FUNCTIONAL CHARACTERIZATION OF FAM26F BASED ON ITS SUBCELLULAR LOCALIZATION AND NOVEL INTERACTING PARTNERS

By

UZMA MALIK

NUST201490168PASAB8014S

Atta-ur-Rahman School of Applied Biosciences (ASAB)

National University of Sciences & Technology (NUST)

Islamabad, Pakistan

2021

FUNCTIONAL CHARACTERIZATION OF FAM26F BASED ON ITS SUBCELLULAR LOCALIZATION AND NOVEL INTERACTING PARTNERS

By

UZMA MALIK

NUST201490168PASAB8014S

A thesis submitted to National University of Sciences and Technology (NUST) in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Applied Biosciences

Supervisor: Dr. Aneela Javed

Atta-ur-Rahman School of Applied Biosciences (ASAB)

National University of Sciences & Technology (NUST)

Islamabad, Pakistan

2021

THESIS ACCEPTANCE CERTIFICATE

Certified that final copy of PhD thesis written by Ms. **Uzma Malik** (Registration No. **NUST201490168PASAB8014S**), of **Atta ur Rahman School of Applied Biosciences,** has been vetted by undersigned, found complete in all respects as per NUST Statutes/Regulations/PhD Policy, is free from plagiarism, errors, and mistakes, and is accepted as partial fulfillment for award of PhD degree. It is further certified that necessary amendments as pointed out by GEC members and foreign/local evaluators of the scholar have also been incorporated in the said thesis.

Signature: _____

Name of Supervisor: **Dr. Aneela Javed**

Date: _____

Dr. Touqeer Ahmed
Head of Department (HoD)
Deptt of Healthcare Biotechnology
Atta-ur-Rahman School of Applied Biosciences (ASAB), NUST, Islamabad

Signature (HOD): _____

Dr. Touqeer Ahmed

Date: 20/9/2021

Dr. Hussnain A. Janjua
Principal
Atta-ur-Rahman School of
Applied Biosciences (ASAB)
NUST, Islamabad

Signature (Dean/Principal): _____

Dr. Hussnain Ahmed Janjua

Date: _____

Annex L
Form PhD-7
DOCTORAL PROGRAMME
OF STUDY
(Must be type written)

National University of Sciences & Technology
REPORT OF DOCTORAL THESIS DEFENCE

Name: Uzma Malik NUST Regn No: NUST201490168PASAB8014S

School/College/Centre: ASAB, NUST

Title: Functional Characterization of FAM26F Based on Its Subcellular Localization and Novel Interacting Partners

DOCTORAL DEFENCE COMMITTEE

Doctoral Defence held on 6th May, 2021

	QUALIFIED	NOT QUALIFIED	SIGNATURE
GEC Member-1: Dr. Saadia Zahid	✓		
GEC Member-2: Dr. Sobia Manzoor	✓		
GEC Member (External): Dr. Saima Zafar	✓		
Supervisor: Dr. Aneela Javed	✓		
Co Supervisor (if appointed):			
External Evaluator-1: Dr. Muhammad Qasim (Local Expert)	✓		
External Evaluator-2: Dr. Sundus Javed (Local Expert)	✓		
External Evaluator-3: Dr. Neelam Sharma-Walia (Foreign Expert)			
External Evaluator-4: Dr. Yiping Li (Foreign Expert)			

FINAL RESULT OF THE DOCTORAL DEFENCE
(Appropriate box to be signed by HOD)

PASS	FAIL
✓	

The student Uzma Malik Regn No NUST201490168PASAB8014S is / is NOT accepted for Doctor of Philosophy Degree.

Dated: 6-5-21

Dr. Hussnain A. Janjua
Principal
Atta-ur-Rahman School of
Applied Biosciences (ASAB)
NUST, Islamabad

Dean/Commandant/Principal/DG

Distribution:
01 x original copy each for PGP Dte, Exam Branch Main Office NUST and Student's dossier at the School/College/Centre.
01 x photocopy each for HoD, Supervisor, Co-Supervisor (if appointed), sponsoring agency (if any) and 05 copies for insertion in Dissertation.
Note:* Decision of External Evaluators (Foreign Experts) will be sought through video conference, if possible, on the same date and their decision wil be intimated (on paper) to HQ NUST at a later date.

Certificate of Approval

This is to certify that the research work presented in this thesis, entitled **Functional Characterization of FAM26F Based on Its Subcellular Localization and Novel Interacting Partners** was conducted by Ms. **Uzma Malik** under the supervision of **Dr. Aneela Javed.** No part of this thesis has been submitted anywhere else for any other degree. This thesis is submitted to the **Atta ur Rahman School of Applied Biosciences** in partial fulfillment of the requirements for the degree of Doctor of Philosophy in field of **Applied Biosciences** Department of **Healthcare Biotechnology** university of **National University of Sciences & Technology (NUST)**

Student Name **Uzma Malik** Signature _____

Examination committee:

a) External examiner 1 : **Dr. Muhammad Qasim** Signature _____
(Associate Professor)
Department of Microbiology,
Kohat University of Science and
Technology, Kohat, Pakistan

b) External Examiner 2 : **Dr. Sundus Javed** Signature _____
(Assistant Professor)
Department of Biosciences,
COMSATS Institute of Information
Technology (CIIT), Islamabad, Pakistan

c) Internal Examiner: **Dr. Saadia Zahid** Signature _____
(Associate Professor)
ASAB (NUST), H-1 2 Campus, Islamabad, Pakistan

Supervisor Name: **Dr. Aneela Javed** Signature _____

Name of Dean / HOD: **Dr. Tougeer Ahmed** Signature _____

Dr. Tougeer Ahmed
Head of Department (HoD)
Deptt of Healthcare Biotechnology
Atta-ur-Rahman School of Applied Biosciences (ASAB), NUST Islamabad

AUTHOR'S DECLARATION

I, **Uzma Malik**, hereby state that my PhD thesis titled **Functional Characterization of FAM26F Based on Its Subcellular Localization and Novel Interacting Partners** is my own work and has not been submitted previously by me for taking any degree from **National University of Sciences and Technology (NUST)** or anywhere else in the country/world.

At any time if my statement is found to be incorrect even after my graduation, the university has the right to withdraw my PhD degree.

Name of student: **Uzma Malik**

Date: **6th May, 2021**

PLAGIARISM UNDERTAKING

I solemnly declare that the research work presented in the thesis titled "**Functional Characterization of FAM26F Based on Its Subcellular Localization and Novel Interacting Partners**" is solely my research work with no significant contribution from any other person. Small contribution/help wherever taken has been duly acknowledged and that complete thesis has been written by me.

I understand the zero tolerance policy of HEC and **National University of Sciences and Technology (NUST)** towards plagiarism. Therefore, I as an Author of the above titled thesis declare that no portion of my thesis has been plagiarized and any material used as refrence is properly referred/cited.

I undertake that if I am found guilty of any formal plagiarism in the above titled thesis even after award of PhD degree, the university reserves the right to withdraw/revoke my PhD degree and that HEC and the University have the right to publish my name on the HEC/University website on which names of students are placed who submitted plagiarized thesis.

Student/Author signature: _____

Name: **Uzma Malik**

Dedicated to Baba Ji and Ammi

ACKNOWLEDGEMENTS

I would like to express my profound gratitude to my supervisor, Dr. Aneela Javed, who is the most responsible person for helping me complete this challenging research. I am extremely grateful to her for her constant guidance, critical insight, constructive suggestions and unwavering support throughout the course of this thesis work. When I started working with her, I had much to learn, and she patiently and dedicatedly took the time to teach me and guide me every step of the way. Without her technical acumen and research insight, along with her meticulous feedback and encouragement, this work would not have been possible.

I am deeply obliged to my thesis committee members, Dr Saadia Zahid, Dr. Sobia Manzoor, and Dr. Saima Zafar for their time, guidance, constructive criticism and feedback which helped improve this work. I would also like to thank Dr Amjad Ali and Dr Muhammad Tahir for assisting me and teaching me skills that I applied in my research.

Amongst colleagues, I owe my sincere and profound gratitude to Ms. Sadia Nazir and Dr. Sajid Mansoor for helping me learn and improve my knowledge by their skills, assistance, co-operation and facilitation throughout the course of my research. I am also extremely thankful to my foreign colleagues Dr. Mohsin Shafiq, Dr. Aneeqa Noor, Dr. Neelam Younas and Dr. Sidra Shahid from University of Medicine, Georg-August University, Gottingen, Germany, for their technical guidance and for teaching me skills that I applied extensively in my research. Above all, I am grateful to them for their unwavering support, friendship, and for providing me with a sound and comfortable environment to work in.

I am specifically thankful to my friends Asma Abro and Sadia Nazir for their caring, wonderful nature and for always being there for me. With their enthusiasm, indispensable moral support, encouragement, sound advice, good company and prayers, they helped to make this thesis interesting and achievable for me.

I offer my regards and blessings to all of those who supported me in any respect during the completion of this work. Alongside that, I thank the faculty members of ASAB, NUST for polishing my knowledgebase, and also my university and lab

fellows for their bundles of prayers, moral support and memorable company during my entire degree.

Finally and above all, no words can express my level of gratitude to my family for their constant and unconditional prayers and unfaltering support throughout this journey that helped me to accomplish my aim. Special thanks to my parents for their hard work and friendship, for taking pride in all my achievements regardless of how small they were, and for making me who I am today. I want to thank my amazing sisters, Ayesha, Sadia, Rabia and Rafia, for adding fun and beauty to my life, and for being a source of inspiration for me. I also want to acknowledge my bundles of joy, my nieces and nephews, for bringing happiness in my life and for alleviating my stress through their adorable company.

CONTENTS

CHAPTER 5

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

%	Percent
3D	Three Dimensional
Aa	Amino Acid
AIDS	Acquired Immunodeficiency Syndrome
APC	Antigen-Presenting Cell
ART	Antiretroviral Treatment
AU	Arbitrary Units
BCR	B Cell Receptor
BLAST	Basic Local Alignment Search Tool
BSA	Bovine Serum Albumin
C^o	Centigrade
Ca_hom_mod	Calcium Homeostasis Modulator
Ca^{+2}	Calcium
CD8	Cluster of Differentiation 8
CDD	Conserved Domain Database
cDNA	Complementary Deoxyribonucleic Acid
CK-II	Casein Kinase II
CO2	Carbon Dioxide
DMEM	Dulbecco's Modified Eagle's Medium
dNTPs	Deoxyribonucleotide Triphosphate
ECL	Enhanced Chemiluminescence
ER	Endoplasmic Reticulum
ESI-QTOF	Electrospray Ionization Quadrupole Time-of-Flight
FAM26F	Family with Sequence Similarity 26, Member F
FBS	Fetal Bovine Serum
GAPDH	Glyceraldehyde 3-Phosphate Dehydrogenase
GC	Glucocorticoids
GFP	Green Fluorescent Protein
GSH	Glutathione
H	Human
H_2O_2	Hydrogen Peroxide

HEK	Human Embryonic Kidney
HIV	Human Immunodeficiency Virus
HMM	Hidden Markov Model
HO^-	Hydroxyl Radicals
IFN	Interferon
Ig	Immunoglobin
IL	Interleukin
INAM	IRF-3–dependent NK-Activating Molecule
IP3	Inositol1,4,5-trisphosphate
IPA	Ingenuity Pathway Analysis
kD	Kilo Dalton
LPS	Lipo Polysaccharide
MA	Micro Array
MSDB	Mascot Database
mDC	Myeloid Dendritic Cell
MgCl2	Magnesium Chloride
Ml	Milliliter
mM	Millimolar
MS	Mass Spectrometer
MSA	Multiple Sequence Alignment
MTS	3-(4,5–dimethylthiazol–2–yl)-5-(3-carboxymethoxyphenyl)-2-(-4-sulfophenyl)-2H-Tetrazolium
NADPH	Nicotinamide Adenine Dinucleotide Phosphate
NCBI	National Centre of Bio-Informatics
NF	Nuclease Free
NK	Natural Killer
Nm	Nanometer
ORF	Open Reading Frame
PAGE	Poly-Acrylamide Gel Electrophoresis
PAMP	Pathogen-Associated Molecular Pattern
PBMC	Peripheral blood mononuclear cell
PBS	Phosphate-buffered saline

PDB	Protein Data Bank
pDCs	Plasmacytoid DCs
PKC	Protein Kinase C
PMA	Phorbol Myristate Acetate
Pmol	Picomole
PMS	Phenazine Methyl Sulphate
PRR	Pattern Recognition Receptors
PVDF	Polyvinylidene Difluoride
qPCR	Quantitative Polymerase Chain Reaction
QRT-PCR	Quantitative-Reverse Transcriptase Polymerase Chain Reaction
RA	Rheumatoid Arthritis
RBC	Red Blood Cell
RLR	RIG-I-Like Receptor
RNA	Ribonucleic Acid
ROS	Reactive Oxygen Species
Rpm	Revolutions Per Minute
RT	Room Temperature
RyR	Ryanodine Receptor
S.D	Standard Deviation
SDS	Sodium Dodecyl Sulphate
SEB	*Staphylococcal Enterotoxin B*
Ser	Serine
SIV	Simian Immunodeficiency Virus
TCR	T Cell Receptor
TGN	Trans Golgi Network
Thr	Threonine
TLR	Toll-Like Receptor
TMD	Transmembrane Domain
TNF	Tumor Necrosis Factor
Trx	Thioredoxin
Tyr	Tyrosine
µl	Microliter

UV	Ultra Violet
VMD	Visual Molecular Dynamics
VUE	Villitis of Unknown Etiology

THESIS PUBLICATIONS

1. **Malik, U.**, Javed, A., Ali, A., & Asghar, K. (2017). Structural and functional annotation of human FAM26F: a multifaceted protein having a critical role in the immune system. *Gene*, *597*, 66-75. https://doi.org/10.1016/j.gene.2016.10.029

2. **Malik U.**, Zafar S., Younas N., Zerr I., & Javed A. (2020). Unveiling the Physical and Functional Niches of FAM26F by Analyzing Its Subcellular Localization and Novel Interacting Partners. *ACS Omega*, 5, 22008-22020. https://doi.org/10.1021/acsomega.0c01249

3. **Malik, U.**, & Javed, A. (2016). FAM26F: an enigmatic protein having a complex role in the immune system. *International Reviews of Immunology*, 1-11. https://doi.org/10.1080/08830185.2016.1206098

ABSTRACT

Human immune system is a multifaceted entity capable of defending the human body against microbial infections by utilizing a network of cells and proteins, destructive enzymes and chemical mediators. About a decade ago, FAM26F was recognized as a protein differentially expressed in various anomalies. At present, it has gained much significance as being a critical modulator in diverse immune responses. Despite the potential role of FAM26F in immune modulation and hence its therapeutic potential, the knowledge of its subcellular localization and interacting partners which can shed light on its specific function is still lacking. The current study was aimed 1) to functionally characterize FAM26F through computational methods; 2) to determine FAM26F's subcellular localization and its interacting partners in order to decipher the particular pathway which is regulated when FAM26F is expressed in a cell; and lastly 3) to analyze its expression with reference to its key interactor in normal and pathological condition to get an insight about its behavior and mechanism of action. Numerous advanced online computational tools were employed for the in silico characterization of FAM26F. Immuno-fluorescence was conducted on HEK293 cells transfected with GFP-tagged Human FAM26F plasmid followed by confocal laser scanning microscopy to determine the subcellular localization of FAM26F. Subsequently, co immunoprecipitation of FAM26F complex was performed using G coupled magnetic Dyna beads. The interactors of FAM26F were identified by ESI-QTOF MS/MS, and its signaling pathway was then determined through Reactome Pathway Database and Ingenuity Pathway Analysis software. Finally, qPCR analysis was performed to determine the expression of FAM26F and its chief interactor in the normal individuals and HIV patients. The *in silico* results revealed FAM26F to be a 315 amino acid long, stable protein that has remained well-conserved throughout evolution. It is a signal peptide deprived transmembrane protein that is secreted through non-classical pathway. The presence of a single well-conserved Ca_hom_mod domain indicated FAM26F to be a cation channel involved in the transport of molecules. A potential N-glycosylation and 14 phosphorylation sites were also predicted. Moreover, the presence of an immunoglobin-like fold in FAM26F emphasized its role in immune responses. The immuno-fluorescence results revealed FAM26F to be largely localized within the Golgi apparatus of the

cell. However, its minor presence in endoplasmic reticulum (ER) pointed towards the probable retrograde transfer of FAM26F from Golgi to ER during adverse conditions. The co-immunoprecipitation and MS/MS results demonstrated a total of 85 proteins, 44 of which significantly co-purified with FAM26F. Interestingly, more than half of these 44 proteins were involved in innate immune system. Further characterizations showed that FAM26F largely interacts with proteins mediating calcium homeostasis of a cell, particularly with Thioredeoxin, which essentially paved the way for depicting its mechanism of action under stress/disease conditions. It is proposed that activation and inhibition of cellular immune response is essentially dependent on whether FAM26F or Thioredoxin considerably interacts with CD30R. Furthermore, the expression analysis of FAM26F along with Thioredoxin in healthy individuals and HIV patients revealed significant downregulation of FAM26F and upregulation of Thioredoxin in HIV patients as compared to controls. As initial stages of the HIV-1 life cycle considerably depend on Trx1, whereas FAM26F is a potent immune regulator and antiviral agent, new generation of anti-HIV therapeutics based on silencing of the Trx1/TrxR1 system and inciting FAM26F expression, as well as targeting the viral surface glycoproteins, may represent a promising approach for HIV treatment. The study essentially and comprehensively demonstrates the presence of FAM26F is the Golgi apparatus, and the involvement of FAM26F in regulating Ca^{2+} signaling of a cell. The future need to study FAM26F in other cell lines and animal models (both normal and diseased) in the context of immune dysfunction is emphasized. Moreover, knockdown studies to identify the specific Ca^{2+} signaling genes regulated by FAM26F expression should be explored. By highlighting the role of FAM26F as a potential therapeutic target in infections, cancers and immune diseases, it is proposed that combination therapies should be employed wile devising the treatment regime for deadly diseases, such as HIV.

Chapter 1

INTRODUCTION

'Immunity' refers to the general capability of the host to resist the attack of microorganisms that would otherwise destroy it. Proteins are considered vital players in controlling the complex defensive networks to generate diverse and plastic immune responses. In the domain of immune system, FAM26F (Family with sequence similarity 26, member F) is a relatively newly identified protein that has gained much significance in the past few years as being critical in modulating diverse immune responses. So far, there are only three studies that provide a brief overview of the FAM26F's function. In 2010, FAM26F was recognized as a TLR signal-derived membrane molecule, which was found to modulate mDC–NK contact-mediated NK activation. Consequently, it was suggested and emphasized that owing to the NK cells activation, FAM26F possesses the capability to serve as a therapeutic for the tumors that are NK sensitive (Ebihara et al., 2010). Another study carried out by the same group revealed that expression of FAM26F on surface of immune cells facilitates the yield of interferon-gamma (IFN-γ) through the NK cells, thus anticipating FAM26F to be a novel target molecule for immunotherapy against IFN-γ suppressible tumors (Kasamatsu et al., 2014). Moreover, investigating its function in SIV-infection showed that pre-infection levels of FAM26F are inversely correlated with general viral load of plasma and thus FAM26F can be regarded as one of the earliest prognostic markers which, in the infection's early stage, can give us information related to the strength and pace of antiviral immune response (Javed et al., 2016).

Apart from these, numerous whole transcriptome analyses have detected FAM26F to be differentially expressed. The examples include a range of clinical studies primarily associated with inflammatory response (Defamie et al., 2008; M. J. Kim et al., 2009; Pankla et al., 2009; Shahzad et al., 2010), in melanoma patients (Ulloa-Montoya et al., 2013) and in hepatitis C virus clearance (Grimes et al., 2013). Upregulation of FAM26F can occur as a result of the interaction among various signaling pathways, including stimulation of TLR3 via poly I:C or TLR4 receptor (Chmielewski et al., 2014; Ebihara et al., 2010; Lee et al., 2014); stimulation of dectin-1 pathway (Chiba et al., 2014); upon exposure to IFN-β (Lee et al., 2014);

upon exposure to IFN-γ alone (J. N. Brown et al., 2010; Chmielewski et al., 2014; Javed et al., 2016) or by the combined stimulation of IFN-γ with either lipopolysaccharide or IFN-β (Chmielewski et al., 2014; S. Zhang et al., 2010); and after infection with murine cytomegalovirus (Manh et al., 2013). Moreover, deletion of mice IFN-α and IFN-β receptors retracted the Poly I:C stimulated induction of FAM26F (Kasamatsu et al., 2014). FAM26F expression in dendritic cells/macrophages was also lost or significantly reduced as a result of deletion of IRF-3 and TICAM-1/TRIF (Ebihara et al., 2010) or IRF-5 (Chiba et al., 2014) which consequently led to inadequate activation of NK cells and thus affected their cytolytic function. Hence it is anticipated that FAM26F may be a significant regulator of immune response whose expression may signify an activated immune system and which may be involved in important immune signaling cascades.

Although there is considerable evidence about the potential role of FAM26F in immune modulation, yet the protein had not been characterized or thoroughly studied to determine its exact function and modulatory pathways. Hence the first phase of the study was to completely annotate the structural and functional features of FAM26F using an innovative *in silico* approach, which was helpful in characterizing the gene and assigning it a probable function.

One of the key tasks of the post genomic period is to functionally characterize the cellular proteins. While *in silico* analysis can give a general insight into the characteristics, proteome analysis can reliably annotate these proteins for determining their interaction partners and functionalities in the cellular environment (Dönnes & Höglund, 2004). Yet, the foremost step in this regard is to determine subcellular localization of each protein in order to demonstrate its working environment within a cell. It impacts protein function by governing the availability and access to various molecular interaction partners (Scott et al., 2005). Hence, the second phase of the study was focused to determine the subcellular localization and to find the interacting partners of FAM26F through experimental procedures for apprehending its function and role in the immune responses.

Dysfunction of the immune system can lead to several diseases which are largely categorized into two classes, immunodeficiency syndromes and immunopathology. Infectious diseases, the main cause of immunodeficiency, are a leading cause of

illness and fatality globally and present a great task for the biomedical sciences. The third phase of the study thus involved the investigation of the association between FAM26F and its chief interactor identified in the second phase in the clinical samples of HIV. The difference in the expression of the two genes was determined and compared among the controls as well as HIV samples to see the trend being followed in the two groups, which helped to delineate their relationship and endorse their relative mode of action as proposed in the second phase.

Utilizing the immune system's countless potential as a therapeutic strategy offers an innovative, unique and powerful treatment strategy for patients with infectious diseases like HIV. Hence, treatment regime of patients with elevated infection must include prompt short-term immune reconstruction to restrain the disease progression as well as aggressive antiviral treatment to attain quick viral suppression. Protein therapeutics has the potential to incite an early innate immune response to immediately defend the host from infection without mounting the long term memory. Through the results, the current study also signifies and proposes the potential of FAM26F to act as a therapeutic against HIV.

Chapter 2

LITERATURE REVIEW

Human immune system is a complex amalgam comprising of a network of cells and molecules, tissues, organs and proteins, all of which have specialized roles that protect the human body against a broad range of threats such as infection by viruses, bacteria, fungi and multi-cellular parasites as well as extraneous substances (thorn, flames etc.). The immune system provides barriers to invasion and uses a network of cells, destructive enzymes and chemical mediators that can be exclusively recruited to destroy the attacking pathogens (Howell & Shepherd, 2018). However, in certain circumstances, such attacks to our defenses can lead to impaired immunity, which is associated with illness or disease.

2.1. Immune Response

Immune system responds to harmful bodies instead of to those that are merely extraneous (Gallucci & Matzinger, 2001). Upon receiving a potential 'stress' or 'danger' stimuli, the immune cells of an organism are activated to mediate an immune response (Rincón & Davis, 2009). This activation, which further initiates primary and secondary immune responses, can be triggered by either the endogenous or the exogenous danger signals. Former signals are released by tissues enduring stress, injury or unusual death, whereas later signals are provoked by pathogens such as bacteria and viruses. Some recently identified endogenous danger signals include reactive oxygen intermediates, heat shock proteins, nucleotides, breakdown products of extracellular-matrix, neuro-mediators and cytokines like the interferons (IFNs) (Gallucci & Matzinger, 2001).

Danger signals can primarily elicit two different kinds of responses. Innate (natural) response is a prompt response which progresses rapidly and precedes the laborious clonal expansion of lymphocytes specific to the antigen (acquired response) (Ismail et al., 2002). The extent of innate response remains the same no matter how many times the stimuli is received, whereas acquired/adaptive responses improve on subsequent exposures to a given stimulus. The innate responses are carried out by phagocytic cells (monocytes, macrophages and neutrophils), cells releasing

mediators of inflammation (mast cells, basophils, and eosinophils), and natural killer (NK) cells, whereas it's molecular modules include acute-phase proteins, complement, and cytokines such as the IFN. Acquired responses encompass the propagation of antigen-specific T and B cells, upon binding of their surface receptors to antigen. Specialized cells, known as antigen-presenting cells or APCs, cooperate with the lymphocytes in response to the antigen by displaying the antigen to them. B cells secrete antigen-specific antibodies called immunoglobulins which account for the elimination of extracellular microorganisms. T cells not only assist B cells to produce antibody, but they also trigger the macrophages and kill the virally infected cells to remove intracellular pathogens (Delves & Roitt, 2000).

Mammalian innate and adaptive immune responses usually work together to eradicate pathogens (Delves & Roitt, 2000). However, activation of these responses must be strictly controlled by intricate mechanisms to regulate their inception and termination (Taganov et al., 2006). The amount and extent of an immune response depends on the combination of responses facilitated by effector and regulatory T cells. For instance, different signaling pathways activated by B-cell receptors (BCRs), Toll-like receptors (TLRs) and cytokine receptors, delivering positive signals to immune cells, each have their own counterbalancing systems. This negative-feedback is crucial for the appropriate homeostasis of an immunological response (Yoshimura et al., 2007).

2.2. Innate Immune System

Innate immune system denotes first line of host defense that intends to distinguish unique pathogen-associated molecular patterns (PAMPs) embodied by conserved components of microorganisms. The immune cells, such as dendritic cells and macrophages, express a limited number of germline-encoded receptors, collectively called as pattern recognition receptors (PRRs), which detect these pathogen molecules (Taganov et al., 2006). There are different sets of PRRs responsible for tailoring innate responses including the transmembrane TLRs and the RIG-I-like receptors (RLRs). TLRs recognize microbial by-products present in the extracellular space while RLRs detect infection within the cytosolic part (Takeuchi & Akira, 2010; Wilkins & Gale Jr, 2010). This recognition of microbial components by different PRRs and subsequent binding of these PRRs to their respective ligands

activates a broad spectrum of signaling cascades, from phagocytosis to the yield of antimicrobial cytokines (Li et al., 2011), which sequentially shapes and augments the inflammatory and acquired immune responses (Taganov et al., 2006).

2.2.1. Components of Innate immunity

Innate immune system is an assortment of various modules, or subsystems, which play diverse but corresponding roles in defending the host against intracellular microbes (Ismail et al., 2002; Medzhitov, 2007). It consists of cellular components including non-phagocytic cells and phagocytic cells present in the tissues and circulation, as well as the molecular components comprising complement and several plasma proteins, and the cytokines (Ismail et al., 2002).

2.2.1.1. Cellular components of Innate immunity

Non phagocytic cells

Epithelial cells, residing at the common entrances like skin and the respiratory, gastrointestinal as well as the genitourinary tract, represent the primary defense against intracellular pathogens. Besides providing physical and chemical obstructions to infection, epithelial cells also yield antimicrobial peptides for killing bacteria. Endothelial cells, which are the main target of Rickettsia, effectively kill the intracellular pathogens by initiating mechanisms analogous to the ones used by specialized phagocytes (Ismail et al., 2002). Examples of non-phagocytic cells include epithelial cells, endothelial cells, cardiac myocytes, hepatocytes, etc.

Phagocytic cells

Phagocytic cells can distinguish between “foreign” and “self ” molecules (Delves & Roitt, 2000) by recognizing the carbohydrates that are usually not exposed on vertebrate cells, consequently engulfing the pathogens through PRRs such as integrin, mannose, scavenger and TLRs, and through receptors for peptides that contain N-formyl methionine. Phagocytes also possess receptors for antibodies and complement, which further increase the phagocytosis by antibody and complement-facilitated exclusion of intracellular pathogens (Ismail et al., 2002). Phagocytes also eliminate the dead or dying cells of the body. In necrotic tissue, dying cells can

discharge materials that elicit an inflammatory response (Delves & Roitt, 2000). Examples of phagocytes include macrophages, monocytes and neutrophils.

Cells secreting inflammatory mediators

Being only weakly phagocytic, the Eosinophils upon activation perhaps release cationic proteins and reactive oxygen metabolites into the extracellular space to primarily kill the parasites. They also release prostaglandins, leukotrienes, and several cytokines (Wardlaw et al., 1995).

Basophils and mast cells are functionally similar (Abraham & Arock, 1998). They both possess high-affinity IgE (Fc*e*R) receptors (Kinet, 1999) and thus get covered with IgE antibodies. These cells have essential role in atopic allergies such as asthma, hay fever, and eczema, in which binding of allergen to IgE results in the cross-linking of Fc*e*R. This event results in the secretion of inflammatory mediators such as prostaglandins, leukotrienes and histamine (Delves & Roitt, 2000).

Dendritic cells

Dendritic cells including the skin Langerhans' cells quietly but consistently endocytose extracellular antigens. They get activated and act as APCs when their surface PRRs recognize unique PAMPs on the microbe surface (Medzhitov & Janeway Jr, 1997). Moreover, endogenous danger signals, for instance interferon-α released from cells infected with virus or an upsurge in heat-shock proteins caused by necrotic cell death, also activate the dendritic cells (Matzinger, 2002).

Natural killer cells

NK cells essentially kill the intracellular pathogens, mostly viruses (Medzhitov, 2007), by either cytotoxically attacking the infected target cells or by triggering the macrophages by the production of IFN-γ. NK usually exerts its cytotoxicity by secreting granules having granzymes and perforin or by inducing death receptor-mediated apoptosis (Ismail et al., 2002).

2.2.1.2. Molecular components of Innate Immunity

Innate responses often contain acute-phase proteins, complement, and cytokines.

Acute-phase proteins are released by hepatocytes when triggered by pro-inflammatory cytokines IL-6 and IL-1β. A crucial element of this response is the secretion of PRRs: collectins, pentraxins and ficolins41–43. The main functions of acute-phase proteins are to opsonize the microbial cells for phagocytosis and activation of complement system (Medzhitov, 2007). They increase resistance to infection and encourage the restoration of injured tissue. Hence serum concentrations of these proteins rise markedly in response to infection, tissue injury, and inflammation. The acute-phase proteins contain some complement components, C-reactive protein (a valuable marker of inflammatory diseases like rheumatoid arthritis), proteinase inhibitors, serum amyloid A protein, and coagulation proteins (Delves & Roitt, 2000).

Complement activation leads to the generation of various immunologically active substances (Delves & Roitt, 2000) which result in a cascade of events, including: opsonization of pathogens aided by covalent binding of C3 fragments; assembly of phagocytes at the site of infection through the chemotactic activity of the released C4 and C5 proteolytic fragments; and direct pathogen killing by the organization of membrane-attack complex, which is the last component of the complement system (Carroll & Fischer, 1997).

Cytokines behave as messengers not only within the immune system, but also amid immune and other body systems, establishing an assimilated network that greatly affects the immune response regulation. A cell possesses specific cytokine receptors that sense the presence of a cytokine. Besides behaving as messengers, some cytokines provide direct protection; for instance, the interferons expelled from virally infected cells cause the surrounding cells to be in a virus resistant state (Delves & Roitt, 2000).

2.3. Acquired immune system

The adaptive immune system represents specific type of immunity in that it develops extremely targeted responses to foreign antigens. Unlike the innate immunity, it experiences a selectivity procedure based on exposure to foreign substance and ‘adapts’ by developing exceedingly precise receptors for the attacking pathogens (Howell & Shepherd, 2018).

After exposure to pathogen, cells and inflammatory mediators from infection sites infiltrate the nearby lymph nodes where they are presented by the APCs. These non-host antigens displayed by the APCs are spotted by host receptors of B and T lymphocyte. After a complex process of clonal selection, a highly selective cell population (clone) is achieved that has the ability to target the original pathogen. Hence the primary response of acquired immunity is a bit slow as compared to innate immunity; nonetheless, immune memory delivers a mechanism for enhanced growth that result in a further swift and significant response to successive contact with the same pathogenic antigen. This mechanism has been exploited clinically in vaccine immunizations (Howell & Shepherd, 2018).

The adaptive immune system comprises of humoral and cellular modules (Howell & Shepherd, 2018) characterized by B and T lymphocytes respectively, that clonally express a huge collection of antigen receptors that are formed by somatic recombination at specific sites, namely antibody/ BCR and TCR. Functionally, naive B and T cells come across antigens in dedicated lymphoid organs and experience the cell division and development prior to carrying out their effector function (Vivier et al., 2011).

2.3.1. Humoral Immunity

Humoral immunity defines the response of B-lymphocytes. Binding of the non-host antigen of an attacking pathogen to the B cell antigen receptor encompasses a molecular signaling cascade that stimulates the activation of B-cells. Activated B cells either segregate into plasma cells or end up becoming memory B cells. Plasma cells produce and secrete specific pathogen binding antibodies. Memory B cells remain inactive until they experience a secondary exposure to the antigen, when they will identify this specific pathogen more rapidly. They can also play a role as APCs (Tangye & Tarlinton, 2009).

An antibody can belong to five immunoglobulin isotypes or classes which, in turn, govern the ensuing immune response. Thus IgM and IgG trigger the activation of complement and cell lysis; IgG, IgM and IgA neutralize the bacterial toxin, IgG, IgA and mast cell cause antiviral activity, and IgE activation leads to basophil degranulation. These are known as class specific functions (Lu et al., 2018).

2.3.2. Cellular Immunity

Cellular immunity defines the response of T cells. Naive T cells emerge within the bone marrow and fetal liver, prior to moving to the thymus for development and maturation of their TCR. These receptors can recognize the variations in MHC molecules present on the cell surface and hence discriminate amongst self and non-self (Chaplin, 2010).

Although, T-cells get activated when the TCR identifies the complex of MHC molecules with non-self antigens on APCs (phagocytes and dendritic cells), however, the fate of this activated T cell depends on the family of co-stimulatory molecule which is simultaneously activated. Effector functions can be of cytotoxic, helper and regulatory in nature, each having distinct impact on immune defense (Howell & Shepherd, 2018).

2.3.2.1. Helper T cells (Th)

These cells possess surface CD4 molecule and regulate both humoral and cellular immunity. Two subcategories of Th cells are known, Th1 and Th2. Th1 type of immune response is the standard immunity against intracellular bacteria and viruses. Th1 encourages the production of B cell IgG by the secretion of interleukin 2 and IFN-γ (amid other cytokines) and also activates macrophages and effector CD8 T cells. Th2 targets parasitic infection and stimulates the production of B cell IgE and the activation of eosinophils, basophils and mast cells (Sallusto & Lanzavecchia, 2009).

