In-silico annotation of innate immune system for identification of cancer prognostic biomarkers

By

Dilraj Kaur

(MT15121)

Under the Supervision of Prof. Gajendra P.S. Raghava

Department of Computational Biology

Indraprastha Institute of Information Technology Delhi

New Delhi - 110020

June, 2022

In-silico annotation of innate immune system for identification of cancer prognostic biomarkers

By

Dilraj Kaur

(MT1521)

A Thesis

Submitted in Partial Fulfillment of the Requirements for the Degree Of

Doctor of Philosophy

Department of Computational Biology

Indraprastha Institute of Information Technology Delhi

New Delhi – 110020

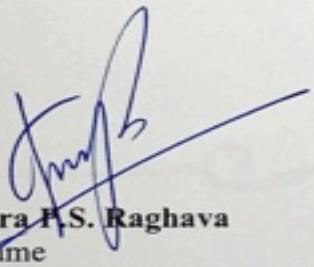
June 2022

Certificate

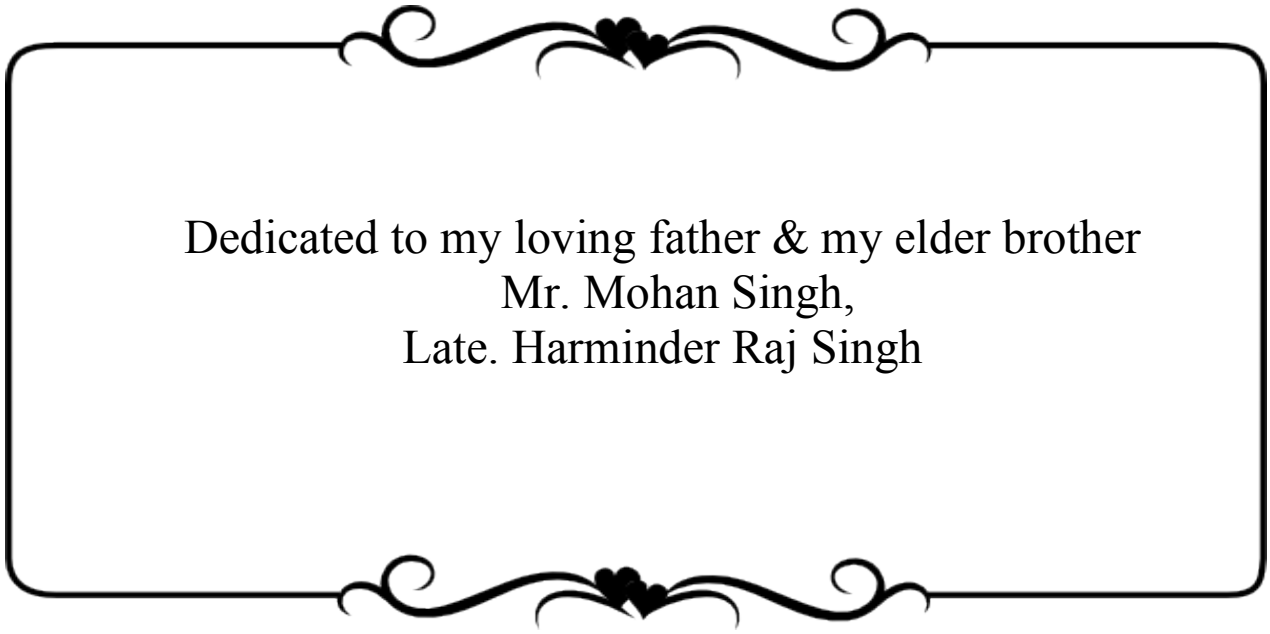
This is to certify that the thesis titled " ***In-silico*** **annotation of innate immune system for identification of cancer prognostic biomarkers**" being submitted by **Mrs. Dilraj Kaur** to the Indraprastha Institute of Information Technology Delhi, for the award of the degree of Doctor of Philosophy, is an original research work carried out by her under my supervision. In my opinion, the thesis has reached the standards fulfilling the requirements of the regulations relating to the degree.

The results contained in this thesis have not been submitted in part or full to any other university or institute for the award of any degree/diploma.

June, 2022
Month, Year



Prof. Gajendra P.S. Raghava
Supervisor Name
Indraprastha Institute of Information Technology Delhi
New Delhi 110 020


Dedicated to my loving father & my elder brother
Mr. Mohan Singh,
Late. Harminder Raj Singh

Acknowledgement

"*Believe in yourself and all that you are. Know that there is something inside you that is greater than any obstacle."* – **Christian D. Larson**

Transitioning from one field to another (experimental lab to *in-silico* lab) was a challenging feat for me. A lot of people made a substantial contribution to making this challenge easier for me. Thus, here by I would like to express my gratitude towards the people who made it possible and who, no matter how difficult the journey seemed, never gave up on me.

I feel so much privileged to have Professor Gajendra Pal Singh Raghava as my doctoral advisor. He provided me the encouragement, support and professional suggestions that were invaluable throughout this roller coaster journey. His friendly nature as a true friend and a matured professional personality at the same time was always helpful to bring enthusiasm in me and my work. His belief in me and providing such supportive environment is really something to admire. His consent guidance and advice on research and career have been priceless. I have always learned something new whenever I have discussion about him. His personality and interest in multiple dynamics of the life teach us to balance our work and personal life so well. I am very much grateful for his patience, motivation, belief, enthusiasm and immense knowledge that made him such great person in human as well. I could not have imagined having a better lab and better advisor for my doctoral research.

