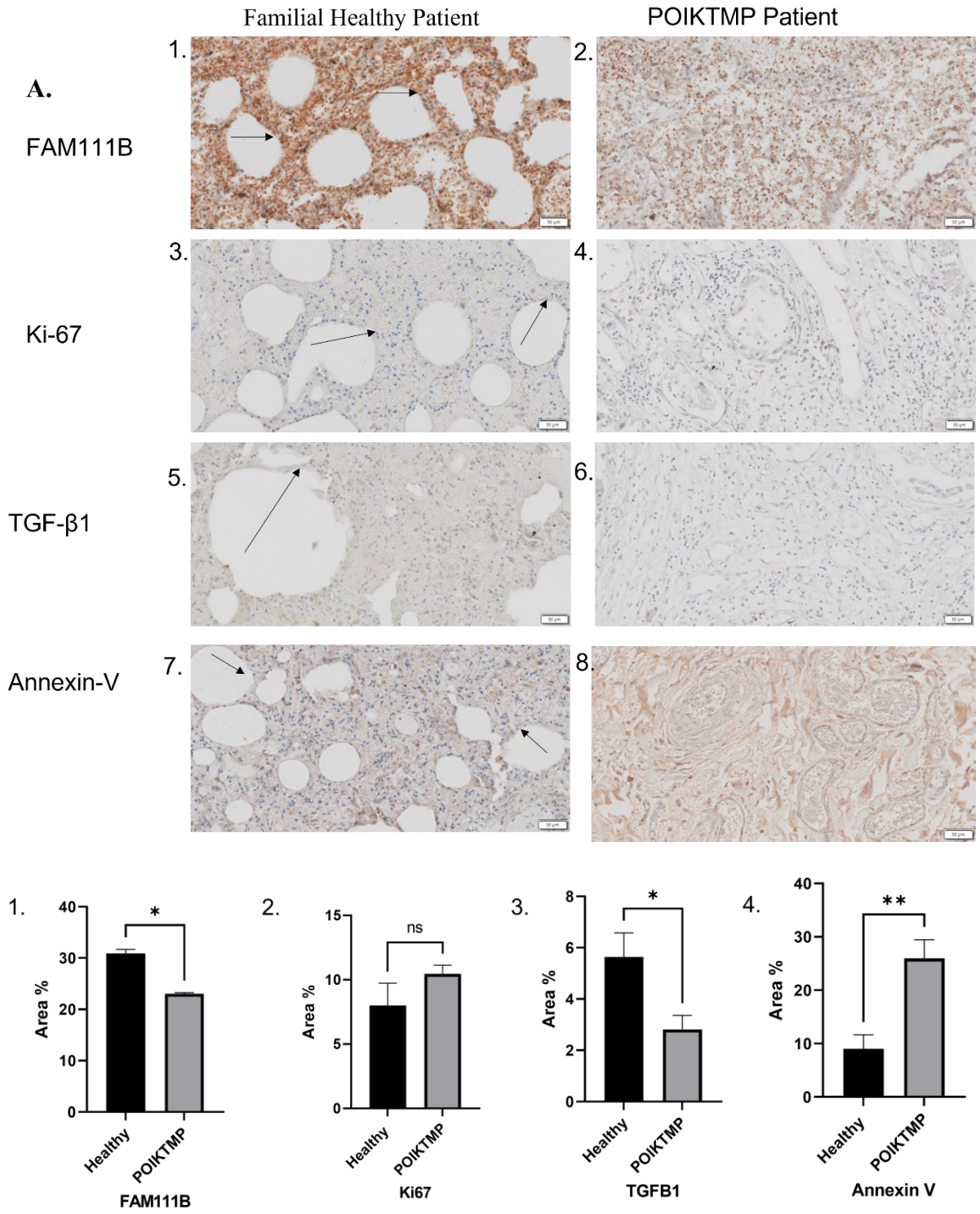
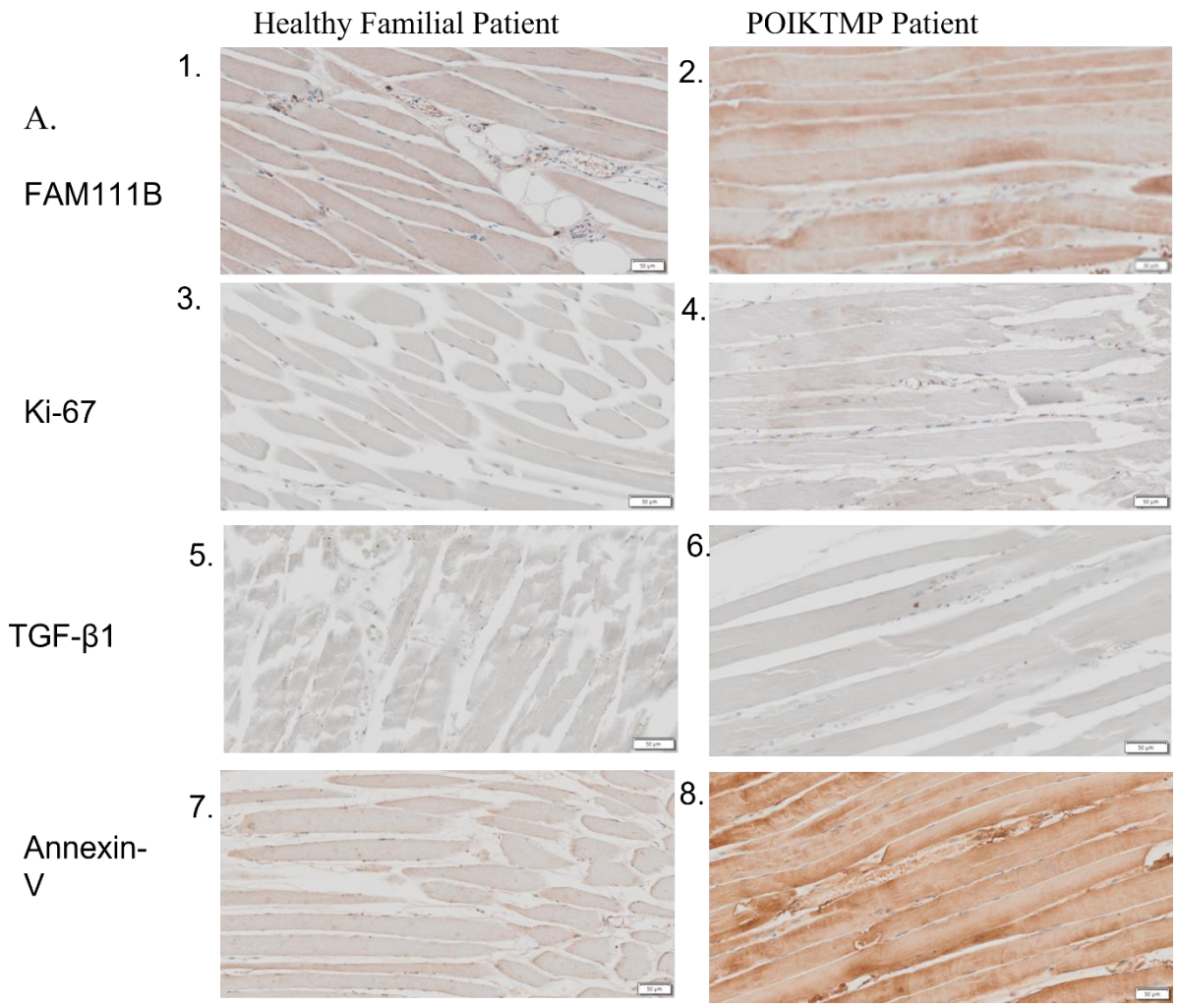


Figure 3-13: Lung Tissue protein expression. A) DAB staining in Lung tissue (Scale bar-50 µM). 1A-Healthy patient-FAM111B marker; 2A-POIKTMP patient-FAM111B marker; 3A-Healthy Familial patient-ki-67 Marker; 4A-POIKTMP patient-ki-67 marker; 5A-Healthy familial patient-TGF-β1; 6A-POIKTMP patient-TGFβ-1 marker; 7A-Healthy familial patient-Annexin V marker; 8A-POIKTMP patient-Annexin V marker. **B)Target Tissue expression**; 1B-FAM111B marker; 2B-Ki-67 marker; 3B-TGFβ-1 marker; 4B-Annexin V marker. *Statistical analysis (*=$p \leq 0.05$, **=$p \leq 0.005$) of three independent experiments.*



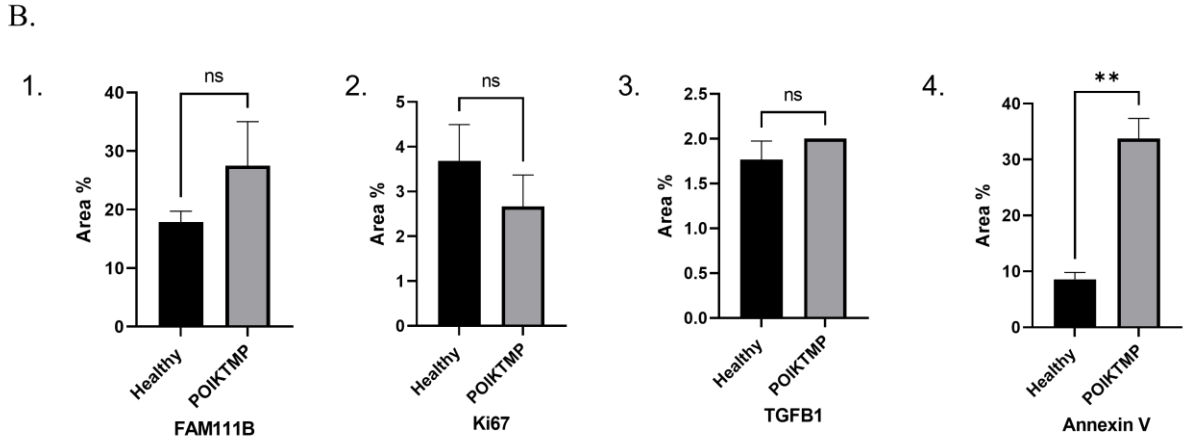


Figure 3-14 Skeletal Muscle tissue protein expression. A) DAB-staining in Skeletal muscle tissue (Scale bar-50 µM). 1A-Healthy patient-FAM111B marker; 2A-POIKTMP patient-FAM111B marker; 3A-Healthy Familial patient-ki-67 Marker; 4A-POIKTMP patient-ki-67 marker; 5A-Healthy familial patient-TGF-β1; 6A-POIKTMP patient-TGFβ-1 marker; 7A-Healthy familial patient-Annexin V marker; 8A-POIKTMP patient-Annexin V marker. **B)** Target Tissue expression; 1B-FAM111B marker; 2B-Ki-67 marker; 3B-TGFβ-1 marker; 4B-Annexin V marker. *Statistical analysis (*=$p \leq 0.05$, **=$p \leq 0.005$) of three independent experiments.*

Chapter 4

4 Discussions and Conclusions

4.1 Discussions

4.2 Confirmation of the *FAM111B Y621D* gene mutation

The study used tissue samples used resected from a post-mortem FFPE organ biopsy emanating from a member of the South African family first diagnosed with POIKTMP. The organ tissues, the lungs, skeletal muscle, and skin are the organs severely affected by the POIKTMP disease.[3,5] Furthermore, this family formed part of the cohorts whose DNA samples were sequenced by Whole-exome sequencing that identified mutations, including the Y621D mutation, in the *FAM111B* gene as the causative gene for the POIKTMP syndrome in a previous study.[3] Hence, the rationale for selecting to confirm the *FAM111B Y621D* mutation in our study samples is that this mutation is unique to this family. Also, this mutation segregated amongst the other family members affected by the disease in a heterozygous dominant manner. Moreover, verifying this mutation was essential to ensure that this mutation was not present in the *FAM111B* gene of the non-familial, gender and age-matched control organ tissue samples used in this study. Therefore, this study used a cost-effective PCR-RFLP method for detecting this *FAM111B* mutation described previously[82,83] and Sanger sequencing in our study samples.