2.3.2.2. Cytotoxic T cells (Tc)

These cells possess CD8 molecule on their surface. They are effector cells and hence destruct the tissue having intracellular pathogens or cancer cells by inducing apoptosis via secretion of perforins, which produce pores in the transmembrane of target cell. Apoptosis is also induced by Tc released granzymes when they enter the cell cytosol (Ismail et al., 2002).

2.3.2.3. Regulatory T cells (TReg)

These cells help in regulating immunity by reducing the activity of both helper and cytotoxic T cells. This family is important in creating tolerance of self-antigens and avoiding auto-immunity. Their action depends on a subgroup of co-stimulatory molecules that trigger co-inhibitory pathways when T cells are activated (Howell & Shepherd, 2018).

2.4. Failure of the Immune System

The immune system sustains an adequate equilibrium in order to deliver ideal protection for the host. Any disturbance to this balance results in immune diseases, categorized into two broad classes:

2.4.1. Immunodeficiency Diseases

These diseases, which may be congenital or acquired, arise due to the lack of single or many components of the immune system. The host will be highly vulnerable to environmental pathogens due to the loss of protective attributes of the immune system. For instance, the elevated incidence of drug allergy associated with HIV infection is thought to arise due to lack of Th regulation of immunity and an increased background level of overall serum IgE (Howell & Shepherd, 2018).

2.4.2. Immune dysfunction (hyper-reactivity or immunopathology)

This dysfunction arises when a specific immune response is harmful to the host. This pathological process might be a reaction to either self or non-self antigens. For instance, autoimmune disorders occur due to the loss of self-tolerance (false recognition of host cells as extraneous) (Howell & Shepherd, 2018).

2.5. FAM26F

As described earlier, the mammalian immune system represents a dynamic organization comprising of various cellular and molecular networks that work together for an effective host defense. Proteins play a vital role in controlling all these complex networks either by providing assistance to the immune system or being involved in immune responses by behaving as signaling molecules, key

transcriptional regulators or surface receptors. This continuous level of functional multiplicity is essentially responsible for generating the diverse and plastic immune responses. The advancement of molecular and structural biology techniques has led to the addition of numerous new proteins in the list. One such relatively recently identified protein involved in various immune modulating responses is FAM26F.

Previously cited as IRF-3–dependent NK-activating molecule (INAM) (Ebihara et al., 2010), FAM26F has gained much significance in the past few years as being a critical player in various infections, stimulation studies, cancer and immune pathogenesis. A few functional and numerous whole transcriptome analyses have detected differential expression of FAM26F. However, its stipulation as an immune protein can be better understood by analyzing the following comprehensive details.

2.5.1. FAM26F Expression on Various Immune Cells

Various studies have described the association of FAM26F with the immune cells. FAM26F was found to be expressed on several Rhesus macaques immune cell populations namely CD4+, CD8+ and CD20+ B cells, with maximum expression on CD4+ cells, along with IP-10, tetherin and MX1 which are well known for their role in innate immune responses (Javed, 2012). Consistent with these results, a microarray analysis and its subsequent validation by qPCR to anticipate the clinical response to glucocorticoids (GC) in Rheumatoid arthritis (RA) therapy revealed that FAM26F was significantly upregulated in $CD4^+$ T-cells of GC responders (Fritsch, 2014). Another study involving TB patients also confirmed the expression of FAM26F on T cells (Matsumiya et al., 2014).

Two other studies conducted on mice models employing expression profiling and analyzing gain/loss-of-function demonstrated that FAM26F is expressed in the NK cells as well as in the accessory cells like CD8α(+), conventional DCs and macrophages (Ebihara et al., 2010; Kasamatsu et al., 2014).

All these studies represent FAM26F as a gene that is substantially expressed in various immune cell populations, implying its role in modulation of diverse immune responses.

2.5.2. FAM26F: An Interferon Responsive Gene

In 2010, Zhang and co-workers studied diverse macrophage activation in response to cytokines and showed that FAM26F was induced by more than one cytokine (IFN-B, IFN-γ and IL-10) (S. Zhang et al., 2010). In another study, FAM26F was reported to be significantly up-regulated in human hepatocytes by IFN-α and IFN-γ at 6 h and 18 h post stimulation respectively (He et al., 2010).

In vitro stimulation studies of SIV infected rhesus macaques derived peripheral blood mononuclear cell (PBMCs) with IFN-γ and IFN-α2 revealed that only IFN-γ stimulation of PBMCs resulted in an increase in RNA expression of FAM26F after 6 to 12 hours post stimulation. Moreover, FAM26F followed the activation and expression kinetics similar to CXCL10, a known interferon type II regulated gene (Javed, 2012).

In another study, the expression of FAM26F was reported to be highly affected by IFNγ stimulation, was mildly responsive to LPS, and highly elevated after joint treatment of vascular cells. In immune cells, STAT1 represents distinctive site of union for antimicrobial and inflammatory synergy amid IFNγ and TLRs. Amalgamation of TLR and IFNγ signaling pathways is brought about by the interaction between transcription factors induced by TLR- and IFNγ. Promoter analysis of FAM26F predicted the presence of ISRE (interferon-stimulated response element) and STAT binding sites in the region, suggesting FAM26F as an IFN-γ responsive gene (Chmielewski et al., 2014).

All these studies indicate that expression of FAM26F can be induced in both auto or paracrine manner by different cytokines/interferons, particularly IFN-γ, i.e. FAM26F can not only instigate the production of IFN-γ, but its expression can be additionally boosted by an auto feedback mechanism. Assuming that FAM26F either assists or that its expression level symbolizes the speed and spread of early IFN-γ-guided immunity (Javed, 2012), future research determining the role of FAM26F expression in the initiation of adaptive immune responses is critical. Besides NK-cells, activated $CD4^+$ T cells and γδ TCR^+ T cell subgroups can also result in the production of early IFN-γ (Devilder et al., 2009; Neves et al., 2010). The potential role of FAM26F in

activation of these cells as well as in locally amplifying the immune response by cell to cell contact shall be investigated further.

2.5.3. FAM26F: A possible Modulator of Inflammation in Infections and Injury

Transcriptome analysis studies have reported differential expression of FAM26F during various infections and other pathophysiological conditions (Table 2.1). These studies include various bacterial, viral and parasitic infections as well as injury and assaults that affect the immune system.

2.5.3.1. Bacterial Infections

Melioidosis, caused by *Burkholderia pseudomallei*, is a serious infectious disease with a 40% death rate even with suitable treatments. The results of a microarray based study identified FAM26F to be a top classifier gene and thus a candidate diagnostic signature discriminating *B. pseudomallei* mediated septicemic melioidosis from sepsis caused by other pathogens (Pankla et al., 2009). Another transcriptome analysis study indicated that treatment of human PBMCs with increasing concentrations of the *Staphylococcus aureus* super antigens induced >10 fold increase in FAM26F transcription along with other gene networks associated with inflammation, corresponding to the domination of highly robust Th1- and Th17 immune response (Grumann et al., 2008). Similarly, a genome-wide study was performed in six tissues for investigating the temporal response of mice when lethally challenged with intranasal administration of *Staphylococcal enterotoxin B* (SEB) causing toxic shock syndrome. FAM26F was identified as one of the 11 unique genes (Irf1, Irf8, Fam26f, Irgm2, Cxcl9, Cxcl10, Cxcl11, Cd274, Parp14, Stat1 and Serpina3g) showing highest expression at 5 hours post- SEB challenge in spleen and/or PBMCs (Ferreyra et al., 2014). Interestingly all other members in the group include genes that release products which induce the IFN pathway and thus are part of a host-wide IFN-response, proposing a similar function for FAM26F as well.

2.5.3.2. Viral infections

In a study conducted in SIV infected Rhesus macaques, FAM26F RNA levels in lymphocytes throughout the acute and post-acute stage of SIV-infection were found

to be a strong prognostic marker for viral replication. Notably, FAM26F expression correlated with total viral load during the acute infection phase, which included lessening of viremia by innate immunity and cytolytic responses of adaptive immunity (Javed, 2012). In HCV patients, FAM26F was among the 91 differentially regulated genes whose enhanced expression was potentially associated with HCV clearance (Grimes et al., 2013).

2.5.3.3. Parasitic Infections

FAM26F was acknowledged as one of the differentially expressed genes in the abomasal Lymph Nodes of ovine Breeds that are known to have varying resistance levels against Gastrointestinal Nematode infections. Interestingly, expression of FAM26F was attributed to the pathways that are well established to represent a more active immune profile which, along with others, include interferon signaling and antiviral innate immunity receptors. (Ahmed et al., 2015).

2.5.3.4. Inflammatory Responses

The gene expression profiles of 21 biopsies from human liver transplants identified FAM26F as one of the top classifiers from among 371 differentially expressed genes that are functionally linked with oxidative stress and inflammation, thus predictive of initially reduced graft function during orthotropic liver transfer (Defamie et al., 2008).

Villitis of unknown etiology (VUE) is a vicious inflammatory wound of villous placenta represented by predominant infiltration of maternal CD8+ T cells into the chorionic villi (Brito et al., 2005; J. S. Kim et al., 2008). A transcriptome analysis of 10 VUE placentas was conducted to determine the association of VUE with systemic inflammatory response(s) of mother and/or fetus. FAM26F was found to be significantly elevated besides numerous chemokines, MHC class I and MHC class II molecules, in placental transcriptome of VUE whose expression also increased corresponding the rigorousness of the inflammatory process (M. J. Kim et al., 2009).

In another study that was designed to test whether burn injury alters the physiology of the tracheal epithelial ultrastructure of a rat model with third degree burn covering 60% of complete body surface, FAM26F was identified as one of the 59 genes

significantly downregulated in the injured animals, thus leading to a decrease in mucociliary clearance (MCC) and cell proliferation, probably due to oxidative injury (Jacob et al., 2015).

All these studies very strongly sustain the belief that FAM26F plays an important role during inflammatory responses that result from infection or injury and thus it can act as a prognostic and diagnostic marker for various disease conditions. Detailed functional studies elucidating the exact role of FAM26F in each disease is a focus of future research.

Table 2.1: Expression of FAM26F during *in vitro* stimulation, various infections and pathophysiological conditions

Study Type	Mode of study	Expression	Sample	Detection Method	Reference
Stimulation studies					
GC therapy	*in vitro*	↑	monocytes and T cells from RA patients	MA	(Fritsch, 2014)
MVA85A vaccine	*in vitro*	↑	blood samples from TB patients	SAM statistical package, GSE analysis	(Matsumiya et al., 2014)
IFN-α and IFN-γ	*in vitro*	↑	human hepatocytes	cDNA MA	(He et al., 2010)
IFN-γ	*in vitro*	↑	PBMCs	qRT-PCR	(Javed, 2012)
IFNγ, LPS, and IFNγ + LPS treatment in human atherosclerosis	*in vitro*	↑ In IFNγ and IFNγ + LPS	VSMCs	MA, qRT-PCR	(Chmielewski et al., 2014)
Bacterial infections					
Septicemic melioidosi	*in vitro*	↑	Whole blood samples from 63 septic patients and 29 uninfected controls	MA	(Pankla et al., 2009)
SAgs treatment	*in vitro*	↑	PBMCs	MA	(Grumann et al., 2008)
SEB challenge	*in vivo*	↑ In spleen	murine model (PBMCs, spleen, lung, liver, kidney, and heart)	oligonucleotide MA	(Ferreyra et al., 2014)
Viral infections					
SIV	*in vivo*	↑	SIV infected rhesus macaques	qRT-PCR	(Javed, 2012)
HCV	*in vitro*	↑	16 anti-HCV antibody-positive individuals	qPCR, MA	(Grimes et al., 2013)
Parasitic infection					
GIN	*ex vivo*	↑	abomasal lymph node	RNAseq, qRT-PCR	(Ahmed et al., 2015)

		In Suffolk lambs	tissue of GIN-free Suffolk and Texel lambs		
Injury/Disease conditions					
Early liver graft failure	*ex vivo*	↑	seven control livers and 21biopsies of human liver transplant	long oligonucleotide MA	(Defamie et al., 2008)
VUE	*ex vivo*	↑	10 VUE placentas	MA, linear modelling	(M. J. Kim et al., 2009)
UC	*in vivo*	↑	Human samples, samples of colon from DSS colitis mouse model	MA	(Julia et al., 2014)
Burn Injury	*in vitro* and *in vivo*	↓	rat model of 60% TBSA third degree scald burn	TEM, IHC, MA	(Jacob et al., 2015)
Immune activation					
Macrophage activation	*in vitro*	↑ In primed macrophages	Elutriated monocytes from healthy hCMV-negative/HIV-negative donors	gel-based, size-fractionation approach, LC-MS/MS	(J. N. Brown et al., 2010; Chmielewski et al., 2014; Javed et al., 2016)
Autocrine signaling of TNF	*in vitro*	↑ In CpG DNA stimulated BMDMs	BMDMs from wild-type and tnf lacking C57BL/6 mice	MA	(Caldwell et al., 2014)
Oviduct development in Chickens	*in vitro* *in vivo*	↑ ↑	MCF-7 or HeLa cell lines DES pallet implanted one week old female chicks	Atlas Data or G2SBC database MA, qRT- PCR, in situ hybridization	(Mosca et al., 2010) (Song et al., 2011)
LPS or PMA-Ionomycin stimulation	*in vivo*	↓	PBMCs from rabbits	MA	(Jacquier et al., 2015)

Cancer studies					
Breast	-	(−) in BRCA2 associated breast tumors	Breast tumors	-	(Stefansson et al., 2009)
breast, mammary gland, cervix and uterus	*in vitro*	Differential expression	Atlas Data	MA	(Mosca et al., 2010)
Metastatic melanoma	*in vitro*	↑	melanoma biopsies from 75 patients with non-resectable MAGE-A3–positive stage III or IV M1a metastatic melanoma	MA, qPCR	(Ulloa-Montoya et al., 2013)
B16D8 tumor	*in vivo*	↑	B16D8 tumor-bearing mice	caliper	(Ebihara et al., 2010)

Abbreviations: ↑: upregulated; ↓ : downregulated; GC: Glucocorticoids; − : Deleted; RA: Rheumatoid Arthritis; MA: Microarray; TB: Tuberculosis; SAM: Significance Analysis of Microarrays; GSE: Gene Set Enrichment; cDNA: Complementary DNA; PBMCs: Peripheral Blood Mononuclear Cells; qPCR or qRT-PCR: Quantitative Real Time Polymerase Chain Reaction; IFN-α: interferon alpha; IFN-γ: interferon gamma; LPS: LipoPolySaccharide; VSMCs: Vascular Smooth Muscle Cells; SAgs: Staphylococcal Superantigens; SEB: *Staphylococcal Enterotoxin B*; GIN: Gastrointestinal Nematode; SIV: Simian Immunodeficiency Virus; VUE: Villitis of Unknown Etiology; UC: Ulcerative Colitis; DSS: Dextran Sodium Sulfate; TBSA: Total Body Surface Area; TEM: Transmission Electron Microscopy; immunohistochemistry (IHC); HIV: Human immunodeficiency virus; hCMV: human Cytomegalo Virus; LC-MS/MS: Liquid Chromatography-Mass Spectrometry; TNF: Tumor Necrosis Factor; BMDMs: Bone Marrow Derived Macrophages; DES: DiethylStilbestrol (estrogen analog); PMA: Phorbol Myristate Acetate.

2.5.4. FAM26F Expression Represents an Activated Immune State

Previously, in the Atlas database for gene expression of unidentified genes (Atlas Data or G2SBC database), FAM26F expression was found to be related to immune responses and was reported to be significantly elevated by estradiol treatment in HeLa or MCF-7 cell lines (Mosca et al., 2010). In 2011, Song and co-workers sought to find the genes and pathways required for the development of chicken oviducts by using female chicks treated with subcutaneous implants of synthetic estrogen analog, diethylstilbestrol (DES). Microarray analysis followed by validation through real-time PCR and in situ hybridization identified FAM26F along with four other genes (CCRN4L, HAS2, NELF, and NTM) to be associated with the action of estrogen in potentially regulating the development, growth, and differentiation of female reproductive tracts (Song et al., 2011).

In a transcriptomic analysis executed on bone marrow derived macrophages, FAM26F was identified as one of the many adaptive immune control genes whose expression was TNF-dependent when stimulated with CpG DNA, in which autocrine TNF prolongs the period of NFkB action and regulates the CpG-induced gene expression programs (Caldwell et al., 2014).

Macrophage activation can result in a wide-range of functional outcomes (Mosser & Edwards, 2008). An activation-specific fingerprint is defined by the unique components possessed by the proteomes that are induced by IFN-γ- primed vs. LPS-activated macrophages. Among numerous other proteins, FAM26F was also identified as activation-specific fingerprint that distinguished primed macrophages from resting (basal) or LPS-triggered macrophages by primarily contributing to protein metabolism, protein transport and immune response (J. N. Brown et al., 2010; Chmielewski et al., 2014; Javed et al., 2016). In another genome-wide expression study carried out in rabbits to examine the response of PBMCs after being stimulated *in vitro* by LPS or phorbol myristate acetate (PMA) and ionomycin, FAM26F came out to be the highest downregulated gene 4 hours post PMA – Ionomycin stimulation of the PBMCs. The result might be due to the limited set of monocytes and macrophages targeted by the LPS as compared to wider number of target cells for PMA-Ionomycin. Nevertheless, the observations reflected that the actions of all the genes combined to give a concerted immune response (Jacquier et al., 2015).

FAM26F is thus differentially expressed during the activation state of immune system, though its exact role in this state is yet to be determined.

2.5.5. FAM26F Contribution in Anti-Tumor Pathways

In 2009, Joosse and group demonstrated that regions of genome carrying RARSL (6q15) and FAM26F (6q22.1) were recurrently deleted in breast tumors associated with BRCA2 (Joosse et al., 2009). In the subsequent year, Atlas Data reported FAM26F to have differential expression in cancer micro-array studies of numerous organs, including the breast, mammary gland, cervix and uterus (Mosca et al., 2010). In metastatic melanoma, FAM26F was identified as one of the 84 genes whose expression was potentially associated with clinical advantage via immune-based MAGE-A3 therapy. The analysis of results revealed that FAM26F expression is mainly driven by IFN-γ with STAT1 pathway (Ulloa-Montoya et al., 2013).

Takashi and colleagues documented FAM26F as a membrane molecule derived by TLR signaling that modulated NK activation in mice through mDC–NK contact and named it INAM (Ebihara et al., 2010). Initially, FAM26F was only marginally present on myeloid dendritic cells (mDC) and NK cells, however co-culture of mDC previously exposed to TLR3 ligand polyI:C and NK cells greatly enhanced FAM26F expression on both cell types. Induction of FAM26F was dependent on the activation of TICAM-1 and IRF-3, as TICAM-1-/-or IRF3-/-knockout mDC were unsuccessful in inducing full NK cytotoxicity. Interestingly, FAM26F also assisted in the reciprocal activation of mDC–NK via its cytoplasmic tail, which was critical for activating NK cells but not for mDC maturation. Subsequently, the adoptive shifting of mDCs expressing INAM into mice that were embedded with NK-sensitive tumors resulted in NK-facilitated reversion of the tumor. Thus it was strongly proposed that by activating NK cells, INAM has medicinal value against NK-sensitive tumors. Additionally, FAM26F activation also triggers the NK cells to produce IFN-γ, making it easier to speculate the pathway by which the tumors may probably be suppressed. Recently, the same group (Kasamatsu et al., 2014) provided further insight into the role of FAM26F by showing that FAM26F is expressed on the surface of CD8α+ cells, conventional macrophages and DCs which facilitate production of IFN-γ from NK cells. FAM26F deficiency leads to decreased IFN-γ production from NK cells and also their accessory cells. Accordingly, it is

anticipated that FAM26F may represent a novel immunotherapeutic target against IFN-γ–suppressible tumors. In a subsequent study, Chiba *et al.* proposed an alternate pathway for NK activation which also included FAM26F. DCs and macrophages harboring a PRR called Dectin-1 activate NK cells for their tumoricidal action. INAM along with other unidentified molecules is portrayed for the DC-facilitated activation of NK cells via Dectin-1-IRF5-INAM pathway (Chiba et al., 2014).

An amalgam of the activation and downstream anti-tumor signaling of FAM26F in the light of above mentioned studies is depicted in Figure 2.1.

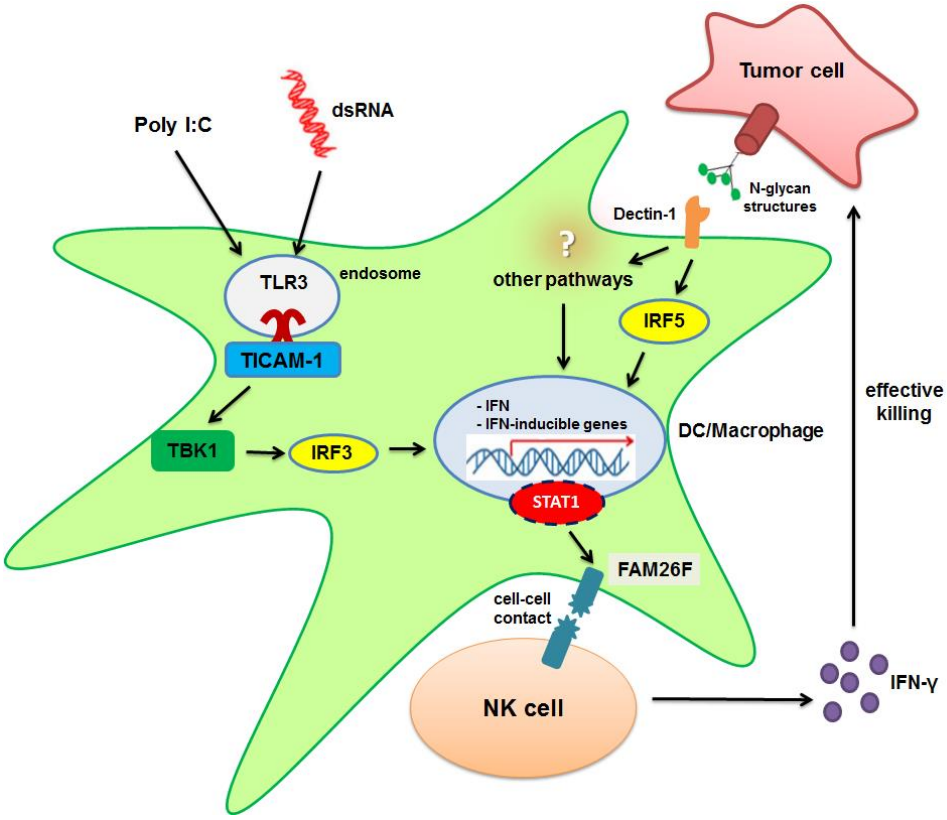


Figure 2.1 Anti-tumor activation pathways of FAM26F. Activation of dendritic cells/macrophages either via the TLR3–TICAM-1 pathway in response to the synthetic analogue polyI:C and viral dsRNA, or via Dectin-1-IRF5 pathway in response to N-glycan structures expressed on tumor cells, leads to enhanced cell surface expression of FAM26F probably through the activation of STAT1, which enables the interaction of mDC with NK cells and their activation via cell–cell contact, which in turn produce IFN-γ for effective killing of the tumor cells (Chiba et al., 2014; Chmielewski et al., 2014; Ebihara et al., 2010; Kasamatsu et al., 2014; Ulloa-Montoya et al., 2013).

2.5.6. An insight into FAM26F signaling mechanism

The cumulative findings of all the studies conducted on FAM26F so far provides us an insight into its transcriptional regulation. Numerous pathways are likely to act synergistically for the upregulation of FAM26F expression. For instance, the expression of FAM26F increases on DCs after TLR3 stimulation by poly I:C (Ebihara et al., 2010), after TLR4 stimulation via LPS exposure (Chmielewski et al., 2014), upon IFN-β exposure (S. Zhang et al., 2010) and by stimulating the dectin-1 pathway (Chiba et al., 2014). The involvement of IRF-3 and TICAM-1/TRIF (Ebihara et al., 2010) or IRF-5 (Chiba et al., 2014) is probable as their deletion causes the loss or significant reduction in FAM26F expression in DCs/macrophages and also impairs the activation and cytolytic role of NK-cells. Moreover, removal of IFNRA1 (receptor of IFN-α and IFN-β) in mice abrogated the induction of FAM26F by polyI:C (Kasamatsu et al., 2014). TNF silencing in macrophages derived from bone marrow also lead to elimination of FAM26F expression after CpG stimulation (Caldwell et al., 2014). All these pathways can jointly activate STAT1 which plausibly drives FAM26F transcription (Chmielewski et al., 2014). Furthermore, direct knock-out of FAM26F in mice considerably decreased the initial IFN-γ production post polyI:C treatment and resulted in decreased eradication of tumor cells (Kasamatsu et al., 2014). While the role of IFN-γ in the transcription of FAM26F in dendritic cells is still unclear, IFN-γ alone (J. N. Brown et al., 2010; Chmielewski et al., 2014; Javed et al., 2016) or together with LPS or IFN-β (Chmielewski et al., 2014; S. Zhang et al., 2010) leads to an increased level of FAM26F RNA in blood lymphocytes (Javed, 2012), vacuolar smooth muscle cells (Chmielewski et al., 2014) and macrophages (J. Brown et al., 2010; S. Zhang et al., 2010). However, the effect of FAM26F expression in immune cells by further promoting an IFN-γ response via FAM26F facilitated cell-cell contact is still uncertain and needs additional exploration.

2.5.7. Significance of Baseline Expression of FAM26F

In an extensive eQTL study, the locus rs2858829 was found to be highly significant in regards to the cis-regulation of FAM26F gene expression (Zeller et al., 2010). In 2014, this locus was reported to correspond to an intergenic region at 6q22.1 (Julia et al., 2014). In humans polymorphisms identified close to the promotor region of

FAM26F are probably linked to basal expression differences and can bear susceptibility to ulcerative colitis (UC). Hence, the expression of FAM26F was further examined in openly accessible microarray studies involving UC and it was found to be constantly and considerably upregulated in colonic mucosa from UC samples having active inflammation (Julia et al., 2014). Furthermore, overlapping studies in UC colonic mucosa between FAM26F gene expression networks in mouse and humans revealed 28 mutual genes, 20 of which possess 'immune response' ontology as the biological process that was most significantly overrepresented. These genes include TNF pathway associated proteins (TNFSF13B and CD40), genes that encode for chemokines (CXCL9 and CXCL10) and also the genes linked to immune cell signaling (LCP1, PTPRC, LAIR1, LYN, SLAMF8, CD84, CIITA, GIMAP4 and CD300A) (Julia et al., 2014).

Although this study remained unsuccessful in assigning functionality to FAM26F, it did effectively reveal the immune and genetic significance for the differential FAM26F expression. It is thus proposed that initial expression of FAM26F may be highly crucial in determining/dictating the immune response of an individual in a diseased condition.

2.5.8. FAM26F: An Immune Regulator

Collectively, the presence of FAM26F on immune cells, its significant response to IFN-γ, its role as an anti-cancer agent as well as its differential expression in various infections, development, proliferation, immunity and pathogen challenge studies emphasize that FAM26F is an important immune regulator whose expression might indicate an active immune system and thus it can competently be regarded as an early diagnostic marker.

2.6. Significance of *In Silico* and Proteome Analysis

The ongoing high-throughput genome sequencing projects along with functional genomic screens has resulted in an exponential increase in the sequence and biological data that researchers have to contend with. Identification of the target proteins from a specific biological study requires the use of bioinformatic tools to sort and prioritize the data. Understanding the biological role of a novel target

protein is only practically possible by transferring the annotation from known proteins. The accurate functional characterization of a protein is fundamental to apprehend life at the molecular level and has vast biomedical and pharmaceutical implications. Due to their natural difficulty and expense, experimental annotation of protein functions cannot cope up with the abundant sequences and structures generated by Genomics Centers. Thus through pragmatic remodeling and theoretical investigations, computational methods have helped in addressing these critical scientific queries.

All biological events in the cell are ruled predominantly by alterations in the expression of key genes that govern all biological functions and activity. Efforts aimed towards understanding the molecular basis of complex disease are reinforced by the accessibility of high throughput strategies for identifying the biomolecules that drive the disease progression. Major technological advancements over the years have given the investigators countless prospects for multidimensional study of biological systems. Inexpensive access and convenience to the technology has significantly increased the volume of generated data. As most biological discrepancies are currently determined at genomic level, a huge amount of expression information is presently accessible via public databases. Moreover, various computational based strategies have been created to exploit the power of these data. The accessibility of massive amounts of sequence data, combined with developments in computational biology offers a perfect framework for *in silico* analysis. This development in accessible data has been coordinated by progress in our capacity to understand and exploit this new information. Biologists now consistently examine enormous microarray datasets, reconstruct biological networks, structurally and functionally explain novel proteins, identify the patterns of protein folding and simulate whole cell activity using computational procedures (Murray et al., 2007).

One of the main tasks of post genomic era is to functionally characterize all of the cellular proteins. Proteome analysis seeks to reliably annotate these proteins for determining their interaction partners and functionalities in the cellular environment (Dönnes & Höglund, 2004). A significant and foremost step in this regard is to determine each protein's subcellular localization in order to demonstrate its operating

environment within a cell. It impacts protein function by governing the availability and access to various molecular interaction partners. Therefore, understanding protein localization along with its interacting partners is essential to characterize the cellular functions of both the hypothetical and newly found proteins (Scott et al., 2005).

Even though ample amount of data is available now about the differential expression of FAM26F based on numerous infections, stimulation and immune related studies, the exact localization of FAM26F as well as its involvement in modulatory pathways that can shed light on its specific function is still unidentified. Thus, current study was aimed to functionally characterize FAM26F through computational and experimental methods, and to analyze its expression with reference to its key interactor in normal and pathological condition to get an insight about its behavior and mechanism of action, as well as its therapeutic potential.

2.7. OBJECTIVES

1. To characterize and understand FAM26F protein using *in silico* approach
2. To determine the subcellular localization of FAM26F within the cell
3. To identify the interacting partners of FAM26F in order to apprehend its molecular function
4. To investigate the expression of FAM26F and its key cellular interactors in healthy individual and HIV patient samples.

Chapter 3

MATERIALS AND METHODS

The study was essentially divided into three phases; first was to analyze and understand FAM26F through *in silico* tools in order to get general idea about this enigmatic protein, second was to identify the subcellular localization and interacting partners of FAM26F using various experimental procedures and 'omics' techniques, and finally to determine the expression of FAM26F along with its chief interactor in HIV patients and healthy individuals.

3.1. *In Silico* Characterization of FAM26F

3.1.1. Sequence Retrieval and Homology Search

Chromosomal localization of human FAM26F was determined by searching the human genome draft sequence on NCBI (http://www.ncbi.nlm.nih.gov/). The FASTA sequence of FAM26F was retrieved from UniProt (http://www.uniprot.org/) by means of its primary accession number. BLASTp (Altschul et al., 1997) was used to search for its homologous sequences with known function.

3.1.2. Multiple Sequence Alignment (MSA) and Phylogenetic Analysis

The human FAM26F protein sequence was aligned with 20 orthologous protein sequences derived from various taxa using Clustal X-v2.0 (Larkin et al., 2007) with gap opening and gap extension multiple alignment penalties set to 10.00 and 0.20 respectively. The names of organisms, their UniProt ID and taxonomic classification have been presented in Table S1. The phylogenetic tree was inferred using Mega7 software (Kumar et al., 2016), using 1000 bootstrap reiterations and Neighbour Joining phylogenic method to determine the evolutionary trend of FAM26F.

3.1.3. Physicochemical Characterization

Physicochemical parameters of FAM26F such as aliphatic index, molecular weight, isoelectric point, instability index, and grand average of hydropathicity (GRAVY) were theoretically computed by ProtParam server of Expasy (http://web.expasy.org/protparam/).

3.1.4. Sub-Cellular Localization

Sub-cellular localization of a protein essentially identifies it as a vaccine or drug target. The proteins residing in the cytoplasm can be probable drug targets, whereas surface membrane proteins can prove to be potential vaccine targets (Vetrivel et al., 2011). CELLO (version 2.0), a two level support vector machine based system, was used to predict sub-cellular localization of FAM26F (Yu et al., 2006). Online prediction tools TMHMM, SOSUI and HMMTOP, based on Hidden Markov Model (HMM), were used for predicting the probability of FAM26F of being a membrane protein (Hirokawa et al., 1998; Krogh et al., 2001; Tusnady & Simon, 2001). PROTTER (Omasits et al., 2013), a web-based tool that allows analysis of interactive protein data was used for visual analysis of FAM26F in the context of protein topology. SingnalP 4.1 (Petersen et al., 2011), the neural network method, was used for the prediction of signal peptide and SecretomeP (Bendtsen et al., 2004) was employed for determining the involvement of FAM26F in non-classical secretory pathway.

3.1.5. Function Prediction and Classification

For predicting precise function of FAM26F, the gene was analyzed using various tools.

The functional domains and family of FAM26F were predicted by Conserved Domain Database (CDD) (Marchler-Bauer et al., 2011), InterProScan (Quevillon et al., 2005), SMART (Letunic et al., 2012), ScanProsite (De Castro et al., 2006), CATH and PANTHER (Mi et al., 2013). Pfam (Finn et al., 2014), SUPERFAMILY (Gough et al., 2001), SVMProt (Cai et al., 2003), CDART (Geer et al., 2002) and Argot2 (Falda et al., 2012). CDD comprises of manually curated domain model which utilizes protein's 3D structure to perceive sequence/structure/function relationship (Marchler-Bauer et al., 2011). SMART and CDART compare the protein sequence with the database on the basis of domain architecture and profiles and searches a sequence with similar domain rather than similar sequence (Geer et al., 2002; Letunic et al., 2012). ScanProsite provides a web interface that identifies PROSITE signature matches in protein sequences on the basis of families, protein domains and functional sites (De Castro et al., 2006). CATH identifies structurally

related proteins despite having little sequence identities among them (Orengo et al., 1997). PANTHER is a comprehensive, curated database of protein families, subfamilies and trees that determines the function of protein by finding its evolutionary relationships (P. D. Thomas et al., 2003). Pfam is a database of curated protein families based on two alignments and HMM profile for the classification of protein into its potential family (Finn et al., 2014). SUPERFAMILY is a database comprising structural and functional annotation for all proteins based on a group of HMMs, which symbolize structural protein domains at the level of SCOP superfamily (Gough et al., 2001). SVMProt is a web-based Support vector machine (SVM) software that utilizes the primary sequence of a protein to classify it into a functional family. A significant classification measurement would return an R-value >2.0 and P-value >60% (Cai et al., 2003). Argot2 is used to rapidly process thousands of sequences against the query sequence for functional inference based on the clustering process of Gene Ontology (GO) and a weighting scheme (Falda et al., 2012).