I sincerely thank Dr. K. Sriram and Dr. Vibhor Kumar for being in my evaluation panel and mentoring me through their constructive support throughout this journey. I always scared of Dr. K. Sriram during my yearly evaluation because of his plethora of questions he asked, and really thankful to him for always providing new and helpful ideas to add up in the work. I would like to thank Dr. K.Sriram, Dr. Vibhor Kumar and ex-faculty Dr. Arnab Bhattacharya for teaching wonderful courses during my M.Tech period in IIITD. I'd also want to thank the professors of the classes for which I was assigned as a teaching assistant, particularly "Dr. Raj Ayyar", whose pleasant demeanour, philosophical insights, and complimentary chocolates, pizza and coffee were

a treat to both my mental and physical hunger. My advisor Professor Raghava, with whom I have worked as TA for 8 semesters. It was a lovely experience. Finally, I'd want to thank the administrative staff at IIITD, particularly “Mrs. Priti Patel”, “Ms. Shipra Jain”, and “Ms. Sheetu Ahuja”, for always being available to answer our questions and handle our academic issues promptly. I am also grateful to IIITD for providing financial assistance as well as first-rate infrastructure and facilities.

I extends my gratitude to the friends and lab colleagues i met at IIITD. This voyage would not have been complete without a large number of people, however it is hard to name everyone. As a result, I'd want to leave out a number of names (ex-Btech and ex-Mtech batch students from IIITD), who were perhaps the most important but are unlikely to ever read this thesis. I'd want to thank my friends and seniors “Dr. Pawan Kumar Raghav, Dr. Akshara, Dr. Sherry, Dr. Piyush, Dr. Salman, Dr. Chakit, Dr. Anjali Lathwal, Dr. Harpreet , Dr. Rajesh, Dr. Vinod and Dr. Lubna” in particular for their help and support whenever it was needed. I also extends my gratitude to my friends and colleagues “Sumeet, Anjali Dhall, Neelam, Dr. Devi, Akanksha, Ritu, Dashleen, Shubham, and Nishant”. I'd like to thank “Neetesh, Raghav, Smriti, Sarita, Krishan, Shreya, Chitrita, Priyadarshini, Aditya, Omkar, Namrata, and Dr. Shiju” for being such wonderful neighbours. The countless encounters I've had with each of them will be remembered for the rest of my life. In brief I have spent more than five years of my life including M.Tech and Phd here only. I have seen old IIITD with only few buildings and only few bunches of the faculties and studen in CB department. A special thanks to ex- director “Dr. Pankaj Jalote” for his vision and making this institute such a wonderful place to learn, experience and achieve. IIITD is not just an institute for me it’s home to me.

Here comes the most difficult part to describe the kind of critical life support and endless love my family have provided to me. I can’t express my heartiest gratitude, indebtedness and regards to my parents and family members for their immense support and love. A very special note to my husband and my senior from the lab “Dr. Chakit Arora” for his indefinite love and understanding. He has been a great friend, a colleague, a mentor, from lab partner to life partner he was there throughout the journey. I am extremely grateful to my loving grandfather Late. Arjan Singh, my parents “Mr. Mohan Singh” and “Mrs. Haripal Kaur” for their love and support. The constant

interrogation by my father about my work was a driving force to perform my studies well. I would also like to thanks my elder brother “Late. Harminder Raj singh” and younger brother “Mr. Gurunam singh” for their faith and love.

I am also extremely thankful to my parents in law “Mr. Bharat Bhushan Arora” and “Mrs. Anju Arora”, for their undaunted trust in me and supporting me in my last 2 years of PhD to make it smooth. They provided all the love and care to me. Always being very encouraging and helpful during my doctoral research. Lastly, I’d like to acknowledge my sister-in-law “Himani Arora” who is like my true sister and a dear friend, and her husband “Alok Nahata” for always being patient listeners and motivating me in every step of my life.

Dilraj Kaur

Abstract

Innate immune system response is the initial/first line of defense against invading pathogens. It is non-specific and involves various cells like macrophages, neutrophils, natural killer cells, dendritic cells. The response is quicker than adaptive immunity. Unlikely there is no antibodies generation and memory after exposure with any type of infection. Innate immunity consist of physical and chemical barriers such as epithelia and antimicrobial chemicals produced at epithelial surfaces. The system is intricate and consists of blood proteins including member of the complement system and other mediators of inflammation. Phagocytic cells like neutrophils and macrophages, dendritic cells, natural killer cells and other lymphoid cells are essential part of the system. The recruitment and activation of neutrophils at the site of infection to eliminate pathogens is a key aspect of the innate response. The innate immune system also expresses a wide range of Pattern Recognition Receptors (PRRs) which are specialised in the recognition of evolutionary conserved structures known as Pathogen Associated Molecular patterns (PAMPs). Toll like Receptors (TLRs), C-lectin Type Receptors (CLRs), Mannose Binding Receptors (MBRs) and Nucleotide-binding Oligomerization Domain (NOD)-Like Receptor (NLRs) are some of the major PRRs. TLRs are being the most extensively studied PRRs. PAMPs are distinguished by being invariant across whole classes of pathogens, required for pathogen survival, and separate from "self." However, PRRs perceive host factors as "danger" signals in some situations, such as when they are present in atypical locations or abnormal molecular complexes as a result of infection, inflammation, or other forms of cellular stress. These are known as Damage Associated Molecular patterns (DAMPs). PRRs are specialised to recognise these DAMPs as well. PRRs present on the cell surface or intracellularly, signal the presence of infection to the host and initiate proinflammatory and antimicrobial responses by activating hundreds of new intracellular signalling pathways that include adaptor molecules, kinases, and transcription factors in response to PAMP recognition. PRR-induced signal transduction pathways eventually result in the activation of gene expression and the synthesis of a diverse range of molecules, including cytokines, chemokines, cell adhesion molecules, and immunoreceptors, which together facilitate the early host response to infection while also serves as a vital link to the adaptive immune response.