4.2.1 Y621D mutation validation using PCR-RFLP

The study results demonstrated by PCR the presence of *FAM111B* in the patient and healthy tissues. Sanger sequencing is commonly used to confirm the presence of *FAM111B* in patients with POIKTMP. However, it is still a stark reality that this sequencing method remains inaccessible or unaffordable for genetic screening labs based in developing countries.[82] Based on this reality, this study sought to explore other methods of mutation screening/confirmation. Moreover, a previous study used this PCR-RFLP method to validate the *FAM111B Y621D* gene mutation using DNA isolated from skin-derived fibroblast of another living family member of the South African family affected by POIKTMP.[84] Therefore, this study attempted to use this PCR-RFLP technique for this mutation using DNA isolated from post-mortem FFPE tissues from the same family.

The results obtained were quite ambiguous, as it was challenging to verify differences between the patient and familial control. We postulate that perhaps the degradation of the

archival material could have had effects on DNA integrity because of the covalent protein-DNA bonds caused by the fixation procedure.[85] Moreover the FFPE tissue has been preserved for more than a decade.

4.2.2 Y621D mutation validation by Sanger sequencing

Sanger sequencing has been historically used as the standard method of detecting DNA mutations.[86] This technique can detect base substitutes, small insertions, and deletions.[86,87] DNA sequencing provides the most reliable method of identifying and confirming variations between species and genotypes. Moreover, this method has often proved successful with DNA samples from various sources including aged biological materials like archival FFPE tissues. Hence, given the inability to effectively validate the mutation of interest using PCR-RFLP, Sanger sequencing was used, which effectively identified the presence of the FAM*111B Y621D* gene mutation using the DNA isolated from the POIKTMP FFPE tissues. Therefore, highlighting the effectiveness of this gene mutation screening and validation method.

4.3 Gene expression studies

Gene expression profiling is effective in discovery analysis experiments like qRT-PCR, microarray, and RNA sequencing that allow the description of genome-wide expression changes in disease and healthy states. One such method, the RT^2 PCR Array Profiler, a technique that combines RT-PCR and microarray techniques are useful in providing a comparison of gene expression levels among conditions.[88,89] Since *FAM111B* mRNA expression has been reported in many tissues like the skin, skeletal muscle, and trachea and recent experimental studies have detected mRNA expression in lung cancer cell lines and skin fibroblast, albeit at a lower level.[3,83] The study sought to assess *FAM111B* expression in tissue types affected by POIKTMP and other key molecules of fibrosis.

4.3.1 Gene expression data using RT-qPCR for FAM111B expression

The RT-qPCR was used to measure *FAM111B* relative gene expression in the lungs, skeletal muscle, and skin tissues. *FAM111B* mRNA expression was shown to be downregulated in the POIKTMP-associated patient particularly the lungs and the skeletal muscle, notable expression in the skin tissue was not significantly expressed. Similar results were seen in the skin fibroblasts of POIKTMP patients' derived tissue, which showed a decrease in *FAM111B* expression.[84] In these studies, mutations in *FAM111B* may lead to induced cytotoxicity due to

an overexpression of the *FAM111B* gene product or upregulation of a serine protease-like domain in the protein.This information implies that the mutation's presence has a significant impact on the gene's function and level of expression. Also, the location of the *FAM111B* gene mutations has been proposed to contribute to disease severity, overexpression/downregulation and phenotypic variations.[3,5,6] Specifically, the South African *FAM111B* mutation, the Y621D mutation, is found to be within the putative protease domain (MWPPD) location and has reported being the cause of death patient's in a POIKTMP-associated case with pronounced pulmonary fibrosis.[6] This may support the hypothesis that *FAM111B* down-regulation leads to lung fibrosis and more advanced stages of the disease.

4.3.2 Gene expression of human fibrosis gene markers in POIKTMP tissues

To profile the gene expression of fibrotic markers in POIKTMP-affected tissues. This study made use of the RT^2 Profiler PCR Array Human Fibrosis system, which contains a panel of 84 fibrosis pathway-associated genes.[64] Since POIKTMP is a multisystem fibrosing disease implicated in fibrosis, the study made use of skin and lungs tissues as they are key organs affected by POIKTMP.[7] The study reported that out of the 84 pathway-focused fibrosis genes assessed, eight were upregulated in the POIKTMP patient-associated tissues. In the POIKTMP lung tissue, pro-fibrotic markers such as platelet-derived growth factor subunit A (PDGFA), Matrix metalloproteinase 3 and 13 (MMP3 & MMP13), Integrin beta-1(ITGB1), transforming growth factor beta-3 (TGFβ-3) and Connective tissue growth factor (CCN2) were upregulated. In the skin tissue, Collagen alpha-type 111 (COL3A1) and Thrombospondin 2 (THBS-2) were shown to be highly upregulated. The few key pro-fibrotic markers belong to categories of growth factors, genes involved in signal transduction, inflammatory cytokines and chemokines, as well as epithelial to mesenchymal transition (EMT) genes, ECM, and cell adhesion molecules of fibrosis.

A key pro-fibrotic mediator upregulated in this experiment is the TGFβ-3; a key marker shown to promote the formation of ECM fibroblast, myofibroblast differentiation, and its resistance to apoptosis. Overexpression of TGFβ-3 *in vivo* has been reported to cause severe fibrosis in experimental animal fibrosis models and human studies of end-stage tissue fibrosis despite the involvement of several complex signalling pathways. Also, previous *FAM111 B*-related studies have implicated TGF-β1 to be associated with POIKTMP,[2] which is also analysed in the current protein expression study.

ECM remodelling enzymes and cell adhesion molecules MMP3 and MMP13 were identified to be highly regulated in the patient lung tissue in this study. Several types of MMP structures have been identified as degraders of a variety of substrates, including ECM components such as collagens, elastin, proteoglycans, and lamins.[90,91] The multiple biological roles of MMPs suggest that they play a role in a wide range of pathological processes, including rheumatoid arthritis, tumour progression, kidney fibrosis, liver fibrosis, heart failure, and lung fibrosis.[91,92] Research suggests that MMP3 and MMP13 contribute to the pathogenesis of pulmonary fibrosis, both in humans and in animal models.[91] Studies have attempted to associate MMP3, 13 and TGF-β in proposed signalling pathways to contribute to end-stage tissue fibrosis as TGF-β has been supported in literature to be established in various fibrosis experimental models.[92] Observations have suggested that MMP-3 specifically may exert a pro-fibrotic response through the TGF-β signalling pathway. Interestingly, this study has observed all these mediators which could support their role in pulmonary fibrotic diseases.

The CCN2 and PDGFA expression was also found to be upregulated in the POIKTMP patient lung tissue, these mediators have both been demonstrated to stimulate fibroblast proliferation *in vitro*. The overexpression of CCN2 *in vivo* induces mild fibrosis in the bleomycin model,[93] which has been extensively studied. Subsequently, PDGFA is upregulated in animals with fibrosis. In association with TGF-β, crosstalk between the two mediators has been proposed to cause fibroblast proliferation.[94] Interestingly, this association between the two growth factors enhances apoptosis in epithelial cells and has been reported to resist apoptosis in mesenchymal cells.

Lastly, COL3A1 and THBS-2 showed high expression in skin POIKTMP-associated patients; both are thought to be ECM molecules, specifically non-structural matricellular proteins like THBS-2.[95] Systemic sclerosis (SSc) fibrosis is influenced profoundly by dysregulated processes that promote fibrosis. SSc fibrosis is a rare chronic heterogeneous disease characterized by inflammation and vasculopathy and has also been associated with THBS-2. As with *FAM111B* in POIKTMP, this condition leads to multisystem tissue fibrosis during the second decade of the life of affected persons, as disease severity progresses. Healthy and fibrotic lungs contain a balanced amount of collagen, a component of lung connective tissue. Idiopathic pulmonary fibrosis, however, is characterized by excessive COL3A1 production, which is facilitated by the presence of myofibroblasts.[96]

By highlighting these key genes, this data supports the association with *FAM111B* with TGF-β, along with other mediators of fibrosis. Since most studies have found *FAM111B* to be downregulated, the current studies focused on the highly expressed genes of interest out of the 84 panel-focused fibrotic genes. Notably, multiple genes were found to be downregulated in the POIKTMP patient samples, however, due to infrastructures, cost-effectiveness and other factors, the study was unable to validate and analyse the downregulated genes, as this might have expanded our knowledge on POIKTMP mediators. Under fibrotic conditions, the upregulated genes analysed are known to influence differentiation, recruitment, synthesis of collagen, and apoptosis evasion in myofibroblasts. In our study, it is demonstrated that the pro-fibrotic markers do more than just affect the ECM turnover during fibrosis; rather, they affect cellular proliferation, survival, gene expression, and inflammatory responses. As a consequence, *FAM111B* can affect myofibroblast activation in POIKTMP-diseased patients, thereby supporting its involvement in fibrosis both directly and indirectly.

4.3.3 Gene-set Enrichment Analysis of Human Fibrosis in association with *FAM111B*

4.3.3.1 GSEA in Biological pathways

Gene-set enrichment analysis (GSEA) was used in this study to further elucidate the association between *FAM111B,* and the list of upregulated genes obtained from the RT^2 Profiler PCR Array Human fibrosis analysis by applying enrichment analysis to annotate the gene sets. GSEA studies observed different categories, specifically the biological pathway; Gene ontologies; Diseases and drugs analysis and lastly the GSEA in Cell types.

The KEGG 2012 Human and REACTOME 2020 biological pathway analysis predicted several associated pathways to be highly enriched in the study gene set, such as prostate cancer, IL-17 signalling pathway, and MET interactions with TNS proteins pathways. Notably, *FAM111B* did not appear to be directly enriched in these libraries, as only the pro-fibrotic terms were enriched in these terms. It can be argued that *FAM111B* contributes to prostate cancer susceptibility through multiple single nucleotide polymorphisms(SNPs) located on 11q12, which is in the same chromosomal position as *FAM111B* and other FAM111 member genes.[7,97] Additionally, the *FAM111B* has been linked to several cancers including pancreatic, breast and cervical cancers.[7,98,99] Moreover, evidence in lung adenocarcinoma patients of *FAM111B* suggests its strong correlation with tumour progression and poor survival rates.[41,42] Also, prostate cancer usually occurs at a late stage,[100] and like cervical cancer is associated with distal progression and metastasis,[7] and a second

possibility is that the cancer cells with *FAM111B*-repair machinery gain an advantage at later stages, thereby promoting cancer development.[7]

IL-17 is a component of cytokines and growth factors linked to the fibrosis of various organs and is known to play a role in the inflammation brought on by fibrosis.[101] A host defence mechanism against extracellular microorganisms that causes persistent inflammation in autoimmune disorders is carried out by this cytokine family. IL-17 is also essential for preserving the integrity of the epithelial layer and promoting wound healing. As a pro-inflammatory cytokine, IL-17 is connected to numerous fibrosis in most of the body, like skin and lung fibrosis in various disease conditions.[102] Lastly, another pathway enriched in the study is the MET interacts with TNS proteins pathway, few studies have explored this pathway, however, it's been associated to promote cell motility, and direct interaction influences MET stability and regulate survival of MET-dependent carcinomas, thus implications of cancer susceptibility[103].