Motifs in a protein sequence are descriptors of the protein function that can frequently be utilized for the estimation of the family of proteins (Peer Bork & Koonin, 1996). InterProScan and MOTIF were used to identify the motifs in FAM26F. InterproScan is an integration platform that scans given protein sequence against a combined resource of various protein function recognition methods of the InterPro affiliated databases for motif discovery (Quevillon et al., 2005). The MOTIF (http://www.genome.jp/tools/motif/) searches the motif-sequence database for assigning function to the protein. A web server PFP-FunDSeqE (http://www.csbio.sjtu.edu.cn/bioinf/PFP-FunDSeqE/) was also used that combines the information of functional domain and evolution to find out the pattern of protein fold.

3.1.6. Prediction of Post Translational Modifications (PTMs)

NetNGlyc was used for the prediction of N-Glycosylation sites in human FAM26F protein. NetNGlyc utilizes artificial neural networks (ANN) which examine the sequence context of Asn-Xaa-Ser/Thr sequons (Gupta et al., 2004). Netphos 2.0 (Blom et al., 1999) was used for anticipating probable phosphorylation sites in FAM26F protein sequences at serine, threonine or tyrosine residues with a least

threshold score of 0.5. The server NetPhosK 1.0 (Blom et al., 2004) was employed for predicting kinase-specific phosphorylation sites in human FAM26F.

3.1.7. Functional Protein Affiliated Networks

The protein function is often controlled by other interacting proteins. Hence, inferring protein-protein relation is essential for predicting a protein's function. STRING (version–10) was used to predict the interacting partners of FAM26F. The interactions consist of direct (physical) as well as indirect (functional) associations, experimental or co-expression (Szklarczyk et al., 2015). Threshold scores greater than 0.400 were considered for results.

3.1.8. Tertiary Structure Prediction

Protein data bank (PDB) (Berman et al., 2000) was explored using BLASTp to check the presence of any entry worthy of being a template (having > 37% similarity) to model the structure of FAM26F. As the results did not yield any suitable template, I-TASSER algorithm (Y. Zhang, 2008) was used for the automated *ab initio* prediction of FAM26F structure and function, which is built on the sequence-to-structure-to-function prototype. The model that returned a greater C-score was chosen as the final model. The generated model was analyzed by PROCHECK (Laskowski et al., 1993) and ProSA (Wiederstein & Sippl, 2007) to assess its stereo chemical quality,, both of which are well-known tools for analyzing the structure of modeled proteins. PROCHECK presents the distribution of residues in different areas on the basis of resolution of >= 2.0 Angstroms and R-factor <= 20.0. ProSA delivers a z-score that computes the difference between the total energy of the model in contrast to an energy distribution deduced from arbitrary conformations to reveal the structure quality. The model with negative z-score is depicted to be of good quality. Finally, VMD program (Humphrey et al., 1996) was employed for visualizing the protein structure.

3.2. *In vitro* characterization of FAM26F

3.2.1. Cell Culture

HEK293 cells were cultured in Dulbecco's modified Eagle's medium (DMEM) (Sigma-Aldrich Chemie, Steinheim, Germany), augmented with 10% fetal bovine

serum (FBS) (Biochrom AG, Berlin, Germany), and 1% penicillin/streptomycin (PS) (Biochrom AG, Berlin, Germany) at 37°C in a humidified atmosphere of 5% CO2.

3.2.2. Transient Transfection

The HEK293 cells were transfected transiently with GFP tagged FAM26F plasmid encoding full length FAM26F gene (Figure 3.1), named FAM26F (NM_001010919) Human Tagged ORF Clone (RG222648, OriGENE) using Lipofectamine 2000 (Invitrogen, Carlsbad, CA, USA) following the manufacturer's instructions. The detailed vector map is given in Figure 3.1. Briefly, the cells in each well of a 6 well plate were allowed to grow to 70-80% confluency. Two Opti-MEM™ I Reduced Serum Medium (Thermo Fisher Scientific) solutions were prepared; Solution A containing 250µl Opti-MEM with 5µl lipofectamine (per well), Solution B having 250µl Opti-MEM with 1.5µg DNA (per well). Both solutions were incubated at room temperature (RT) for 5 minutes, after which they were mixed (Solution C) and further incubated for 20 minutes at RT. Meanwhile, DMEM media from each well was removed, and cells were washed with 1X PBS carefully along the side of the well, preventing detachment of the cells. 1.5 ml Opti-MEM and 500µl of Solution C were added to each well and plates were placed at 37°C. The media was removed after 6-8 hours and replaced with DMEM media. The cells were harvested at 3h, 6h, 12h, 24h, 36h and 48h time intervals post transfection.

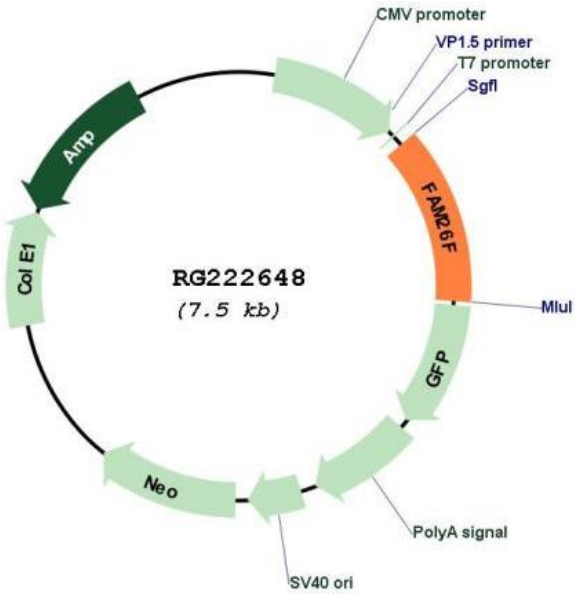


Figure 3.1: Detailed vector map of FAM26F (NM_001010919) Human Tagged ORF Clone (RG222648, OriGENE), the GFP tagged FAM26F plasmid encoding full length FAM26F gene along with multiple restriction sites. **(https://www.origene.com/drawmapbysku?SKU=RG222648)**

3.2.3. Cell Viability Assay

Viable cells were detected by MTS proliferation assay (Promega, Madison, WI, USA), which quantifies the reduction of [3-(4,5–dimethylthiazol–2–yl)-5-(3-carboxymethoxyphenyl)-2-(-4-sulfophenyl)-2H-tetrazolium] (MTS) to a water soluble formazan salt, a phenomenon only displayed by metabolically active cells (Cory et al., 1991). Briefly, $1x10^5$ cells per well were seeded in 24-well plates (Nunc, Roskilde, Denmark) and permitted to grow for 12 h at 37°C. Thereafter, the cells were transfected with C terminus GFP tagged FAM26F plasmid for variable times (3 h, 6 h, 12 h, 24 h, 36 h and 48 h). Quantitative assessment of cell viability was achieved by using MTS reagent in the presence of phenazine methyl sulphate (PMS). The culture media was changed before treating with MTS. Once combined MTS/PMS solution was added to each well, plates were placed in a humidified atmosphere having 5% CO_2 for 1 h at 37°C for colour development. Multiscan plate reader (Labsystems, Manassas, VA, USA) and Accent software 2.6 were used to record the values of absorbance at 490 nm. The final absorbance value was achieved by subtracting the background absorbance of the cell-free medium incubated with the MTS reagent from the sample wells. All MTS assays were executed in triplicates.

3.2.4. Caspase-3 Activity Assay

The Caspase-3-activity assay quantitatively measures alterations in the protease activity of caspase-3 (DEVDase), which is an initial event in apoptosis (Gurtu et al., 1997). Caspase-3 activity assay kit was used as per the manufacturer's protocol. Briefly, both untreated control cells and C-terminus GFP tagged FAM26F transfected cells were lysed inside the cell lysis buffer for 15 minutes at 4°C. A centrifugation at 10000 x g was conducted next. Thereafter, protein concentration in the supernatants was assessed. Then, 50 μg of the total cell lysate was incubated with 50 μM caspase-3 specific substrate DEVD-pNA for a period of 4 to 5 h at 37 °C. The Caspase-3 facilitated pNA release was evaluated through absorbance at 405 nm. The background absorbance from the untreated controls was then deducted from the ultimate absorbance value attained for the samples.

3.2.5. Antibodies and Fluorescent Probes

The primary antibodies used include rabbit anti-FAM26F (Abcam, Cambridge, U.K.), rabbit anti-VCP (Abcam, Cambridge, U.K.), rabbit anti-Rab9 (Cell Signaling technology, Frankfurt, Germany), mouse IgG actin cytoplasmic 1 (Sigma, Steinheim, Germany), rabbit anti-Syntaxin 6 (Abcam, Cambridge, U.K.), rabbit anti-Golgin (Abcam, Cambridge, U.K.), mouse anti mu-Calpain (Invitrogen, Carlsbad, CA, USA), mouse anti-S100-A7 (Sigma-Aldrich,Steinheim, Germany), mouse anti_vinculin (Sigma-Aldrich,Steinheim, Germany), rabbit anti-Thioredoxin (Epitomics, Abcam, UK), mouse anti-Peroxiredoxin (Abcam, Cambridge, U.K.) and rabbit anti-Calmodulin (Abcam, Cambridge, U.K.). The secondary antibodies used consisted of anti-rabbit pAb (Santa Cruz Biotechnology, Santa Cruz, CA, USA), HRP-conjugated rabbit anti-mouse pAb (IBA, Goettingen, Germany), goat anti-rabbit (Alexa 555-conjugated), goat anti-mouse cy3-conjugated (Dianova, Hamburg, Germany), and anti-mouse (Alexa 555-conjugated).

3.2.6. Co-Immunofluorescence and Confocal Laser Scanning Microscopy

HEK293 cells grown on glass cover slips in 24-well plates (Nunc, Roskilde, Denmark) were transfected with FAM26F plasmid. After 24 h of transfection, the cells were immobilized using 4% paraformaldehyde solution for 20 minutes,

followed by 3 washes with PBS of 5 minutes each. Thereafter, cell permeabilization was achieved with 0.2% Triton X-100 in 1x PBS for 10 minutes, and then blocked for 30 minutes in a solution containing 10% FBS in 1% BSA solution in PBS. Cells were incubated with the primary antibody at 1:100 dilution in 1% BSA in PBS and kept at 4°C overnight. Cells were then treated with the secondary antibody at 1:200 dilution with 1% BSA in PBS and finally counterstained with TOPRO-3 iodide for 1 min to stain the nuclei. The coverslips were then mounted onto the glass slides using mounting media Fluoromount (DAKO, Hamburg, Germany). All the mentioned steps were conducted in dark and each reaction was terminated by washing the coverslips thrice with 1xPBS. The slides were then kept in dark at 4°C until visualized. Confocal laser scanning microscopy was performed using LSM510 laser-scanning microscope (Zeiss, Gottingen, Germany; 488 nm Argon, 543 and 633 nm Helium-Neon excitation wavelengths). Individual images were analyzed separately for co-localization using LSM 5 (Zeiss) or ImageJ (WCIF plugin) software.

3.2.7. Immunoblot Analysis

HEK293 cells in a concentration of 4.8 x 10^5 cells per well were plated in 6 well plates and permitted to grow for 24 h at 37°C till they were 70-80% confluent. Thereafter, the cells were transfected with C terminus GFP tagged FAM26F plasmid and again kept at 37°C for 24 h, after which they were harvested and lysed using Tris-Triton lysis buffer (50 mM Tris-HCl pH 8.0, 1 % Triton X-100, 0.5 % CHAPS, 1 mM DTT) supplemented with phosphatase and protease inhibitors (Roche). The cell debris was cleared from the lysates by centrifugation at 14000 rpm for 30 minutes at 4 °C. The Supernatant was collected and Bradford Assay (Bio-Rad) was utilized to check the concentration of protein. Equivalent protein concentrations in the lysates were subjected to 12% SDS polyacrylamide gel electrophoresis. Standard protocol was used to electro-transfer the proteins from the gel onto a polyvinylidene difluoride (PVDF) membrane (Millipore) (Towbin et al., 1979). A blocking solution of 5% non-fat milk diluted in Phosphate buffered saline supplemented with 0.05% Tween20 (PBS-T) was then used to soak the blot in. Primary antibody anti-FAM26F (1:1000) was added to the blot which was then incubated overnight at 4 °C. Thereafter, the membranes were washed in 1× PBS-T and incubated with the rabbit horseradish peroxidase-conjugated secondary antibody (1:10,000) for 1 h at room

temperature. Immunoreactivity was then detected by incubating the membranes in Enhanced Chemiluminescence (ECL) solution and images were visualized in ChemiDoc (Bio-Rad). The intensities of bands were checked by densitometry using ImageLab^{TM} (Bio-Rad) data analyzer software. Then membrane was reblotted with glyceraldehyde-3-phosphate dehydrogenase as a loading control.

3.2.8. Co-Immunoprecipitation

Cell lysis and protein extraction was carried out in the manner described above. Immunoprecipitation was performed using protein G Magnetic Dynabeads® (Invitrogen) following the manufacturer's instructions. Typically, 6 µg of the anti-FAM26F antibody diluted to 1:50 in PBS was added to 30 µl of Dynabeads and incubated for 30 mins at 4°C. 500 µg of protein lysate was then added to antibody-Dynabead complex and left overnight at 4°C. The next day, beads were rinsed thrice with 0.3% CHAPS in water, and 20 µl of 2x Laemmli buffer (doi:10.1101/pdb.rec10878Cold Spring Harb Protoc2007.) was used to elute the immunoprecipitated proteins from the beads-antibody-antigen complex. The elute was then cooked for 5 mins at 95°C and run onto 12% SDS-PAGE, followed by immunoblot analysis as described above.

3.2.9. In-Gel Tryptic Digestion and MS/MS Analysis

The eluates were run on 12% SDS-PAGE for 5-10 mins to get a 5cm run window. The gel was then stained with Coomassie Blue; the stained blue protein spots corresponding to the labeled proteins in the western blot were manually excised from the gel and washed with distilled water for 15 min. The gel pieces were de-stained by washing twice for 10 min with a solution containing 100 mmol/L ammonium bicarbonate/acetonitrile (1:1, v/v) and then for a third time until all visible blue dye was removed. In-gel digestion with trypsin was carried out according to a protocol described previously (Asif et al., 2007). The extracted peptides were then dissolved in 0.1% formic acid (FA) for ESI-QTOF MS/MS. One microliter of tryptic digested peptide solution was introduced using a CapLC auto sampler (Waters) onto a µ-precolumn cartridge C18 pepMap (300 µm 5 mm; 5 µm partical size) and further separated through a C18 pepMap100 nano Series (75 µm 15 cm; 3 µm partical size) analytical column (LC Packings). The mobile phase consisted of solution A (0.1%

FA in 5% ACN) and solution B (0.1% FA in 95% ACN). The single sample run time was set for 60 min. The chromatographically separated peptides were then analyzed on a Q-TOF Ultima Global (Micromass, Manchester, U.K.) mass spectrometer equipped with a nanoflow ESI Z-spray source in positive ion mode. The data acquisition was performed using MassLynx (v 4.0) software on a Windows NT PC and data were further processed on Protein-Lynx-Global-Server (v 2.1), (Micromass, Manchester, UK). Processed data were searched against MSDB and Swiss-Prot databases through the Mascot search engine using a peptide mass tolerance of ±0.5 Da and fragment mass tolerance of ±0.5 Da. The search criteria were set up to maximum one missed cleavage allowed by trypsin and protein modifications set to methionine oxidation and carbamidomethylcysteine, when appropriate.

3.2.10. Computational Analysis

To determine the signaling pathways in which the identified proteins or interacting partners of FAM26F were involved, Reactome Pathway Database (https://reactome.org/) and Ingenuity Pathway Analysis (IPA®) software (https://www.qiagenbioinformatics.com/products/ingenuity-pathway-analysis/) were used. Reactome is a pathway database which allows the visualization, interpretation and analysis of pathway knowledge by employing intuitive bioinformatics tools (Croft et al., 2010). IPA on the other hand is a powerful analysis and search tool that is capable of revealing the significance of 'omics data by identifying the specific biological system in which the query proteins are involved (Krämer et al., 2013).

3.2.11. Statistical and Image Analysis

All the results from this study were acquired on the basis of four individual experiment sets. Descriptive statistics was used to express the results as mean ± S.D. All the confocal images were quantitatively analyzed and assessed using Imagej (WCIF plugin). ImageLab^{TM} (Bio-Rad) software was used to perform the densitometric analysis of the 1-DE gels. All the graphs were prepared by GraphPad PRISM (GraphPad Inc.).

3.3. Expression analysis of FAM26F

3.3.1. Ethical Statement

All Patients recruited in this study gave a written signed consent after they were well informed about the research being carried out. The study did not commence until written approval was granted by the Ethical Review Board of Shaheed Zulfiqar Ali Bhutto Medical University (SZABMU) Islamabad, Pakistan and institutional review board (IRB) of Atta-ur-Rahman School of Applied Biosciences (ASAB), National University of Sciences & Technology (NUST) Islamabad, Pakistan, after all the research protocols were thoroughly assessed.

3.3.2. HIV Blood Sampling

Since HIV is a biohazardous agent, all sampling was conducted under strict Biosafety level-2 conditions and all standard precautionary measures were undertaken. A total of 50 HIV-1 positive blood samples (5 women and 45 men), with an average age of 34±11.8 years were collected at the Pakistan Institute of Medical Sciences (PIMS) referral lab after informed written consent was taken from the patients. The inclusion criteria were only those HIV-1 patients who were either treatment naïve or those who had failed to respond to the antiretroviral treatment (ART). The demographic details including age, sex, viral load, any co-infection and other relevant clinical data was noted down from the patient's card ensuring anonymity at all stages. Control samples amounting to 40 (10 males and 30 females) with 25±10 years of average age were collected from students at ASAB, NUST, Islamabad after verbal consent, and some of the control samples were provided by courtesy of Dr. Yasmeen Badshah. It was ensured that the control samples were HIV-negative and had no prior history of HIV/AIDS.

3.3.3. RNA Extraction

RNA was extracted from the control and HIV-1 positive blood samples and cDNA was created for further QRT-PCR expression analysis. For this purpose, first 5ml of peripheral venous blood was drawn from each patient and transferred from the sterile syringe into a 5ml blood K3EDTA vacutainer. The blood was then shifted into a 50ml polypropylene conical falcon. A 1X RBC Lysis Buffer was added and the final

mixture volume was bought up to 25ml. The blood-Lysis buffer mixture was allowed to stand for 10 minutes at RT. Thereafter, the mixture was centrifuged at approximately 600 x g for 10 minutes at RT. The supernatant was cautiously removed and the pellet was re-suspended in 1ml RBC-Lysis buffer to remove any remaining traces of RBCs. The contents were shifted to a new 1.5ml Eppendorf tube and left for 5 minutes, succeeded by centrifugation at 3000 rpm for 2 minutes at RT. The supernatant was discarded, and the pellet was re-suspended in 1ml of PBS. Again, the cells were centrifuged at 3000 rpm for 2 minutes at RT. Supernatant was discarded, and the pellet was thoroughly re-suspended in 250µl of PBS. To this, 750µl of Trizol® LS solution (Thermo Fisher Scientific Inc) was added and the resultant solution was vigorously mixed to ensure complete homogenization. Then 200µl of chloroform was added. The mixture was thoroughly hand shaken for a minimum of 15 seconds and then left on ice for 10 minutes. The tubes were occasionally inverted to assure proper mixing. The samples were then centrifuged at 14000 rpm for 20 minutes at 4°C. After centrifugation, the mixture was separated into 3 layers: an upper aqueous layer holding RNA, a white interface enclosing DNA and a lower reddish phenol-chloroform layer containing proteins. Almost 500 µl of the aqueous layer was carefully removed while taking care not to remove any portion of the underlying interphase and shifted to a fresh 1.5ml sterile Eppendorf tube. To the aqueous layer, 500 µl of ice-chilled isopropanol was added and the solution was mixed thoroughly and kept at -20°C for 10 minutes. Subsequently, the solution was centrifuged at 14000 rpm for 20 minutes at 4°C. The RNA forms a gel-like pellet at the base of the tube. The supernatant was carefully discarded, and the pellet was resuspended in 1ml of ice-chilled 75% Ethanol. The sample was then centrifuged at 14000 rpm for 10 minutes at 4°C. The supernatant was cautiously discarded, and remnant ethanol was removed using a pipette. The RNA pellet was left to air dry for 10 minutes and was then dissolved in 25µL of Nuclease-free (NF) water. The RNA quantification was achieved using NanoDrop 2000 UV-Vis Spectrophotometer (Thermo Scientific). A 260/280 of >1.8 suggested pure RNA. The RNA was also run on an agarose gel and discrete 28S and 18S rRNA bands ensured that RNA had good integrity.

3.3.4. cDNA Synthesis

cDNA was synthesized using Moloney Murine Leukemia Virus Reverse Transcriptase (M-MLV RT) (Invitrogen, Cat No: 28025013). Concisely, 1 µL (100 pmol) of Random Hexamer, 1 µL of 10mM dNTPs and 1µg of the extracted RNA were added to a sterile, RNA-free tube placed on ice and volume was raised to 15µl using sterile NF water. The mixture was briefly centrifuged and heated at 65°C for 5 minutes. The tube was then immediately cooled down on ice. Thereafter, 4µl of 5X RT-buffer was incorporated in the tube, gently mixed, and incubated at 37°C for 2 minutes. Then 1µl of M-MLV RT was added and mixture was kept at 42°C for 60 minutes. Finally, the reverse transcriptase enzyme was deactivated by heating the mixture to 70°C for 10 minutes. The cDNA was confirmed using Beta-actin PCR and was kept aside at -20°C until further needed.

3.3.5. Primer Designing

The cDNA templates used for primer design were obtained from Genbank of National Centre of Bio-Informatics (NCBI). Primers for the three genes FAM26F, Beta actin and Thioredoxin, were designed by the online software Primer3 (Untergasser et al. 2012) (http://bioinfo.ut.ee/primer3/). Each primer pair was carefully examined for its specificity and precautions were taken to avoid secondary structures and dimers before it was used in further experiments. Primer sequence properties were determined by the online tool 'OligoCalc: Oligonucleotide Properties Calculator' that checked the GC content, Tm, formation of hairpins and capacity of self-dimerization (Kibbe 2007) (URL: http://biotools.nubic.northwestern.edu/OligoCalc.html). The specificity of each primer pair was further affirmed by UCSC Genome browser (https://genome.ucsc.edu/) and NCBI Primer-BLAST tool (Ye et al. 2012). Amplicons lengths were kept between 100-250bp. The list of primer pairs used in the experiment along with their Tm and product size is given in Table 3.1.

Table 3.1: List of genes along with their primer sequences, Tm and amplicon size

No.	Gene	Sequence	Product size (bp)	Tm (°C)
1	Beta Actin F	CATGTACGTTGCTATCCAGGC	250	62
	Beta Actin R	CTCCTTAATGTCACGCACGAT		
2	FAM26F F	CACCCGATGCCTATCTCCAG	194	62
	FAM26F R	TTTGCTGCCACTCTTTCATGC		
3	TRX F	CTCTGTTTGGTGCTTTGGATCC	135	62
	TRX R	CAAGTTTATCACCTGCAGCGTC		

3.3.6. Primer Optimization

Gradient PCR was used to optimize the PCR conditions for each primer pair considering a series of annealing temperatures (54°C to 64°C). The PCR mix of 20µl had 2µl of 10X PCR Buffer, 2µl of 25mM $MgCl_2$, 1.5µl of 10mM dNTPs, 1µl each of 10µM reverse and forward primer, 0.5µl (5U/µl) of *Taq* polymerase (0.125U/20µl) and 2µl of cDNA.

The reaction was carried out in thermocycler GeneAmp® PCR System 9700. The cycling conditions consisted of initial denaturation of cDNA at 95°C for 5 minutes, succeeded by 35 cycles of three steps; 95°C for 45 seconds, 62°C or annealing Tm for 45 seconds and 72°C for 45 seconds. Finally, the last extension was given at 72°C for 10 minutes.

All of the PCR products were run on 2% agarose gel for 40 minutes at 120V. The gel was then visualized in the gel-doc system under UV to analyze the gene bands.

3.3.7. Quantitative Real Time PCR (QRT-PCR)

The expression level of selected target genes in healthy individuals and HIV patients was quantified through QRT-PCR carried out on ABI prism 7500 Fast SDS (Applied Biosystems, California USA). The reactions were set in a clean environment to avoid any contamination. Each 20µl reaction mixture contained 2µl Sybr Green (5.0X) (Thermo fisher scientific), 13µl RNase free water, 1.5 µl each of 10µM reverse and forward primers, 2µl of 1:3 diluted cDNA. The reaction mixture was primarily

subjected to 50°C for 20 seconds and 95°C for 10 minutes. This was trailed by 40 PCR cycles, each consisting of 95°C for 15 seconds and 62°C for 1 minute. Additionally, the amplification specificity was monitored by including a melting/dissociation step. Housekeeping gene Beta Actin was measured simultaneously as internal control for normalization of target genes expression. Results were examined and relativistic expression (rE) of target gene was calculated using the formula: $rE = 100 \times {}^{-\Delta Ct}$ where ΔCt was computed by subtracting the average Ct of Beta Actin from average Ct of target gene. Relative quantification of gene expression was calculated by Livak method i.e. as $2^{-\Delta\Delta Ct}$.

3.3.8. Statistical Analysis of QRT-PCR Results

Graph-Pad Prism 6.0 (Graph-Pad Software, San Diego, CA USA) was used for statistical analysis, where p-value < 0.05 was taken as significant. Student t-test (unpaired) at confidence interval of 95% was used for comparing the sample groups.

Chapter 4

RESULTS

4.1. *In silico* characterization of FAM26F

FAM26F is located on human chromosome 6 reference genomic contig NC_000006.12 between KRT18P22 and TRAPPC3L genes, mapping to the chromosomal location 6q22.1. It consists of three exons, which make up a 1141 bp coding region, encoding a 315 amino acid polypeptide. The neighboring genes along with their position with respect to FAM26F is shown in Figure 4.1.

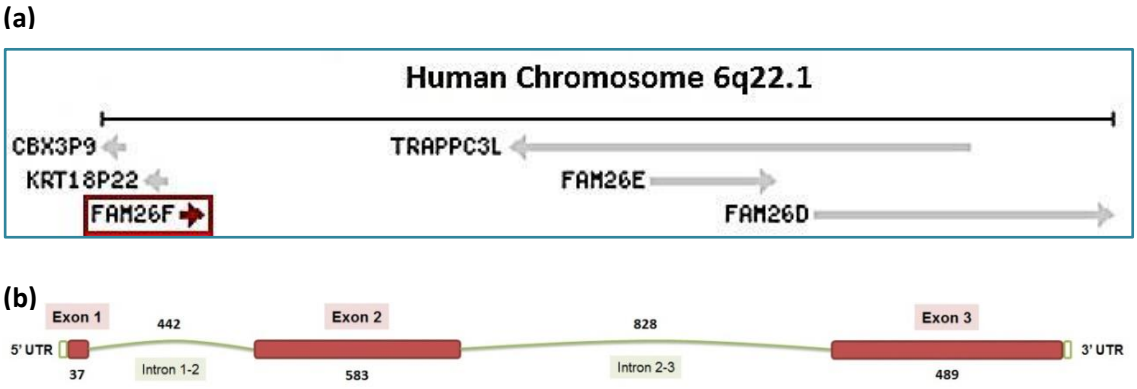


Figure 4.1: Chromosomal localization and gene structure of FAM26F. (a) Human FAM26F is located in forward orientation between KRT18P22 and TRAPPC3L genes, mapping to the chromosomal location 6q22.1**. (b)** The figure represents the arrangement of exons and introns within FAM26F sequence together with the number of base pairs covered by each.

The UniProt ID of human FAM26F is Q5R3K3. This ID was used to recover its sequence for further systematic sequence analysis by means of different tools. The MSA of FAM26F from 20 individual species including humans disclosed 38 perfect, 37 high and 25 weakly conserved residue groups. The conserved residues together with their precise locations are illustrated in the MSA presented in supplementary Figure S1. The phylogenetic analysis of the rooted Neighbor Joining tree revealed that FAM26F is evolutionary conserved. The human FAM26F sequence is most closely related to that of primates, specifically chimpanzees as human and chimpanzees share a common cluster (Figure 4.2). The next close neighbour sequences are that of the elephant, squirrel, bovine and carnivores, followed by

rodents, amphibians and reptiles. Aves appeared to be the farthest orthologs of human FAM26F.

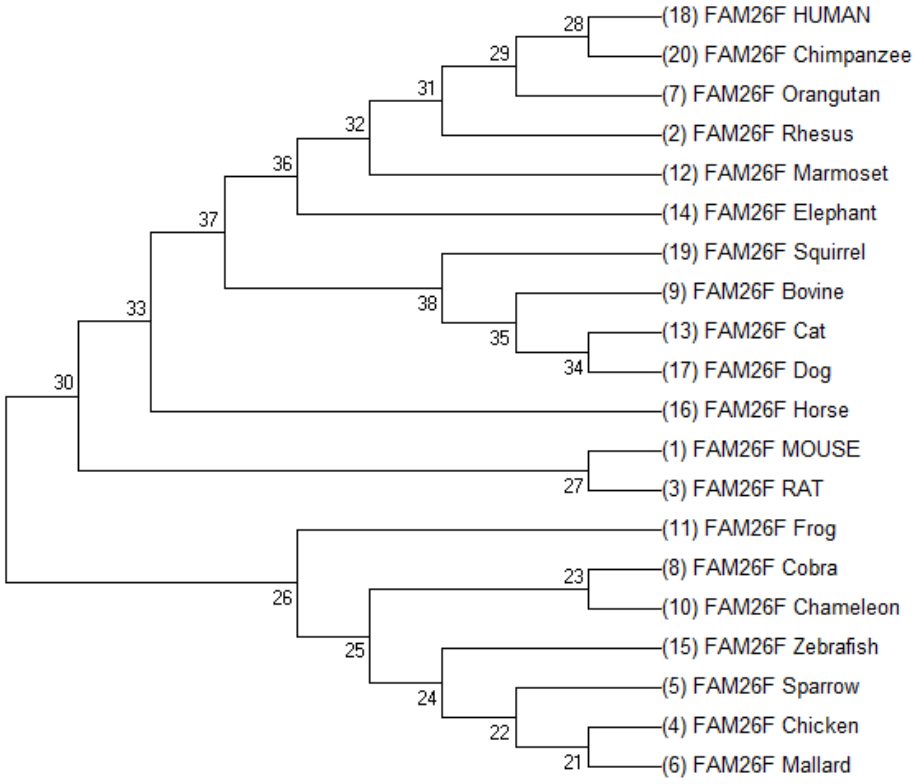


Figure 4.2: Phylogenetic Tree of FAM26F. The phylogenetic tree was constructed using Mega7 software from complete alignment of 20 FAM26F protein sequences by a Neighbor Joining method with bootstrapping of 1000 reiterations. The analysis revealed that FAM26F has remained well conserved throughout evolution. The human FAM26F sequence is most closely related to that of primates, sharing a common cluster with chimpanzees. The next closest neighbours are elephant, squirrel, bovine and carnivores, followed by rodents, amphibians and reptiles. Aves appear as the farthest orthologs of human FAM26F.

The physicochemical parameters of FAM26F as computed by Expasy's ProtParam tool are compiled in Table 4.1.

Table 4.1. Physicochemical parameters of FAM26F predicted by Expasy's ProtParam tool

UNIPROT ID	No. of Amino Acids	Molecular weight, M_w (Da)	Theoretical PI	Extinction coefficient (M^{-1} cm^{-1})	Instability Index	Aliphatic index	Grand average of hydropathicity (GRAVY)
					Computed	Classification	
Q5R3K3	315	34458.2	8.87	48650	35.10	stable	0.280

The subcellular localization and nature of FAM26F was predicted by using several tools. The outcomes of these predictions anticipated FAM26F to be a membrane protein having several transmembrane helices which is devoid of a signal peptide (Table 4.2).

Table 4.2: Sub-cellular localization of FAM26F

UNIPROT ID	Sub-cellular localization (CELLO)	Signal Peptide (SignalP)	Secretory Protein (SecretomeP)	Transmembrane (TM) helices prediction			
				SOSUI	TMHMM	HMMTOP	PROTTER
Q5R3K3	Plasma membrane	No	Yes	3 TM helices (50-72, 99-121, 178-200)	4 TM helices (17-39, 49-71, 102-124, 176-195)	5 TM helices (18-39 52-71 90-109 114-133 176-195)	4 TM helices (19-39, 52-72, 104-124, 176-196)

Visual representation of FAM26F TM helices as obtained by PROTTER is given in Figure 4.3.

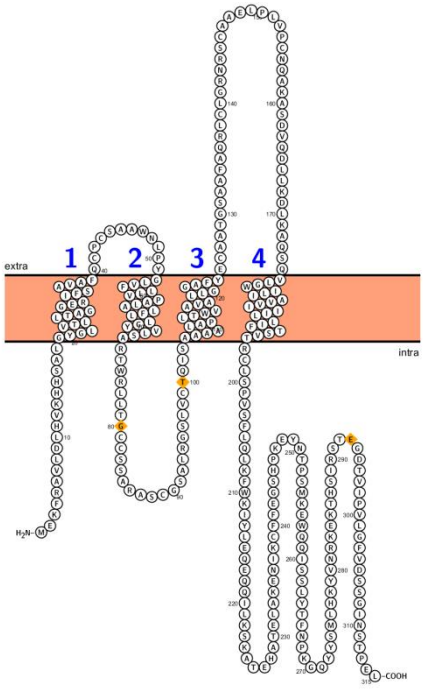


Figure 4.3: Visual illustration of FAM26F TM helices. Although different number of TM helices were predicted by the various softwares used (Table 3.1), the regions occupied by the helices were almost similar. In this figure, the numbers in blue represent the four consensus TM helices covering the amino acid positions 19-39, 52-72, 104-124 and 176-196.

4.1.1. Functional Characterization and Classification

Virtually all the tools used for domain prediction, including CDD, SMART, ScanProsite, Pfam and CDART showed that FAM26F contains a single Ca_hom_mod (calcium homeostasis modulator) domain ranging from position 1-248 (Figure 4.4). However, some tools like InterProScan, CATH and SUPERFAMILY did not give any hit at all. PANTHER also did not define the family of FAM26F or provide any signaling pathway information. The results are clustered in Table 4.3.

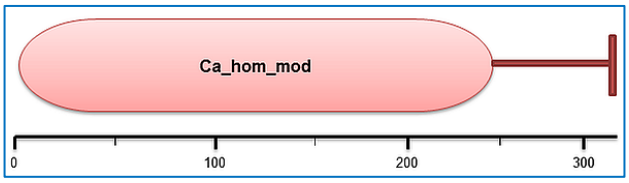


Figure 4.4 Predicted domain in FAM26F by various tools. The single domain Ca_hom_mod covers the amino acids ranging from position 1-248.