Innate immunity is rapidly evolving, with novel cell types and molecular pathways being discovered and paradigms changing continuously. Over the last decade, our understanding of the

processes by which pathogens are identified has improved significantly. This field has previously been thoroughly investigated. Still, the appropriate annotation of data, as well as the development of more efficient computing resources and diverse methodologies, remains a significant problem. To handle this, we have created a comprehensive knowledge base on PRRs and their corresponding ligands Pattern Recognition Receptor Database 2.0 (PRRDB2.0), which is an updated version of PRRDB. The database consists of more than 2700 entries data from 2008-2018. It provides a user-friendly all-device compatible webserver known as PRRDB2.0. This webserver includes detailed information on numerous classes of PRRs as well as their respective ligands/agonists. The database contains information such as the name, source, origin, role, sequence, length, and assay utilised for both elements. Proper annotation and adequate computational resources can help to understand and design the immune cells, the inflammasome, and DNA sensing. All of these are crucial for the activation and orchestration of innate immunity, which might lead to new treatment options for autoimmune, autoinflammatory, and infectious diseases. We developed "PRRpred" and "DefPred" tools that will help in the annotation of the innate immune system molecules. 'PRRpred' is an in-silico prediction of PRRs. It can predict whether the given protein sequence is PRR or not. It consists of two modules for prediction the first one based on composition of the protein sequence and the other one is based on evolutionary information. The best performing model is a hybrid model of both with Basic Local Alignment Search Tool (BLAST). User can download the prediction result in csv format with the result whether the provided input is PRR/Non-PRR. It is also accompanied by a user friendly, all device compatible web server. Whereas "DefPred" is an *in-silico* tool for scanning, predicting, and designing defensins. Defensins are host specific defense molecules, and are one of the class of Anti-microbial Peptides (AMPs). In this study, we described a reliable method developed for predicting defensins with high precision. We systematically collected defensins, AMPs and non-defensins from various resources to create the largest possible datasets. Developing new defensins can be a very effective alternative to drug resistance, and they are less toxic since they are host specific and produced in the host body.

Each year, cancer alone claims the lives of millions of people all over the world. Despite advancements in cancer treatments, patient survival rates are still below average. The study of the innate immune system has led to the identification of key regulators and the development of chemo-therapeutics that can target them and reverse the state of a cancer patient. We tried to find

out the relation between the gene expression of PRRs and the survival of patients with cancer. Firstly, we identified the prognostic gene signature from the expression profile of PRRs genes in case of endometrial cancer. Later on, we identified the most effective drugs from existing drugs using prognostic gene signature and did repurposing of FDA approved drugs.

Our next goal was to design a universal biomarker corresponding to all types of cancer-based on PRR gene expressions. We tried and developed a 12 gene biomarkers. Although, the biomarker signatures's efficiency is seen to differ among different cancer types, a substantial stratification is achieved in all cases. Lastly to check our hypothesis when there is a change in biological insight, is their any change in the performance of the prognostic biomarker across multiple cancer. For this we have compared two major pathways apoptotic and PRR biomarker genes in case of THCA, MESO and SKCM. We found both the pathways are highly interlinked and there is dependency of their genes in case of cancer.

Altogether, the work discussed here in this thesis recommends some novel approach for the proper annotation of innate system molecules. Also, these molecules related signaling genes were utilized to create prognostic biomarker in various cancer. We anticipate that clinicians and researchers will use the findings of our investigations to develop advanced cancer treatment approaches.

List of Publications

Thesis Related Publications

- **Kaur, D.,** Arora, C., & Raghava, G. P. (2020). A hybrid model for predicting pattern recognition receptors using evolutionary information. *Frontiers in immunology*, *11*, 71.

- **Kaur, D.,** Patiyal, S., Sharma, N., Usmani, S.S., & Raghava, G.P. (2019). PRRDB2.0 : a comprehensive database of pattern recognition receptors and their ligands. *Database, 2019*

- **Kaur, D.,** Arora, C., & Raghava, G. P. S. (2021). Prognostic Biomarker-Based Identification of Drugs for Managing the Treatment of Endometrial Cancer. *Molecular Diagnosis & Therapy*, 1-18.

- **Kaur, D.,** Patiyal, S., Arora, C., Singh, R., Lodhi, G., & Raghava, G. (2021). *In-Silico* Tool for Predicting, Scanning, and Designing Defensins. *Frontiers in immunology*, *12*, 780610. https://doi.org/10.3389/fimmu.2021.780610

Other Publications

- Patiyal, S., Kaur, **D., Kaur**, H., Sharma, N., Dhall, A., Sahai, S., ... & Raghava, G. P. (2020). A web-based platform on Coronavirus Disease-19 to maintain predicted diagnostic, drug, and vaccine candidates. *Monoclonal antibodies in immunodiagnosis and immunotherapy*, *39*(6), 204-216.