4.3.3.2 GSEA in ontologies

Ontology analysis of our gene set analysed biological processes, molecular function and cellular components implicated in the study data set. Once more, *FAM111B* did not appear in the enriched terms of the ontology libraries, the study records only pro-fibrotic genes to be enriched. Nevertheless, the biological processes library of the gene set established several enriched terms such as regulation of glomerular mesangial cell proliferation, neuroinflammatory response, cell-cell adhesion mediated by integrins and regulation of phosphatidylinositol synthetic process. Other key terms include the regulation of inward rectifier potassium channel activity and the regulation of hydrogen peroxide metabolic activity.

In Metabolic functions, the enriched term was the platelet-derived growth factor binding with PDGF subunits serving as a fibrosis mediator and in the cellular components saw the glial cell projection was highly enriched. As previously discussed PDGF is a common mediator of fibrosis, therefore it's not surprising that we see its appearance. The glial cell is key effector cell in neurological disorders, as it contributes to fibrotic scarring in response to injuries.[104]

FAM111B functional studies have associated the gene with apoptosis with concrete evidence of involvement in apoptosis and DNA repair pathways mediated by p53's direct or indirect downstream activation of *FAM111B*.[7,37,41,105] It has been found that the *FAM111B* mutation might cause epithelial apoptosis, which makes POIKTMP-affected fibroblasts resistant to

apoptosis.[7] This postulates biological processes, and metabolic functional and cellular components of *FAM111B* to be involved in DNA repair, genome instability, chronic inflammation, and unusual apoptosis of the epithelial cells and fibroblasts of the tissues and organs affected by this disease.[7]

4.3.3.3 GSEA in Diseases and Drugs

The gene-set library category in diseases and drugs reported the input gene library to be enriched in COVID-19-related gene set 2021, in this library the enriched terms were SARS perturbation, 112 and 262 down genes from GEN3VA.These were enriched based on high combined scoring. Interestingly, the *FAM111B* gene appeared to be a commonly enriched gene in this library, overlapping with the other fibrotic genes. This can be expected as SARS-CoVID-19 is a respiratory disease, with observed lung infections reported. Additionally, in the GTEx Aging signatures library, enriched terms were associated with small intestines in the 20-29 vs 40-49 range; Breasts of 20-29 vs 50-59 age range then finally in the Esophagus in the 20-29 vs 30-39 age range. The GTEx lung tissue term saw *FAM111B* overlapping with common fibrotic genes. *FAM111B* has been associated to be an age-related disease, studies have established the gene to manifest in the early months of an affected individual, with phenotypes severity intensifying in the second decade of the affected individual.[3,5,35] Moreover, *FAM111B* has been reported to be expressed in the progression and development of breast cancer.[98]

4.3.3.4 GSEA in Cell types

The results in this category explored libraries analysed based on *FAM111B* overlap with pro-fibrotic genes in the gene-set library. The GTEx Tissue Expression downregulated library, here high enrichment based on score ranking was detected in male and female thyroid-associated cells. However, the *FAM111B* association overlapped in the female brain cells, which would explain the glial cell projection observed in the metabolic function category of the ontologies. The Human gene atlas 2021 library observed high enrichment in smooth muscle cells, and once more *FAM111B* enrichment was detected in the salivary gland cells. Lastly, the Descartes Human Cell types library expressed high enrichment in stromal cells in the stomach, *FAM111B* however, was detected in the lymphoid cells of the liver, this is not surprising, as *FAM111B* is associated to cause liver disease and other liver involvement.[34]

Findings in the Gene-set enrichment analysis were quite informative in establishing and confirming the proposed characteristics of *FAM111B*. While *FAM111B* was not directly

highlighted in the pathway and ontologies category, its association with prostate cancer and biological association in proliferation and differentiation. The enrichment analysis was also effective in establishing the direct involvement of *FAM111B* in lung and liver cancer association, as well as pulmonary fibrosis disease based on its involvement in SARS-COVID-19 diseases and Aging signatures libraries.

4.4 Tissue expression of *FAM111B* and fibroproliferative biomarker proteins

4.4.1 Immunohistochemistry

Immunohistochemistry studies to access the protein expression of fibroproliferative specific markers was carried out in POIKTMP and control tissues. The markers were primarily the *FAM111B* marker, which is the gene of interest and the causative gene for the POIKTMP disease. To analyse cell proliferation and mitotic activity in tissue samples, ki-67 a nuclear marker associated with cellular proliferation was used to detect cellular activity in the respective tissues. For the detection of cell death and degeneration, the tissues were subjected to an early apoptosis marker, Annexin V. Lastly, as POIKTMP disease is a multisystem fibrotic disease, the study explored the role of TGF-β which is a mediator of fibrogenesis.

Since *FAM111B* relative gene expression was found to be downregulated at the mRNA level using RT-PCR analysis, experimental analysis was further extended by immunohistochemistry analysis to assess FAM111B protein in the lung, skin and skeletal FFPE sections. The immunohistochemistry findings confirmed decreased expression in the POIKTMP-associated patient in the lungs where inflammation and fibrosis are prominent in comparison with the control samples. The skin and skeletal muscle demonstrated upregulation of *FAM111B* in the patient tissue than the control. Findings in immunohistochemistry corroborated what was observed in gene expression studies using RT-PCR.

The ki-67 is a marker that is shown to be expressed during cell proliferation and can be detected in all the stages of the cell cycle excluding the G_0 phase. Thus, this biomarker is well-suited to elucidate the tissue alteration processes in fibrosis.[106] Recent research indicates that the *FAM111B* mutation is connected to the control of proliferation in POIKTMP, however, according to research, the proliferation rate was found to be reduced in cells deficient in *FAM111B* and Y621D mutant cells, but not in cells that overexpressed *FAM111B*.[84] This supports the observed results for the ki-67 protein expression which was non-significant in the lung and skeletal muscle tissue of the patient. However, the skin tissue

observed a high protein expression for the POIKTMP patient. Previous studies have been conducted using immunohistochemistry to investigate the association between the expression pattern of *FAM111B* and ki-67 which revealed a significant association in LUAD.[42] They were able to conclude that although expression was associated, other factors influence this correlation, such as tumour size, T classification, frequency of lymphovascular invasion and N classification.[41,42] The present study used FFPE tissue, and we now understand that POIKTMP and *FAM111B* seem more critical in some cell types than others, which would explain the proliferation activity observed in the skin tissue.

The TGF-β1 was also studied, as a regulator of fibrosis; our study aimed at correlating its expression based on the high expression seen by the RT^2 Profiler PCR Array analysis. The marker was found to be non-significantly expressed in the lung and skeletal tissues. However, the skin tissue demonstrated up-regulated expression of TGF-β1. No statistical difference in staining intensity was observed between ki-67 and TGFβ-1 in the patient's lung and skeletal samples. On the contrary, TGFβ-1 and ki-67 were almost absent within the POIKTMP patient lung, including areas of active fibrosis. Moreover, it seems TGF-β1, seems to be down-regulated in POIKTMP, as it was observed in the RT^2 Profiler PCR Array, which saw TGF-β3 to be upregulated and TGFβ-1 downregulated. Given that fibrosis is frequently characterized by the activation of numerous profibrotic pathways, a multi-pathway extended strategy may explain other mediators involved in the advancement of fibrosis.

It is important to acknowledge that an array of stimuli can trigger apoptosis, and not all of them trigger the same type of programmed cell death. Various diseases including fibrosis have been linked to dysregulation of apoptosis.[107] Researchers found that epithelial apoptosis played an important role in pulmonary fibrosis. In contrast, apoptosis of myofibroblasts was inhibited in the lungs of IPF patients. These findings suggest a link between dysregulated apoptosis and the onset of pulmonary fibrosis.[107] These findings are consistent with the present study findings, which showed that the POIKTMP patient tissue showed an increased rate of apoptosis. Accordingly, as previously mentioned, it is plausible that the *FAM111B* mutations' reduced production or quick degradation of *FAM111B* could increase epithelial apoptosis and, as a result, in the resistance to apoptosis of nearby fibroblasts in tissues impacted by POIKTMP.[7] Consistent with this information, the mutation's existence prevented the tissues' ability to undergo apoptosis, which accelerated the fibrosis process.

4.5 Conclusion

This study sought to extend the knowledge on the role of the human *FAM111B* gene and its mutation in fibrosis, specifically POIKTMP, by accessing the expression profiles of *FAM111B* and known human fibrosis markers in the organ/tissues affected by the disease. From the results shown in our experimental study, the study was able to confirm the presence of the *FAM111B* Y621D mutation in the tissues of a deceased family member of the South African family first diagnosed with POIKTMP using Sanger sequencing. In line with the confirmation results, the *FAM111B* relative gene expression levels were found downregulated in the lungs and skeletal patients compared to familial healthy subjects and upregulated in the skin tissue of the POIKTMP patient. The study concludes that *FAM111B* could be downregulated in more advanced forms of pulmonary fibrosis and its gene expression tissue specific.

Furthermore, this study analysed the association between *FAM111B* and several fibrotic pathways-associated genes and found some pro-fibrotic related genes: PDGFA, COL3A1, CCN2, MMP3, MMP13, THBS-2, ITGB1 and TGF-β3 upregulated in POIKTMP. These genes were then used to create a gene-set list and inputted on Enrichr for Gene-set enrichment analysis (GSEA). GSEA in the categories of Disease and Drugs, as well as GSEA with Cell types analysis, observed *FAM111B* enrichment in the terms. The libraries associated with SARS-COVID-19 and GTEx Aging signatures libraries reported *FAM11B* to be expressed in these diseases. Similarly, in the Cell types category *FAM111B* enrichment was reported to appear in libraries for GTEx Tissue Expression downregulated; the Human Gene Atlas and the Descartes Cell types and tissue 2021 libraries. Among these, liver, lung, and breast tissues were tissues/organs associated with *FAM111B* enrichment.

Finally, to support the premise that *FAM111B* downregulation may contribute to lung fibrosis and lead to more progressive disease stages, this study demonstrated that *FAM111B* protein expression levels correlate with the gene expression data, which observed *FAM111B* to be downregulated in the lungs than the skin tissue. Proliferation and key pro-fibrotic marker Ki-67and TGF-β1 was assessed, and the findings suggest the skin tissue to be associated with proliferation activity, as well as pro-fibrotic responses. Overall, apoptosis was highly associated with POIKTMP patient tissues.