Table 4.3: Domain identified in FAM26F and its function

UNIPROT ID	Domains identified by InterProScan, CATH, PANTHER & SUPERFAMILY	Domains identified by CDD, SMART, ScanProsite, Pfam & CDART	Domain start	Domain end	Function
Q5R3K3	Nil	Ca_hom_mod	1	248	Control cytosolic calcium concentration

SVMProt predicted FAM26F to belong to the transmembrane family, having P-value of 2.7 and R-value of 92.1%. Argot2 successfully assigned the molecular, biological and cellular function to FAM26F. According to the results, FAM26F is involved in cation channel activity at the molecular level. Biologically, FAM26F performs the transport activity, including ion/cation transport, ion/cation transmembrane transport and general transport. At the cellular level, it is predicted to be a membrane or an integral component of the plasma membrane and/or membrane.

The MOTIF identified a single significant motif 'Ca_hom_mod' when comparing FAM26F sequence with different motif libraries, including Pfam, NCBI-CDD, PROSITE pattern etc. The insignificant motifs were not considered in the results. PFP-FunDSeqE identified FAM26F to possess an immunoglobulin-like fold.

4.1.2. Predicted Post Translational Modifications

N-linked glycosylation is the most common type of glycosylation essential for cellular attachment and for the folding of some eukaryotic proteins. Additionally, it can also regulate a protein's function, for example, by acting as an on-off switch in some cases (Maverakis et al., 2015). NetNGlyc predicted only one potential N-glycosylation site in FAM26F at position 143 as illustrated in Figure 4.5(a).

For the estimation of probable phosphorylation of Ser, Thr and Tyr residue, NetPhos 2.0 server was used which predicted overall 14 sites (Ser: 8, Thr: 3, Tyr: 3) having eminent phosphorylation potential. These include Ser residues at positions 88, 96, 200, 253, 288, 291, 307 and 311; Thr residues at positions 79, 251 and 292; Tyr residues at positions 264, 272 and 279. Graphical illustration can be seen in Figure 4.5(b). Moreover, kinase-specific phosphorylation sites in FAM26F were predicted

by NetPhosK. The results of general and kinase specific phosphorylation sites are compiled in Table 4.4.

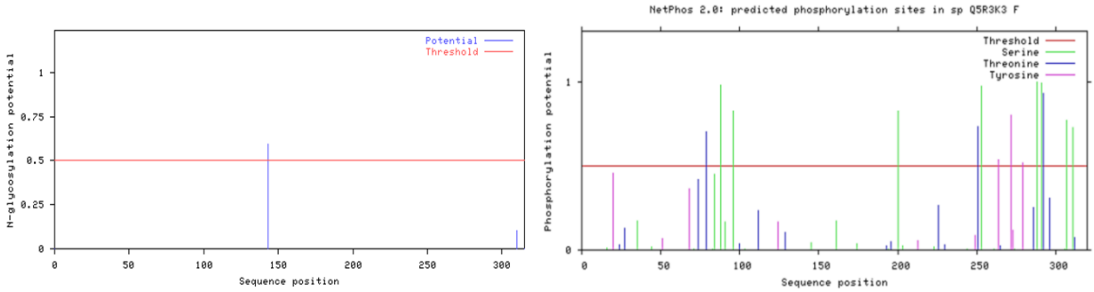


Figure 4.5: Post translational modifications of FAM26F. (a) Predicted potential N-glycosylation site in FAM26F sequence. **(b)** Predicted phosphorylation sites in FAM26F sequence by NetPhos 2.0. As illustrated in Figure 4.5, 8 Serine, 3 Threonine and 3 Tyrosine residues (total 14) have high phosphorylation probability.

Table 4.4: General and kinase specific phosphorylation sites predicted in FAM26F

Name of residue	Position	Phosphorylation prediction	Kinase
Serine (S)	83	No	PKC
	84	No	PKC, cdc2
	88	Yes	PKC
	96	Yes	PKA, cdc2
	131	No	cdc2
	145	No	PKA
	174	No	DNAPK
	200	Yes	p38MAPK, cdk5
	223	No	PKC, cdc2
	253	Yes	PKC
	262	No	DNAPK
	274	No	cdc2
	288	Yes	PKC, PKA
	291	Yes	-
	306	No	CKI
	307	Yes	PKA
	311	Yes	CKII

Threonine (Thr)	79	Yes	PKC
	100	No	cdc2
	129	No	cdc2
	230	No	CKII
	251	Yes	-
	265	No	PKC
	292	Yes	-
	312	No	p38MAPK
Tyrosine (Tyr)	124	No	EGFR
	249	No	EGFR
	264	Yes	-
	272	Yes	INSR
	279	Yes	-

PKC: Protein Kinase C; cdc: Cell division control protein; PKA: Protein Kinase A; DNAPK: DNA-dependent protein kinase; MAPK: Mitogen-activated protein kinase; cdk: Cyclin-dependent kinase; ck: Casein kinase; EGFR: Epidermal Growth Factor Receptor; INSR: Insulin receptor

4.1.3. FAM26F interaction network

Interaction of proteins can help us in predicting the function of a protein with unknown function. For this purpose, STRING (Szklarczyk et al., 2015) was used to determine the interacting partners of FAM26F. A total of ten genes/proteins were found to interact with FAM26F with a medium confidence score (>0.400). These include PHF20L1, HEXB, NMT1, ANKMY2, IRF5, SPPL2A, LMAN2L, TSPAN13, SLC8B1 and USP25. The interaction network is illustrated in Figure 4.6, whereas the proteins along with their respective sizes, function, score and type of association with FAM26F are detailed in Table 4.5.

Except for IRF5, all the proteins were also found to interact with one another for their functional activity. IRF5 is the only protein found to co-express with FAM26F in humans, whereas putative homologs of SLC8B1 and USP25 are co-expressed with FAM26F in other species. All the remaining proteins were identified through text-mining.

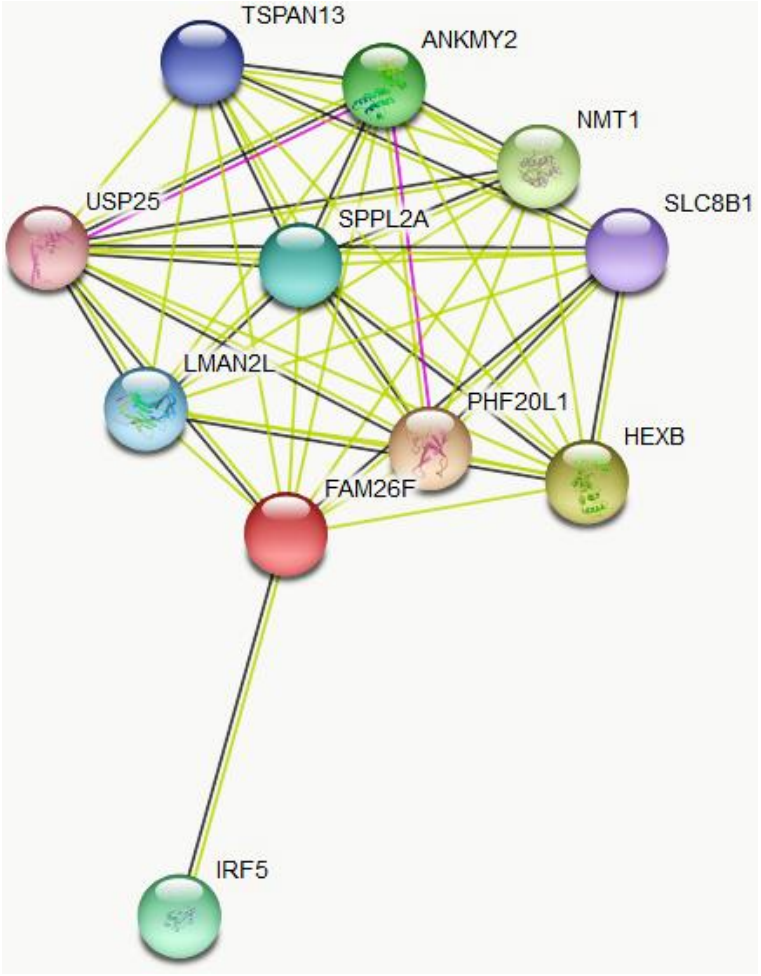


Figure 4.6: FAM26F interaction network. The FAM26F interaction network was generated by STRING. Ten genes/proteins, namely PHF20L1, HEXB, NMT1, ANKMY2, IRF5, SPPL2A, LMAN2L, TSPAN13, SLC8B1 and USP25 were predicted to interact with FAM26F with medium stringency. While all these proteins interact with FAM26F, they were also found to interact with one another (with the exception of IRF5) for their functional activity.

Table 4.5: STRING based interactors of FAM26F

No.	Gene	Protein	Protein size	Function (Uniprot)	STRING Score	Type of association
1	PHF20L1	PHD finger protein 20 like 1	1017 aa	-	0.705	Text mining
2	HEXB	Beta-hexosaminidase subunit beta	556 aa	Degrades GM2 gangliosides, and several other molecules containing terminal N-acetyl hexosamines, in the brain and other tissues	0.677	Text mining
3	NMT1	Glycylpeptide N-tetradecanoyltransferase 1	496 aa	Adds a myristoyl group to the N-terminal glycine residue of certain cellular and viral proteins	0.663	Text mining
4	ANKMY2	Ankyrin repeat and MYND domain-containing protein 2	441 aa	May be involved in the trafficking of signaling proteins to the cilia	0.612	Text mining
5	IRF5	Interferon regulatory factor 5	514 aa	Transcription factor that induces interferons IFNA and INFB and inflammatory cytokines upon virus infection. Activated by TLR7 or TLR8 signaling	0.612	Co-expression
6	SPPL2A	Signal peptide peptidase-like 2A	520 aa	Intramembrane-cleaving aspartic protease (I-CLiP) that cleaves type II membrane signal peptides in the hydrophobic plane of	0.566	Text mining

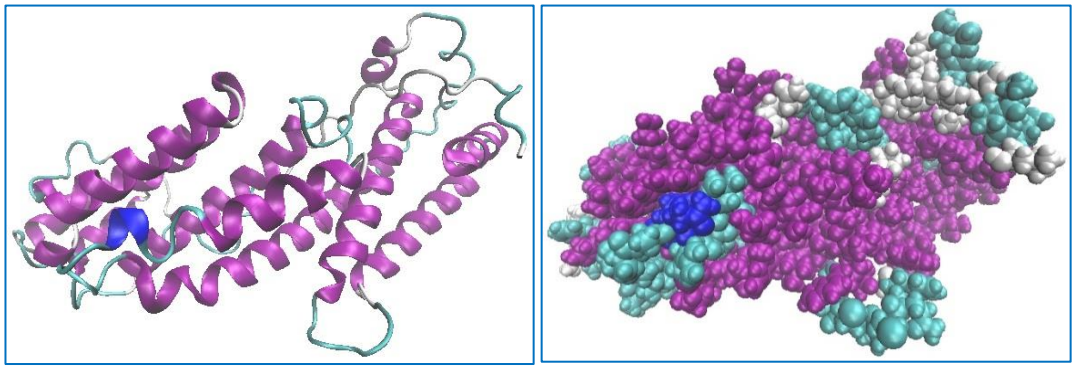
				the membrane. Functions in FASLG, ITM2B and TNF processing		
7	LMAN2L	VIP36-like protein	359 aa	Might regulate the export from endoplasmic reticulum of a subset of glycoproteins. Might regulate ERGIC-53	0.545	Text mining
8	TSPAN13	Tetraspanin 13	204 aa	-	0.544	Text mining
9	SLC8B1	Mitochondrial sodium/calcium exchanger protein	584 aa	Mitochondrial sodium/calcium antiporter that mediates sodium-dependent calcium efflux from mitochondrion	0.529	Homologous co-expression in other species
10	USP25	Ubiquitin carboxyl-terminal hydrolase 25	1125 aa	Deubiquitinating enzyme that hydrolyzes ubiquitin moieties conjugated to substrates and prevents proteasomal degradation of substrates	0.519	Homologous co-expression in other species

4.1.4. 3D Model of Human FAM26F

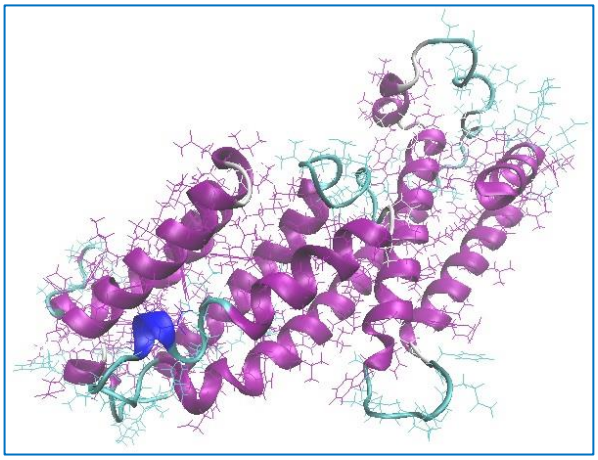
Understanding the structural characteristics of FAM26F required the prediction of both secondary and tertiary structures of the protein. I-TASSER server provided the automated prediction of FAM26F structure and function. The secondary structure predicted by I-TASSER is depicted in supplementary figure 1 (Figure S1) highlighting the helix, strand and coil conformation analogous to each residue in FAM26F sequence.

For the tertiary structure, the server returned five full length models for FAM26F based on C-score. C-score represents a confidence score that reflects the reliability of

assessment and quality of approximation of the models predicted. Usually, this value lies between -5 to 2, where a greater C value indicates a better model quality (Y. Zhang, 2008). The 3D view of the best selected model of FAM26F on the basis of C score and good evaluation results is given in Figure 4.7.



(A)



(B)

Figure 4.7: 3D structure of FAM26F predicted by I-TASSER is visualized by VMD. (A) The figure visibly demonstrates different conformations present in FAM26F. Major part of the protein is in α-helical conformation (purple), with a more condensed 3_{10} *helix* also visible (blue), and a few turns (cyan) and coils (white). (B) The single conserved domain Ca_hom_mod identified in FAM26F is highlighted, covering the positions 1-248 of the total 315 amino acids.

4.2. *In vitro* characterization of FAM26F

4.2.1. Determining Optimal Time Period for Maximum FAM26F Expression

Before proceeding with the localization or immunoprecipitation experiments, it is important to determine the optimal time post transfection at which minimum cell death and maximum protein expression is obtained. For this purpose, HEK293 cells were transiently transfected with FAM26F plasmid. Cellular cytotoxicity was determined at 3h, 6h, 12h, 24h, 36h and 48h post transfection by performing biochemical tests including MTS assay and Caspase-3 activity assay. The results of both the assays showed maximum cell viability and minimum cell death at 24h post transfection (Figure 1a and Figure 1b). The cell viability increased with increasing time and peaked at 24h post transfection, after which it declined and remained almost constant thereafter. In case of cell death assay, minimum number of dead cells were observed at 24h, the time point which was then selected for further experiments.

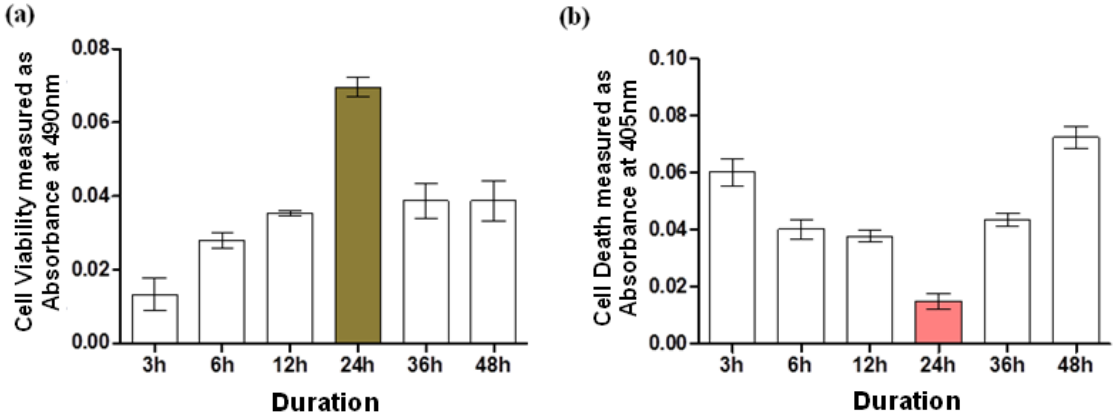
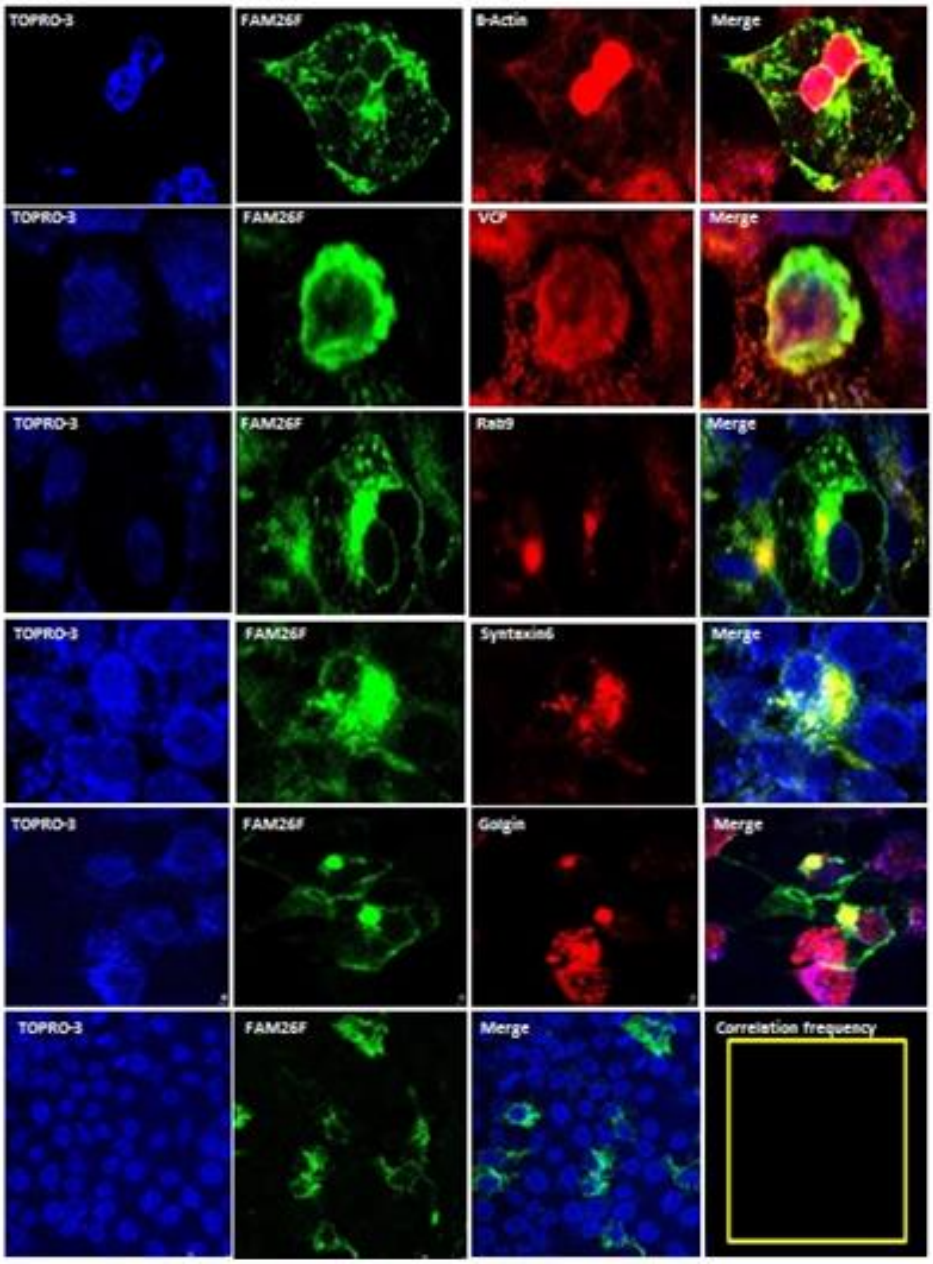


Figure 4.8: Relationship between the absorbance measured with MTS and Caspase assay and the time course post transfection. The results of (a) MTS assay and (b) Caspase-3 activity assay confirmed that HEK293 cells had maximum viability and minimum cell death at 24 h post transfection. The OD values of samples were normalized to the values obtained with untreated control cells and then divided by 10 to adjust/fit with the standard range.

4.2.2. Subcellular Localization of FAM26F

To determine the localization of FAM26F within the cell, HEK293 cells were transiently transfected with GFP-tagged FAM26F plasmid (NM_001010919), followed by co-staining with antibodies specific for respective endogenous markers of various cellular organelles and compartments. The localization of FAM26F was checked 24h post transfection in Endoplasmic reticulum (ER), Golgi apparatus, cytoskeleton, endosomes and in the cell nucleus. Antibody staining revealed FAM26F to be majorly localized within the Golgi apparatus of the cell, whereas its fair presence could also be detected in the ER. The localization within Golgi apparatus was confirmed by using two different Golgi specific antibodies, staining different portions of the Golgi apparatus. No co-localization of FAM26F was observed with the endosomes or nucleus, and it was insignificant in case of cytoskeleton as well. The results are illustrated in Figure 4.9.

(a)


TOPRO-3
FAM26F
β-Actin
Merge
TOPRO-3
FAM26F
VCP
Merge
TOPRO-3
FAM26F
Rab9
Merge
TOPRO-3
FAM26F
Syntaxin6
Merge
TOPRO-3
FAM26F
Golgin
Merge
TOPRO-3
FAM26F
Merge
Correlation frequency

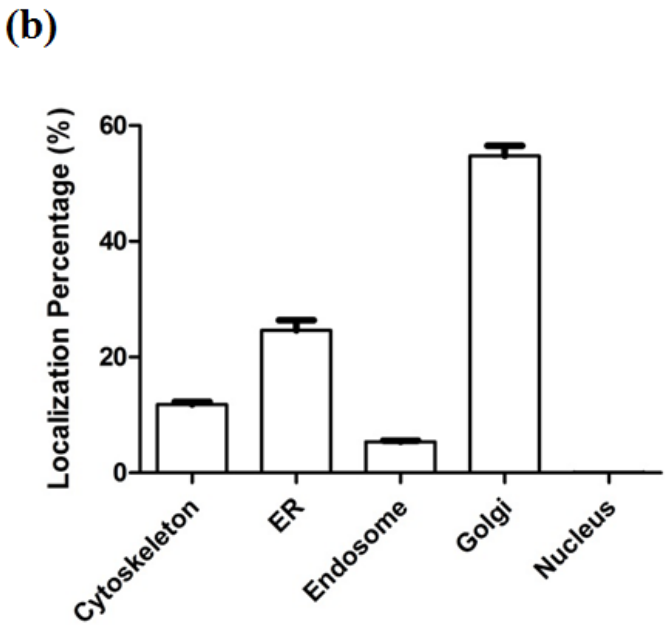


Figure 4.9: Localization of FAM26F in Golgi apparatus of HEK293 cells. (a) Confocal laser scanning microscopy images of HEK293 cells depicting transfected FAM26F (green) co-stained with various organelle specific markers (red channel), including antibodies against Beta actin (cytoskeleton marker), VCP (ER marker), Rab9 (endosomal marker), Syntaxin 6 and Golgin (Golgi apparatus marker). Nucleus was stained using TOPRO3 iodide (blue channel). Scale bar:10 µm. **(b)** Densitometric analysis from 25 different images clearly reveal FAM26F to be majorly localized in the Golgi apparatus of the cell.

4.2.3. Co-Immunoprecipitation of FAM26F and MS/MS Analysis

To determine the interacting partners of FAM26F, control (untransfected) and transfected HEK293 cell lysates (24h) were immunoprecipitated with rabbit anti-FAM26F antibody. Negative controls were also used in order to identify and eliminate the non- specifically bound samples/antibody. Eluates were resolved on SDS-PAGE, electrotransferred onto PVDF membrane, and detected with FAM26F antibody (Figure 4.10).

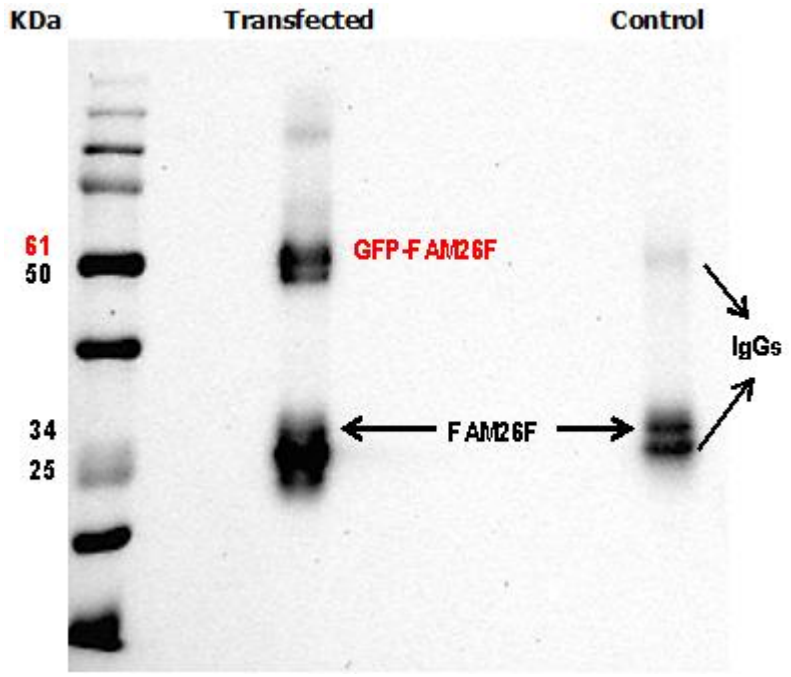


Figure 4.10: Co-immunoprecipitation and Western blot analysis. FAM26F control (untransfected) and transfected HEK293 cell lysates (24h) were immunoprecipitated with rabbit anti-FAM26F antibody. Eluates were resolved on SDS-PAGE, electrotransferred onto PVDF membrane, and detected with FAM26F antibody. The figure shows successful elution of FAM26F complex (transfected) at 61 kD. The endogenous protein expression can be observed at 34 kD. The IgG bands at 25 and 50 kD respectively are also visible.

Once positive confirmation was attained, the residual eluate was 1-DE resolved and then stained with Coomassie Blue. The entire lane from the eluates was cut, in-gel digested, and proteins were classified through the Q-TOF MS/MS analysis. The proteins present abundantly in the negative sample lanes were taken as background contaminants. Therefore, they were removed from proteins list obtained from control and FAM26F transfected eluates. Likewise, the trypsin digested products were also removed as background hits. Moreover, exclusion criteria was also applied so that only those proteins were displayed which had a total calculated probability of $\geq 95\%$, and had no fewer than 2 identified unique peptides that have at least 99% identification probability. Out of the protein list that remained, those proteins were considered for further analysis which were differentially expressed between the control and transfected eluate, the peptide count being higher for transfected eluate as compared to the control eluate. Total amount of such proteins came out to be 44.

Each of these 44 proteins was then manually checked for its function through Uniprot (http://www.uniprot.org/) (Table S2).

These 44 proteins were then analyzed using Reactome database in order to predict the pathways in which they were involved. Reactome results showed that above half of the proteins (almost 52%) identified in the study were involved in innate immune system, 38.6% in neutrophil degranulation, and the remaining 10-13% were either involved in phosphorylation, degradation or regulation of apoptosis (Figure 4.11).

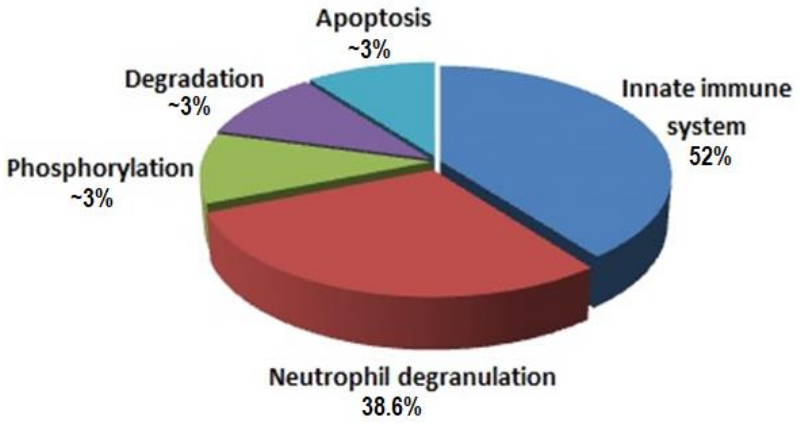


Figure 4.11: Distribution of identified proteins into functional groups. The largest group comprising 52% of the proteins is constituted by innate immune system, followed by neutrophil degranulation with 38.6% of proteins. A low percentage (10%) of proteins was involved either in phosphorylation, degradation or in regulation of apoptosis.

To get further insight into these broader categories displayed by Reactome, the refined protein set was analyzed with IPA software which revealed that majority of the identified proteins fall in the category of calcium-binding proteins, and may hence be involved in maintaining calcium homeostasis of the cell (Figure 4.12). The six proteins involved in calcium regulatory mechanism were finally selected for further validation. Table 4.5 enlists the selected proteins along with their calcium specific/dependent functions.

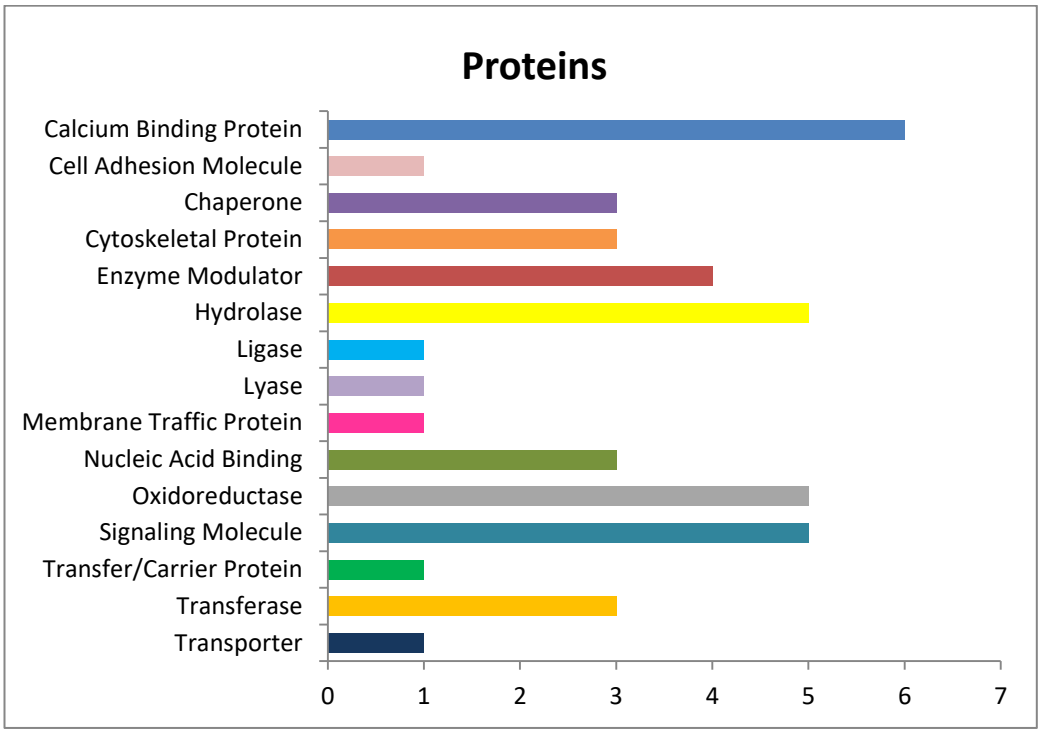


Figure 4.12: IPA distribution of identified proteins into functional groups. IPA revealed that majority of the identified FAM26F interacting proteins fall in the category of calcium-binding proteins, and hence FAM26F may be involved in maintaining calcium homeostasis of the cell. The other major interacting groups included oxidoreductase, signaling molecules and hydrolases.