- Kaur, H., Bhalla, S., **Kaur, D**., & Raghava, G. P. (2020). CancerLivER: a database of liver cancer gene expression resources and biomarkers. *Database*, *2020*.

- Pande, A., Patiyal, S., Lathwal, A., Arora, C., **Kaur, D.,** Dhall, A., ... & Raghava, G. P. (2019). Computing wide range of protein/peptide features from their sequence and structure. *bioRxiv*, 599126.

- Arora, C., **Kaur, D**., Lathwal, A., & Raghava, G. P. (2020). Risk prediction in cutaneous melanoma patients from their clinico-pathological features: superiority of clinical data over gene expression data. *Heliyon*, *6*(8), e04811.

- Arora, C., **Kaur, D**., Naorem, L. D., & Raghava, G. (2021). Prognostic biomarkers for predicting papillary thyroid carcinoma patients at high risk using nine genes of apoptotic pathway. *PloS one*, *16*(11), e0259534. https://doi.org/10.1371/journal.pone.0259534

- Arora, C., **Kaur, D**., & Raghava, G. (2021). Universal and cross-cancer prognostic biomarkers for predicting survival risk of cancer patients from expression profile of apoptotic pathway genes. *Proteomics*, e2000311. Advance online publication. https://doi.org/10.1002/pmic.202000311

- Lathwal A, Kumar R, **Kaur D**, Raghava GPS. In silico model for predicting IL-2 inducing peptides in human ***(Submitted)***

- Kumar, V., Patiyal, S., Kumar, R., Sahai, S., **Kaur, D.,** Lathwal, A., & Raghava, G. (2021). B3Pdb: an archive of blood-brain barrier-penetrating peptides. *Brain structure & function*, *226*(8), 2489–2495. https://doi.org/10.1007/s00429-021-02341-5

Table of contents

List of Abbreviations

PRR	Pattern recognition receptor
PAMP	Pathogen associated molecular patterns
DAMP	Damage associated molecular patterns
APC	Antigen presenting cells
BLAST	Basic local alignment search tool
CD-HIT	Cluster Database at High Identity with Tolerance
E-value	Expect value
LPS	Lipopolysaccharides
PSI-BLAST	Position-specific iterative basic local alignment search tool
AAC	Amino acid composition
DPC	Dipeptide composition
PSSM	Position specific scoring matrix
FDA	Food and drug adminstration
MCC	Matthews correlation coefficient
AUROC	Area under the receiver operating characteristic
SVM	Support vector machine
RF	Random forest
DT	Decision tree
LR	Logistic regression
ACC	Adrenocortical carcinoma
BLCA	Bladder Urothelial Carcinoma
LGG	Brain Lower Grade Glioma
BRCA	Breast invasive carcinoma
CESC	Cervical squamous cell carcinoma and endocervical adenocarcinoma

CHOL	Cholangiocarcinoma
LCML	Chronic Myelogenous Leukemia
COAD	Colon adenocarcinoma
ESCA	Esophageal carcinoma
GBM	Glioblastoma multiforme
HNSC	Head and Neck squamous cell carcinoma
KICH	Kidney Chromophobe
KIRC	Kidney renal clear cell carcinoma
KIRP	Kidney renal papillary cell carcinoma
LAML	Acute Myeloid Leukemia
LIHC	Liver hepatocellular carcinoma
LUAD	Lung adenocarcinoma
LUSC	Lung squamous cell carcinoma
DLBC	Lymphoid Neoplasm Diffuse Large B-cell Lymphoma
MESO	Mesothelioma
MISC	Miscellaneous
OV	Ovarian serous cystadenocarcinoma
PAAD	Pancreatic adenocarcinoma
PCPG	Pheochromocytoma and Paraganglioma
PRAD	Prostate adenocarcinoma
READ	Rectum adenocarcinoma
SARC	Sarcoma
SKCM	Skin Cutaneous Melanoma
STAD	Stomach adenocarcinoma
TGCT	Testicular Germ Cell Tumors
THYM	Thymoma
THCA	Thyroid carcinoma
UCS	Uterine Carcinosarcoma
UCEC	Uterine Corpus Endometrial Carcinoma
UVM	Uveal Melanoma
CRC	Colo-rectal cancer
TCGA	The cancer genome atlas
GEP	Gene expression profile
GO	Gene ontology

HR	Hazard ratio
GPM	Good prognostic marker
BPM	Bad prognostic marker
C	Concordance index
CI	Confidence interval
RS	Risk score
RG	Risk grade
Cox-PH	Cox proportional hazard
OS	Overall survival
AUROC	Area under receiver operating curve
PI	Prognostic index
RF	Random forest
SVR	Support vector regressor
KNN	k-nearest neighbours
DT	Decision Trees
MLR	Machine learning regression
TP	True positive
FP	False negative
TN	True negative
FN	False negative
RSEM	RNA-Seq by Expectation-Maximization
TLR	Toll like receptor
NLR	Nucleotide-binding oligomerization domain (NOD)-like receptor
PGRP	
AMP	Anti microbial peptides
CLR	C-type lectin receptors
RLR	Retinoic acid inducible gene-1 like receptors
MBL	Mannose binding lectin
ODN	Oligodeoxynucleotide
PCC	Pearson correlation coefficient