Collectively this study's data demonstrates for the first time, the downregulation of *FAM111B* in different forms of POIKTMP patient-associated tissues and highlights upregulated key fibrotic-specific markers which could in future help delineate the molecular mechanisms behinds *FAM111B* gene or its mutation leads to POIKTMP and perhaps fibrosis in general. Therefore, mutations in the *FAM111B* gene may represent a useful biomarker for POIKTMP and other fibrotic disorders, and more so a viable therapeutic targets in treating these diseases. Hence, *FAM111B* mutations, in conjunction with experimental data generated by cutting-edge functional genomic technologies, will pave the way for uncovering the functional mechanism of *FAM111B*.

Our data exhibit several limitations that should be treated with caution. First, based on our findings, it is somewhat ambiguous whether *FAM111B* down-regulation within the fibrotic lung is a primary event or just a consequence of the fibrogenic cascade. Second, it is imperative to consider the downregulated fibrotic genes identified in this study' s array experiments and the role they may play in POIKTMP and other fibrotic diseases. Nonetheless, our study represents the first attempt to correlate POIKTMP within the human fibrosis pathway and to associate this pathway with *FAM111B* dysregulation.

5 References

1 Wynn, T. A. & Ramalingam, T. R. Mechanisms of fibrosis: therapeutic translation for fibrotic disease. *Nat Med* **18**, 1028-1040 (2012). https://doi.org:10.1038/nm.2807

2 Khumalo, N. P. *et al.* Poikiloderma, tendon contracture and pulmonary fibrosis: a new autosomal dominant syndrome? *Br J Dermatol* **155**, 1057-1061 (2006). https://doi.org:10.1111/j.1365-2133.2006.07473.x

3 Mercier, S. *et al.* Mutations in FAM111B cause hereditary fibrosing poikiloderma with tendon contracture, myopathy, and pulmonary fibrosis. *Am J Hum Genet* **93**, 1100-1107 (2013). https://doi.org:10.1016/j.ajhg.2013.10.013

4 <1013-Article Text-3506-1-10-20090630.pdf>.

5 Mercier, S. *et al.* 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**, 1-16 (2015).

6 Arowolo, A., Rhoda, C. & Khumalo, N. Mutations within the putative protease domain of the human FAM111B gene may predict disease severity and poor prognosis: A review of POIKTMP cases. *Exp Dermatol* **31**, 648-654 (2022). https://doi.org:10.1111/exd.14537

7 Arowolo, A., Malebana, M., Sunda, F. & Rhoda, C. Proposed Cellular Function of the Human FAM111B Protein and Dysregulation in Fibrosis and Cancer. *Frontiers in oncology* **12** (2022).

8 Hoeger, P. H., Koehler, L. M., Reipschlaeger, M. & Mercier, S. Hereditary fibrosing poikiloderma (POIKTMP syndrome) report of a new mutation and review of the literature. *Pediatric Dermatology* (2022).

9 Wynn, T. A. Common and unique mechanisms regulate fibrosis in various fibroproliferative diseases. *J Clin Invest* **117**, 524-529 (2007). https://doi.org:10.1172/JCI31487

10 Wynn, T. A. Cellular and molecular mechanisms of fibrosis. *J Pathol* **214**, 199-210 (2008). https://doi.org:10.1002/path.2277

11 Huang, E. *et al.* The roles of immune cells in the pathogenesis of fibrosis. *International journal of molecular sciences* **21**, 5203 (2020).

12 Baues, M. *et al.* Fibrosis imaging: Current concepts and future directions. *Adv Drug Deliv Rev* **121**, 9-26 (2017). https://doi.org:10.1016/j.addr.2017.10.013

13 Rosenbloom, J., Castro, S. V. & Jimenez, S. A. Narrative review: fibrotic diseases: cellular and molecular mechanisms and novel therapies. *Annals of internal medicine* **152**, 159-166 (2010).

14 Usunier, B., Benderitter, M., Tamarat, R. & Chapel, A. Management of fibrosis: the mesenchymal stromal cells breakthrough. *Stem cells international* **2014** (2014).

15 Rosenbloom, J., Mendoza, F. A. & Jimenez, S. A. Strategies for anti-fibrotic therapies. *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease* **1832**, 1088-1103 (2013).

16 Wick, G. *et al.* The immunology of fibrosis. *Annu Rev Immunol* **31**, 107-135 (2013). https://doi.org:10.1146/annurev-immunol-032712-095937

17 Khalil, N. & Greenberg, A. in *Clinical applications of TGF-b.* 194-211.

18 Van Linthout, S., Miteva, K. & Tschöpe, C. Crosstalk between fibroblasts and inflammatory cells. *Cardiovascular research* **102**, 258-269 (2014).

19 Ihn, H. Pathogenesis of fibrosis: role of TGF-β and CTGF. *Current opinion in rheumatology* **14**, 681-685 (2002).

20 Aschner, Y. & Downey, G. P. Transforming growth factor-β: master regulator of the respiratory system in health and disease. *American journal of respiratory cell and molecular biology* **54**, 647-655 (2016).
21 Gu, L. *et al.* Effect of TGF-β/Smad signaling pathway on lung myofibroblast differentiation 4. *Acta pharmacologica Sinica* **28**, 382-391 (2007).
22 Meng, X.-m., Nikolic-Paterson, D. J. & Lan, H. Y. TGF-β: the master regulator of fibrosis. *Nature Reviews Nephrology* **12**, 325-338 (2016).
23 Walker, E. J., Heydet, D., Veldre, T. & Ghildyal, R. Transcriptomic changes during TGF-β-mediated differentiation of airway fibroblasts to myofibroblasts. *Scientific reports* **9**, 1-14 (2019).
24 Chung, J. Y.-F. *et al.* Tgf-β signaling: From tissue fibrosis to tumor microenvironment. *International Journal of Molecular Sciences* **22**, 7575 (2021).
25 Rayinda, T., van Steensel, M. & Danarti, R. Inherited skin disorders presenting with poikiloderma. *Int J Dermatol* **60**, 1343-1353 (2021). https://doi.org:10.1111/ijd.15498
26 Kaviarasan, P., Prasad, P. & Viswanathan, P. Kindler syndrome. *Indian Journal of Dermatology, Venereology and Leprology* **71**, 348-350 (2005). https://doi.org:10.4103/0378-6323.16788
27 Chasseuil, E. *et al.* Dermatological manifestations of hereditary fibrosing poikiloderma with tendon contractures, myopathy and pulmonary fibrosis (POIKTMP): a case series of 28 patients. *Br J Dermatol* **181**, 862-864 (2019). https://doi.org:10.1111/bjd.17996
28 Bythell-Douglas, R. & Deans, A. J. A structural guide to the bloom syndrome complex. *Structure* **29**, 99-113 (2021).
29 Dokic, Y. *et al.* Hereditary fibrosing poikiloderma with tendon contractures, myopathy, and pulmonary fibrosis: Hepatic disease in a child with a novel pathogenic variant of FAM111B. *JAAD Case Rep* **6**, 1217-1220 (2020). https://doi.org:10.1016/j.jdcr.2020.09.025
30 Zhang, Z. *et al.* Family of hereditary fibrosing poikiloderma with tendon contractures, myopathy and pulmonary fibrosis caused by a novel FAM111B mutation. *J Dermatol* **46**, 1014-1018 (2019). https://doi.org:10.1111/1346-8138.15045
31 Mercier, S. *et al.* FAM111B mutation is associated with pancreatic cancer predisposition. *Pancreas* **48**, e41-e42 (2019).
32 Goussot, R. *et al.* 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**, 143-150 (2017).
33 Takeichi, T. *et al.* Syndromic inherited poikiloderma due to a de novo mutation in FAM111B. *Br J Dermatol* **176**, 534-536 (2017). https://doi.org:10.1111/bjd.14845
34 Macchiaiolo, M. *et al.* Expanding phenotype of FAM111B-related disease focusing on liver involvement: Literature review, report of a case with end-stage liver disease and proposal for a new acronym. *American Journal of Medical Genetics Part A* (2022).
35 Kury, S. *et al.* CUGC for hereditary fibrosing poikiloderma with tendon contractures, myopathy, and pulmonary fibrosis (POIKTMP). *Eur J Hum Genet* **24** (2016). https://doi.org:10.1038/ejhg.2015.205
36 Kazlouskaya, V., Feldman, E. J., Jakus, J., Heilman, E. & Glick, S. 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**, e443-e445 (2018).

37 Hoffmann, S. *et al.* FAM111 protease activity undermines cellular fitness and is amplified by gain-of-function mutations in human disease. *EMBO Rep* **21**, e50662 (2020). https://doi.org:10.15252/embr.202050662

38 Takimoto-Sato, M. *et al.* Case Report: Hereditary Fibrosing Poikiloderma With Tendon Contractures, Myopathy, and Pulmonary Fibrosis (POIKTMP) Presenting With Liver Cirrhosis and Steroid-Responsive Interstitial Pneumonia. *Frontiers in genetics* **13** (2022).

39 Chen, F. *et al.* Mutation in FAM111B Causes Hereditary Fibrosing Poikiloderma with Tendon Contracture, Myopathy, and Pulmonary Fibrosis. *Acta Derm Venereol* **99**, 695-696 (2019). https://doi.org:10.2340/00015555-3186

40 Seo, A. *et al.* FAM111B Mutation Is Associated With Inherited Exocrine Pancreatic Dysfunction. *Pancreas* **45**, 858-862 (2016). https://doi.org:10.1097/MPA.0000000000000529

41 Sun, H. *et al.* FAM111B, a direct target of p53, promotes the malignant process of lung adenocarcinoma. *Onco Targets Ther* **12**, 2829-2842 (2019). https://doi.org:10.2147/OTT.S190934

42 Kawasaki, K. *et al.* FAM111B enhances proliferation of KRAS-driven lung adenocarcinoma by degrading p16. *Cancer science* **111**, 2635-2646 (2020).

43 Roversi, G. *et al.* 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**, e20200332 (2021). https://doi.org:10.1590/1678-4685-GMB-2020-0332

44 Mercier, S., Küry, S. & Barbarot, S. Hereditary Fibrosing Poikiloderma with Tendon Contractures, Myopathy, and Pulmonary Fibrosis Synonym: POIKTMP.

45 Bunnik, E. M. & Le Roch, K. G. An introduction to functional genomics and systems biology. *Advances in wound care* **2**, 490-498 (2013).

46 Gasperskaja, E. & Kučinskas, V. The most common technologies and tools for functional genome analysis. *Acta Medica Lituanica* **24**, 1-11 (2017).

47 Nakashima, H., Akahoshi, M. & Tanaka, Y. in *PCR Protocols* 319-322 (Springer, 2003).

48 Yang, W., Kang, X., Yang, Q., Lin, Y. & Fang, M. Review on the development of genotyping methods for assessing farm animal diversity. *Journal of animal science and biotechnology* **4**, 1-6 (2013).

49 Nangru, A., Maharana, B., Vohra, S. & Kumar, B. Molecular identification of Theileria species in naturally infected sheep using nested PCR–RFLP. *Parasitology Research* **121**, 1487-1497 (2022).