Table 4.6: The selected MS identified proteins and their functional involvement in calcium regulation/homeostasis

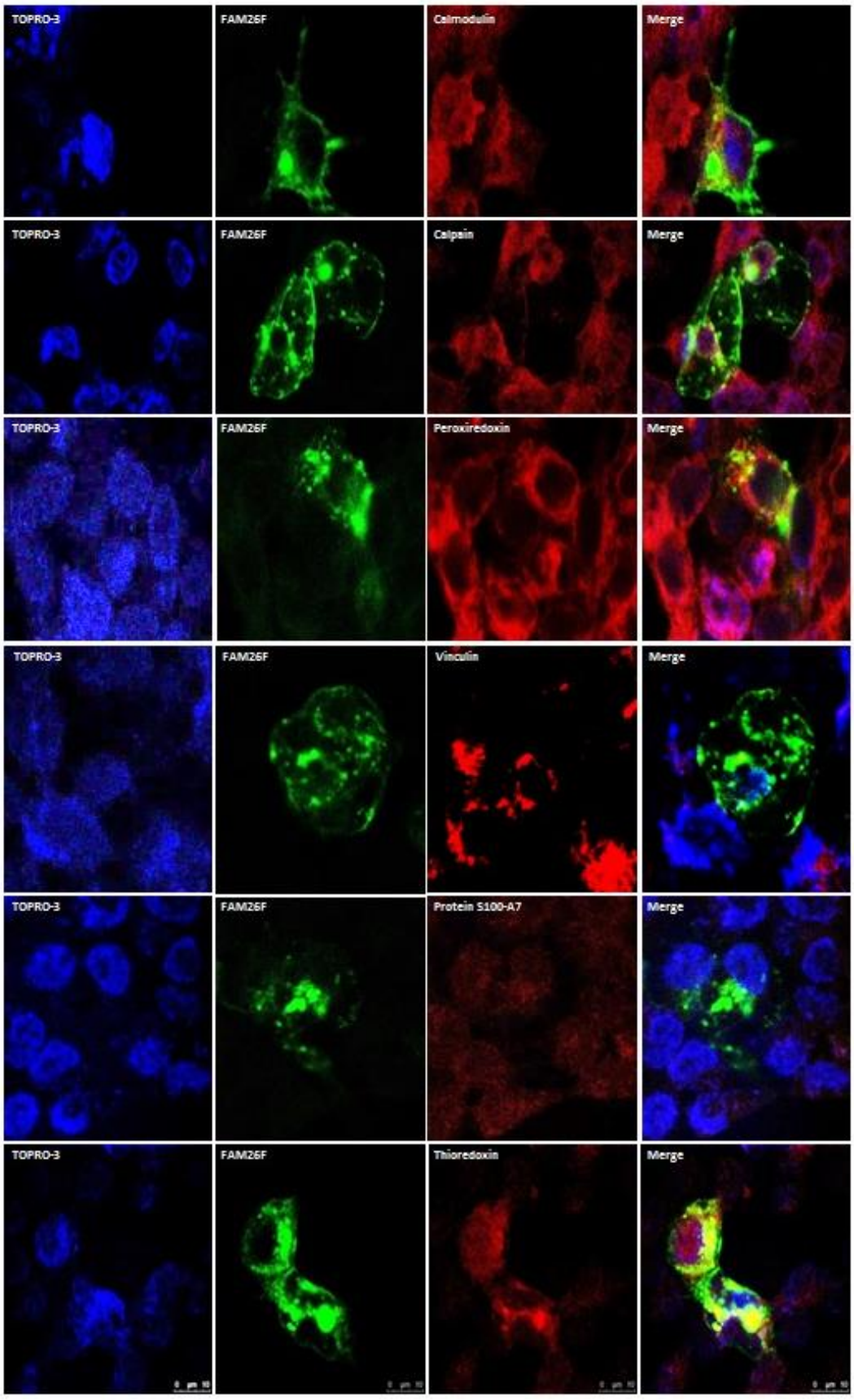
No.	Accession Number	Protein Name	Function	Localization (Uniprot)	Refs
1.	P07384	Calpain	A non-lysosomal thiol-protease regulated by Calcium that catalyzes partial proteolysis of substrates required for signal transduction and cytoskeletal remodeling	Cytoplasm; Translocates to the plasma membrane upon Ca2+ binding	(Sorimachi et al., 1990) (Hsu et al., 2011)
2.	P18206	Vinculin	A key platelet protein which experiences calcium dependent tyrosine phosphorylation during the activation of platelet	Plasma membrane; cytoskeleton Other: adherens junction; focal adhesion	(Vostal & Shulman, 1993)
3.	P31151	Protein S100-A7	Calcium-binding protein containing the EF hand motif that displays antimicrobial activities against bacteria and triggers immunomodulatory activities	Extracellular region or secreted by a non-classical secretory pathway; Other: cytoplasm	(Gläser et al., 2005)
4.	P10599	Thioredoxin	Calcium-dependent oxidation of thioredoxin occurs during the initiation of cellular growth and stress conditions	Extracellular region or secreted by a leaderless secretory pathway; cytoplasm; nucleus	(Gitler et al., 2002)

5.	P32119	Peroxiredoxin-2	Prx2 plays a role in calcium-activated potassium transport through the Gardos channel, and calcium has been reported to increase membrane binding of Prx2	cytoplasm	(Moore et al., 1991) (Bayer et al., 2016)
6.	Q9NZT1	Calmodulin-like protein 5	A calcium-binding regulatory protein transmitting a momentary increase in the concentration of intracellular calcium to activation of specific enzymes	Extracellular region or secreted	(Linse et al., 1991)

4.2.4. Functional affiliation of MS/MS Identified Interactors with FAM26F

To further determine the extent of functional relatedness of each of the selected proteins with FAM26F, proteins were immuno-stained and their co-localization with FAM26F was visualized using confocal laser scanning microscopy as described earlier. The ImageJ (WCIF plugin) software was used to determine the extent of co-localization. Co-localization in the fluorescence imaging is characterized by the amount of overlap displayed by two dissimilar fluorescent labels having dissimilar emission wavelengths. If the fluorescent signal from two separately labelled proteins is detected within the same 3D pixel, it means that the two proteins are present at the same physical location or are extremely close to each other. In this study, it was demonstrated that FAM26F majorly interacts with Thioredoxin (Trx), as was evident from its high co-localization frequency. Moreover, some co-localization was also seen with Peroxiredoxin, whereas the remaining proteins did not show significant results (Figure 4.13). Interestingly, both Thioredoxin and Peroxiredoxin have significant roles in Thioredoxin system which detoxifies the reactive oxygen species (ROS) and hence maintains the cells in a reduced environment (Dunn et al., 2010).

(a)


TOPRO-3
FAM26F
Calmodulin
Merge
TOPRO-3
FAM26F
Calpain
Merge
TOPRO-3
FAM26F
Peroxiredoxin
Merge
TOPRO-3
FAM26F
Vinculin
Merge
TOPRO-3
FAM26F
Protein S100-A7
Merge
TOPRO-3
FAM26F
Thioredoxin
Merge

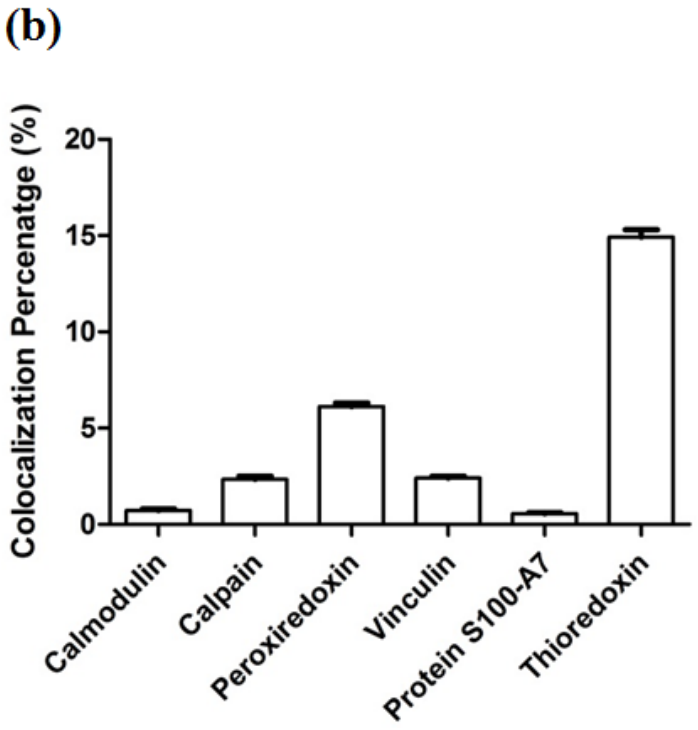


Figure 4.13: Co-localization of FAM26F with its various identified interactors. (a) Confocal laser scanning microscopy images of HEK293 cells depicting transfected FAM26F (green) co-stained with various identified interactors (red channel), including Calmodulin, Calpain, Peroxiredoxin, Vinculin, Protein S-100 A7 and Thioredoxin. Nucleus was stained using TOPRO3 iodide (blue channel). Scale bar: 10 µm. **(b)** Densitometric analysis from 4 independent images (±SD). The results showed FAM26F to be majorly interacting with Thioredoxin protein, and to a lesser extent with Peroxiredoxin protein, both of which are critical proteins of Thioredoxin system. Co-localization with other interactors was not significant.

4.3. Expression Analysis of FAM26F

As co-immunoprecipitation and immunofluorescence studies revealed FAM26F to predominantly interact with Trx, hence the expression of FAM26F and Trx was next determined in the PBMCs of healthy individuals and Human Immunodeficiency Virus-1 (HIV-1) infected Acquired Immunodeficiency Syndrome (AIDS) patients, to know whether these genes had association at mRNA level as well. For this, qPCR analysis was performed utilizing the cDNA prepared from the blood extracted RNA from these two group of individuals. The results revealed that FAM26F was significantly downregulated in HIV patients as compared to the controls (p value:

0.0491). On the other hand, Trx was significantly upregulated in HIV patients in comparison to the healthy individuals (p value: 0.0010). When both these genes were simultaneously analyzed, it was shown that Trx was significantly upregulated than FAM26F in both the controls (p value: 0.0367) as well as patient samples (p value: <0.0001), the upregulation being many folds higher in the patients than in the controls (Figure 4.14).

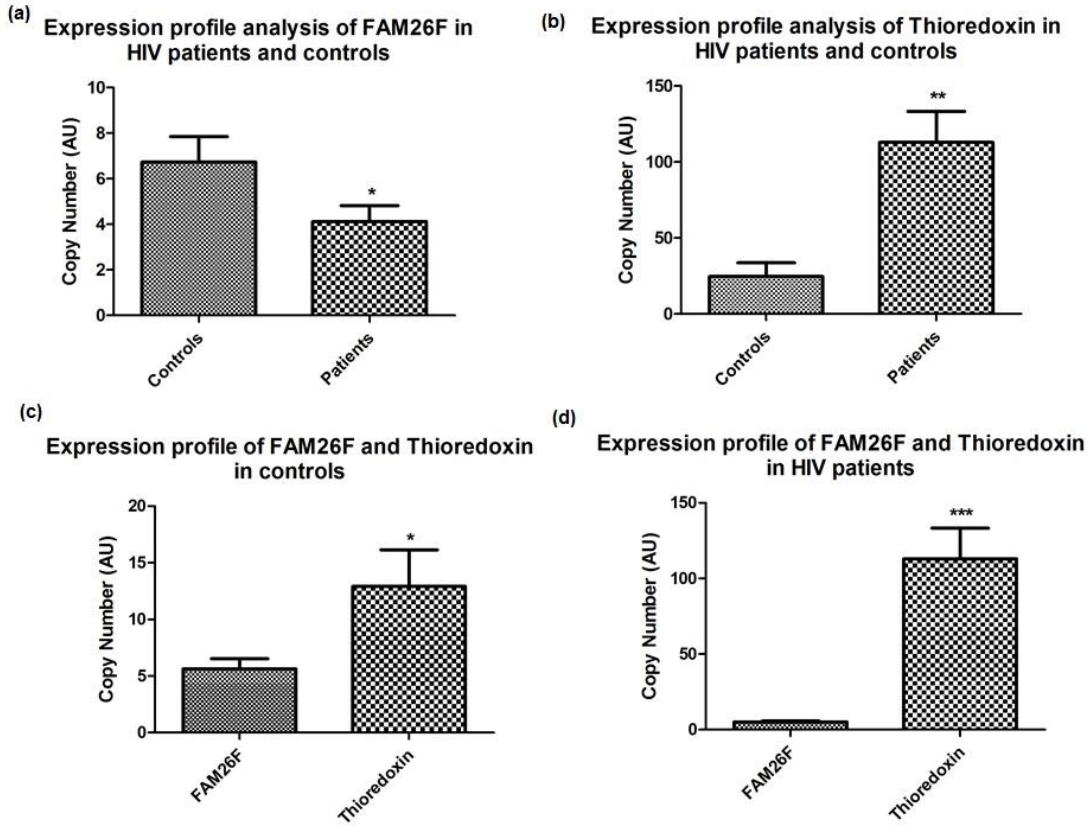


Figure 4.14: Expression analysis of FAM26F and Thioredoxin in healthy individuals and HIV patients. (a) FAM26F appears to be significantly downregulated in HIV patients as compared to controls (p value: 0.0491). **(b)** On the other hand, Trx is significantly upregulated in HIV patients in comparison to the healthy individuals (p value: 0.0010). **(c&d)** Simultaneous analysis of FAM26F and Trx showed Trx to be significantly upregulated than FAM26F in both the controls (p value: 0.0367) as well as patient samples (p value: <0.0001), the upregulation of gene being many folds higher in the patients than in the controls. The significance was calculated by Student's *t* test (*P < 0.05, **P < 0.01, ***P < 0.001). *AU*: arbitrary units

Chapter 5

DISCUSSION

Numerous studies have reported FAM26F to play a critical role in immunity, infection, cell differentiation and as an antitumor agent, and have demonstrated it to be regulated by different cytokines/interferons like IFN-γ, TNF etc. (Malik & Javed, 2016), the comprehensive protein annotation, precise function and modulatory pathways of FAM26F are yet unknown. Hence in the current study, FAM26F was completely analyzed using an innovative *in silico* approach which provided a simple and precise method for rapid identification of the hidden structure, function, evolutionary significance, localization and post translational modifications of FAM26F. Moreover, functional characterization of FAM26F was accomplished by analyzing its subcellular localization and identifying its interacting partners in HEK293 cells. Molecular techniques like transient transfection, immunofluorescence, western blot analysis and co immunoprecipitation, as well as advanced tools like confocal laser scanning microscopy and IPA were employed to increase the precision and significance of the outcomes. Finally, once the interactors were identified, the expression of FAM26F and its chief interactor was studied in the normal individuals as well as in HIV patient samples in order to better comprehend their association in a diseased environment.

5.1. *In silico* characterization of FAM26F

Human FAM26F was found to be located on human chromosome 6 reference genomic contig NC_000006.12, mapping to the chromosomal position 6q22.1. It encodes a 1141 bp long mRNA possessing 3 exons that translates into a 315 amino acid long, stable protein with a molecular weight of 34.258 kD. The multiple sequence alignment and phylogenetic analysis revealed that the FAM26F has remained relatively conserved during evolution, showing greatest homology with primates, especially with Chimpanzee. FAM26F is probably located within the plasma membrane of a cell having no signal peptide attached to it, thereby eliminating the chances of being classically secreted. However, on the basis of the NN score, which came out to be higher (0.66) than the normal threshold of 0.5 for

the non-secreted proteins, SecretomeP revealed that FAM26F is a non-classically secreted protein. The susceptibility of FAM26F to be a membrane protein was predicted by four different softwares, namely TMHMM, SOSUI, HMMTOP and PROTTER, all of which gave a positive result. However, each software returned a different number of transmembrane helices in the protein sequence. SOSUI, TMHMM, HMMTOP and PROTTER identified 3, 4, 5 and 4 TM helices respectively. Although the number of predicted transmembrane helices was different, the positions covered by TM helices were synchronized as can be seen in Table 4.2.

Generally, genes contain one or more structural and/or functional units, called domains, which arrange in different patterns to represent the wide range of protein families found in nature (Punta et al., 2012). In FAM26F, only a single well conserved domain Ca_hom_mod was observed to be significant, covering 248 residues of the protein (Figure 4.4). Ca_hom_mod is a family of transmembrane proteins which control the concentration of cytosolic calcium and thus might be pore-forming ion channels (Dreses-Werringloer et al., 2008). The molecular function of FAM26F was strengthened by Argot2 results which demonstrated that FAM26F is a cation channel that assists the transport of substances (such as macromolecules, small molecules, ions) into, out of or within a cell, or between cells.

Moreover, FAM26F was predicted to have an immunoglobin-like fold. The immunoglobulin (Ig) fold is one of the most common protein modules that consist of a pair of β sheets having antiparallel β strands, bridged together by a single disulphide bond and surrounding a central hydrophobic core. Two key features determine the function of this structure. First, three hypervariable loops existing on one side of the structure serving as a potential surface for binding. These loops encompass the hypervariable sequences occurring in the immune cells like antibodies and in T-cell receptors. Second, the amino terminus is at the opposite end from the carboxyl terminus, which unites the structural domains to form chains, similar to the L and H chains of antibodies (P. Bork et al., 1994). The immunoglobulin folds/domains often interact with other immunoglobin-like domains via their beta-sheets (Potapov et al., 2004). The possession of this specific fold by FAM26F emphasizes its role in the immune responses in diseased conditions.

Post translational modifications can affect the physical and chemical attributes of proteins, making their manifestation vital for protein heterogeneity. Glycosylation is the process of adding a carbohydrate moiety (the glycosyl group) to a protein molecule like Asparagine, hydroxyline, serine or threonine (Gupta et al., 2004). In FAM26F, a potential N-glycosylation site was observed. However, proteins lacking a signal peptide are probably not visible to the N-glycosylation machinery and hence may not undergo glycosylation (in vivo) despite the presence of potential motifs (Gupta et al., 2004). Thus FAM26F may also be deprived of this modification as it does not contain any signal peptide. Another important PTM is phosphorylation which plays a vital role in cellular signal pathways, cellular regulation, membrane transport, metabolism, growth and differentiation (Wong et al., 2007). In FAM26F, there are 14 potential phosphorylation sites which include phosphorylation at 8 Serine, 3 Threonine and 3 Tyrosine residues (Figure 4.5(b)). Some of these sites are phosphorylated by kinases whereas there are some other sites/residues in FAM26F which although do not appear to be phosphorylated but are predicted to undergo kinase specific phosphorylation (Table 4.3). The results unveiled Ser 96 to hold the highest probability to be phosphorylated by protein kinase C (PKC) and thus warrant experimental studies in this regard. However, when the alignment of FAM26F protein sequences was analysed in comparison to the predicted phosphorylated sites, none of the sites except one (Tyrosine at position 264) appeared to be conserved among the species (Figure S2). Hence, this Tyrosine residue may have some significant role in the regulation of FAM26F.

Although functional features determined by domain detection and subcellular localization are highly essential, the precise prediction of biological and cellular functions needs the assistance of protein-protein interactions as these interactions help us to reveal the involvement of a protein in various metabolic pathways. FAM26F was found to interact with ten proteins, namely PHF20L1, HEXB, NMT1, ANKMY2, IRF5, SPPL2A, LMAN2L, TSPAN13, SLC8B1 and USP25 (Figure 4.6). Except for IRF5, all the other nine proteins were among the several proteins which were identified to play roles in the defense and immunity in swine (Dawson et al., 2013). However, no direct interaction between the proteins was demonstrated in the study. Moreover, no such association or categorization between FAM26F and these nine proteins has been found in humans also. On the other hand, the association of

FAM26F with IRF5 is already discussed in the section of literature review. In response to N-glycan structures expressed on tumor cells, Dectin-1-IRF5 pathway is triggered, resulting in the activation of dendritic cells/macrophages which leads to enhanced cell surface expression of FAM26F, which in turn enhances the tumoricidal activity of NK cells (Chiba et al., 2014). Despite of this, no known (from curated databases or experiments) or predicted (through gene neighborhood, gene fusions or gene co-occurrence) interaction has been observed between FAM26F and IRF5. So although IRF5 is essential for FAM26F expression as deletion of IRF5 results in loss of or markedly reduced *FAM26F* expression in dendritic cells/macrophages, they may interact indirectly to bring about the immune activity attributed to them. This may also be the reason why IRF5 was not identified as one of the interactors during the co-immunoprecipitation experiments. Hence, the interacting partners of FAM26F were then reliably determined through experimental procedures.

Knowledge of 3D structure provides invaluable insights for estimating the molecular foundation of protein function. In this study, owing to the absence of any PDB template for structure prediction through homology modelling, *ab-initio* approach was undertaken to generate a high quality 3D structure of FAM26F as analyzed by multiple structure analysis tools (Figure 4.7). However, this model should be validated by X-ray crystallography or Nuclear Magnetic Resonance (NMR) before using it for further analysis like active site prediction, potential ligand binding sites etc.

5.2. *In vitro* characterization of FAM26F

For the *in vitro* studies, all experiments were performed on HEK cells. Mammalian cells such as HEK-293 and the Chinese hamster ovary (CHO) cells are the common hosts used to express recombinant proteins to study their biophysical, structural, and pharmacological properties (Baldi et al., 2007; Dalton & Barton, 2014). However, HEK-293 cells are preferred over CHO for the expression of membrane proteins (Bollin et al., 2011) primarily because of their possession of post-translational modification machineries that aid in the proper folding and/or optimal biological activity of target proteins. Moreover, they exhibit high transfection efficiency,

faithful translation, and processing of proteins (Wurm, 2004) leading to higher protein yields (Backliwal et al., 2008). These attributes along with their morphology and cell size, rapid rate of division, easy maintenance, and their ability to express ion channels and transgenic receptor proteins with high fidelity (P. Thomas et al., 2005), made them the most suitable choice for the experiments.

The immunofluorescence results revealed that FAM26F is majorly expressed inside a cell's Golgi apparatus. Golgi apparatus has two types of protein populations, Golgi transient proteins and Golgi resident proteins. The proteins that undergo post translational modifications while passing through the Golgi apparatus and are then selectively targeted to various organelles are known as Golgi transient proteins. Whereas, the population of Golgi proteins performing these functions are the Golgi resident proteins (Munro, 1998). FAM26F comes in the latter category. This is evident from the fact that FAM26F does not contain any 'leader sequence' (signal peptide) attached to it and also does not undergo N-glycosylation (Malik et al., 2017), omitting its chance to be targeted to other organelles or to the extracellular space. The Golgi resident proteins may in turn either be integral membrane proteins (embedded within the membrane) or peripheral membrane proteins present on the Golgi's cytoplasmic face. FAM26F contains 3-5 transmembrane helices (Malik et al., 2017), which readily makes it an integral membrane protein.

There are various retention signals/mechanisms which are accountable for the localization of resident proteins within the Golgi apparatus. The resident proteins may possess either one or several of these attributes within their sequence or structure which can contribute to their steady-state retention in a particular Golgi sub-compartment (Banfield, 2011). One of the key retention signals which on its own is adequate to confer Golgi localization of a protein is the presence of a single transmembrane domain (TMD) with a small portion of N-terminal cytoplasm (Munro, 1998). The significance of TMD with reference to Golgi localization has previously been observed in some glycosyltransferases, Golgi-resident SNARE proteins e.g. SedSp and Sftlp3r4, and with certain viral proteins destined to the Golgi apparatus (Banfield et al., 1994; Rayner & Pelham, 1997). Interestingly, FAM26F has also been found to possess a single transmembrane 'calcium homeostasis

modulator' domain (Malik et al., 2017), preceded by a short cytoplasmic portion at N-terminal, which confers its localization to the Golgi apparatus.

Moreover, some membrane proteins have Golgi localization signals in their cytoplasmic domains or within the sequences flanking the TMD. Some well-known examples are TGN38, Furin, and cation-dependent and cation-independent mannose 6-phosphate receptors (MPRs), in which cytoplasmic domain not only aids their localization in the Golgi (Conibear & Pearse, 1994; Kornfeld, 1992; Kornfeld & Mellman, 1989; J. P. Luzio & Banting, 1993; Mauxion et al., 1995; Stanley & Howell, 1993; Takahashi et al., 1995), but also has the ability to interact with cytosolic transport proteins (Glickman et al., 1989; S. M. Jones et al., 1993; Le Borgne et al., 1993). In these protein, there are certain tyrosine or serine-containing signals which are responsible for their trans Golgi network (TGN) localization (Takahashi et al., 1995). In MPR, a serine (Ser) residue in the cytoplasmic tail having the tendency to get phosphorylated by casein kinase II (CK-II), is responsible for Golgi localization of the protein and is also closely linked with release from the TGN (Körner et al., 1994; Le Borgne et al., 1993; Méresse & Hoflack, 1993; Meresse et al., 1990). Consistently, FAM26F also has a Serine (Ser) residue in its cytoplasmic tail at position 311 which is predicted to be phosphorylated by CKII (Malik et al., 2017) and which might also aid in the transport activities of FAM26F. Hence in view of the results, FAM26F is affirmed to be localized in the Golgi apparatus, owing to its possession of TMD and the Ser-containing sequence, just like some of the other Golgi retained proteins possessing both of these signals (Fenteany & Colley, 2005; Zerfaoui et al., 2002).

Intriguingly, FAM26F was previously proposed to exist on the plasma membrane of the immune cells (Ebihara et al., 2010). Similar results were obtained from the *in silico* prediction of FAM26F localization through CELLO. However, the results of *in vitro* experiments indicated FAM26F to be a Golgi resident protein. The difference in the proposed/predicted and experimental results can be due to the non-classical secretion of FAM26F from Golgi to the plasma membrane, which might be aided by its Ser-containing sequence in the cytoplasmic domain. This phenomenon of protein cycling between organelles/compartments has long been observed in case of other Golgi retained proteins as well. Furin and TGN38 are both primarily located in the

TGN but they cycle between TGN and the plasma membrane (Ladinsky & Howell, 1992; J. Luzio et al., 1990; Reaves et al., 1993; Takahashi et al., 1995). MPRs are predominantly TGN and endosome localized proteins and they also recycle between these compartments and the cell surface (Kornfeld, 1992; Kornfeld & Mellman, 1989).

In order to perform its function in vivo, a protein hardly operates alone (Yanagida, 2002). In fact, more than 80% of the proteins have been found to be acting in complexes (Berggård et al., 2007). Moreover, the proteins carrying out similar cellular processes often interact with each other to perform their function (Von Mering et al., 2002). Thus, a protein with known function can serve as a tool to determine the function of its interacting proteins whose function is yet unidentified. Current study identified the interacting partners of FAM26F using co-immunoprecipitation and immunofluorescence techniques in order to determine the cellular process regulated by this protein. HEK293 cells were transiently transfected with FAM26F plasmid, cells were lysed and FAM26F along with its interacting partners were co-immunoprecipitated using Dyna beads. The eluates were subjected to MS analysis to identify the interacting proteins, which were then scrutinized to determine the cellular processes they belong to. It was observed that majority of the proteins were involved in innate immune system. Further analysis through IPA software showed that most proteins were regulating the calcium homeostasis pathway of the cell. Hence proteins linked with calcium signaling were selected and subsequently analyzed by inspecting their co-localization with FAM26F using immunofluorescence and confocal laser scanning microscopy.

Co-localization of FAM26F was visualized with various identified interactors including Calmodulin, Calpain, Peroxiredoxin, Vinculin, Protein S-100 A7 and Thioredoxin. The results demonstrated FAM26F to be highly colocalizing with Trx, whereas co-localization with Peroxiredoxin was also seen to some extent. Trx is a small dithiol-disulfide oxidoreductase that exists in all living cells (Arnér & Holmgren, 2000). Trx belongs to the Thioredoxin system, one of the principal antioxidant systems in mammalian cells which ensure a cell's reduced environment by detoxifying ROS (Dunn et al., 2010). Peroxiredoxin, on the other hand, is thiol-specific thioredoxin-dependent peroxidase that traps the hydrogen peroxide and thus

protects the cell against apoptosis (Wood et al., 2003). Overall, peroxiredoxins react with hydrogen peroxide and get oxidized; Trx then reduces these peroxidases to enable the entrapment of ROS (Dunn et al., 2010).

FAM26F is considered as a pore-forming component of a voltage-operated ion channel having a calcium homeostasis modulator domain, and is expected to assist in calcium homeostasis though its transport activity. The question is how do FAM26F and Trx interact to regulate the calcium homeostasis of a cell and contribute to the innate immune system? This is dependent on the mutual interaction between Calcium and ROS. Calcium (Ca^{+2}) is an essential second messenger modulating both intra- and extracellular signaling mechanisms to regulate a variety of cellular functions through the action of pumps, buffers and exchangers present on the plasma membrane and in the internal organelles (Görlach et al 2015). Among the various signaling pathways with which Ca^{+2} interacts is ROS, which includes hydrogen peroxide (H_2O_2), hydroxyl radicals (HO•) and superoxide anion (O^{2-}•). ROS interacts with signaling molecules that are sensitive to redox reactions such as protein tyrosine phosphatases, protein kinases, transcription factors and ion channels thereby altering their biological activity and resulting in the regulation of cellular processes including hypoxic signal transduction, growth factor signaling, autophagy, stem cell proliferation and differentiation and immune responses (Cordeiro & Jacinto, 2013; Sies, 2015). There is bidirectional interplay among ROS and Ca^{+2} signaling, where ROS can regulate cellular Ca^{+2} signaling by modulating the activity of several Ca^{+2} pumps, channels and exchangers, while Ca^{+2} signaling is necessary for the production of ROS (Gordeeva et al., 2003). Thus, enhanced levels of Ca trigger ROS-generating enzymes to form free radicals (Görlach et al., 2015). Malfunctioning in either of the systems might consequently disturb the functioning of the other system, resulting in detrimental outcomes that may lead to the pathogenesis of a wide range of diseases (Görlach et al., 2015).

5.2.1. Proposed Signaling of FAM26F (Mechanism of Action)

Upon environmental stimulation or under stress conditions, numerous cell types respond by using Ca^{+2} signals to process the information intracellularly and consequently induce appropriate biological responses by activating the expression of specific genes (Berridge et al., 2000; Clapham, 2007). In case of immune cells e.g.

neutrophils, the adhering of particles, fungi, bacteria, or soluble inflammatory mediators to particular cell surface receptors induce the entry of extracellular Ca^{+2} as a primary signaling response to trigger the effector functions (Nunes et al., 2013). This Ca^{+2} activates protein kinase C which then phosphorylates certain cytosolic subunits like p40phox, p47phox, p67phox, and the Rac GTPase in order to translocate them to the plasma membrane where they bind with NADPH oxidase and activates it (Cathcart, 2004). This activated NOX (NADPH oxidase) is known as the 'respiratory burst' enzyme and being part of the innate immunity, mediates the release of large amounts of ROS (Nunes et al., 2013). This happens by the generation of inositol1,4,5-trisphosphate (IP_3) by the membrane receptors, which in turn activates IP_3Rs and Ca^{2+} is released from the intracellular stores (ER and Golgi) through either IP_3Rs or ryanodine receptors (RyR) (Pinton et al., 1998; Steinckwich et al., 2011). Thus Ca^{2+}-induced ROS and subsequent synergistic effect of both on Ca^{2+} release set up a self-amplifying loop (Chaudhari et al., 2014). Increase in intracellular Ca^{2+} is important for immune response(Feske, 2007; R. S. Lewis, 2001; Steinckwich et al., 2011). However, extreme ER stress can result in overload of mitochondrial Ca^{2+}, accumulation of ROS, and depletion of ATP, thereby activating mitochondria-dependent apoptosis (Raturi & Simmen, 2013).

This perturbation in ER function, mediated by rapid decrease in Ca^{2+} concentrations and increase in ROS in the ER lumen constitutes cellular ER stress and the Unfolded Protein Response (UPR) activation (Görlach et al., 2006). The cytosolic Ca^{2+} employs dual ways to exert cellular oxidative response; one by activating NADPH oxidase respiratory burst and secondly by inducing oxidation of Trx (Gitler et al., 2002). As discussed earlier, this Trx detoxifies the ROS and maintains the reduced environment of the cell. Moreover, under stress condition, the mammalian cell Trx1/Tx1R system, normally present in the cytoplasm, can migrate into the nucleus, therein inducing the transcription of certain genes such as p53, AP-1, NF-kB, HIFa and the glucocorticoid receptor, or it can be secreted into the extracellular environment where it contributes to the immune system network (Holmgren & Lu, 2010; D. T. Jones et al., 2006; Lillig & Holmgren, 2007; Saraiva et al., 2002). Similar to FAM26F, Trx also has no signal sequence and hence it gets secreted through unconventional secretory route to extracellular environment (Wollman et al., 1988). This unconventional mode of secretion holds true for several other proteins as

well, including proangiogenic fibroblast growth factor 2 (FGF2) (Seelenmeyer et al., 2008), tumor-mediated immune suppressive galectin 1 (gal-1), inflammatory cytokines like IL-1β (Braddock & Quinn, 2004), IL-1α (Cohen et al., 2010), IL-33 (Kakkar & Lee, 2008), protein high-mobility group box 1 (HMGB1) (Wang et al., 1999) and macrophage migration inhibitory factor (MIF) (Lue et al., 2002). Nevertheless, a protein can employ any of the several different pathways for their unconventional secretion (detailed by (Pompa et al., 2017)).

The results of this study have concluded FAM26F to be localized within a cell's Golgi apparatus. However, it can be translocated back to the ER following the retrograde transport. Retrograde transport is the process by which certain fusion and export proteins like v-SNAREs or Vma21 (Ballensiefen et al., 1998; Malkus et al., 2004), misfolded proteins (Todorow et al., 2000; Valkova et al., 2011), escaped ER proteins (M. J. Lewis et al., 1990; Semenza et al., 1990), or even Golgi resident proteins (Zaal et al., 1999) are recycled back from Golgi apparatus to ER, either for maintaining the organelles' steady state composition (Ivessa et al., 1995) or under stress conditions (Rhee et al., 2005). In case of recycling Golgi resident proteins, one can speculate that the Golgi protein residency is generally distributed amid the Golgi and ER (Jiang & Storrie, 2005; Young et al., 2005). A well-known example of recycling Golgi resident protein is Golgi glycosyltransferases (Rhee et al., 2005). Hence under stress situation or stimulation, FAM26F may recycle to the ER probably through tubular extensions that emerge from the Golgi cisternae, as is the case when retrograde transport occurs following treatment of cells with BFA (Lippincott-Schwartz, 1993), instead of the distinct vesicular carriers similar to the ones mediating anterograde transport. This justifies the apportioned presence of FAM26F in ER visualized during the localization experiments of FAM26F. Further, it has been previously reported that this retrograde transport is dependent on the Ca2+ gradient present between the cytosol and the lumen of the Golgi apparatus and ER (Ivessa et al., 1995).

Once in ER, the stress and discomposure of ER caused by regulated Ca^{2+} and ROS levels and activation of Trx may lead to the secretion of FAM26F from the ER by unconventional means as a part of innate immune response. This secretion from ER maybe through direct transportation to the plasma membrane, or directly to the

extracellular space with or without the use of secretory lysosomes (Pompa et al., 2017). Although both FAM26F and Trx are secreted due to immune activation and by the immune cells, both have entirely different functions. Trx1 is a powerful cell survival and growth factor which has been observed to be considerably elevated in numerous types of cancers (Ceccarelli et al., 2008; Chaiswing et al., 2007; Lincoln et al., 2003; Powis et al., 2000; Rubartelli et al., 1992) and is usually associated with tumor aggressiveness, immune system inhibition (S. J. Kim et al., 2005) and decreased survival in tumor patients (Welsh et al., 2002). On the contrary, FAM26F although upregulated in cancers has been shown to have therapeutic potential against NK sensitive (Ebihara et al., 2010) and IFN-γ –suppressible (Ebihara et al., 2010; Kasamatsu et al., 2014) tumors and has also been associated with clinical benefits in metastatic melanoma (Ulloa-Montoya et al., 2013).

Interestingly, in the extracellular space, Trx1 has been demonstrated to catalytically interact with a single target protein, the CD30 cell membrane receptor (CD30R), expressed on immune cells (B, T, monocytes, NK cells, DCs, granulocytes and eosinophils), which holds great clinical significance (Eichenauer et al., 2007; JJ van derVliet et al., 2007; Schwertassek et al., 2007). This is because the resulting CD30R's redox state governs its capability to bind to its cognate ligand CD30L and also to transduce signals (Schwertassek et al., 2007). Although CD30 is not a well-known molecule, yet it performs diverse roles including signal transduction that regulates the physiological homeostasis of the T helper cell (Th1/Th2/Th3/Th17) network functions and therefore contributes towards an efficient immune response (Contasta et al., 2010; Pellegrini et al., 2003; Zeiser et al., 2007). The functional activity of CD30R is modulated by both Trx1 and CD30's soluble (s) component, sCD30 (Eichenauer et al., 2007; Schwertassek et al., 2007), and abnormally elevated levels of both result in a deficit of Th1-cell function (Contasta et al., 2010; S.-H. Kim et al., 2008; Pellegrini et al., 2003; Saraiva et al., 2002) and have also been observed in cancers (Del Beato et al., 1997; Kuljaca et al., 2009). In line with these findings, it can be suggested that FAM26F may also have the ability to interact with CD30R either directly or indirectly (by binding with Trx and blocking its activity) to bring about the diverse immune responses and tumor regression attributed to it. Nonetheless, whether the interaction leads to the inhibition or activation of immune response depends on whether FAM26F or Trx will associate with CD30R. This

proposition evidently requires thorough investigation through experimental procedures before it can be held true or significant. Graphical representation of the proposed mechanism of action of FAM26F is illustrated in Figure 5.1.