List of genes and their description

Gene	Description
BUB1B	BUB1 Mitotic Checkpoint Serine/Threonine Kinase B
CCNB1	Cyclin B1
CDC20	Cell Division Cycle 20
NCAPG	Non-SMC Condensin I Complex Subunit G
CLDN9	Claudin 9
AK4	Adenylate Kinase 4
PC	Pyruvate Carboxylase
GPC1	Glypican 1
SRD5A3	Steroid 5 Alpha-Reductase 3
B4GALT1	Beta-1,4-Galactosyltransferase 1
GMPPB	GDP-Mannose Pyrophosphorylase B
CHST6	Carbohydrate Sulfotransferase 6
B4GALT4	Beta-1,4-Galactosyltransferase 4
CACNA2D2	Calcium Voltage-Gated Channel Auxiliary Subunit Alpha2delta 2
CTSW	Cathepsin W
NOL4	Nucleolar Protein 4
SIGLEC1	Sialic Acid Binding Ig Like Lectin 1
TMEM150B	Transmembrane Protein 150B
TRPM5	Transient Receptor Potential Cation Channel Subfamily M Member 5

CLEC1B	C-Type Lectin Domain Family 1 Member B
CLEC3A	C-Type Lectin Domain Family 3 Member A
MRC1	Mannose Receptor C-Type 1
IRF7	Interferon Regulatory Factor 7
CTSB	Cathepsin B
FCN1	Ficolin 1
RIPK2	Receptor Interacting Serine/Threonine Kinase 2
CLEC3B	C-Type Lectin Domain Family 3 Member B
CLEC12B	C-Type Lectin Domain Family 3 Member B
TLR4	Toll Like Receptor 4
NLRP10	NLR Family Pyrin Domain Containing 10
MAPKAPK2	MAPK Activated Protein Kinase 2
TNIP1	TNFAIP3 Interacting Protein 1
SARM1	Sterile Alpha And TIR Motif Containing 1
BTK	Bruton Tyrosine Kinase
ITGB2	Integrin Subunit Beta 2
HAVCR2	Hepatitis A Virus Cellular Receptor 2
FCRL3	Fc Receptor Like 3
CD163	CD163 Molecule
CD300LF	CD300 Molecule Like Family Member F

CD68	CD68 Molecule
CTSS	Cathepsin S
CLEC10A	C-Type Lectin Domain Containing 10A
CLEC12A	C-Type Lectin Domain Family 12 Member A
NR1H3	Nuclear Receptor Subfamily 1 Group H Member 3
CLEC4E	C-Type Lectin Domain Family 4 Member E
CD209	CD209 Molecule
ITGAM	Integrin Subunit Alpha M
TLR8	Toll Like Receptor 8
NLRP10	NLR Family Pyrin Domain Containing 10
NLRP9	NLR Family Pyrin Domain Containing 9
FCN1	Ficolin 1
RIPK2	Receptor Interacting Serine/Threonine Kinase 2
FUNDC1	FUN14 Domain Containing 1
HSP90AA1	Heat Shock Protein 90 Alpha Family Class A Member 1
UNC93B1	Unc-93 Homolog B1, TLR Signaling Regulator

ALPK1	Alpha Kinase 1
APPL1	Adaptor Protein, Phosphotyrosine Interacting With PH Domain And Leucine Zipper 1
CASP8	Caspase 8
CD5	CD5 Molecule
CLEC2D	C-Type Lectin Domain Family 2 Member D
HMGB1	High Mobility Group Box 1
HSP90B1	Heat Shock Protein 90 Beta Family Member 1
IKBKG	Inhibitor Of Nuclear Factor Kappa B Kinase Regulatory Subunit Gamma
IRF1	Interferon Regulatory Factor 1
KLRB1	Killer Cell Lectin Like Receptor B1
NCAN	Neurocan

List of Software/Databases

Name	Availability
sklearn	https://scikit-learn.org/stable/
caret	https://github.com/topepo/caret
survival	https://cran.r-project.org/web/packages/survival/index.html
survminer	https://cran.r-project.org/web/packages/survminer/index.html
survMisc	https://cran.r-project.org/web/packages/survMisc/index.html
TCGA-Assembler 2	https://github.com/compgenome365/TCGA-Assembler-2
Connectivity Map 2	https://clue.io/cmap
dendextend	https://cran.r-project.org/package=dendextend
survivalROC	https://cran.r-project.org/web/packages/survivalROC/index.html
Firehose	https://gdac.broadinstitute.org
randomForestSRC	https://cran.r-project.org/package=randomForestSRC
TCGA-GDC	https://portal.gdc.cancer.gov
AJCC Individualized	http://www.melanomaprognosis.net
Melanoma Prediction Tool	
StringDB	https://string-db.org
GEO	https://www.ncbi.nlm.nih.gov/geo/
GeneCards	https://www.genecards.org
CCGD	http://ccgd-starrlab.oit.umn.edu
GTEx	https://gtexportal.org/home/
GEPIA	http://gepia.cancer-pku.cn
SurvExpress	http://bioinformatica.mty.itesm.mx:8080/Biomatec/SurvivaX.jsp
HPA	https://www.proteinatlas.org
Programming environments	Python, R, HTML, PHP, CSS and Javascript