50 Crossley, B. M. *et al.* Guidelines for Sanger sequencing and molecular assay monitoring. *Journal of Veterinary Diagnostic Investigation* **32**, 767-775 (2020).

51 Morozova, O. & Marra, M. A. Applications of next-generation sequencing technologies in functional genomics. *Genomics* **92**, 255-264 (2008).

52 Verma, S. & Gazara, R. K. in *Translational Bioinformatics in Healthcare and Medicine* 29-47 (Elsevier, 2021).

53 Wheeler, D. A. *et al.* The complete genome of an individual by massively parallel DNA sequencing. *nature* **452**, 872-876 (2008).

54 Buxbaum, J. & Hof, P. R. *The neuroscience of autism spectrum disorders*. (Academic Press, 2012).

55 Nair, S. V., Thomas, G. & Ankathil, R. Next-generation sequencing in cancer. *Journal of Maxillofacial and Oral Surgery* **20**, 340-344 (2021).
56 Simon, M. M. *et al.* Current strategies for mutation detection in phenotype-driven screens utilising next generation sequencing. *Mammalian Genome* **26**, 486-500 (2015).
57 Kahvejian, A., Quackenbush, J. & Thompson, J. F. What would you do if you could sequence everything? *Nature biotechnology* **26**, 1125-1133 (2008).
58 Sahu, P. K. *et al.* Next generation sequencing based forward genetic approaches for identification and mapping of causal mutations in crop plants: A comprehensive review. *Plants* **9**, 1355 (2020).
59 Dong, Z. & Chen, Y. Transcriptomics: advances and approaches. *Science China Life Sciences* **56**, 960-967 (2013).
60 Valdés, A., Ibáñez, C., Simó, C. & García-Cañas, V. Recent transcriptomics advances and emerging applications in food science. *TrAC Trends in Analytical Chemistry* **52**, 142-154 (2013).
61 Celis, J. E. *et al.* Gene expression profiling: monitoring transcription and translation products using DNA microarrays and proteomics. *FEBS letters* **480**, 2-16 (2000).
62 King, H. C. & Sinha, A. A. Gene expression profile analysis by DNA microarrays: promise and pitfalls. *Jama* **286**, 2280-2288 (2001).
63 Bumgarner, R. Overview of DNA microarrays: types, applications, and their future. *Current protocols in molecular biology* **101**, 22.21. 21-22.21. 11 (2013).
64 Arikawa, E., Quellhorst, G., Han, Y., Pan, H. & Yang, J. RT2 Profiler PCR Arrays: Pathway-focused gene expression profiling with qRT-PCR. *SA Biosciences Technical Article* **11** (2010).
65 Kukurba, K. R. & Montgomery, S. B. RNA sequencing and analysis. *Cold Spring Harbor Protocols* **2015**, pdb. top084970 (2015).
66 Han, Y., Gao, S., Muegge, K., Zhang, W. & Zhou, B. Advanced applications of RNA sequencing and challenges. *Bioinformatics and biology insights* **9**, BBI. S28991 (2015).
67 Haque, A., Engel, J., Teichmann, S. A. & Lönnberg, T. A practical guide to single-cell RNA-sequencing for biomedical research and clinical applications. *Genome medicine* **9**, 1-12 (2017).
68 Ozsolak, F. & Milos, P. M. RNA sequencing: advances, challenges and opportunities. *Nature reviews genetics* **12**, 87-98 (2011).
69 Rai, M. F., Tycksen, E. D., Sandell, L. J. & Brophy, R. H. Advantages of RNA-seq compared to RNA microarrays for transcriptome profiling of anterior cruciate ligament tears. *Journal of Orthopaedic Research®* **36**, 484-497 (2018).
70 Wang, Z., Gerstein, M. & Snyder, M. RNA-Seq: a revolutionary tool for transcriptomics. *Nature reviews genetics* **10**, 57-63 (2009).
71 Aslam, B., Basit, M., Nisar, M. A., Khurshid, M. & Rasool, M. H. Proteomics: technologies and their applications. *Journal of chromatographic science* **55**, 182-196 (2017).
72 Mitulović, G. & Mechtler, K. HPLC techniques for proteomics analysis—a short overview of latest developments. *Briefings in Functional Genomics* **5**, 249-260 (2006).
73 Pillai-Kastoori, L., Schutz-Geschwender, A. R. & Harford, J. A. A systematic approach to quantitative Western blot analysis. *Analytical biochemistry* **593**, 113608 (2020).

74 Taylor, S. C., Berkelman, T., Yadav, G. & Hammond, M. A defined methodology for reliable quantification of Western blot data. *Molecular biotechnology* **55**, 217-226 (2013).
75 Heidebrecht, F., Heidebrecht, A., Schulz, I., Behrens, S.-E. & Bader, A. Improved semiquantitative Western blot technique with increased quantification range. *Journal of immunological methods* **345**, 40-48 (2009).
76 Mahmood, T. & Yang, P.-C. Western blot: technique, theory, and trouble shooting. *North American journal of medical sciences* **4**, 429 (2012).
77 Kuang, J., Yan, X., Genders, A. J., Granata, C. & Bishop, D. J. An overview of technical considerations when using quantitative real-time PCR analysis of gene expression in human exercise research. *PloS one* **13**, e0196438 (2018).
78 Chen, E. Y. *et al.* Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC bioinformatics* **14**, 1-14 (2013).
79 Xie, Z. *et al.* Gene set knowledge discovery with enrichr. *Current protocols* **1**, e90 (2021).
80 Kuleshov, M. V. *et al.* Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic acids research* **44**, W90-W97 (2016).
81 Crowe, A. R. & Yue, W. Semi-quantitative determination of protein expression using immunohistochemistry staining and analysis: an integrated protocol. *Bio-protocol* **9**, e3465-e3465 (2019).
82 Arowolo, A., Rhoda, C., Mbele, M., Oluwole, O. G. & Khumalo, N. A cost-effective method for detecting mutations in the human FAM111B gene associated with POIKTMP syndrome. *Egyptian Journal of Medical Human Genetics* **23**, 1-7 (2022).
83 Rhoda, C., Sunda, F., Kidzeru, E., Khumalo, N. P. & Arowolo, A. FAM111B dysregulation promotes malignancy in fibrosarcoma and POIKTMP and a low-cost method for its mutation screening. *Cancer Treatment and Research Communications*, 100679 (2023).
84 Rhoda, C. *A study of the expression and cellular function of the human FAM111B gene*, Faculty of Health Sciences, (2021).
85 Hedegaard, J. *et al.* Next-generation sequencing of RNA and DNA isolated from paired fresh-frozen and formalin-fixed paraffin-embedded samples of human cancer and normal tissue. *PloS one* **9**, e98187 (2014).
86 Arsenic, R. *et al.* Comparison of targeted next-generation sequencing and Sanger sequencing for the detection of PIK3CA mutations in breast cancer. *BMC clinical pathology* **15**, 1-9 (2015).
87 Tsiatis, A. C. *et al.* Comparison of Sanger sequencing, pyrosequencing, and melting curve analysis for the detection of KRAS mutations: diagnostic and clinical implications. *The Journal of Molecular Diagnostics* **12**, 425-432 (2010).
88 Tarca, A. L., Romero, R. & Draghici, S. Analysis of microarray experiments of gene expression profiling. *American journal of obstetrics and gynecology* **195**, 373-388 (2006).
89 Boone, D. R. *et al.* Pathway-focused PCR array profiling of enriched populations of laser capture microdissected hippocampal cells after traumatic brain injury. *PloS one* **10**, e0127287 (2015).
90 Ra, H.-J. & Parks, W. C. Control of matrix metalloproteinase catalytic activity. *Matrix biology* **26**, 587-596 (2007).
91 Giannandrea, M. & Parks, W. C. Diverse functions of matrix metalloproteinases during fibrosis. *Disease models & mechanisms* **7**, 193-203 (2014).

92 Yamashita, C. M. *et al.* Matrix metalloproteinase 3 is a mediator of pulmonary fibrosis. *The American journal of pathology* **179**, 1733-1745 (2011).
93 Liu, S., Shi-wen, X., Abraham, D. J. & Leask, A. CCN2 is required for bleomycin-induced skin fibrosis in mice. *Arthritis & Rheumatism* **63**, 239-246 (2011).
94 Todd, N. W., Luzina, I. G. & Atamas, S. P. Molecular and cellular mechanisms of pulmonary fibrosis. *Fibrogenesis & tissue repair* **5**, 1-24 (2012).
95 Feng, D. & Gerarduzzi, C. Emerging roles of matricellular proteins in systemic sclerosis. *International Journal of Molecular Sciences* **21**, 4776 (2020).
96 Zhang, X., Liu, H., Hock, T., Thannickal, V. J. & Sanders, Y. Y. Histone deacetylase inhibition downregulates collagen 3A1 in fibrotic lung fibroblasts. *International Journal of Molecular Sciences* **14**, 19605-19617 (2013).
97 Akamatsu, S. *et al.* Common variants at 11q12, 10q26 and 3p11. 2 are associated with prostate cancer susceptibility in Japanese. *Nature genetics* **44**, 426-429 (2012).
98 Li, W., Hu, S., Han, Z. & Jiang, X. YY1-Induced Transcriptional Activation of FAM111B Contributes to the Malignancy of Breast Cancer. *Clin Breast Cancer* **22**, e417-e425 (2022). https://doi.org:10.1016/j.clbc.2021.10.008
99 Gong, Q. *et al.* Clinicopathological features, prognostic significance, and associated tumor cell functions of family with sequence similarity 111 member B in pancreatic adenocarcinoma. *Journal of Clinical Laboratory Analysis*, e24784 (2022).
100 Dejous, C. & Krishnan, U. M. Sensors for diagnosis of prostate cancer: Looking beyond the prostate specific antigen. *Biosensors and Bioelectronics* **173**, 112790 (2021).
101 Ahmed, S., Misra, D. P. & Agarwal, V. Interleukin-17 pathways in systemic sclerosis-associated fibrosis. *Rheumatology International* **39**, 1135-1143 (2019).
102 Ramani, K. & Biswas, P. S. Interleukin-17: Friend or foe in organ fibrosis. *Cytokine* **120**, 282-288 (2019).
103 Muharram, G. *et al.* Tensin-4-dependent MET stabilization is essential for survival and proliferation in carcinoma cells. *Developmental cell* **29**, 421-436 (2014).
104 Fernández-Klett, F. & Priller, J. The fibrotic scar in neurological disorders. *Brain pathology* **24**, 404-413 (2014).
105 Ruggiano, A. & Ramadan, K. DNA–protein crosslink proteases in genome stability. *Communications Biology* **4**, 1-11 (2021).
106 Neumann, S. & Kaup, F. α-SMA and Ki-67 Immunohistochemistry as Indicators for the Fibrotic Remodeling Process in the Liver of Dogs. *Journal of Advanced Veterinary Research* **2**, 42-46 (2012).
107 Uhal, B. D. The role of apoptosis in pulmonary fibrosis. *European Respiratory Review* **17**, 138 (2008). https://doi.org:10.1183/09059180.00010906