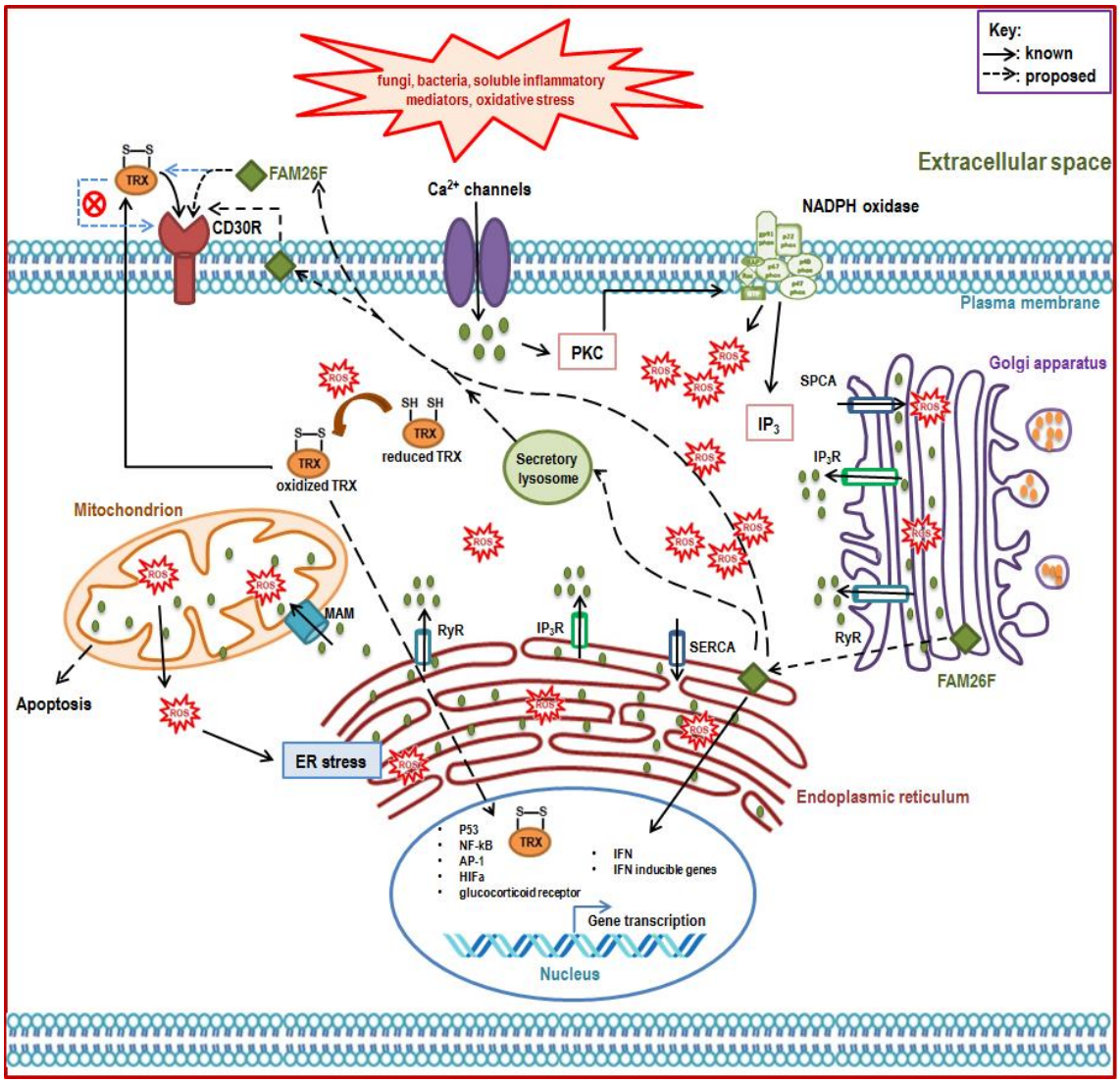


Figure 5.1: Proposed signaling mechanism of FAM26F. Environmental stimulation or stress conditions induce the entry of extracellular Ca2+ into the cell to activate the effector functions. This Ca2+ activates protein kinase C which in turn activates NADPH oxidase. This activated NOX is known as the 'respiratory burst' enzyme and mediates the discharge of surplus amounts of ROS by generating inositol1,4,5-trisphosphate(IP3) by the membrane receptors, which in turn activates IP3Rs and releases Ca2+ from the ER and Golgi through either IP3Rs or ryanodine receptors (RyR). Increased cytosolic Ca2+ induces oxidation of Trx which scavenges the ROS and maintains the reduced environment of the cell. Moreover, the

mammalian cell Trx1/Tx1R system normally present in the cytoplasm can migrate into the nucleus where it induces the transcription of certain genes, or it can be secreted into the extracellular environment where it contributes to the immune system network. On the other hand, the stress and discomposure of ER caused by regulated Ca2+ and ROS levels can induce retrograde transfer of FAM26F back to the ER from Golgi and its subsequent secretion from ER either directly to the plasma membrane, or to the extracellular space with or without the use of secretory lysosomes. Trx1 catalytically interacts with CD30R expressed on immune cells. It is suggested that FAM26F may also have the ability to interact with CD30R either directly or indirectly to bring about the diverse immune responses and tumor regression attributed to it. Nonetheless, whether the interaction leads to the inhibition or activation of immune response depends on whether FAM26F or Trx will associate with CD30R.

5.3. Expression analysis of FAM26F

Numerous microarray studies have shown the differential expression of FAM26F in various cancers and infections; however, none of the studies determined the functional aspect of the protein. One of the three studies that investigated the functional regulation of FAM26F was conducted in SIV, in which FAM26F was proposed as an important prognostic marker of the infection (Javed et al., 2016). This study provided a rational for the future functional studies to explore the nature of FAM26F interaction with other proteins and hence the signaling cascades regulated by their association. It was thus most relevant to analyze the expression of FAM26F in a much similar diseased state and the counterpart of SIV in humans, that is in the HIV patients.

The expression of FAM26F and Trx was determined in the WBCs of healthy individuals and HIV infected AIDS patients using qPCR technique to determine whether these genes had any association at mRNA level and secondly to see their co-relation in a diseased state.

HIV is a crucial health issue worldwide that has taken above 39 million lives since it was first discovered in 1983 (Elliott & Hoyne, 2015), This makes AIDS as one of the top ravaging infectious diseases that is upsetting humanity, with probably 37.9

million people having HIV infection as per 2018 assessments (https://www.unaids.org/en). Humans can contract HIV infection by two distinct yet related viruses: HIV-1 and HIV-2. As HIV-2 is less contagious, hence 98% of AIDS cases arise from HIV-1 infection (Lever, 2009). HIV is an obligate intracellular pathogen that can only replicate after entering the host and hijacking its cellular machinery (Collins & Collins, 2014). Being a small RNA virus, HIV comprises of nine genes; five genes aid in viral replication, whereas remaining four genes named as accessory proteins (Lever, 2009) are primarily responsible for the evasion of host immune response (Collins & Collins, 2014).

The results revealed that FAM26F was significantly downregulated in the control samples than in the HIV patient samples. This was surprising and somewhat inconsistent with the general trend where FAM26F is majorly reported to be increased in the diseased state as compared to normal state (summarized in Table 2.1), and this is also expected of it as FAM26F is speculated as an immune protein. However, this decrease in FAM26F expression in the patient samples can be explained by the fact that HIV is characterized to cause an immunodeficiency state within the infected individuals, which the virus achieves by using various methods. One of these means is decreasing the amount or functionality of immune cells of the body.

CD4+ T-cells are the chief mediators of both cellular and humoral immunity in humans against exogenous antigens and are maintained at a constant ratio with cytotoxic T cells (McCune, 2001). HIV causes both qualitative and quantitative imperfections in the CD4+ T-cell compartment, leading to progressive decline of the cells from circulation as well as from the entire body (Vidya Vijayan et al., 2017). Although gradual loss of CD4+ T-cells number and function is the hallmark of HIV pathogenesis, the magnitude of HIV infection also extends to the innate immune system (Bernstein et al., 2009). NK cells facilitate the innate immunity, and HIV-infected individuals demonstrate loss of NK cell activity (Ratcliffe et al., 1994). Another study reported that HIV infection also influences DC/NK cell interactions. Moreover, CD4+ NK cells attack the vulnerable target cells by increasing production of the cytokines TNF-α and IFN-γ. This function is altered in HIV infected individuals. Plasmacytoid DCs (pDCs) play a central role in innate immune response

against viral pathogens by releasing enormous amount of IFN (Guha & Ayyavoo, 2013). HIV-1 inhibits pDCs activity by reducing pDC cell counts in peripheral blood (Müller-Trutwin & Hosmalin, 2005). Moreover, exposure to HIV-1 gp120 suppresses activation of pDC and the production of pro-inflammatory cytokines (Martinelli et al., 2007).

As FAM26F is reported to be expressed on all of these immune cell populations namely CD4+, NK and DC cells, majorly on CD4+ cells, and its expression is documented to be influenced by IFN-γ, a decline in all these cell types and in IFN-γ production as a result of HIV infection can lead to significant decrease in FAM26F expression, as observed in the qPCR results.

In contrast to FAM26F expression, qPCR results showed Trx to be significantly upregulated in the HIV patient samples in comparison to control samples. Trx1 is secreted from activated B and T cells and resides on the cell surface. Earlier studies have indicated a multifaceted role of Trx1 in HIV infection, involving both promoting and inhibitory activity against HIV infection (Moolla et al., 2016; Newman et al., 1994; Okamoto et al., 1992; Reiser et al., 2016).

At the inception, Trx neutralizes oxidative stress associated with virus attack and blocks HIV replication (Gromer et al., 2004). However, ultimately, Trx contributes to immunosuppression in HIV infected individuals (Nakamura et al., 2001). HIV entry begins by the attachment of viral glycoprotein gp120 to the CD4 receptor of host target cell, trailed by the reduction of structural disulfdes of gp120 and CD4. This causes a conformational modification of gp120 which then interacts with the chemokine co-receptors of the cell i.e. CXCR4 or CCR5, ultimately leading to the fusion of viral envelope with the host cell (Ryser & Flückiger, 2005). Earlier studies have described Trx-1 to be one of the main redox systems efficiently catalyzing gp120 and CD4 reduction (Azimi et al., 2010), hence favoring HIV pathogenesis. In fact, the plasma concentration of Trx is elevated in later stages of HIV infection as viruses can induce Trx expression for their own benefit (Nakamura et al., 2002). Moreover, higher concentrations of Trx have been reported to be associated with lower CD4 counts (Nakamura et al., 1996). As detailed in section 5.2.1, Trx1 is also responsible for the reduction of protein receptor CD30R which then interacts with CD30 and causes downstream signaling. Interestingly, CD30 appears to have a

significant role in AIDS, as CD30 signaling augments HIV replication (Romagnani et al., 1996) and higher levels of serum sCD30 levels are associated with disease progression (Pizzolo et al., 1994).

Overall, the results of *in vivo* experiments showed an antagonistic relation between FAM26F and Thioredoxin. The simultaneous regulation of both the proteins points towards the fact that a single signaling pathway might regulate them. The ultimate immune response (whether activation or repression) will thus depend on the role of the dominant protein.

Although use of antioxidants for HIV/AIDS treatment has prevailed for decades, latest data shows that the main antioxidant systems of the cell including Trx and glutathione (GSH) systems essentially encourage HIV infection while inhibiting effective immune defenses. Thus, inhibiting Trx and GSH may result in effective HIV regression. Interestingly, a recent study showed that anti-Trx1 antibodies used for specific inhibition of Trx1 repressed HIV-1 entry by >80%, signifying a central role for Trx-1 in HIV-1 entry (Moolla et al., 2016). Similar results have been stated by other studies as well in which different drugs targeting Trx-1 or TrxR1 significantly inhibited HIV entry and/or replication, decreased plasma HIV-RNA counts and increased the T-cell counts in patients (Balzarini, 2007; Becker et al., 2000; Fenouillet et al., 2001; Gallina et al., 2002; Lundström & Holmgren, 1990; Reiser et al., 2012; Reiser et al., 2016; Shapiro & Masci, 1996; Trono et al., 2010).

As initial stages of the HIV-1 life cycle considerably depend on Trx1, new generation of anti-HIV therapeutics based on silencing of the Trx1/TrxR1 system and inciting FAM26F expression, as well as targeting the viral surface glycoproteins, may represent a promising approach for HIV treatment.

5.4. Conclusions

The knowledge of a protein's subcellular localization and interacting partners are crucial for elucidating its cellular function and associated regulatory networks. Current study is the first to focus on functional characterization of FAM26F by analyzing its subcellular localization and identifying its novel interacting partners using advanced computational tools and proteome approaches.

In the present study, FAM26F was completely analyzed using an innovative *in silico* approach, which provided a simple and accurate method for rapid identification of the hidden structure, function, evolutionary significance, localization and post translational modifications of FAM26F. Moreover, high throughput proteomics and interactomics techniques were employed to shed light on the physical and functional niches of FAM26F by analyzing its subcellular distribution and complexes under native conditions in HEK293 cells. The computational procedures, confocal laser scanning imaging, immuno flourescence and co-immunoprecipitation experiments provided valuable insights into: (1) apportioned distribution of FAM26F within the Golgi apparatus and ER where it controls the transport of Ca^{2+} ions to regulate calcium signaling of a cell, thereby governing and regulating the immune responses; and (2) identification of the interaction of FAM26F with calcium homeostasis proteins, particularly with Thioredoxin. Moreover, the differential and antagonistic expression of FAM26F relative to Thioredoxin in controls and HIV clinical samples opened new perspectives of a possible mechanistic link between FAM26F and Trx, regulating the immune response.

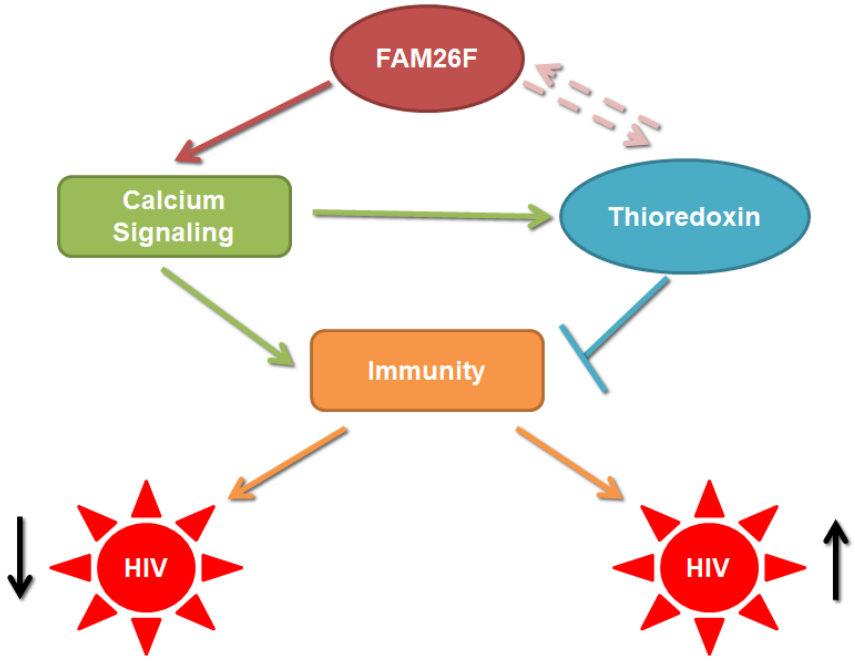


Figure 5.2: Overall summary of the study. FAM26F is localized within a cell's Golgi Apparatus where it controls the transport of Ca^{2+} ions to regulate Calcium signaling of a cell, thereby governing and regulating the immune system. The regulation of Ca^{2+} homeostasis requires FAM26F to associate with certain calcium binding proteins. Among the calcium binding proteins, FAM26F chiefly interacts with Thioredoxin in an inverse association, which signifies that the ultimate immune response depends on the role of the dominant protein. FAM26F dominance will administer a positive immune response, thereby limiting the HIV infection. On the other hand, Thioredoxin will inhibit an appropriate immune response, thus promoting the HIV infection.

5.5. Future prospects

The future studies shall focus on determining the localization and interaction of FAM26F in various other cell lines (both normal and diseased cell lines) to assess whether there is any difference depending on the type of cell line or disease. Moreover, the experiments detailed in the study should be further validated in animal models (*in vivo*) and the interaction pattern should be observed between the normal and diseased conditions to get a clear picture of the impact of disease on regulation of FAM26F and on the physiology of the cells. Knockdown experiments should be conducted to identify the specific genes in the Ca^{2+} signaling pathway that are regulated by the downregulation of FAM26F, and also to identify other pathways which are governed by FAM26F expression.

The potential of FAM26F and Trx for use as targets and biomarkers for various pathological conditions and diseases including HIV have been described and can be inferred from literature, as can be the alterations in the physiological pathways regulating the redox and Ca^{2+} and hence the immunological systems. However, we believe that targeting FAM26F and Trx simultaneously would be more effective and beneficial for optimizing redox regulation and the functioning of immune system. Moreover, further knowledge of the mechanisms that regulate the ROS and Ca^{2+} levels in different cell organelles and the subsequent regulation of FAM26F may result in novel therapeutic strategies for the different diseases that are inflicted by the dysfunctional balance between Ca^{2+} and ROS. Hence, these aspects should be investigated further to enhance our understanding of this enigmatic protein and the ways in which it regulates different signaling pathways or combat the diseases.

REFERENCES

Abraham, S. N., & Arock, M. (1998). *Mast cells and basophils in innate immunity.* Paper presented at the Seminars in immunology.

Ahmed, A. M., Good, B., Hanrahan, J. P., McGettigan, P., Browne, J., Keane, O. M., et al. (2015). Variation in the Ovine Abomasal Lymph Node Transcriptome between Breeds Known to Differ in Resistance to the Gastrointestinal Nematode.

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., et al. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res, 25*(17), 3389-3402.

Arnér, E. S., & Holmgren, A. (2000). Physiological functions of thioredoxin and thioredoxin reductase. *European Journal of Biochemistry, 267*(20), 6102-6109.

Asif, A. R., Armstrong, V. W., Voland, A., Wieland, E., Oellerich, M., & Shipkova, M. (2007). Proteins identified as targets of the acyl glucuronide metabolite of mycophenolic acid in kidney tissue from mycophenolate mofetil treated rats. *Biochimie, 89*(3), 393-402.

Azimi, I., Matthias, L. J., Center, R. J., Wong, J. W., & Hogg, P. J. (2010). Disulfide bond that constrains the HIV-1 gp120 V3 domain is cleaved by thioredoxin. *Journal of Biological Chemistry, 285*(51), 40072-40080.

Backliwal, G., Hildinger, M., Chenuet, S., Wulhfard, S., De Jesus, M., & Wurm, F. M. J. N. a. r. (2008). Rational vector design and multi-pathway modulation of HEK 293E cells yield recombinant antibody titers exceeding 1 g/l by transient transfection under serum-free conditions. *36*(15), e96-e96.

Baldi, L., Hacker, D. L., Adam, M., & Wurm, F. M. J. B. l. (2007). Recombinant protein production by large-scale transient gene expression in mammalian cells: state of the art and future perspectives. *29*(5), 677-684.

Ballensiefen, W., Ossipov, D., & Schmitt, H. D. (1998). Recycling of the yeast v-SNARE Sec22p involves COPI-proteins and the ER transmembrane proteins Ufe1p and Sec20p. *Journal of cell science, 111*(11), 1507-1520.

Balzarini, J. (2007). Targeting the glycans of glycoproteins: a novel paradigm for antiviral therapy. *Nature Reviews Microbiology, 5*(8), 583.

Banfield, D. K. (2011). Mechanisms of protein retention in the Golgi. *Cold Spring Harbor perspectives in biology*, a005264.

Banfield, D. K., Lewis, M. J., Rabouille, C., Warren, G., & Pelham, H. (1994). Localization of Sed5, a putative vesicle targeting molecule, to the cis-Golgi network involves both its transmembrane and cytoplasmic domains. *The Journal of cell biology, 127*(2), 357-371.

Bayer, S. B., Low, F. M., Hampton, M. B., & Winterbourn, C. C. (2016). Interactions between peroxiredoxin 2, hemichrome and the erythrocyte membrane. *Free radical research, 50*(12), 1329-1339.

Becker, K., Gromer, S., Schirmer, R. H., & Müller, S. (2000). Thioredoxin reductase as a pathophysiological factor and drug target. *European Journal of Biochemistry, 267*(20), 6118-6125.

Bendtsen, J. D., Jensen, L. J., Blom, N., Von Heijne, G., & Brunak, S. (2004). Feature-based prediction of non-classical and leaderless protein secretion. *Protein Eng Des Sel, 17*(4), 349-356.

Berggård, T., Linse, S., & James, P. (2007). Methods for the detection and analysis of protein–protein interactions. *Proteomics, 7*(16), 2833-2842.

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., et al. (2000). The Protein Data Bank. *Nucleic Acids Res, 28*(1), 235-242.

Bernstein, H. B., Wang, G., Plasterer, M. C., Zack, J. A., Ramasastry, P., Mumenthaler, S. M., et al. (2009). CD4+ NK cells can be productively infected with HIV, leading to downregulation of CD4 expression and changes in function. *Virology, 387*(1), 59-66.

Berridge, M. J., Lipp, P., & Bootman, M. D. (2000). The versatility and universality of calcium signalling. *Nature reviews Molecular cell biology, 1*(1), 11.

Blom, N., Gammeltoft, S., & Brunak, S. (1999). Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. *J Mol Biol, 294*(5), 1351-1362.

Blom, N., Sicheritz-Ponten, T., Gupta, R., Gammeltoft, S., & Brunak, S. (2004). Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence. *Proteomics, 4*(6), 1633-1649.

Bollin, F., Dechavanne, V., Chevalet, L. J. P. e., & purification. (2011). Design of Experiment in CHO and HEK transient transfection condition optimization. *78*(1), 61-68.

Bork, P., Holm, L., & Sander, C. (1994). The immunoglobulin fold. Structural classification, sequence patterns and common core. *J Mol Biol, 242*(4), 309-320.

Bork, P., & Koonin, E. V. (1996). Protein sequence motifs. *Current opinion in structural biology, 6*(3), 366-376.

Braddock, M., & Quinn, A. (2004). Targeting IL-1 in inflammatory disease: new opportunities for therapeutic intervention. *Nature Reviews Drug Discovery, 3*(4), 330.

Brito, H., Juliano, P., Altemani, C., & Altemani, A. (2005). Is the immunohistochemical study of the inflammatory infiltrate helpful in distinguishing villitis of unknown etiology from non-specific infection villitis? *Placenta, 26*(10), 839-841.

Brown, J., Wallet, M. A., Krastins, B., Sarracino, D., & Goodenow, M. M. (2010). Proteome bioprofiles distinguish between M1 priming and activation states in human macrophages. *J Leukoc Biol, 87*(4), 655-662.

Brown, J. N., Wallet, M. A., Krastins, B., Sarracino, D., & Goodenow, M. M. (2010). Proteome bioprofiles distinguish between M1 priming and activation states in human macrophages. *Journal of leukocyte biology, 87*(4), 655-662.

Cai, C. Z., Han, L. Y., Ji, Z. L., Chen, X., & Chen, Y. Z. (2003). SVM-Prot: Web-based support vector machine software for functional classification of a protein from its primary sequence. *Nucleic Acids Res, 31*(13), 3692-3697.

Caldwell, A. B., Cheng, Z., Vargas, J. D., Birnbaum, H. A., & Hoffmann, A. (2014). Network dynamics determine the autocrine and paracrine signaling functions of TNF. *Genes Dev, 28*(19), 2120-2133.

Carroll, M. C., & Fischer, M. B. (1997). Complement and the immune response. *Current opinion in immunology, 9*(1), 64-69.

Cathcart, M. K. (2004). Regulation of superoxide anion production by NADPH oxidase in monocytes/macrophages: contributions to atherosclerosis. *Arteriosclerosis, thrombosis, and vascular biology, 24*(1), 23-28.

Ceccarelli, J., Delfino, L., Zappia, E., Castellani, P., Borghi, M., Ferrini, S., et al. (2008). The redox state of the lung cancer microenvironment depends on the levels of thioredoxin expressed by tumor cells and affects tumor progression and response to prooxidants. *International Journal of Cancer, 123*(8), 1770-1778.

Chaiswing, L., Bourdeau-Heller, J. M., Zhong, W., & Oberley, T. D. (2007). Characterization of redox state of two human prostate carcinoma cell lines with different degrees of aggressiveness. *Free Radical Biology and Medicine, 43*(2), 202-215.

Chaplin, D. D. (2010). Overview of the immune response. *Journal of Allergy and Clinical Immunology, 125*(2), S3-S23.

Chaudhari, N., Talwar, P., Parimisetty, A., Lefebvre d'Hellencourt, C., & Ravanan, P. (2014). A molecular web: endoplasmic reticulum stress, inflammation, and oxidative stress. *Frontiers in cellular neuroscience, 8*, 213.

Chiba, S., Ikushima, H., Ueki, H., Yanai, H., Kimura, Y., Hangai, S., et al. (2014). Recognition of tumor cells by Dectin-1 orchestrates innate immune cells for anti-tumor responses. *Elife, 3*, e04177.

Chmielewski, S., Olejnik, A., Sikorski, K., Pelisek, J., Błaszczyk, K., Aoqui, C., et al. (2014). STAT1-dependent signal integration between IFNγ and TLR4 in vascular

cells reflect pro-atherogenic responses in human atherosclerosis. *PloS one, 9*(12), e113318.

Clapham, D. E. (2007). Calcium signaling. *Cell, 131*(6), 1047-1058.

Cohen, I., Rider, P., Carmi, Y., Braiman, A., Dotan, S., White, M. R., et al. (2010). Differential release of chromatin-bound IL-1α discriminates between necrotic and apoptotic cell death by the ability to induce sterile inflammation. *Proceedings of the National Academy of Sciences, 107*(6), 2574-2579.

Collins, D. R., & Collins, K. L. (2014). HIV-1 accessory proteins adapt cellular adaptors to facilitate immune evasion. *PLoS pathogens, 10*(1), e1003851.

Conibear, E., & Pearse, B. (1994). A chimera of the cytoplasmic tail of the mannose 6-phosphate/IGF-II receptor and lysozyme localizes to the TGN rather than prelysosomes where the bulk of the endogenous receptor is found. *Journal of cell science, 107*(4), 923-932.

Contasta, I., Totaro, R., Berghella, A., Pellegrini, P., Del Beato, T., Carolei, A., et al. (2010). Soluble CD30: a biomarker for evaluating the clinical risk versus benefit of IFNβ1A treatment in multiple sclerosis patients. *International journal of immunopathology and pharmacology, 23*(1), 213-226.

Cordeiro, J. V., & Jacinto, A. (2013). The role of transcription-independent damage signals in the initiation of epithelial wound healing. *Nature Reviews Molecular Cell Biology, 14*(4), 249.

Cory, A. H., Owen, T. C., Barltrop, J. A., & Cory, J. G. J. C. c. (1991). Use of an aqueous soluble tetrazolium/formazan assay for cell growth assays in culture. *3*(7), 207-212.

Croft, D., O'kelly, G., Wu, G., Haw, R., Gillespie, M., Matthews, L., et al. (2010). Reactome: a database of reactions, pathways and biological processes. *Nucleic acids research, 39*(suppl_1), D691-D697.

Dalton, A. C., & Barton, W. A. J. P. s. (2014). Over-expression of secreted proteins from mammalian cell lines. *23*(5), 517-525.

Dawson, H. D., Loveland, J. E., Pascal, G., Gilbert, J. G., Uenishi, H., Mann, K. M., et al. (2013). Structural and functional annotation of the porcine immunome. *14*(1), 1-16.

De Castro, E., Sigrist, C. J., Gattiker, A., Bulliard, V., Langendijk-Genevaux, P. S., Gasteiger, E., et al. (2006). ScanProsite: detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. *Nucleic acids research, 34*(suppl 2), W362-W365.

Defamie, V., Cursio, R., Le Brigand, K., Moreilhon, C., Saint-Paul, M. C., Laurens, M., et al. (2008). Gene expression profiling of human liver transplants identifies an early

transcriptional signature associated with initial poor graft function. *Am J Transplant, 8*(6), 1221-1236.

Del Beato, T., Berghella, A. M., Pellegrini, P., Domenico, A., & Casciani, C. U. (1997). The role of the soluble CD30 serum level in colorectal cancer: a possible marker for a patient subset which could benefit from IL-2 biotherapy. *Cancer biotherapy & radiopharmaceuticals, 12*(5), 297-304.

Delves, P. J., & Roitt, I. M. (2000). The immune system. *New England journal of medicine, 343*(1), 37-49.

Devilder, M. C., Allain, S., Dousset, C., Bonneville, M., & Scotet, E. (2009). Early triggering of exclusive IFN-gamma responses of human Vgamma9Vdelta2 T cells by TLR-activated myeloid and plasmacytoid dendritic cells. *J Immunol, 183*(6), 3625-3633.

Dönnes, P., & Höglund, A. (2004). Predicting protein subcellular localization: past, present, and future. *Genomics, proteomics & bioinformatics, 2*(4), 209-215.

Dreses-Werringloer, U., Lambert, J. C., Vingtdeux, V., Zhao, H., Vais, H., Siebert, A., et al. (2008). A polymorphism in CALHM1 influences Ca2+ homeostasis, Abeta levels, and Alzheimer's disease risk. *Cell, 133*(7), 1149-1161.

Dunn, L. L., Buckle, A. M., Cooke, J. P., & Ng, M. K. (2010). The emerging role of the thioredoxin system in angiogenesis. *Arteriosclerosis, thrombosis, and vascular biology, 30*(11), 2089-2098.

Ebihara, T., Azuma, M., Oshiumi, H., Kasamatsu, J., Iwabuchi, K., Matsumoto, K., et al. (2010). Identification of a polyI:C-inducible membrane protein that participates in dendritic cell-mediated natural killer cell activation. *J Exp Med, 207*(12), 2675-2687.

Eichenauer, D. A., Simhadri, V. L., von Strandmann, E. P., Ludwig, A., Matthews, V., Reiners, K. S., et al. (2007). ADAM10 inhibition of human CD30 shedding increases specificity of targeted immunotherapy in vitro. *Cancer research, 67*(1), 332-338.

Elliott, H., & Hoyne, G. F. (2015). The role of the Nef protein in MHC-I downregulation and viral immune evasion by HIV-1. *Journal of Clinical & Cellular Immunology, 6*(6).

Falda, M., Toppo, S., Pescarolo, A., Lavezzo, E., Di Camillo, B., Facchinetti, A., et al. (2012). Argot2: a large scale function prediction tool relying on semantic similarity of weighted Gene Ontology terms. *BMC Bioinformatics, 13 Suppl 4*, S14.

Fenouillet, E., Barbouche, R., Courageot, J., & Miquelis, R. (2001). The Catalytic Activity of Protein Disulfide Isomerase Is Involved in Human Immunodeficiency Virus Envelope–Mediated Membrane Fusion after CD4 Cell Binding. *The Journal of infectious diseases, 183*(5), 744-752.

Fenteany, F. H., & Colley, K. J. (2005). Multiple signals are required for α2, 6-sialyltransferase (ST6Gal I) oligomerization and Golgi localization. *Journal of Biological Chemistry, 280*(7), 5423-5429.

Ferreyra, G. A., Elinoff, J. M., Demirkale, C. Y., Starost, M. F., Buckley, M., Munson, P. J., et al. (2014). Late multiple organ surge in interferon-regulated target genes characterizes staphylococcal enterotoxin B lethality. *PLoS One, 9*(2), e88756.

Feske, S. (2007). Calcium signalling in lymphocyte activation and disease. *Nature Reviews Immunology, 7*(9), 690.

Finn, R. D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R. Y., Eddy, S. R., et al. (2014). Pfam: the protein families database. *Nucleic Acids Res, 42*(Database issue), D222-230.

Fritsch, R. (2014). Maturation, reactivity and therapy induced changes of memory T-cells in rheumatic autoimmune diseases.

Gallina, A., Hanley, T. M., Mandel, R., Trahey, M., Broder, C. C., Viglianti, G. A., et al. (2002). Inhibitors of protein-disulfide isomerase prevent cleavage of disulfide bonds in receptor-bound glycoprotein 120 and prevent HIV-1 entry. *Journal of Biological Chemistry, 277*(52), 50579-50588.

Gallucci, S., & Matzinger, P. (2001). Danger signals: SOS to the immune system. *Current opinion in immunology, 13*(1), 114-119.

Geer, L. Y., Domrachev, M., Lipman, D. J., & Bryant, S. H. (2002). CDART: protein homology by domain architecture. *Genome Res, 12*(10), 1619-1623.

Gitler, C., Zarmi, B., Kalef, E., Meller, R., Zor, U., & Goldman, R. (2002). Calcium-dependent oxidation of thioredoxin during cellular growth initiation. *Biochemical and biophysical research communications, 290*(2), 624-628.

Gläser, R., Harder, J., Lange, H., Bartels, J., Christophers, E., & Schröder, J.-M. (2005). Antimicrobial psoriasin (S100A7) protects human skin from Escherichia coli infection. *Nature immunology, 6*(1), 57.

Glickman, J., Conibear, E., & Pearse, B. (1989). Specificity of binding of clathrin adaptors to signals on the mannose-6-phosphate/insulin-like growth factor II receptor. *The EMBO Journal, 8*(4), 1041-1047.

Gordeeva, A., Zvyagilskaya, R., & Labas, Y. A. (2003). Cross-talk between reactive oxygen species and calcium in living cells. *Biochemistry (Moscow), 68*(10), 1077-1080.

Görlach, A., Bertram, K., Hudecova, S., & Krizanova, O. (2015). Calcium and ROS: a mutual interplay. *Redox biology, 6*, 260-271.

Görlach, A., Klappa, P., & Kietzmann, D. T. (2006). The endoplasmic reticulum: folding, calcium homeostasis, signaling, and redox control. *Antioxidants & redox signaling, 8*(9-10), 1391-1418.

Gough, J., Karplus, K., Hughey, R., & Chothia, C. (2001). Assignment of homology to genome sequences using a library of hidden Markov models that represent all proteins of known structure. *J Mol Biol, 313*(4), 903-919.

Grimes, C. Z., Hwang, L. Y., Wei, P., Shah, D. P., Volcik, K. A., & Brown, E. L. (2013). Differentially regulated gene expression associated with hepatitis C virus clearance. *J Gen Virol, 94*(Pt 3), 534-542.

Gromer, S., Urig, S., & Becker, K. (2004). The thioredoxin system—from science to clinic. *Medicinal research reviews, 24*(1), 40-89.

Grumann, D., Scharf, S. S., Holtfreter, S., Kohler, C., Steil, L., Engelmann, S., et al. (2008). Immune cell activation by enterotoxin gene cluster (egc)-encoded and non-egc superantigens from Staphylococcus aureus. *J Immunol, 181*(7), 5054-5061.