List of Figures

List of Tables

1

Introduction

1.1 Immune System

The immune system is a complex network of cells and proteins protecting the body from infection. It is an intricated network of organs, white blood cells, proteins (antibodies), lymphoid organs, humoral factors, cells, cytokines, and other biomolecules. This system works collectively to defend you from external invaders like bacteria, fungi, viruses, and other parasites that cause infection, sickness, and disease. The immune system is critical to our existence. Our body would be vulnerable if we did not have an immune system. Our immune system is pledged to keep us healthy while moving through a sea of germs. The immune system's overall job is to prevent or restrict infection. When your immune system performs correctly, it can distinguish which cells belong to you and are foreign to your body. It stimulates, mobilizes, fights, and eliminates foreign invader germs that might damage you. The role of the immune system is best understood when it fails. Its misfunctioning results in severe infections, immunodeficiency, autoimmune disorders, hyper allergy, and tumors (Parkin and Cohen 2001). These specialized cells and immune system components defend the body against disease. This protection is called immunity. It can also be described as a perplexing biological system that recognizes and accepts what belongs to the self while also acknowledging and rejecting what is foreign (non-self). A detailed representation of the human immune system is demonstrated in Figure 1.1.

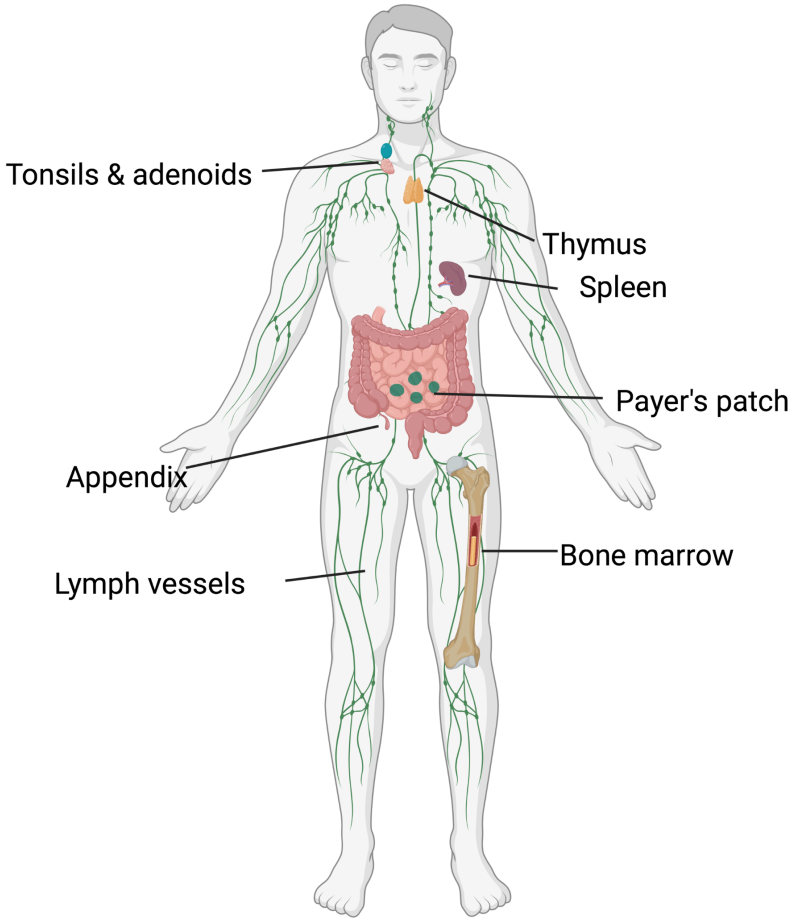


Figure 1.1 Representation of the organs playing significant role in human immune system

1.2 Types of Immunity

There are mainly three types of immunity; innate, adaptive, and passive immunity. Innate immunity is also known as inherent immunity, aka non-specific immune response. People are born with natural immunity or innate immunity, which serves as a general defense. For example, the skin works as a barrier to keep pathogens out of the body. Additionally, the immune system knows when to act against foreign and potentially dangerous invaders. It often refers to a physical, chemical, and biological barrier that provides the first line of defense. Innate immune components like neutrophils, monocytes, cytokines, macrophages, complement receptors, and acute-phase proteins provide an immediate defense to the host. It is non-inclusive or generic and non-specific in action. Adaptive immunity takes over when any pathogen surpasses the innate one. Adaptive immunity develops during the

lifespan. When any person is exposed to the illness or is immunized against them using the vaccine, he/she acquire the adaptive immunity. The immune system of higher animals is distinguished by adaptive immunity. This response is made up of antigen-specific responses mediated by T cells and B lymphocytes. Although the innate response is quick, it can be harmful to the tissue as it is non-specific. At the same time, the adaptive one is precise and nonharmful but is relatively slower. Because the adaptive response remembers, further exposure results in a more powerful and quicker reaction, although this is not immediate. Because our immune system remembers former enemies, this is also referred to as immunological memory. Passive immunity is known to be the one that is borrowed from a different source and is temporary only. Likewise, antibodies present in a mother's milk provide a newborn baby interim protection against the illness to which the mother has been exposed. A more detailed representation of categories of immunity is shown in Figure 1.2.