A. Appendix

A.0 RNA Quantification using the Bioanalyzer

A. Lane 1 | B. Lane 2 | C. Lane 3

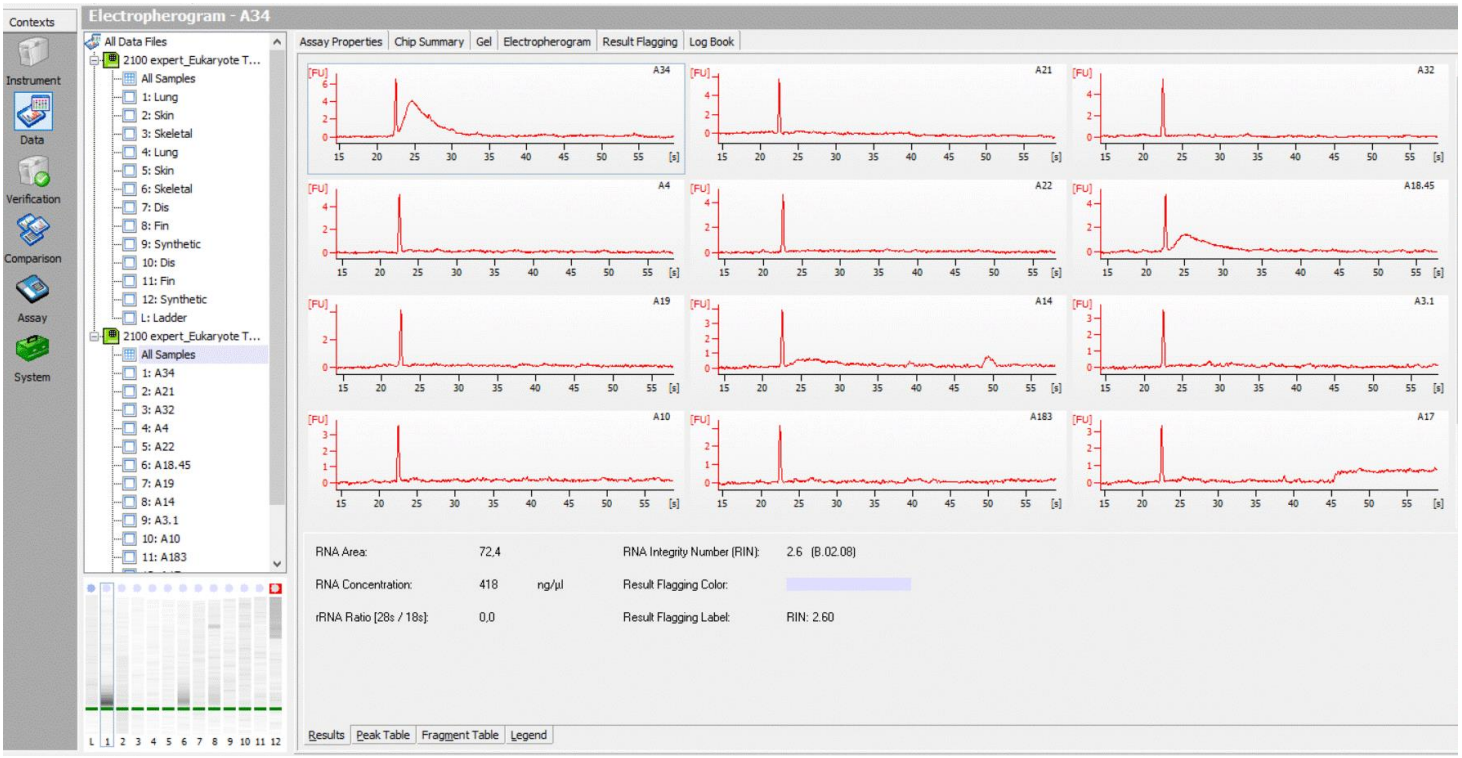


Figure A-0 Schematic Diagram of RNA Quantification of FFPE samples. Schematic diagram depicting the RNA quantification of the FFPE samples following RNA extraction to measure the quality and quantity of the samples prior to downstream applications. A) Lane 1 is a visualization of the lung samples from the patient at the top to the three familial controls below. B) Lane 2 is the skin samples, with the same sequence of patient then controls. C) Lane 3 is the visualization of the skeletal muscle samples following the same sequence.

A.1 Scatter plot diagram RT^2 Profiler PCR Array

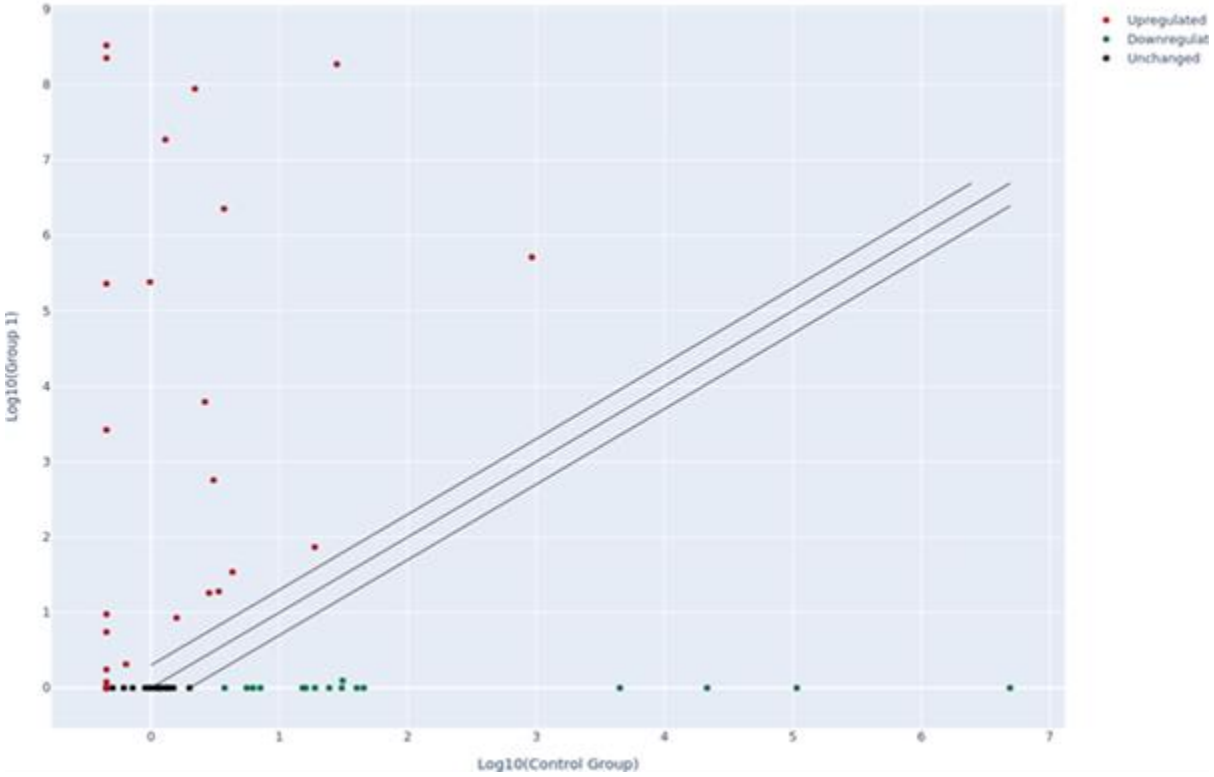


Figure A-1:Lung RT^2 Profiler PCR Array analysis report. (A) Scatterplot diagram comparing the normalized expression of every gene between the patient (Group1) and healthy (Control group) against one another to observe large gene expression changes. The central diagonal line indicates unchanged gene expression, while the outer diagonal lines indicate the selected fold regulation threshold. Genes with data points beyond the outer lines in the upper left and lower right corners are up-regulated or down-regulated, respectively, by more than the fold regulation threshold in the y-axis Group relative to the x-axis Group.

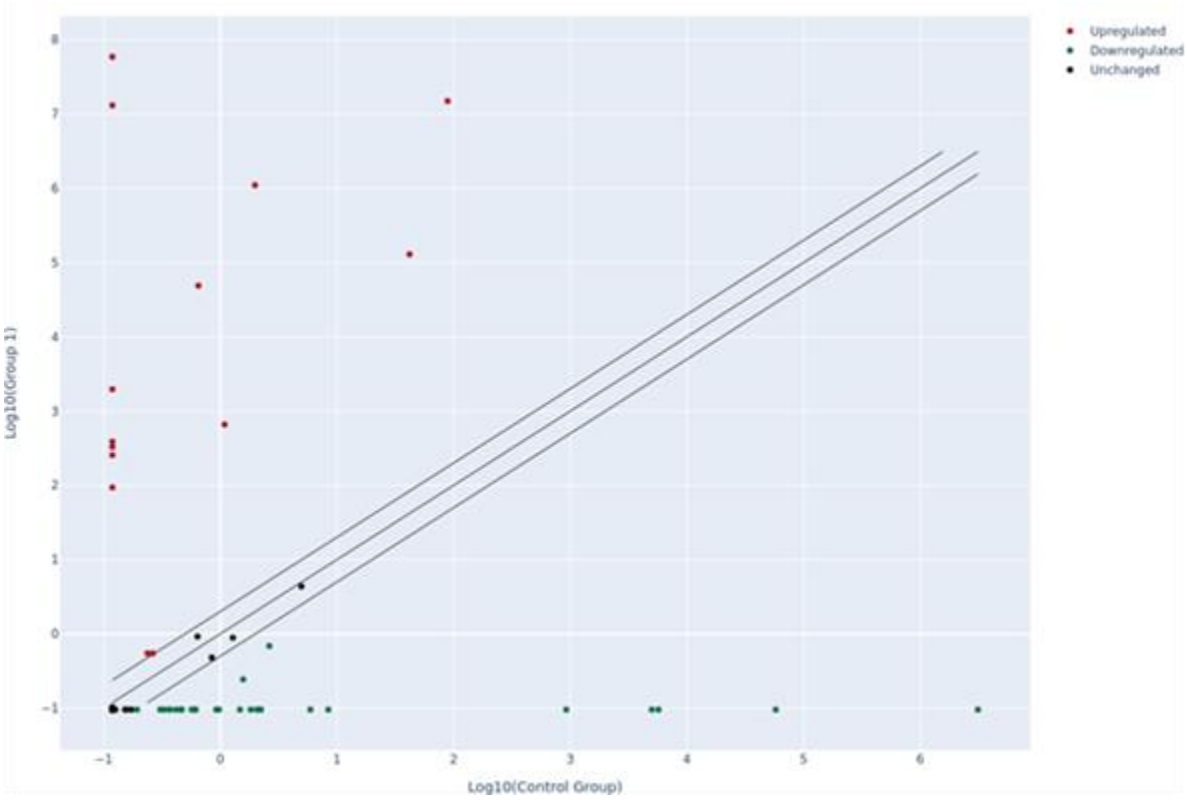


Figure A-2 Skin RT^2 Profiler PCR Array analysis report. (A) Scatterplot diagram comparing the normalized expression of every gene between the patient (Group1) and healthy (Control group) against one another to observe large gene expression changes. The central diagonal line indicates unchanged gene expression, while the outer diagonal lines indicate the selected fold regulation threshold. Genes with data points beyond the outer lines in the upper left and lower right corners are up-regulated or down-regulated, respectively, by more than the fold regulation threshold in the y-axis Group relative to the x-axis Group.