Guha, D., & Ayyavoo, V. (2013). Innate immune evasion strategies by human immunodeficiency virus type 1. *Isrn Aids, 2013*.

Gupta, R., Jung, E., & Brunak, S. (2004). Prediction ofN-glycosylation sites in human proteins.

Gurtu, V., Kain, S. R., & Zhang, G. J. A. b. (1997). Fluorometric and colorimetric detection of caspase activity associated with apoptosis. *251*(1), 98-102.

He, X.-S., Nanda, S., Ji, X., Calderon-Rodriguez, G. M., Greenberg, H. B., & Liang, T. J. (2010). Differential transcriptional responses to interferon-α and interferon-γ in primary human hepatocytes. *Journal of Interferon & Cytokine Research, 30*(5), 311-320.

Hirokawa, T., Boon-Chieng, S., & Mitaku, S. (1998). SOSUI: classification and secondary structure prediction system for membrane proteins. *Bioinformatics, 14*(4), 378-379.

Holmgren, A., & Lu, J. (2010). Thioredoxin and thioredoxin reductase: current research with special reference to human disease. *Biochemical and biophysical research communications, 396*(1), 120-124.

Howell, M., & Shepherd, M. (2018). The immune system. *Anaesthesia & Intensive Care Medicine*.

Hsu, C.-Y., Henry, J., Raymond, A.-A., Méchin, M.-C., Pendaries, V., Nassar, D., et al. (2011). Deimination of human filaggrin-2 promotes its proteolysis by calpain 1. *Journal of Biological Chemistry*, jbc. M110. 197400.

Humphrey, W., Dalke, A., & Schulten, K. (1996). VMD: visual molecular dynamics. *J Mol Graph, 14*(1), 33-38, 27-38.

Ismail, N., Olano, J. P., Feng, H.-M., & Walker, D. H. (2002). Current status of immune mechanisms of killing of intracellular microorganims. *FEMS microbiology letters, 207*(2), 111-120.

Ivessa, N. E., De Lemos-Chiarandini, C., Gravotta, D., Sabatini, D. D., & Kreibich, G. (1995). The Brefeldin A-induced retrograde transport from the Golgi apparatus to the endoplasmic reticulum depends on calcium sequestered to intracellular stores. *Journal of Biological Chemistry, 270*(43), 25960-25967.

Jacob, S., Zhu, Y., Kraft, R., Cotto, C., Carmical, J. R., Wood, T. G., et al. (2015). Physiologic and molecular changes in the tracheal epithelium of rats following burn injury. *Int J Burns Trauma, 5*(1), 36-45.

Jacquier, V., Estelle, J., Schmaltz-Panneau, B., Lecardonnel, J., Moroldo, M., Lemonnier, G., et al. (2015). Genome-wide immunity studies in the rabbit: transcriptome variations in peripheral blood mononuclear cells after in vitro stimulation by LPS or PMA-Ionomycin. *BMC Genomics, 16*, 26.

Javed, A. (2012). Gene Expression Pattern and Functional Analysis of CD8 T Cells from individuals with or without anti HIV/SIV noncytolytic activity.

Javed, A., Leuchte, N., Salinas, G., Opitz, L., Stahl-Hennig, C., Sopper, S., et al. (2016). Pre-infection transcript levels of FAM26F in peripheral blood mononuclear cells inform about overall plasma viral load in acute and post-acute phase after simian immunodeficiency virus infection. *Journal of General Virology, 97*(12), 3400-3412.

Jiang, S., & Storrie, B. (2005). Cisternal rab proteins regulate Golgi apparatus redistribution in response to hypotonic stress. *Molecular biology of the cell, 16*(5), 2586-2596.

JJ van derVliet, H., Koon, H. B., Yue, S. C., Uzunparmak, B., Seery, V., Gavin, M. A., et al. (2007). Effects of the administration of high-dose interleukin-2 on immunoregulatory cell subsets in patients with advanced melanoma and renal cell cancer. *Clinical Cancer Research, 13*(7), 2100-2108.

Jones, D. T., Pugh, C. W., Wigfield, S., Stevens, M. F., & Harris, A. L. (2006). Novel thioredoxin inhibitors paradoxically increase hypoxia-inducible factor-α expression but decrease functional transcriptional activity, DNA binding, and degradation. *Clinical cancer research, 12*(18), 5384-5394.

Jones, S. M., Crosby, J. R., Salamero, J., & Howell, K. E. (1993). A cytosolic complex of p62 and rab6 associates with TGN38/41 and is involved in budding of exocytic vesicles from the trans-Golgi network. *The Journal of Cell Biology, 122*(4), 775-788.

Joosse, S. A., van Beers, E. H., Tielen, I. H., Horlings, H., Peterse, J. L., Hoogerbrugge, N., et al. (2009). Prediction of BRCA1-association in hereditary non-BRCA1/2 breast carcinomas with array-CGH. *Breast Cancer Res Treat, 116*(3), 479-489.

Julia, A., Domenech, E., Chaparro, M., Garcia-Sanchez, V., Gomollon, F., Panes, J., et al. (2014). A genome-wide association study identifies a novel locus at 6q22.1 associated with ulcerative colitis. *Hum Mol Genet, 23*(25), 6927-6934.

Kakkar, R., & Lee, R. T. (2008). The IL-33/ST2 pathway: therapeutic target and novel biomarker. *Nature reviews Drug discovery, 7*(10), 827.

Kasamatsu, J., Azuma, M., Oshiumi, H., Morioka, Y., Okabe, M., Ebihara, T., et al. (2014). INAM plays a critical role in IFN-gamma production by NK cells interacting with polyinosinic-polycytidylic acid-stimulated accessory cells. *J Immunol, 193*(10), 5199-5207.

Kim, J. S., Romero, R., Kim, M. R., Kim, Y. M., Friel, L., Espinoza, J., et al. (2008). Involvement of Hofbauer cells and maternal T cells in villitis of unknown aetiology. *Histopathology, 52*(4), 457-464.

Kim, M. J., Romero, R., Kim, C. J., Tarca, A. L., Chhauy, S., LaJeunesse, C., et al. (2009). Villitis of unknown etiology is associated with a distinct pattern of chemokine up-regulation in the feto-maternal and placental compartments: implications for conjoint maternal allograft rejection and maternal anti-fetal graft-versus-host disease. *J Immunol, 182*(6), 3919-3927.

Kim, S.-H., Oh, J., Choi, J.-Y., Jang, J.-Y., Kang, M.-W., & Lee, C.-E. (2008). Identification of human thioredoxin as a novel IFN-gamma-induced factor: mechanism of induction and its role in cytokine production. *BMC immunology, 9*(1), 64.

Kim, S. J., Miyoshi, Y., Taguchi, T., Tamaki, Y., Nakamura, H., Yodoi, J., et al. (2005). High thioredoxin expression is associated with resistance to docetaxel in primary breast cancer. *Clinical Cancer Research, 11*(23), 8425-8430.

Kinet, J.-P. (1999). The high-affinity IgE receptor (FcεRI): from physiology to pathology. *Annual review of immunology, 17*(1), 931-972.

Körner, C., Herzog, A., Weber, B., Rosorius, O., Hemer, F., Schmidt, B., et al. (1994). In vitro phosphorylation of the 46-kDa mannose 6-phosphate receptor by casein kinase II. Structural requirements for efficient phosphorylation. *Journal of Biological Chemistry, 269*(24), 16529-16532.

Kornfeld, S. (1992). Structure and function of the mannose 6-phosphate/insulinlike growth factor II receptors. *Annual review of biochemistry, 61*(1), 307-330.

Kornfeld, S., & Mellman, I. (1989). The biogenesis of lysosomes. *Annual review of cell biology, 5*(1), 483-525.

Krämer, A., Green, J., Pollard Jr, J., & Tugendreich, S. (2013). Causal analysis approaches in ingenuity pathway analysis. *Bioinformatics, 30*(4), 523-530.

Krogh, A., Larsson, B., von Heijne, G., & Sonnhammer, E. L. (2001). Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol, 305*(3), 567-580.

Kuljaca, S., Liu, T., Dwarte, T., Kavallaris, M., Haber, M., Norris, M. D., et al. (2009). The cyclin-dependent kinase inhibitor, p21 WAF1, promotes angiogenesis by repressing gene transcription of thioredoxin-binding protein 2 in cancer cells. *Carcinogenesis, 30*(11), 1865-1871.

Kumar, S., Stecher, G., & Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Mol Biol Evol*.

Ladinsky, M. S., & Howell, K. E. (1992). The trans-Golgi network can be dissected structurally and functionally from the cisternae of the Golgi complex by brefeldin A. *European journal of cell biology, 59*(1), 92-105.

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., et al. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics, 23*(21), 2947-2948.

Laskowski, R. A., MacArthur, M. W., Moss, D. S., & Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *Journal of applied crystallography, 26*(2), 283-291.

Le Borgne, R., Schmidt, A., Mauxion, F., Griffiths, G., & Hoflack, B. (1993). Binding of AP-1 Golgi adaptors to membranes requires phosphorylated cytoplasmic domains of the mannose 6-phosphate/insulin-like growth factor II receptor. *Journal of Biological Chemistry, 268*(30), 22552-22556.

Lee, M. N., Ye, C., Villani, A. C., Raj, T., Li, W., Eisenhaure, T. M., et al. (2014). Common genetic variants modulate pathogen-sensing responses in human dendritic cells. *Science, 343*(6175), 1246980.

Letunic, I., Doerks, T., & Bork, P. (2012). SMART 7: recent updates to the protein domain annotation resource. *Nucleic Acids Res, 40*(Database issue), D302-305.

Lever, A. M. (2009). HIV: the virus. *Medicine, 37*(7), 313-316.

Lewis, M. J., Sweet, D. J., & Pelham, H. R. (1990). The ERD2 gene determines the specificity of the luminal ER protein retention system. *Cell, 61*(7), 1359-1363.

Lewis, R. S. (2001). Calcium signaling mechanisms in T lymphocytes. *Annual review of immunology, 19*(1), 497-521.

Li, S., Wang, L., Berman, M., Kong, Y.-Y., & Dorf, M. E. (2011). Mapping a dynamic innate immunity protein interaction network regulating type I interferon production. *Immunity, 35*(3), 426-440.

Lillig, C. H., & Holmgren, A. (2007). Thioredoxin and related molecules–from biology to health and disease. *Antioxidants & redox signaling, 9*(1), 25-47.

Lincoln, D. T., Ali, E. E., Tonissen, K. F., & Clarke, F. M. (2003). The thioredoxin-thioredoxin reductase system: over-expression in human cancer. *Anticancer research, 23*(3B), 2425-2433.

Linse, S., Helmersson, A., & Forsen, S. (1991). Calcium binding to calmodulin and its globular domains. *Journal of Biological Chemistry, 266*(13), 8050-8054.

Lippincott-Schwartz, J. (1993). Bidirectional membrane traffic between the endoplasmic reticulum and Golgi apparatus. *Trends in cell biology, 3*(3), 81-88.

Lu, L. L., Suscovich, T. J., Fortune, S. M., & Alter, G. (2018). Beyond binding: antibody effector functions in infectious diseases. *Nature Reviews Immunology, 18*(1), 46.

Lue, H., Kleemann, R., Calandra, T., Roger, T., & Bernhagen, J. (2002). Macrophage migration inhibitory factor (MIF): mechanisms of action and role in disease. *Microbes and infection, 4*(4), 449-460.

Lundström, J., & Holmgren, A. (1990). Protein disulfide-isomerase is a substrate for thioredoxin reductase and has thioredoxin-like activity. *Journal of Biological Chemistry, 265*(16), 9114-9120.

Luzio, J., Brake, B., Banting, G., Howell, K., Braghetta, P., & Stanley, K. (1990). Identification, sequencing and expression of an integral membrane protein of the trans-Golgi network (TGN38). *Biochemical Journal, 270*(1), 97-102.

Luzio, J. P., & Banting, G. (1993). Eukaryotic membrane traffic: retrieval and retention mechanisms to achieve organelle residence. *Trends in biochemical sciences, 18*(10), 395-398.

Malik, U., Javed, A., Ali, A., & Asghar, K. (2017). Structural and functional annotation of human FAM26F: A multifaceted protein having a critical role in the immune system. *Gene, 597*, 66-75.

Malik, U., & Javed, A. J. I. r. o. i. (2016). FAM26F: an enigmatic protein having a complex role in the immune system. 1-11.

Malkus, P., Graham, L. A., Stevens, T. H., & Schekman, R. (2004). Role of Vma21p in assembly and transport of the yeast vacuolar ATPase. *Molecular biology of the cell, 15*(11), 5075-5091.

Manh, T. P. V., Alexandre, Y., Baranek, T., Crozat, K., & Dalod, M. (2013). Plasmacytoid, conventional, and monocyte-derived dendritic cells undergo a profound and convergent genetic reprogramming during their maturation. *European journal of immunology, 43*(7), 1706-1715.

Marchler-Bauer, A., Lu, S., Anderson, J. B., Chitsaz, F., Derbyshire, M. K., DeWeese-Scott, C., et al. (2011). CDD: a Conserved Domain Database for the functional annotation of proteins. *Nucleic Acids Res, 39*(Database issue), D225-229.

Martinelli, E., Cicala, C., Van Ryk, D., Goode, D. J., Macleod, K., Arthos, J., et al. (2007). HIV-1 gp120 inhibits TLR9-mediated activation and IFN-α secretion in plasmacytoid dendritic cells. *Proceedings of the National Academy of Sciences, 104*(9), 3396-3401.

Matsumiya, M., Harris, S. A., Satti, I., Stockdale, L., Tanner, R., O'Shea, M. K., et al. (2014). Inflammatory and myeloid-associated gene expression before and one day after infant vaccination with MVA85A correlates with induction of a T cell response. *BMC Infect Dis, 14*, 314.

Matzinger, P. (2002). An innate sense of danger. *Annals of the New York Academy of Sciences, 961*(1), 341-342.

Mauxion, F., Schmidt, A., Le, R. B., & Hoflack, B. (1995). Chimeric proteins containing the cytoplasmic domains of the mannose 6-phosphate receptors codistribute with the endogenous receptors. *European journal of cell biology, 66*(2), 119-126.

Maverakis, E., Kim, K., Shimoda, M., Gershwin, M. E., Patel, F., Wilken, R., et al. (2015). Glycans in the immune system and The Altered Glycan Theory of Autoimmunity: a critical review. *J Autoimmun, 57*, 1-13.

McCune, J. M. (2001). The dynamics of CD4+ T-cell depletion in HIV disease. *Nature, 410*(6831), 974.

Medzhitov, R. (2007). Recognition of microorganisms and activation of the immune response. *Nature, 449*(7164), 819.

Medzhitov, R., & Janeway Jr, C. A. (1997). Innate immunity: impact on the adaptive immune response. *Current opinion in immunology, 9*(1), 4-9.

Méresse, S., & Hoflack, B. (1993). Phosphorylation of the cation-independent mannose 6-phosphate receptor is closely associated with its exit from the trans-Golgi network. *The Journal of cell biology, 120*(1), 67-75.

Meresse, S., Ludwig, T., Frank, R., & Hoflack, B. (1990). Phosphorylation of the cytoplasmic domain of the bovine cation-independent mannose 6-phosphate receptor. Serines 2421 and 2492 are the targets of a casein kinase II associated to the Golgi-derived HAI adaptor complex. *Journal of Biological Chemistry, 265*(31), 18833-18842.

Mi, H., Muruganujan, A., Casagrande, J. T., & Thomas, P. D. (2013). Large-scale gene function analysis with the PANTHER classification system. *Nat Protoc, 8*(8), 1551-1566.

Moolla, N., Killick, M., Papathanasopoulos, M., & Capovilla, A. (2016). Thioredoxin (Trx1) regulates CD4 membrane domain localization and is required for efficient CD4-dependent HIV-1 entry. *Biochimica et Biophysica Acta (BBA)-General Subjects, 1860*(9), 1854-1863.

Moore, R. B., Mankad, M. V., Shriver, S. K., Mankad, V. N., & Plishker, G. A. (1991). Reconstitution of Ca (2+)-dependent K+ transport in erythrocyte membrane vesicles requires a cytoplasmic protein. *Journal of Biological Chemistry, 266*(28), 18964-18968.

Mosca, E., Alfieri, R., Merelli, I., Viti, F., Calabria, A., & Milanesi, L. (2010). A multilevel data integration resource for breast cancer study. *BMC Syst Biol, 4*, 76.

Mosser, D. M., & Edwards, J. P. (2008). Exploring the full spectrum of macrophage activation. *Nat Rev Immunol, 8*(12), 958-969.

Müller-Trutwin, M., & Hosmalin, A. (2005). Role for plasmacytoid dendritic cells in anti-HIV innate immunity. *Immunology and cell biology, 83*(5), 578-585.

Munro, S. (1998). Localization of proteins to the Golgi apparatus. *Trends in cell biology, 8*(1), 11-15.

Murray, D., Doran, P., MacMathuna, P., & Moss, A. C. (2007). In silico gene expression analysis–an overview. *Molecular cancer, 6*(1), 50.

Nakamura, H., De Rosa, S., Roederer, M., Anderson, M. T., Dubs, J. G., Yodoi, J., et al. (1996). Elevation of plasma thioredoxin levels in HIV-infected individuals. *International immunology, 8*(4), 603-611.

Nakamura, H., De Rosa, S. C., Yodoi, J., Holmgren, A., Ghezzi, P., Herzenberg, L. A., et al. (2001). Chronic elevation of plasma thioredoxin: inhibition of chemotaxis and curtailment of life expectancy in AIDS. *Proceedings of the National Academy of Sciences, 98*(5), 2688-2693.

Nakamura, H., Masutani, H., & Yodoi, J. (2002). Redox imbalance and its control in HIV infection. *Antioxidants and Redox Signaling, 4*(3), 455-464.

Neves, P. C., Rudersdorf, R. A., Galler, R., Bonaldo, M. C., de Santana, M. G., Mudd, P. A., et al. (2010). CD8+ gamma-delta TCR+ and CD4+ T cells produce IFN-gamma at 5-7 days after yellow fever vaccination in Indian rhesus macaques, before the induction of classical antigen-specific T cell responses. *Vaccine, 28*(51), 8183-8188.

Newman, G. W., Balcewicz-Sablinska, M. K., Guarnaccia, J. R., Remold, H. G., & Silberstein, D. S. (1994). Opposing regulatory effects of thioredoxin and eosinophil cytotoxicity-enhancing factor on the development of human immunodeficiency virus 1. *Journal of Experimental Medicine, 180*(1), 359-363.

Nunes, P., Demaurex, N., & Dinauer, M. C. (2013). Regulation of the NADPH oxidase and associated ion fluxes during phagocytosis. *Traffic, 14*(11), 1118-1131.

Okamoto, T., Ogiwara, H., Hayashi, T., Mitsui, A., Kawabe, T., & Yodoi, J. (1992). Human thioredoxin/adult T cell leukemia-derived factor activates the enhancer binding protein of human immunodeficiency virus type 1 by thiol redox control mechanism. *International immunology, 4*(7), 811-819.

Omasits, U., Ahrens, C. H., Müller, S., & Wollscheid, B. (2013). Protter: interactive protein feature visualization and integration with experimental proteomic data. *Bioinformatics*, btt607.

Orengo, C. A., Michie, A., Jones, S., Jones, D. T., Swindells, M., & Thornton, J. M. (1997). CATH–a hierarchic classification of protein domain structures. *Structure, 5*(8), 1093-1109.

Pankla, R., Buddhisa, S., Berry, M., Blankenship, D. M., Bancroft, G. J., Banchereau, J., et al. (2009). Genomic transcriptional profiling identifies a candidate blood biomarker signature for the diagnosis of septicemic melioidosis. *Genome Biol, 10*(11), R127.

Pellegrini, P., Berghella, A. M., Contasta, I., & Adorno, D. (2003). CD30 antigen: not a physiological marker for TH2 cells but an important costimulator molecule in the regulation of the balance between TH1/TH2 response. *Transplant immunology, 12*(1), 49-61.

Petersen, T. N., Brunak, S., von Heijne, G., & Nielsen, H. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods, 8*(10), 785-786.

Pinton, P., Pozzan, T., & Rizzuto, R. (1998). The Golgi apparatus is an inositol 1, 4, 5-trisphosphate-sensitive Ca2+ store, with functional properties distinct from those of the endoplasmic reticulum. *The EMBO journal, 17*(18), 5298-5308.

Pizzolo, G., Vinante, F., Morosato, L., Nadali, G., Chilosi, M., Gandini, G., et al. (1994). High serum level of the soluble form of CD30 molecule in the early phase of HIV-1 infection as an independent predictor of progression to AIDS. *AIDS (London, England), 8*(6), 741-745.

Pompa, A., De Marchis, F., Pallotta, M. T., Benitez-Alfonso, Y., Jones, A., Schipper, K., et al. (2017). Unconventional transport routes of soluble and membrane proteins and their role in developmental biology. *International journal of molecular sciences, 18*(4), 703.

Potapov, V., Sobolev, V., Edelman, M., Kister, A., & Gelfand, I. (2004). Protein--protein recognition: juxtaposition of domain and interface cores in immunoglobulins and other sandwich-like proteins. *J Mol Biol, 342*(2), 665-679.

Powis, G., Mustacich, D., & Coon, A. (2000). The role of the redox protein thioredoxin in cell growth and cancer. *Free Radical Biology and Medicine, 29*(3-4), 312-322.

Punta, M., Coggill, P. C., Eberhardt, R. Y., Mistry, J., Tate, J., Boursnell, C., et al. (2012). The Pfam protein families database. *Nucleic Acids Res, 40*(Database issue), D290-301.

Quevillon, E., Silventoinen, V., Pillai, S., Harte, N., Mulder, N., Apweiler, R., et al. (2005). InterProScan: protein domains identifier. *Nucleic Acids Res, 33*(Web Server issue), W116-120.

Ratcliffe, L., Lukey, P., MacKenzie, C., & Ress, S. (1994). Reduced NK activity correlates with active disease in HIV patients with multidrug-resistant pulmonary tuberculosis. *Clinical & Experimental Immunology, 97*(3), 373-379.

Raturi, A., & Simmen, T. (2013). Where the endoplasmic reticulum and the mitochondrion tie the knot: the mitochondria-associated membrane (MAM). *Biochimica Et Biophysica Acta (BBA)-Molecular Cell Research, 1833*(1), 213-224.

Rayner, J. C., & Pelham, H. R. (1997). Transmembrane domain-dependent sorting of proteins to the ER and plasma membrane in yeast. *The EMBO Journal, 16*(8), 1832-1841.

Reaves, B., Horn, M., & Banting, G. (1993). TGN38/41 recycles between the cell surface and the TGN: brefeldin A affects its rate of return to the TGN. *Molecular biology of the cell, 4*(1), 93-105.

Reiser, K., François, K. O., Schols, D., Bergman, T., Jörnvall, H., Balzarini, J., et al. (2012). Thioredoxin-1 and protein disulfide isomerase catalyze the reduction of similar disulfides in HIV gp120. *The international journal of biochemistry & cell biology, 44*(3), 556-562.

Reiser, K., Mathys, L., Curbo, S., Pannecouque, C., Noppen, S., Liekens, S., et al. (2016). The Cellular Thioredoxin-1/Thioredoxin Reductase-1 Driven Oxidoreduction Represents a Chemotherapeutic Target for HIV-1 Entry Inhibition. *PloS one, 11*(1), e0147773.

Rhee, S. W., Starr, T., Forsten-Williams, K., & Storrie, B. (2005). The steady-state distribution of glycosyltransferases between the Golgi apparatus and the endoplasmic reticulum is approximately 90: 10. *Traffic, 6*(11), 978-990.

Rincón, M., & Davis, R. J. (2009). Regulation of the immune response by stress-activated protein kinases. *Immunological reviews, 228*(1), 212-224.

Romagnani, S., Annunziato, F., Manetti, R., Almerigogna, F., Biagiotti, R., Giudizi, M. G., et al. (1996). Role for CD30 in HIV expression. *Immunology letters, 51*(1-2), 83-88.

Rubartelli, A., Bajetto, A., Allavena, G., Wollman, E., & Sitia, R. (1992). Secretion of thioredoxin by normal and neoplastic cells through a leaderless secretory pathway. *Journal of Biological Chemistry, 267*(34), 24161-24164.

Ryser, H. J.-P., & Flückiger, R. (2005). Keynote review: progress in targeting HIV-1 entry. *Drug discovery today, 10*(16), 1085-1094.

Sallusto, F., & Lanzavecchia, A. (2009). Heterogeneity of CD4+ memory T cells: functional modules for tailored immunity. *European journal of immunology, 39*(8), 2076-2082.

Saraiva, M., Smith, P., Fallon, P. G., & Alcami, A. (2002). Inhibition of type 1 cytokine–mediated inflammation by a soluble CD30 homologue encoded by ectromelia (mousepox) virus. *Journal of Experimental Medicine, 196*(6), 829-839.

Schwertassek, U., Balmer, Y., Gutscher, M., Weingarten, L., Preuss, M., Engelhard, J., et al. (2007). Selective redox regulation of cytokine receptor signaling by extracellular thioredoxin-1. *The EMBO journal, 26*(13), 3086-3097.

Scott, M. S., Calafell, S. J., Thomas, D. Y., & Hallett, M. T. (2005). Refining protein subcellular localization. *PLoS computational biology, 1*(6), e66.

Seelenmeyer, C., Stegmayer, C., & Nickel, W. (2008). Unconventional secretion of fibroblast growth factor 2 and galectin-1 does not require shedding of plasma membrane-derived vesicles. *FEBS letters, 582*(9), 1362-1368.

Semenza, J. C., Hardwick, K. G., Dean, N., & Pelham, H. R. (1990). ERD2, a yeast gene required for the receptor-mediated retrieval of luminal ER proteins from the secretory pathway. *Cell, 61*(7), 1349-1357.

Shahzad, K., Cadeiras, M., Memon, S., Zeeberg, B., Klingler, T., Sinha, A., et al. (2010). Gene expression signatures of peripheral blood mononuclear cells during the early post-transplant period in patients developing cardiac allograft vasculopathy. *Journal of transplantation, 2010*.

Shapiro, D., & Masci, J. (1996). Treatment of HIV associated psoriatic arthritis with oral gold. *The Journal of rheumatology, 23*(10), 1818-1820.

Sies, H. (2015). Oxidative stress: a concept in redox biology and medicine. *Redox biology, 4*, 180-183.

Song, G., Seo, H. W., Choi, J. W., Rengaraj, D., Kim, T. M., Lee, B. R., et al. (2011). Discovery of candidate genes and pathways regulating oviduct development in chickens. *Biol Reprod, 85*(2), 306-314.

Sorimachi, H., Ohmi, S., Emori, Y., Kawasaki, H., Saido, T., Ohno, S., et al. (1990). A novel member of the calcium-dependent cysteine protease family. *Biological chemistry Hoppe-Seyler, 371*, 171-176.

Stanley, K. K., & Howell, K. E. (1993). TGN38/41: a molecule on the move. *Trends in cell biology, 3*(8), 252-255.

Stefansson, O. A., Jonasson, J. G., Johannsson, O. T., Olafsdottir, K., Steinarsdottir, M., Valgeirsdottir, S., et al. (2009). Genomic profiling of breast tumours in relation to BRCA abnormalities and phenotypes. *Breast Cancer Res, 11*(4), R47.

Steinckwich, N., Schenten, V., Melchior, C., Bréchard, S., & Tschirhart, E. J. (2011). An essential role of STIM1, Orai1, and S100A8–A9 proteins for Ca2+ signaling and FcγR-mediated phagosomal oxidative activity. *The Journal of Immunology*, 1001338.

Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., et al. (2015). STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res, 43*(Database issue), D447-452.

Taganov, K. D., Boldin, M. P., Chang, K.-J., & Baltimore, D. (2006). NF-κB-dependent induction of microRNA miR-146, an inhibitor targeted to signaling proteins of innate immune responses. *Proceedings of the National Academy of Sciences, 103*(33), 12481-12486.

Takahashi, S., Nakagawa, T., Banno, T., Watanabe, T., Murakami, K., & Nakayama, K. (1995). Localization of furin to the trans-Golgi network and recycling from the cell surface involves Ser and Tyr residues within the cytoplasmic domain. *Journal of Biological Chemistry, 270*(47), 28397-28401.

Takeuchi, O., & Akira, S. (2010). Pattern recognition receptors and inflammation. *Cell, 140*(6), 805-820.

Tangye, S. G., & Tarlinton, D. M. (2009). Memory B cells: Effectors of long-lived immune responses. *European journal of immunology, 39*(8), 2065-2075.

Thomas, P., Smart, T. G. J. J. o. p., & methods, t. (2005). HEK293 cell line: a vehicle for the expression of recombinant proteins. *51*(3), 187-200.

Thomas, P. D., Campbell, M. J., Kejariwal, A., Mi, H., Karlak, B., Daverman, R., et al. (2003). PANTHER: a library of protein families and subfamilies indexed by function. *Genome research, 13*(9), 2129-2141.

Todorow, Z., Spang, A., Carmack, E., Yates, J., & Schekman, R. (2000). Active recycling of yeast Golgi mannosyltransferase complexes through the endoplasmic reticulum. *Proceedings of the National Academy of Sciences, 97*(25), 13643-13648.

Towbin, H., Staehelin, T., & Gordon, J. (1979). Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. *Proceedings of the National Academy of Sciences, 76*(9), 4350-4354.

Trono, D., Van Lint, C., Rouzioux, C., Verdin, E., Barré-Sinoussi, F., Chun, T.-W., et al. (2010). HIV persistence and the prospect of long-term drug-free remissions for HIV-infected individuals. *Science, 329*(5988), 174-180.

Tusnady, G. E., & Simon, I. (2001). The HMMTOP transmembrane topology prediction server. *Bioinformatics, 17*(9), 849-850.

Ulloa-Montoya, F., Louahed, J., Dizier, B., Gruselle, O., Spiessens, B., Lehmann, F. F., et al. (2013). Predictive gene signature in MAGE-A3 antigen-specific cancer immunotherapy. *J Clin Oncol, 31*(19), 2388-2395.

Valkova, C., Albrizio, M., Röder, I. V., Schwake, M., Betto, R., Rudolf, R., et al. (2011). Sorting receptor Rer1 controls surface expression of muscle acetylcholine receptors by ER retention of unassembled α-subunits. *Proceedings of the National Academy of Sciences, 108*(2), 621-625.

Vetrivel, U., Subramanian, G., & Dorairaj, S. (2011). A novel in silico approach to identify potential therapeutic targets in human bacterial pathogens. *Hugo J, 5*(1-4), 25-34.

Vidya Vijayan, K., Karthigeyan, K. P., Tripathi, S. P., & Hanna, L. E. (2017). Pathophysiology of CD4+ T-cell depletion in HIV-1 and HIV-2 infections. *Frontiers in immunology, 8*, 580.

Vivier, E., Raulet, D. H., Moretta, A., Caligiuri, M. A., Zitvogel, L., Lanier, L. L., et al. (2011). Innate or adaptive immunity? The example of natural killer cells. *science, 331*(6013), 44-49.

Von Mering, C., Krause, R., Snel, B., Cornell, M., Oliver, S. G., Fields, S., et al. (2002). Comparative assessment of large-scale data sets of protein–protein interactions. *Nature, 417*(6887), 399.

Vostal, J. G., & Shulman, N. R. (1993). Vinculin is a major platelet protein that undergoes Ca2+-dependent tyrosine phosphorylation. *Biochemical Journal, 294*(3), 675-680.

Wang, H., Bloom, O., Zhang, M., Vishnubhakat, J. M., Ombrellino, M., Che, J., et al. (1999). HMG-1 as a late mediator of endotoxin lethality in mice. *Science, 285*(5425), 248-251.

Wardlaw, A. J., Moqbel, R., & Kay, A. B. (1995). Eosinophils: biology and role in disease *Advances in immunology* (Vol. 60, pp. 151-266): Elsevier.

Welsh, S. J., Bellamy, W. T., Briehl, M. M., & Powis, G. (2002). The redox protein thioredoxin-1 (Trx-1) increases hypoxia-inducible factor 1α protein expression: Trx-1 overexpression results in increased vascular endothelial growth factor production and enhanced tumor angiogenesis. *Cancer research, 62*(17), 5089-5095.

Wiederstein, M., & Sippl, M. J. (2007). ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic acids research, 35*(suppl 2), W407-W410.

Wilkins, C., & Gale Jr, M. (2010). Recognition of viruses by cytoplasmic sensors. *Current opinion in immunology, 22*(1), 41-47.

Wollman, E., d'Auriol, L., Rimsky, L., Shaw, A., Jacquot, J., Wingfield, P., et al. (1988). Cloning and expression of a cDNA for human thioredoxin. *Journal of Biological Chemistry, 263*(30), 15506-15512.

Wong, Y. H., Lee, T. Y., Liang, H. K., Huang, C. M., Wang, T. Y., Yang, Y. H., et al. (2007). KinasePhos 2.0: a web server for identifying protein kinase-specific phosphorylation sites based on sequences and coupling patterns. *Nucleic Acids Res, 35*(Web Server issue), W588-594.

Wood, Z. A., Schröder, E., Harris, J. R., & Poole, L. B. (2003). Structure, mechanism and regulation of peroxiredoxins. *Trends in biochemical sciences, 28*(1), 32-40.

Wurm, F. M. J. N. b. (2004). Production of recombinant protein therapeutics in cultivated mammalian cells. *22*(11), 1393-1398.

Yanagida, M. (2002). Functional proteomics; current achievements. *Journal of Chromatography B, 771*(1-2), 89-106.

Yoshimura, A., Naka, T., & Kubo, M. (2007). SOCS proteins, cytokine signalling and immune regulation. *Nature Reviews Immunology, 7*(6), 454.

Young, J., Stauber, T., del Nery, E., Vernos, I., Pepperkok, R., & Nilsson, T. (2005). Regulation of microtubule-dependent recycling at the trans-Golgi network by Rab6A and Rab6A'. *Molecular biology of the cell, 16*(1), 162-177.

Yu, C. S., Chen, Y. C., Lu, C. H., & Hwang, J. K. (2006). Prediction of protein subcellular localization. *Proteins, 64*(3), 643-651.

Zaal, K. J., Smith, C. L., Polishchuk, R. S., Altan, N., Cole, N. B., Ellenberg, J., et al. (1999). Golgi membranes are absorbed into and reemerge from the ER during mitosis. *Cell, 99*(6), 589-601.

Zeiser, R., Nguyen, V. H., Hou, J.-Z., Beilhack, A., Zambricki, E., Buess, M., et al. (2007). Early CD30 signaling is critical for adoptively transferred CD4+ CD25+ regulatory T cells in prevention of acute graft-versus-host disease. *Blood, 109*(5), 2225-2233.