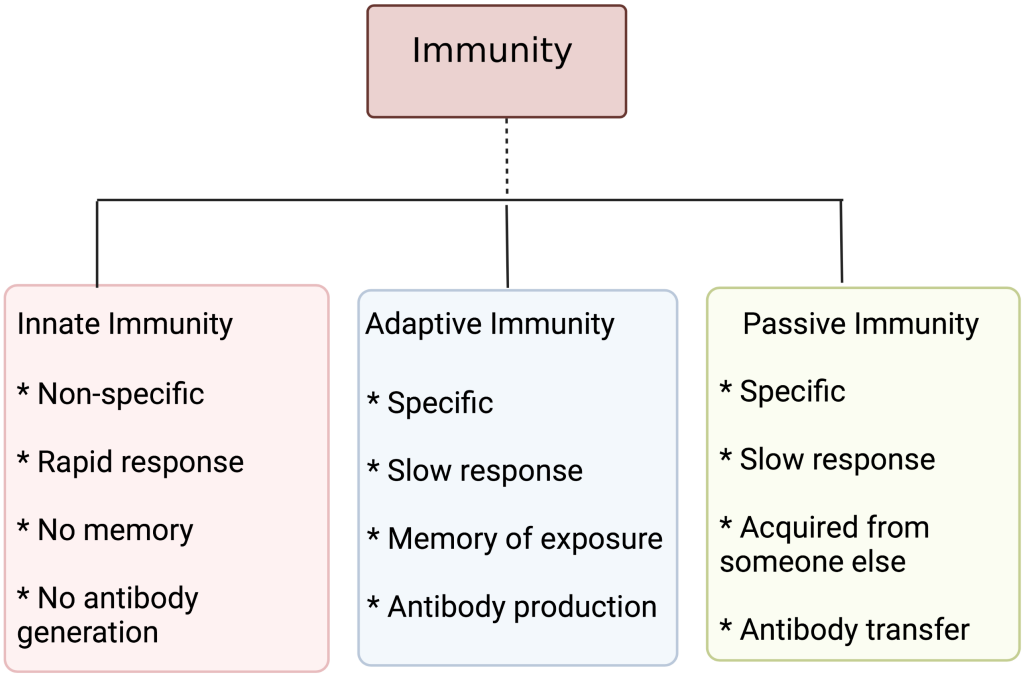


Figure1.2 Types of immunity

1.3 Components of the Immune System

The immune system is composed of cells, tissue, organs, and numerous chemicals that combat infections or illnesses. The fundamental elements of this system have white blood cells, the lymphatic system, the antibodies, the spleen, the thymus, and the bone marrow. These are the immune system components that actively combat infection. The entire immune cells arise from a precursor in the bone marrow and the mature cells through sequential modification from different body parts. The representation of components of both the immunity is shown in Figure 1.3.

Skin: Usually, the skin provides initial protection against microbes. The immune cells can be found in different layers of skin and these cells expel the antimicrobial substances.

Bone marrow: It involves the stem cells, which further grow into various cell types. The common myeloid progenitor cells present in the bone marrow are the originator or precursor of the innate cells like neutrophils, macrophages, monocytes, mast cells, dendritic cells, basophils, and others that are important for the response. The adaptive cells like B and T cells arise from common lymphoid progenitor cells. These cells help in the immunological memory against the pathogens that acted in the past. Natural killer (NK) cells have the same precursor and possess features of both the arms of the immunity. They provide instant responses like innate defense and the memory cells preserve the immunological memory.

Bloodstream: The immune cells invigilate the bloodstream regularly for any abnormalities. A blueprint of the immune system is taken when a blood test reports the white blood cells. The rare or abundance of these cells in the bloodstream indicates a problem.

Thymus: The thymus, a tiny organ in the upper chest, is where T lymphocytes develop.
The *lymphatic system* is a network of veins and tissues made up of lymph, an extracellular fluid, and lymphoid organs like lymph nodes. The lymphatic system serves as a route for communication and transit between tissues and the circulation. Immune cells travel via the lymphatic system and congregate in lymph nodes located throughout the body.

Lymph nodes serve as a communication gateway for immune cells to sample information from the body. For example, if adaptive immune cells in the lymph node detect fragments of a bacterium brought in from afar, they will activate, reproduce, and leave the lymph node to circulate and treat the pathogen. As a result, physicians may examine patients for enlarged lymph nodes, which may signal an active immune response.

Spleen: The spleen is a digestive organ found behind the stomach. Although it is not directly related to the lymphatic system, it is essential for processing information from the circulation. Immune cells are abundant in certain parts of the spleen, and when blood-borne infections are recognized, they activate and respond appropriately.

Tissue of the mucosa: Pathogens like to enter through mucosal surfaces, and specific immunological hubs are strategically situated in mucosal tissues such as the respiratory tract and the gut.