A2. Gene-set Enrichment Analysis

A2.1 GSEA with gene-set libraries in Biological Pathways

Table A-1: KEGG 2021 Human-Enriched Terms

Index	Name	P-value	Adjusted p-value	Odds Ratio	Combined score
1	IL-17 signaling pathway	0.0007701	0.01839	61.80	443.03
2	Prostate cancer	0.0008197	0.01839	59.84	425.24
3	Platelet activation	0.001334	0.01839	46.53	308.01
4	Relaxin signaling pathway	0.001443	0.01839	44.69	292.32
5	Transcriptional misregulation in cancer	0.003157	0.02581	29.78	171.45
6	Focal adhesion	0.003454	0.02581	28.42	161.07
7	Rap1 signaling pathway	0.003763	0.02581	27.17	151.70
8	Regulation of actin cytoskeleton	0.004048	0.02581	26.16	144.11
9	Melanoma	0.03194	0.08526	35.07	120.78
10	Glioma	0.03325	0.08526	33.64	114.51

Table A-2:REACTOME 2022 enriched terms

Index	Name	P-value	Adjusted p-value	Odds Ratio	Combined score
1	MET Interacts With TNS Proteins R-HSA-8875513	0.002248	0.02423	624.59	3808.55
2	Fibronectin Matrix Formation R-HSA-1566977	0.002697	0.02616	499.65	2955.69
3	Assembly Of Collagen Fibrils And Other Multimeric Structures R-HSA-2022090	0.000001821	0.00008833	184.60	2439.70
4	EGFR Transactivation By Gastrin R-HSA-2179392	0.003595	0.02682	356.86	2008.48
5	RUNX2 Regulates Genes Involved In Cell Migration R-HSA-8941332	0.003595	0.02682	356.86	2008.48
6	RUNX3 Regulates YAP1-mediated Transcription R-HSA-8951671	0.003595	0.02682	356.86	2008.48
7	CHL1 Interactions R-HSA-447041	0.004043	0.02802	312.23	1720.62
8	Activation Of Matrix Metalloproteinases R-HSA-1592389	0.00009436	0.002288	183.96	1705.05
9	Extracellular Matrix Organization R-HSA-1474244	7.569e-8	0.000007342	86.12	1412.13
10	Collagen Formation R-HSA-1474290	0.000007258	0.0002347	114.39	1353.63

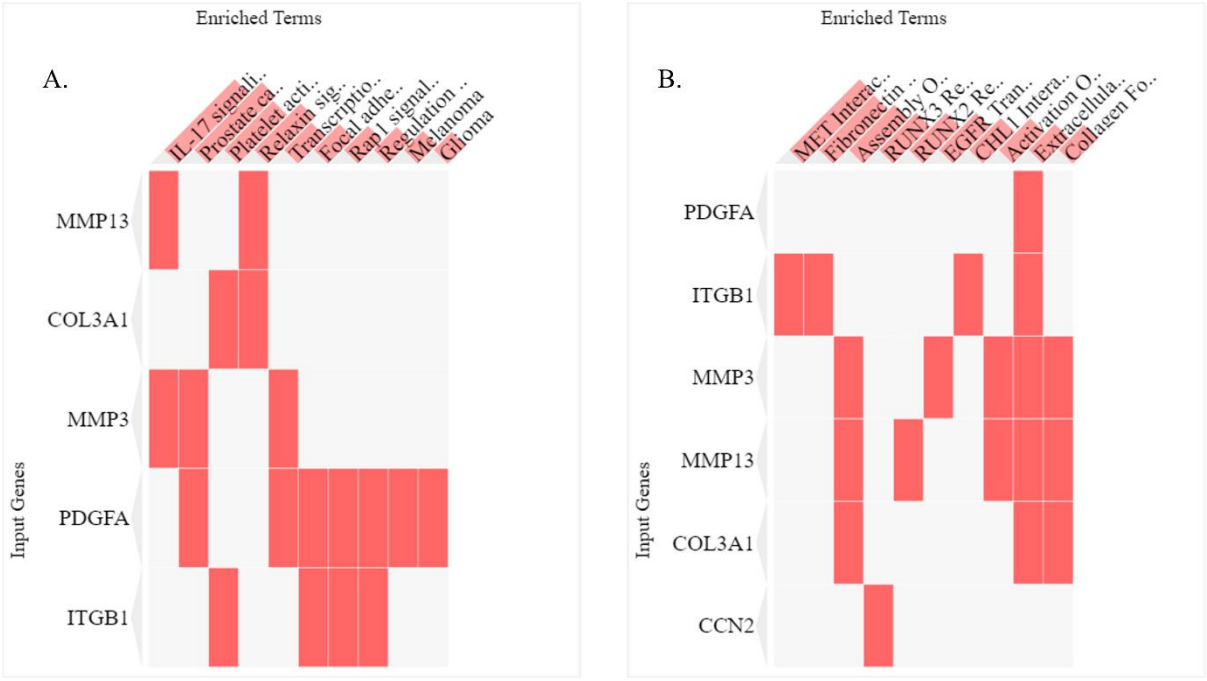


Figure A-3:Biological Pathway Clustergram of enriched terms in association with the input genes in the gene-set library. ***A)*** KEGG 2021 Human pathway**,** MMP3 and MMP13 seemed to overlap in the IL-17 signalling pathways. In the prostate cancer MMP3and PDGFA were highly enriched. **B**) Reactome 2022 pathway- where 6 of the 9 genes overlapped with each other. ITGB1 seemed to be enriched in the MET interacts with TNS proteins.

A.2.2 GSEA with gene-set libraries in Ontologies

Table A-3 GSEA of Biological processes terms

Index	Name	P-value	Adjusted p-value	Odds Ratio	Combined score
1	regulation of glomerular mesangial cell proliferation (GO:0072124)	0.002248	0.01850	624.59	3808.55
2	neuroinflammatory response (GO:0150076)	0.002248	0.01850	624.59	3808.55
3	cell-cell adhesion mediated by integrin (GO:0033631)	0.002248	0.01850	624.59	3808.55
4	regulation of phosphatidylinositol biosynthetic process (GO:0010511)	0.002248	0.01850	624.59	3808.55
5	regulation of inward rectifier potassium channel activity (GO:1901979)	0.002248	0.01850	624.59	3808.55
6	negative regulation of hydrogen peroxide metabolic process (GO:0010727)	0.002248	0.01850	624.59	3808.55
7	regulation of hydrogen peroxide metabolic process (GO:0010310)	0.002697	0.01915	499.65	2955.69
8	negative regulation of neuron migration (GO:2001223)	0.002697	0.01915	499.65	2955.69
9	positive regulation of oxidative stress-induced cell death (GO:1903209)	0.003146	0.02109	416.35	2398.85
10	extracellular structure organization (GO:0043062)	1.706e-8	0.000001580	117.18	2095.93

\

Table A-4 GSEA library in Molecular Function enriched terms

Index	Name	P-value	Adjusted p-value	Odds Ratio	Combined score
1	platelet-derived growth factor binding (GO:0048407)	0.000009879	0.0002272	634.35	7310.93
2	C-X3-C chemokine binding (GO:0019960)	0.002248	0.008618	624.59	3808.55
3	collagen binding involved in cell-matrix adhesion (GO:0098639)	0.002248	0.008618	624.59	3808.55
4	cell-matrix adhesion mediator activity (GO:0098634)	0.003146	0.009045	416.35	2398.85
5	platelet-derived growth factor receptor binding (GO:0005161)	0.005836	0.01491	208.11	1070.48
6	metalloendopeptidase activity (GO:0004222)	0.0005867	0.006747	71.11	529.13
7	chemokine binding (GO:0019956)	0.01431	0.02992	80.48	341.79
8	protease binding (GO:0002020)	0.001209	0.007308	48.95	328.85
9	metallopeptidase activity (GO:0008237)	0.001271	0.007308	47.71	318.14
10	protein heterodimerization activity (GO:0046982)	0.003030	0.009045	30.42	176.43

Table A-5 Enriched terms of cellular components

Index	Name	P-value	Adjusted p-value	Odds Ratio	Combined score
1	glial cell projection (GO:0097386)	0.006284	0.05000	192.10	973.89
2	actin-based cell projection (GO:0098858)	0.0006011	0.01322	70.23	520.88
3	neuromuscular junction (GO:0031594)	0.01697	0.07096	67.41	274.78
4	sarcolemma (GO:0042383)	0.02316	0.07096	48.87	184.02
5	microvillus (GO:0005902)	0.02536	0.07096	44.50	163.50
6	filopodium (GO:0030175)	0.02580	0.07096	43.71	159.87
7	platelet alpha granule lumen (GO:0031093)	0.02975	0.07273	37.74	132.64
8	endoplasmic reticulum lumen (GO:0005788)	0.006819	0.05000	19.90	99.25
9	platelet alpha granule (GO:0031091)	0.03979	0.08105	27.95	90.12
10	Golgi lumen (GO:0005796)	0.04412	0.08105	25.12	78.38

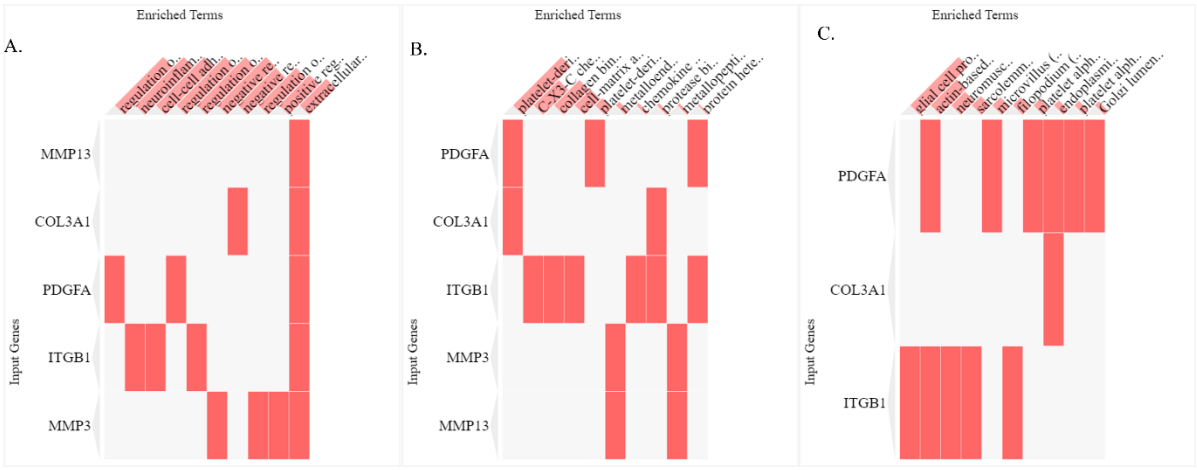


Figure A-4 Enriched terms of input genes in GSEA in gene ontologies. A) Biological processes, B)Molecular function, C) Cellular component