Zeller, T., Wild, P., Szymczak, S., Rotival, M., Schillert, A., Castagne, R., et al. (2010). Genetics and beyond--the transcriptome of human monocytes and disease susceptibility. *PLoS One, 5*(5), e10693.

Zerfaoui, M., Fukuda, M., Langlet, C., Mathieu, S., Suzuki, M., Lombardo, D., et al. (2002). The cytosolic and transmembrane domains of the β1, 6 N-

acetylglucosaminyltransferase (C2GnT) function as a cis to medial/Golgi-targeting determinant. *Glycobiology, 12*(1), 15-24.

Zhang, S., Kim, C. C., Batra, S., McKerrow, J. H., & Loke, P. (2010). Delineation of diverse macrophage activation programs in response to intracellular parasites and cytokines. *PLoS Negl Trop Dis, 4*(3), e648.

Zhang, Y. (2008). I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics, 9*, 40.

APPENDIX

Table S1: The names, UniProt ID and taxonomic classification of 20 different organisms whose protein sequences were used for multiple sequence alignment and phylogenetic analysis

No.	Organism	UniProt ID	Taxonomic Classification
1.	Homo Sapiens (Human)	Q5R3K3	Kingdom: Animalia Phylum: Chordata Class: Mammalia Order: Primates
2.	Mus Musculus (Mouse)	Q8C9E8	Kingdom: Animalia Phylum: Chordata Class: Mammalia Order: Rodentia
3.	Rattus norvegicus (Rat)	Q561R8	Kingdom: Animalia Phylum: Chordata Class: Mammalia Order: Rodentia
4.	Anas platyrhynchos (Mallard)	R0JJ38	Kingdom: Animalia Phylum: Chordata Class: Aves Order: Anseriformes
5.	Zonotrichia albicollis (White-throated sparrow)	D8KW56	Kingdom: Animalia Phylum: Chordata Class: Aves Order: Passeriformes
6.	Ophiophagus hannah (King cobra)	V8NJ84	Kingdom: Animalia Phylum: Chordata Class: Reptilia Order: Squamata
7.	Danio rerio (Zebrafish)	A0PJS6	Kingdom: Animalia Phylum: Chordata Class: Actinopterygii Order: Cypriniformes
8.	Pan troglodytes (Chimpanzee)	H2QTL7	Kingdom: Animalia Phylum: Chordata Class: Mammalia Order: Primates

9.	Gallus gallus (Chicken)	E1C3Z9	Kingdom: Animalia Phylum: Chordata Class: Aves Order: Galliformes
10.	Canis lupus familiaris (Dog)	F6XP55	Kingdom: Animalia Phylum: Chordata Class: Mammalia Order: Carnivora
11.	Bos taurus (Bovine)	E1B9Q9	Kingdom: Animalia Phylum: Chordata Class: Mammalia Order: Cetartiodactyla
12.	Xenopus tropicalis (Western clawed frog)	F7A1D8	Kingdom: Animalia Phylum: Chordata Class: Amphibia Order: Anura
13.	Felis catus (Cat)	M3X031	Kingdom: Animalia Phylum: Chordata Class: Mammalia Order: Carnivora
14.	Macaca mulatta (Rhesus macaque)	F6SXL3	Kingdom: Animalia Phylum: Chordata Class: Mammalia Order: Primates
15.	Anolis carolinensis (Green anole) (American chameleon)	G1KP70	Kingdom: Animalia Phylum: Chordata Class: Reptilia Order: Squamata
16.	Equus caballus (Horse)	F6WLN6	Kingdom: Animalia Phylum: Chordata Class: Mammalia Order: Perissodactyla
17.	Loxodonta africana (African elephant)	G3TIV7	Kingdom: Animalia Phylum: Chordata Class: Mammalia Order: Proboscidea
18.	Pongo abelii (Sumatran orangutan)	H2PK50	Kingdom: Animalia Phylum: Chordata

			Class: Order:	Mammalia Primates
19.	Callithrix jacchus (White-tufted-ear marmoset)	F6YD15	Kingdom: Phylum: Class: Order:	Animalia Chordata Mammalia Primates
20.	Ictidomys tridecemlineatus (Thirteen-lined ground squirrel)	I3M8E7	Kingdom: Phylum: Class: Order:	Animalia Chordata Mammalia Rodentia

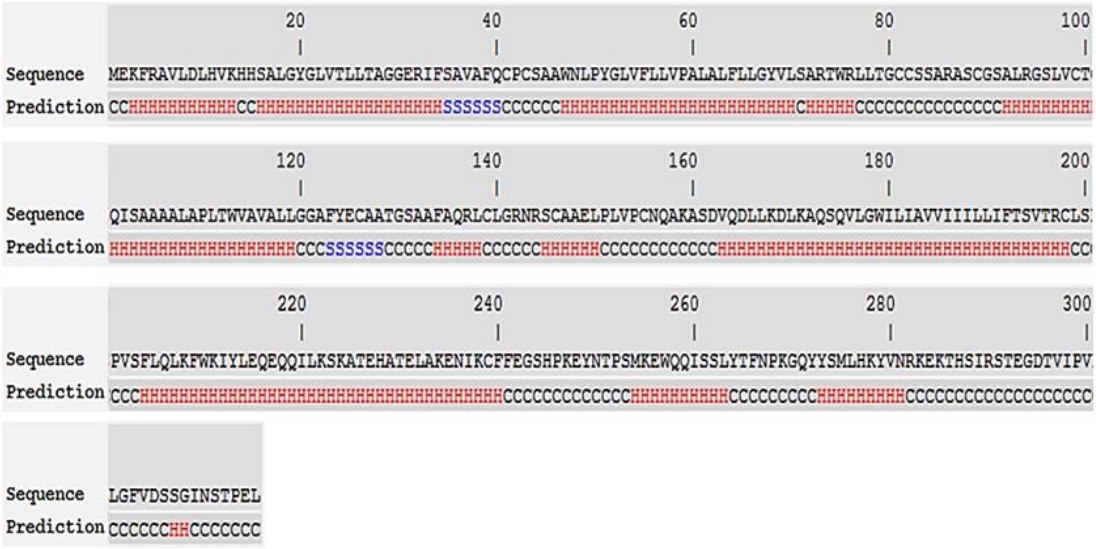


Figure S1: FAM26F secondary structure predicted by I-TASSER. The prediction highlights the helix, strand and coil confirmation corresponding to each residue in the protein sequence.

```
FAM26F_Zebrafish      MDKLKSLFSFAQKQETSLGFGLISIITVGSEHVFSVFAFKCPCN-DWNFVYGNVCLLVPA
FAM26F_Cobra          MEKFRPVLDFIVNHQKVLGYGAVSLLTAGSERIFSVVVFKCPCN-SWNMLYGGMFLLMPA
FAM26F_Chameleon      MEKLRTVLDFCLSHQKALGYGAVSLLTLGSERIFSVVVFKCPCN-SWNMLYGTVFLLAPA
FAM26F_Sparrow        MDTLQKAVDFCIRHQTTLSFSIVSLLTAASERVFSYVVFKCPCN-SENLVYGYSFLLAPA
FAM26F_Mallard        MEKLRTAMDFCIRHQKFLGYSIVSLLTAASEYIFSSVVFKCPCN-SWNTLYGCSFIIAPA
FAM26F_Chicken        MEKLHTVVNFCIRYQKTLGYSIVSLLTIASEQIFSSVAFRCPCN-SWNTLYGCVFLLVPA
FAM26F_Frog           MEYFKTVLNLGVKHQAVLGYGALSLLAAVGEKVFSTVLFQCPCN-GWNHVYGMVFLLVPA
FAM26F_Horse          IARFGTVLATYLKHHSKLGCGLASLLTASGELIFSTVVFQCPCSAAWNLPYSLVFLLLPA
FAM26F_MOUSE          MEKFKAVLDLQRKHRNALGYSLVTLLTAGGEKIFSSVVFQCPCTATWNLPYGLVFLLVPA
FAM26F_RAT            MEKFKAVLDLQIKHRSALGYGLVTLLTAGGEKIFSTVVFQCPCTATLNLTYGLVFLLVPA
FAM26F_Cat            MDKFQAVLNLHLQHPHALGYGLVTLLTAGGERLFSVAVFQCPCSAAWNLPYGLVFLLVPA
FAM26F_Bovine         MEKFREVLNLHLKHRNALGYCLVSLLTAGGERIFSTTVFQCPCSVAWNLPYGLVFLLVPA
FAM26F_Dog            MEKLQTVLDLQLKQHRALGYGLVSLLTAGGERLFSAAVFQCPCSAAWNLPYGLVFLLVPA
FAM26F_Squirrel       MEKFKTLLDLHLRHSSALGYGLVTLLTASGERIFSTVVFQCPCSATWNLPYGLVFLLVPA
FAM26F_Elephant       MEKFQKVLDLYQKHHSALGYGLVTLLTAGGERLFSTVVFSCPCSATWNLPYGLVFLLVPA
FAM26F_Marmoset       MEKFRAVLDLHLKHHSALGYGLVTLLTAGGERIFSTVVFQCPCSATWNLPYGLVFLLVPA
FAM26F_Rhesus         MEKFRAVLDLHLKHHSALGYGLVTLLTAGGERIFSTVVFQCPCSAAWNLPYGLVFLLVPA
FAM26F_Orangutan      MEKFRAVLDLHLKHHSALGYGLVTLLTAGGERIFSAVVFQCPCSAAWNLPYGLVFLLVPA
FAM26F_HUMAN          MEKFRAVLDLHVKHHSALGYGLVTLLTAGGERIFSAVAFQCPCSAAWNLPYGLVFLLVPA
FAM26F_Chimpanzee     MEKFRAVLDLHVKHHSALGYGLVTLLTAGGERIFSAVAFQCPCSAAWNLPYGLVFLLVPA
                      :  :   .         *.    ::::  .* :**   * ***.   *  *.   :: **

FAM26F_Zebrafish      AALLILSYMLSNKTWKLFTGLCYRRSR----LCRF-NYTFGFLCVFLQITVTAMVAPLSW
FAM26F_Cobra          LILLLLGSLLSVRSWKVLTGCCSKGRP---CRCPRGNRLQRHLQVMGLAILSAAVAPLTW
FAM26F_Chameleon      LILFLLGLLLNTRSWKVLTGCCAPGKL---CLCPHANRFWRYVQVLWLVMVTAAVAPITW
FAM26F_Sparrow        FVLLLLGYMMNARTWRLFTGMCSPEKH---PQYCSWRTWAHFCQLFVPMTAKASVAPLTW
FAM26F_Mallard        FVLFLLGYMTNARVWLLVTGRCSPKNQ---CSCDS---CGHFHKVLVPVTASALVAPFTW
FAM26F_Chicken        LLLFLLGYMVNARTWLLLTGSCPQEKK---HCCGFGEKGCFYLKVLASVTASTLVAPLTW
FAM26F_Frog           VILFLLGYMLNFPLWKHMTGCCNYEKQGH-RVCRR---GLRCLQVFWQMTFVSALAPLIW
FAM26F_Horse          LVFYLLGCVLRTRSLCLLCSCCRTRNAGI-NRC-------DLCDECCRVSGPAVVSSVTW
FAM26F_MOUSE          LALFLLGYALSARTWRLLTGCCSRSA-----RFSS---GLRSAFVCAQLSMTAAFAPLTW
FAM26F_RAT            LALFLLGYALSARTWRLLTGCCSRSAS---TRSSS---GLRSTLVCAQVSAVAALAPLTW
FAM26F_Cat            LALFLLGYVLSARTWRLLTGCCARGAR--PGGCGP---RLRAALVCVQLSATAAVAPLTW
FAM26F_Bovine         LALFLLGYVLSARTWRLLTGCCA-------RSCGS---GLRGALVCAQISATAAVAPLTW
FAM26F_Dog            LALFLLGYVLNARTWRLLTGCCARSSRSRARGCGA---GLRGAVVCAQLGASAAVAPLTW
FAM26F_Squirrel       LVLFLLGYMLSARTWRLLTGCCAPGAR---VRCSS---GLRGALVFTQIAVAAALAPLTW
FAM26F_Elephant       LALFLLGYVLSARTWRLLTGCCAPGAR---RGCGA---GLRCTLVCTQLSAAAALAPLTW
FAM26F_Marmoset       LALFLLGYMLSARTWRLLTGCCAPSAR---GCCRS---GLRGYLVCAQLSAAAALAPLTW
FAM26F_Rhesus         LALFLLGYVLSARTWRLITGCGC-RAR---ASCGS---GLRGSLVCAQLSAAAALAPLTW
FAM26F_Orangutan      LALFLLCYVLSARTWRLLTGCCS-SAR---ASCGS---ALRASCVCMQLSAAATLAPLTW
FAM26F_HUMAN          LALFLLGYVLSARTWRLLTGCCS-SAR---ASCGS---ALRGSLVCTQISAAAALAPLTW
FAM26F_Chimpanzee     LALFLLGYVLSARTWRLLTGCCS-SAR---ASCGS---ALRGSLVCAQISAAAALAPLTW
                        : :*          . .                                 : .: . *

FAM26F_Zebrafish      IAVALLKGEFFECSMTGANFTLFRRHICSDKYP--HCRTDLEKFPCASTATGIPQSE---
FAM26F_Cobra          ISVALLGGSFYECTATGTPI--LQKYVCKGEGEEEECLKTLVKVPCLSPTSPSSE-----
FAM26F_Chameleon      IAVALLGGSFYECAATGSAI--LQNYMCQDKG--EECFKKVLQVPCQSSLSSQEM-----
FAM26F_Sparrow        IAVALLGANFYECAASGSNM--TAQLFCKNKGN--YSQEQLYKMPCDEELAAAMS-----
FAM26F_Mallard        IAVALLSASFYACAASGNSF--IRKLVCKDIKE--YCNASLEKIPCDEELSKKI------
FAM26F_Chicken        IAVALLSASFYECAASGSSL--IRHRVCRDIDKLTACRELLEKIPCDEKVVGQLLNENSA
FAM26F_Frog           IALALLNGTFYVCIVSGLPW--KHA---CDYKE--TCIKELPHIPCPGSYTSDLSKE--D
FAM26F_Horse          VAVALLGGAVYECCASGSTF--KADRLCVGRNS--SCAAQLPLVPCRQAQDPLVQ-D--L
FAM26F_MOUSE          VAVALLEGSFYQCAVSGSAR--LAPYLCKGRDP--NCNATLPQAPCNKQKV-EMQ-E---
FAM26F_RAT            VAVALLGGSFYQCAVSGSTR--LASYLCKDRNH--SCIAKLPQVPCNKQEA-EMQ-E---
FAM26F_Cat            VAVALLGGAFYECAASGSAT--ATISVKRTKRR--EGTCSVPDT--AP---GRSA-P--V
FAM26F_Bovine         VAVALLGGAFYECAASGSEV--LARYLCVGRDP--RCAAQLPLVPCQQAQAPDVK-Q--L
FAM26F_Dog            VAVALLGGAFYECAAAGSAP--LARRLCRGRAP--ACEAQLPLAPCLPAQGPDAQ-G--L
FAM26F_Squirrel       VAVALLGGSFYECAASGSKV--LGPYLCRNR-T--GCTDQLPLVPCSKDQKSDLQ-D--L
FAM26F_Elephant       VAVALLGGAFYECAVSGSAP--FARFLCQGRDP--SCVAQLPLVPCNQAKESEVQ-N--L
FAM26F_Marmoset       VAVALLGGAFYECAASGSAA--FAQRLCSGRHS--SCADELPLVPCHQAKASDVQ-D--L
FAM26F_Rhesus         VAVALLGGAFYECAASGSAV--LAQRLCLDRDH--NCAAELPLVPCHEAKASDVQ-D--L
FAM26F_Orangutan      VAVALLGGAFYECAASGSAA--FAQRLCLGRDR--NCAAELPLVPCHEAKASDVQ-D--L
FAM26F_HUMAN          VAVALLGGAFYECAATGSAA--FAQRLCLGRNR--SCAAELPLVPCNQAKASDVQ-D--L
FAM26F_Chimpanzee     VAVALLGGAFYECAATGSAA--FAQRLCLGRDR--SCAAELPLVPCNQAKASDVQ-D--L
                      :::*** . .: *  :*                      :

FAM26F_Zebrafish      REAVLSLIRAESQVLGWTLIASVMLFTFLLTCMARCYSPISYMQLKFWKMYTQKESDFLD
FAM26F_Cobra          -LEILATLRAQSQVFGWILIASIFTVALLTTCIAHCRSPVSVLQLAFWKVYLQKEQQLFE
FAM26F_Chameleon      -QDMLTNLRAQSQVMGWVLIASIFTLALAATCISRCRSPVSILQLTFWKMYLEKEQQLFE
FAM26F_Sparrow        --SVCLSFHAQSQLIGWFLIVTIMALALISTCVTHCFSPVSYLQFKFWKIYSRKEHKLFE
FAM26F_Mallard        --GAFHSLQAQSQMVGWLLIAIIMTAALISTCFSYCCSPVSHFQLKFWKIYLKKEQEVFE
FAM26F_Chicken        SKVGLVSFRAQSQILGWLLIVTIIIVALISTCISRCFSPVSYLQLKFWKIYLEKEREHFE
FAM26F_Frog           INEVQRFMRAESQVLGWTVMCCVLVIIVLSTCISRCMSPVSFLQLKFWKMYIEREQELFD
```

```
FAM26F_Horse          Q----KVLRAESQVVGWSLIAAVILLFLIFTCVSHHRAGGSSLQLKFWKIYSEEEHQCLR
FAM26F_MOUSE          ---ILSQLKAQSQVFGWILIAAVIILLLLVKSVTRCFSPVSYLQLKFWEIYWEKEKQILQ
FAM26F_RAT            ---ILSQLKAQSQVLGWVLIAAVIFLLLVFKCVSRCFSPVSYLQLKFWEIYLEKEKQILQ
FAM26F_Cat            LPYLLSFLLCFPQVLGWILIAVVIIFLLIFTSIIRCLSPVSFLQLKFWKIYLEQEQQILK
FAM26F_Bovine         QKELQKELTAHSQVLGWVLIAVVIILLLICTSITRCLSPVSYLQLKFWKIYLQQEQQILK
FAM26F_Dog            L----RELKAHSQVLGWILIALVIIFLLIFTSVSRCLSPVSFLQLKFWKIYLEQEQQILK
FAM26F_Squirrel       L----KELKAQSQVLGWIIIAVVIIAVVIFTSVCRCLSPVSFLQLKFWKIYLENEQEIFK
FAM26F_Elephant       L----KELRAQSQVLGWILIAIVIIVLLIFTSLSRCLSPVSFLQLKFWKIYLEQEQKILK
FAM26F_Marmoset       L----KDLKAQSQVLGWILIAVVIIVLLIFTSVSRCLSPVSFLQLKFWKIYLEQEQHILK
FAM26F_Rhesus         L----KDLKAQSQVLGWILIAAVIIVLLIFTCVSRCLSPVSFLQLKFWKIYLEQEQQILK
FAM26F_Orangutan      L----KDLKAQSQVLGWILIAVVIIILLIFTSVIRCLSPVSFLQLKFWKIYLEQEQQILK
FAM26F_HUMAN          L----KDLKAQSQVLGWILIAVVIIILLIFTSVTRCLSPVSFLQLKFWKIYLEQEQQILK
FAM26F_Chimpanzee     L----KDLKAQSQVLGWILIAVVIIILLIFTSVTRCLSPVSFLQLKFWKIYLEQEQQILK
                             : .  *:.** ::  ::   .  ...    :  * :*: **::* ..* . :

FAM26F_Zebrafish      SYTSQHAENLAKRNITSFFELTKPIPIKSPPRQAWEKVSRFYKYQNMNQYYSILHKYVCT
FAM26F_Cobra          TMAKEHASKLAERNLKSFFDSTELEPFQTPSAKAWDNISSLFAFNPKDNCYSMIHKYVSK
FAM26F_Chameleon      VKAKEHAAQLAERNLGCFFDSTHLEPIHTPSAKAWRGISSMFAFNPEEHYYSMIHKYVSS
FAM26F_Sparrow        TKAKEHANKLAERNTNCFFEATDPAPFPTPSNEDWQKVSVSYTFNSQSQYYSVLHKYVNT
FAM26F_Mallard        IKAKDHAAKLAERNVHLFFEPADPAPFCTPSNEDWQKISFPYGFSTKEQHYSMIHKYVNT
FAM26F_Chicken        TKAKEHAAQLAERNINCFFEATNPPPFQTPSNDNWQKISFPYAFSKNNQYYSMIHKYANA
FAM26F_Frog           IRCKEHATKLAERNIKIFFDHTKVEPFITPNNKEWNQISSTYTFNKKKQYYSMLHRFVEL
FAM26F_Horse          TQVTERATKLADENVRCFFEGSRPTGCNIPSMERWQEISSPYPFNPEDKYLSALHRSVNE
FAM26F_MOUSE          NQAAENATQLAEENVRCFFECSKPKECNTPSSKDWQEISALYTFNPKNQFYSMLHKYVSR
FAM26F_RAT            SQAAEHATQLARENIRSFFECSKPKECNTPSRKDWQQISALYTFNSKNQFYSMLHKYVSR
FAM26F_Cat            SQATEHATELAKENVKCFFECSHPKEYNTPSIRDWQQISSLYTFNPKEQYYSMLHKYVHR
FAM26F_Bovine         SQATEHAMQLAKENIKCFFECSHPQEYNTPSIKDWQQISSLYTFNPKEQYYSTLHKYVNK
FAM26F_Dog            AQATEHAMELAKENVKCFFECLHPEECNTPSMKDWQQISSLYTFNPKEQYYSMLHKYVNR
FAM26F_Squirrel       SQAKEHATDLAKENVKCFFEHSHPKEYNTPSIKDWQQISTLYTFNPKDQYYSMLHKYVNR
FAM26F_Elephant       TQATEHATILAEENVKSFFKGSCPKEHYTPGIKDWQQISSLYTFNPKEQYYSILHKYVNK
FAM26F_Marmoset       SEATEHATELAKENVKCFFEGSRPKECNTPSMKEWQQISSLYTFNPKDQYYSLLHKYVNK
FAM26F_Rhesus         SKATEHATELAKENVKCFFEGSHPKECNTPSVKEWQQISSLYTFNQKDPYYSMLHKYVNR
FAM26F_Orangutan      SKATEHATELAKENIKCFFEGSHPKEYNTPSIEEWQQISSLYTFNPKGQYYSMLHKYVNR
FAM26F_HUMAN          SKATEHATELAKENIKCFFEGSHPKEYNTPSMKEWQQISSLYTFNPKGQYYSMLHKYVNR
FAM26F_Chimpanzee     GKATEHATELAKENIKCFFEGSHPKEYNTPSVKEWQQISSLYTFNPKGQYYSMLHKYVNR
                          :.*  ** .*   **.         *    *  :*  : :.      * :*: .

FAM26F_Zebrafish      CEDLENPASRGSVRSENDFSNPAALAFVDESKMVL------
FAM26F_Cobra          KSQ------SGSIKSAEGDIYPSCLQFVDGANVEVQIL---
FAM26F_Chameleon      KTS------SGSIRSAEGDTFPTCLGFVDGVGVADSQVL--
FAM26F_Sparrow        NRG------NDAEF-QEEGQDLNVIEFVDEAQPSVSGL---
FAM26F_Mallard        NRG------KTS---SGTDQIHNVLAFVDEAHGIES-----
FAM26F_Chicken        SRG------NISTV-SEGDQICSALEFVDEAHTSEQAF---
FAM26F_Frog           SDR------NQSFISLEGDMVPPALFFSIISGLLSSPNL--
FAM26F_Horse          IQN------RHTMKSPSGD----------------------
FAM26F_MOUSE          EEM------SGSVRSVEGDAVIPALGFVDDMSMTNTHEL--
FAM26F_RAT            KEV------SSSLHSVEGDVVVPVLGFVDDAAMANTHGV--
FAM26F_Cat            TEK------SESIKSKEGDTVIPILGFVDTPGMHTTADL--
FAM26F_Bovine         KQK------NQSSTSSEEDAMVPVLGFVDSSDMNSTTDL--
FAM26F_Dog            KEK------THSIISKEGDAVIPVLGFVDTPGININTAAEL
FAM26F_Squirrel       KEE------SDSIKSSEGDVMLPVLGFVDSPGINSSTGL--
FAM26F_Elephant       SEK------SQSIRSAEGDALFPVLGFVDSSGINITAEI--
FAM26F_Marmoset       KEK------THSIRSTEGDTVIPILGFVDSTGINSTPGL--
FAM26F_Rhesus         KEK------THSIRSTEGDTVIPVLGFVDSSSINSTPGL--
FAM26F_Orangutan      KEK------THSIRSTEGDTVIPVLGFVDSSGINSTPGL--
FAM26F_HUMAN          KEK------THSIRSTEGDTVIPVLGFVDSSGINSTPEL--
FAM26F_Chimpanzee     KEK------THSIRSTEGDTVIPVLGFVDSSGINSTPGL--
```

Figure S2: Multiple sequence alignment of FAM26F protein sequence from 20 different organisms using Clustal X.

Asterisks below the alignment indicate perfectly conserved residue, the colons indicate the residues variation found within strongly conserved groups and dots indicate the residues variation present within weaker conserved residue groups. The arrow indicates a single conserved Tyrosine residue which has also been predicted to be phosphorylated during post translational modifications.

Table S2: Differentially expressed proteins between the control and transfected eluates, with the MS peptide count being higher for transfected eluate as compared to the control eluate.

No.	Protein Name	Accession Number	Molecular weight	Function (Uniprot)
1	Protein S100-A9	P06702	13 kDa	A calcium- and zinc-binding protein which plays a prominent role in the regulation of inflammatory processes and immune response
2	Caspase-14	P31944	28 kDa	Non-apoptotic caspase involved in epidermal differentiation
3	Cathepsin D	P07339	45 kDa	An acid protease active in intracellular protein breakdown.
4	Annexin A2	P07355	39 kDa	Calcium-regulated membrane-binding protein might be involved in heat-stress response. Inhibits PCSK9-enhanced LDLR degradation, probably reduces PCSK9 protein levels but also competes with LDLR for binding with PCSK9
5	Bleomycin hydrolase	Q13867	53 kDa	Normal physiological role of BLM hydrolase is unknown, but it catalyzes the inactivation of the antitumor drug BLM thus protecting normal and malignant cells from BLM toxicity
6	Histidine ammonia-lyase	P42357	73 kDa	A cytosolic enzyme catalyzing the first reaction in histidine catabolism, the nonoxidative deamination of L-histidine to *trans*-urocanic acid
7	Cluster of Serpin B3	P29508	45 kDa	May act as a papain-like cysteine protease inhibitor to modulate the host

				immune response against tumor cells. Also functions as an inhibitor of UV-induced apoptosis via suppression of the activity of c-Jun NH_2-terminal kinase (JNK1).
8	Protein S100-A7	P31151	11 kDa	Calcium-binding protein containing the EF hand motif that displays antimicrobial activities against bacteria and triggers immunomodulatory activities
9	Protein S100-A8	P05109	11 kDa	A calcium- and zinc-binding protein which plays a prominent role in the regulation of inflammatory processes and immune response.
10	Glyceraldehyde-3-phosphate dehydrogenase	P04406	36 kDa	Plays a role in glycolysis and nuclear functions. Component of the GAIT (gamma interferon-activated inhibitor of translation) complex which mediates interferon-gamma-induced transcript-selective translation inhibition in inflammation processes
11	Protein-glutamine gamma-glutamyltransferase K	P22735	90 kDa	Catalyzes the cross-linking of proteins and the conjugation of polyamines to proteins
12	Alpha-2-macroglobulin-like protein 1	A8K2U0	161 kDa	Is able to inhibit all four classes of proteinases by a unique 'trapping' mechanism
13	Protein-glutamine	Q08188	77 kDa	Catalyzes the calcium-dependent formation of isopeptide cross-links

	gamma-glutamyltransferase E			between glutamine and lysine residues in various proteins, as well as the conjugation of polyamines to proteins
14	Arginase-1	P05089	35 kDa	Involved in an antimicrobial effector pathway in polymorphonuclear granulocytes (PMN). Upon PMN cell death is liberated from the phagolysosome and depletes arginine in the microenvironment leading to suppressed T cell and natural killer (NK) cell proliferation and cytokine secretion
15	Gamma-glutamylcyclotransferase	O75223	21 kDa	Induces release of cytochrome c from mitochondria with resultant induction of apoptosis. May play a significant role in glutathione homeostasis
16	Polyubiquitin-B	P0CG47	26 kDa	Involved in: DNA repair; in ERAD (endoplasmic reticulum-associated degradation) and in cell-cycle regulation; in lysosomal degradation; in kinase modification; in protein degradation via the proteasome; in endocytosis, in DNA-damage responses as well as in signaling processes leading to activation of the transcription factor NF-kappa-
17	Peroxiredoxin-2	P32119	22 kDa	Plays a role in cell protection against oxidative stress by detoxifying peroxides and as sensor of hydrogen peroxide-mediated signaling events. Might participate in the signaling cascades of growth factors and tumor necrosis factor-alpha by regulating the intracellular concentrations of H_2O_2

18	Filaggrin	P20930	435 kDa	Essential for the regulation of epidermal homeostasis and is responsible for the skin barrier function
19	Zinc-alpha-2-glycoprotein	P25311	34 kDa	Stimulates lipid degradation in adipocytes and causes the extensive fat losses associated with some advanced cancers. Also involved in fertilization and immuno-regulation.
20	Cystatin-A	P01040	11 kDa	An intracellular thiol proteinase inhibitor having an important role in desmosome-mediated cell-cell adhesion in the lower levels of the epidermis.
21	Gasdermin-A	Q96QA5	49 kDa	May promote pyroptosis. Also binds to bacterial and mitochondrial lipids, including cardiolipin, and exhibits bactericidal activity
22	Galectin-7	P47929	15 kDa	Could be involved in cell-cell and/or cell-matrix interactions necessary for normal growth control. Pro-apoptotic protein that functions intracellularly upstream of JNK activation and cytochrome c release.
23	Calpain-1 catalytic subunit	P07384	82 kDa	A non-lysosomal thiol-protease regulated by Calcium that catalyzes partial proteolysis of substrates required for signal transduction and cytoskeletal remodeling
24	Fructose-bisphosphate aldolase A	P04075	39 kDa	Plays a key role in glycolysis and gluconeogenesis. May also function as scaffolding protein. Proposed to participate in the control of host redox homeostasis and the inflammatory

				immune response.
25	Catalase	P04040	60 kDa	Serves to protect cells from the toxic effects of hydrogen peroxide. Promotes growth of cells including T-cells, B-cells, myeloid leukemia cells, melanoma cells, mastocytoma cells and normal and transformed fibroblast cells.
26	Thioredoxin	P10599	12 kDa	Participates in various redox reactions through the reversible oxidation of its active center dithiol to a disulfide (calcium dependent) and catalyzes dithiol-disulfide exchange reactions.
27	F-box only protein 50	Q6ZVX7	31 kDa	Promotes cell proliferation.
28	78 kDa glucose-regulated protein	P11021	72 kDa	Endoplasmic reticulum chaperone that plays a key role in protein folding and quality control in the endoplasmic reticulum lumen
29	Calmodulin-like protein 5	Q9NZT1	16 kDa	A calcium-binding regulatory protein transmitting a momentary increase in the concentration of intracellular calcium to activation of specific enzymes
30	Protein POF1B	Q8WVV4	68 kDa	Plays a key role in the organization of epithelial monolayers by regulating the actin cytoskeleton. May be involved in ovary development
31	Ganglioside GM2 activator	P17900	21 kDa	Exhibits some calcium-independent phospholipase activity. Binds gangliosides and stimulates ganglioside GM2 degradation; stimulates only the breakdown of ganglioside GM2 and

				glycolipid GA2 by beta-hexosaminidase A
32	Cluster of Plectin	Q15149	532 kDa	Interlinks intermediate filaments with microtubules and microfilaments and anchors intermediate filaments to desmosomes or hemidesmosomes.
33	Fatty acid-binding protein, epidermal	Q01469	15 kDa	Selectively delivers specific fatty acids from the cytosol to the nucleus, wherein they activate nuclear receptors. Controls retrograde endocannabinoid signaling. Modulates inflammation
34	Carboxypeptidase A4	Q9UI42	47 kDa	Metalloprotease that functions in neuropeptide processing and regulation in the extracellular environment
35	Proteasome subunit alpha type-4	P25789	29 kDa	Component of the 20S core proteasome complex involved in the proteolytic degradation of most intracellular proteins.
36	Proteasome subunit alpha type-3	P25788	28 kDa	Component of the 20S core proteasome complex involved in the proteolytic degradation of most intracellular proteins. Binds to the C-terminus of CDKN1A and thereby mediates its degradation. Negatively regulates the membrane trafficking of the cell-surface thromboxane A2 receptor (TBXA2R) isoform 2
37	Malate dehydrogenase, mitochondrial	P40926	36 kDa	Essential for the conversion of malate to oxaloacetate as part of the proper functioning of the Krebs cycle

38	Vinculin	P18206	124 kDa	A key platelet protein which experiences calcium dependent tyrosine phosphorylation during the activation of platelet
39	Proteasome subunit alpha type-7	O14818	28 kDa	Component of the 20S core proteasome complex involved in the proteolytic degradation of most intracellular proteins. Inhibits the transactivation function of HIF-1A under both normoxic and hypoxia-mimicking conditions. Plays a role in hepatitis C virus internal ribosome entry site-mediated translation. Mediates nuclear translocation of the androgen receptor (AR) and thereby enhances androgen-mediated transactivation. Promotes MAVS degradation and thereby negatively regulates MAVS-mediated innate immune response.
40	Eukaryotic translation initiation factor 6	P56537	27 kDa	Involved in ribosome biogenesis. Behaves as a stimulatory translation initiation factor downstream insulin/growth factors. Associates with pre-60S subunits in the nucleus and is involved in its nuclear export. Required for ROS-dependent megakaryocyte maturation and platelets formation, controls the expression of mitochondrial respiratory chain genes involved in reactive oxygen species (ROS) synthesis (By similarity). Involved in miRNA-mediated gene silencing by the RNA-induced silencing complex (RISC).

				Modulates cell cycle progression and global translation of pre-B cells
41	Proteasome subunit beta type-1	P20618	26 kDa	Component of the 20S core proteasome complex involved in the proteolytic degradation of most intracellular proteins.
42	Lysosomal protective protein	P10619	54 kDa	Associates with both beta-galactosidase and neuraminidase and exerts a protective function necessary for their stability and activity; a carboxypeptidase that can deamidate tachykinins.
43	Lysosome-associated membrane glycoprotein 1	P11279	45 kDa	Presents carbohydrate ligands to selectins. Also implicated in tumor cell metastasis. Acts as a receptor for Lassa virus protein.
44	Proteasome subunit alpha type-6	P60900	27 kDa	Component of the 20S core proteasome complex involved in the proteolytic degradation of most intracellular proteins.