A2.3 GSEA with gene-set libraries in Disease and Drugs

Table A-6 Covid-19 Related gene sets 2021 enriched terms

Index	Name	P-value	Adjusted p-value	Odds Ratio	Combined score
1	SARS perturbation; 112 Down Genes from GEN3VA; Human airway epithelium (HAE) cells; Accession: GSE47961 Platform: GPL6480; Entry 4	0.0007865	0.04208	61.13	436.96
2	SARS perturbation; 264 Down Genes from GEN3VA; Human airway epithelium (HAE) cells; Accession: GSE47961 Platform: GPL6480; Entry 2	0.0001108	0.01185	44.73	407.38
3	COVID19-Orf8 protein host PPI from Krogan	0.02096	0.1495	54.20	209.49
4	SARS perturbation; 114 Down Genes from GEN3VA; Human airway epithelium (HAE) cells; Accession: GSE47961 Platform: GPL6480; Entry 6	0.03413	0.2050	32.75	110.63
5	500 genes up-regulated by MHV-A59 in murine liver cells from GSE146074 5d	0.01272	0.1493	14.28	62.35
6	500 genes down-regulated by SARS-CoV-2 in mouse Lung cells at 7 dpi from GSE162113	0.01516	0.1493	13.00	54.45
7	Top 500 downregulated genes in mouse lung with SARS-CoV-2 infection (Day 7) from GEO GSE162113	0.01516	0.1493	13.00	54.45
8	SARS Perturbation; 134 Down Genes from GEN3VA Mouse Lung; Accession: GSE19137 Platform: GPL1261 Entry 3	0.05787	0.2050	18.95	54.00
9	500 top upregulated genes from SARS-CoV-2 infection at 72 HPI from GSE157852	0.01569	0.1493	12.75	52.99
10	500 genes down-regulated by MHV-A59 in murine spleen cells from GSE146074 5d	0.01624	0.1493	12.52	51.59

Table A-7 GTEx Aging Signatures 2021 enriched terms

Index	Name	P-value	Adjusted p-value	Odds Ratio	Combined score
1	GTEx SmallIntestine 20-29 vs 40-49 Down	0.005287	0.1040	22.75	119.24
2	GTEx Breast 20-29 vs 50-59 Down	0.005287	0.1040	22.75	119.24
3	GTEx Esophagus 20-29 vs 30-39 Down	0.005287	0.1040	22.75	119.24
4	GTEx Uterus 20-29 vs 40-49 Up	0.1071	0.1071	9.91	22.14
5	GTEx Nerve 20-29 vs 40-49 Down	0.1071	0.1071	9.91	22.14
6	GTEx Skin 20-29 vs 50-59 Down	0.1071	0.1071	9.91	22.14
7	GTEx Kidney 20-29 vs 30-39 Up	0.1071	0.1071	9.91	22.14
8	GTEx Liver 20-29 vs 70-79 Down	0.1071	0.1071	9.91	22.14
9	GTEx SmallIntestine 20-29 vs 60-69 Up	0.1071	0.1071	9.91	22.14
10	GTEx SmallIntestine 20-29 vs 60-69 Down	0.1071	0.1071	9.91	22.14

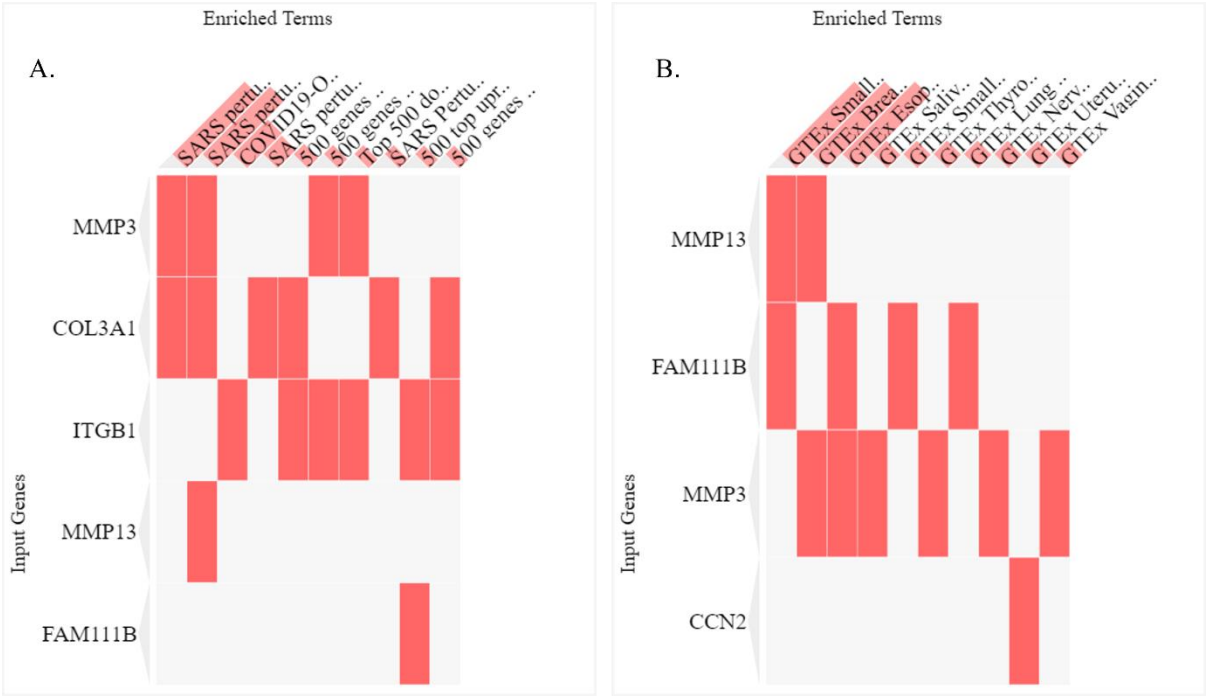


Figure A-5 Clustergram of enriched terms of the gene-set library list in the Disease and drugs category. **A)** COVID-19 Related Gene sets 2021, *FAM111B* appeared in the 500 top upregulated genes from SARS-CoV-2 infection. **B)** GTEx Aging Signatures 2021, *FAM111B* enrichment occurred in the GTEx Small intestines, esophagus, and lungs.

A2.4 GSEA in Cell types

Table A-8 GTEx Tissue Expression Down enriched terms

Index	Name	P-value	Adjusted p-value	Odds Ratio	Combined score
1	GTEX-WYVS-0326-SM-3NM9V thyroid female 40-49 years	0.01786	0.5318	63.95	257.41
2	GTEX-QLQ7-0726-SM-2I5G2 thyroid male 60-69 years	0.01919	0.5318	59.37	234.73
3	GTEX-P78B-0526-SM-2I5F7 thyroid female 40-49 years	0.02360	0.5318	47.93	179.56
4	GTEX-S341-1326-SM-4AD72 cervix uteri female 40-49 years	0.02536	0.5318	44.50	163.50
5	GTEX-XGQ4-1826-SM-4AT6F prostate male 50-59 years	0.02712	0.5318	41.52	149.79
6	GTEX-PX3G-2626-SM-2I3EG thyroid female 20-29 years	0.02800	0.5318	40.18	143.66
7	GTEX-XV7Q-0326-SM-4BRVM thyroid female 40-49 years	0.02975	0.5318	37.74	132.64
8	GTEX-PLZ6-1126-SM-3P5ZR prostate male 30-39 years	0.03107	0.5318	36.09	125.29
9	GTEX-QDT8-2926-SM-32PKC brain female 30-39 years	0.0006178	0.5039	15.40	113.82
10	GTEX-RN64-1626-SM-48FD7 kidney male 50-59 years	0.03413	0.5393	32.75	110.63

Table A-9 Enriched terms in the Human Gene Atlas library

Index	Name	P-value	Adjusted p-value	Odds Ratio	Combined score
1	SmoothMuscle	0.0004592	0.004133	27.27	209.56
2	Salivarygland	0.02536	0.1102	44.50	163.50
3	OlfactoryBulb	0.03674	0.1102	30.35	100.27
4	PancreaticIslet	0.05402	0.1118	20.36	59.41
5	Kidney	0.06213	0.1118	17.60	48.90
6	Uterus	0.08691	0.1304	12.37	30.22
7	Wholebrain	0.1030	0.1324	10.33	23.48
8	Colorectaladenocarcinoma	0.1240	0.1395	8.46	17.66
9	SkeletalMuscle	0.1819	0.1819	5.55	9.47

Table A-10 Descartes Cell Types and Tissues 2021 enriched terms

Index	Name	P-value	Adjusted p-value	Odds Ratio	Combined score
1	Stromal cells in Stomach	0.02272	0.1091	49.85	188.66
2	Stromal cells in Muscle	0.04152	0.1091	26.74	85.09
3	Stromal cells in Eye	0.04195	0.1091	26.46	83.91
4	Lymphoid cells in Liver	0.05445	0.1091	20.19	58.77
5	Stromal cells in Adrenal	0.06340	0.1091	17.23	47.52
6	Stromal cells in Kidney	0.07270	0.1091	14.93	39.13
7	Stromal cells in Thymus	0.1200	0.1543	8.77	18.59
8	Vascular endothelial cells in Cerebrum	0.2169	0.2388	4.55	6.95
9	Vascular endothelial cells in Cerebellum	0.2388	0.2388	4.07	5.83

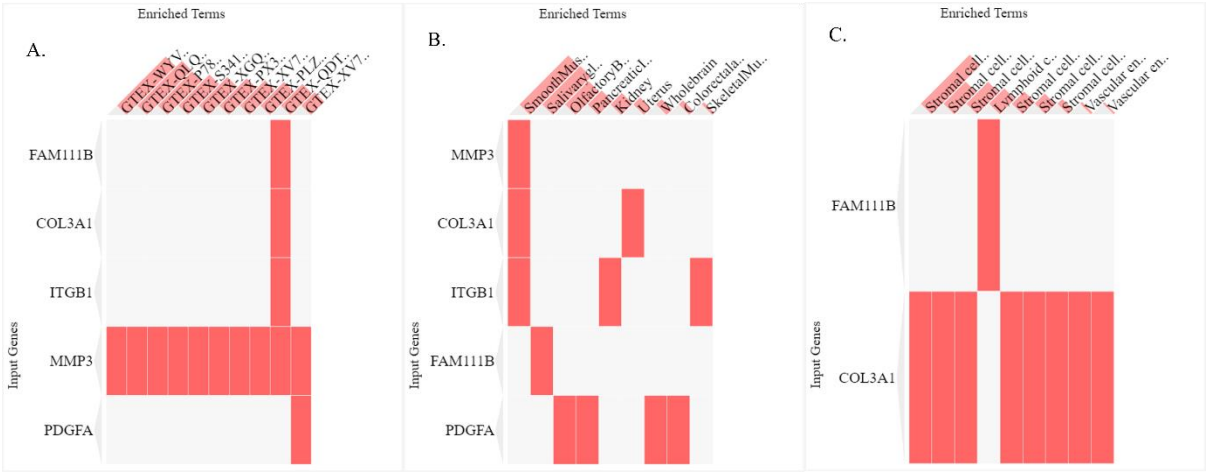


Figure A-6 Clustergram of the enriched terms associated with the input genes of the Cell types category. **A)** GTEx Tissue Expression Down, FAM*111B* enrichment was observed in the GTEx-QDT8-2926-SM-32PKC brain female 30-39. **B)** In the Human Gene Atlas *FAM111B* was enriched in salivary glands. **C)** Descartes Cell types and Tissues 2021, *FAM111B* was enriched in the lymphoid cells in liver