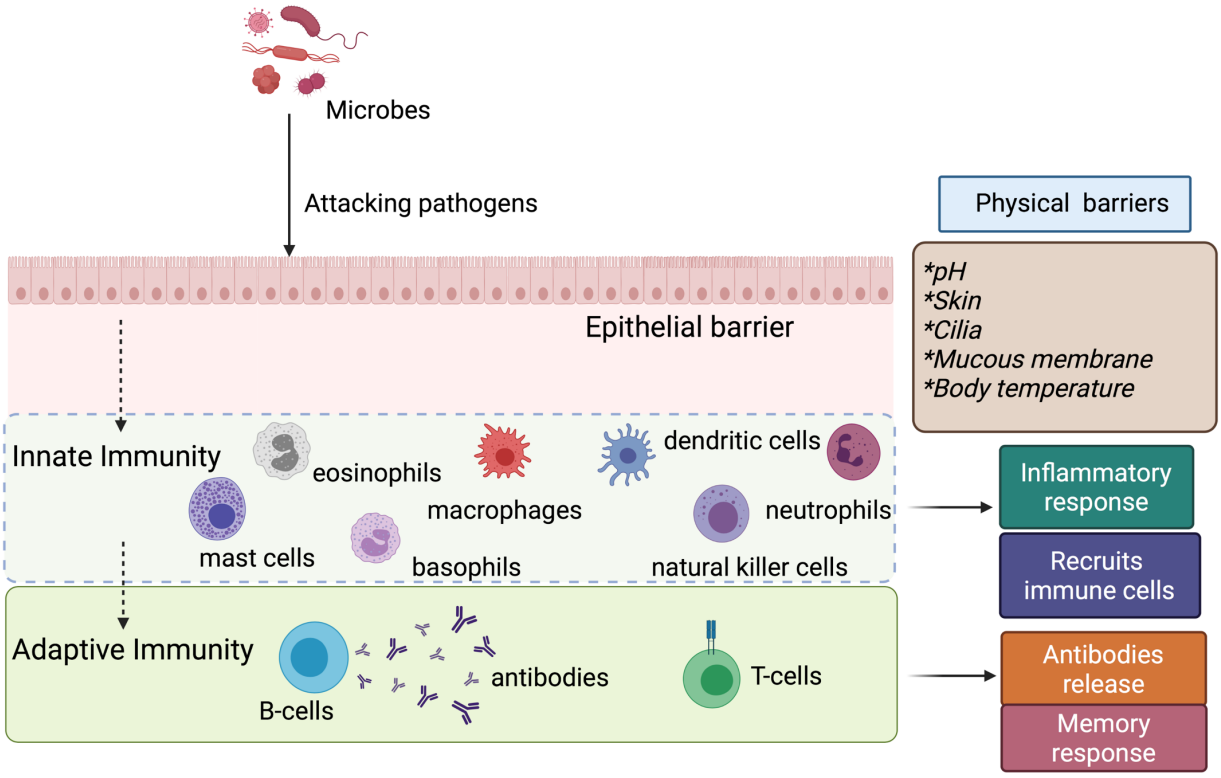


Figure 1.3 The diagrammatic representation of the components of innate and adaptive immune system.

1.4 How Innate Immune System Works

The innate immune system is a host defense mechanism that has evolved over time, with significant aspects shared by plants, invertebrates, and mammals (Buchmann 2014; Bryant and Monie 2012). The word 'Innate' comes from the Latin word 'Innatus,' which means 'inborn. It consists of cells and mechanisms that act as a nonspecific first line of defense against invading pathogens. Innate immune responses rely on the body's ability to recognize pathogens that have conserved features not found in the uninfected host. In animals, innate immune defenses cover almost all tissues, especially barrier surfaces like the skin and mucosal surfaces of the respiratory and gastrointestinal tracts. Non-hematopoietic cells, as well as specialized myeloid and lymphoid sensor and effector cells, can start and exert innate defense mechanisms and become activated in response to tissue injury, infection, or genotoxic stress. (Galli, Borregaard, and Wynn 2011). Through germline-encoded receptors, the innate immune system may "detect" such situations.

Pattern Recognition Receptors (PRRs) such as toll-like receptors (TLRs) are proteins capable of recognizing molecules frequently found in pathogens (so-called Pathogen-Associated Molecular Patterns—PAMPs), or molecules released by damaged cells (Damage-Associated Molecular Patterns—DAMPs). They emerged phylogenetically prior to the appearance of the adaptive immunity and, therefore, are considered part of the innate immune system. Antimicrobial peptides (AMPs) (Hilchie, Wuerth, and Hancock 2013; Lees et al. 2019), complement factors (Degn and Thiel 2013; de Cordoba et al. 2012), alarmins (Chan et al. 2012; D. Yang et al. 2009), cytokines/chemokines (Paterson et al. 2021), chitinases/chitinase-like proteins(Lee et al. 2008), acute-phase proteins, proteases, and other less-categorised molecules are examples of innate immune responses mediated by cell-dependent mechanisms (e.g. phagocytosis and cytotoxicity).

Different PRR families have been studied in the past, with transmembrane proteins like Toll-like receptors (TLRs) and C-type lectins receptors (CLRs) being the most studied, as well as cytoplasmic proteins like nucleotide-binding oligomerization domain (NOD)-like receptors (NLRs) and retinoic acid-inducible gene-I-like receptors (RLRs). TLRs are type-1 transmembrane proteins that detect PAMPs associated with invading pathogens both outside and within the cell, as well as in intracellular endosomes and lysosomes (Tartey and Takeuchi 2017; Kawai and Akira 2010; Hoving, Wilson, and Brown 2014b; Franchi et al. 2009; Loo and Gale 2011). CLRs are signaling transmembrane receptors that are important

in antifungal immunity. The fundamental role of all PRRs is to detect PAMPs or DAMPs, which are important microbial components. The interaction of PRRs with PAMPs causes a variety of effects, including immune cell maturation, migration, and activation, as well as cytokine and chemokine production (Taghavi et al. 2017). The transcription of genes controlling proteins implicated in the inflammatory response, such as type I interferons (IFNs), proinflammatory cytokines, chemokines, antimicrobial proteins, and so on, is upregulated by most PRRs.

The innate immune system efficiently distinguishes pathogen types based on PRR, and hence recruits the most efficient adaptive immune response to eliminate infections and their toxic molecules (Jain and Pasare 2017; Palm and Medzhitov 2009). The engagement of PRRs in response to PAMPs causes the activation of various cell death mechanisms in order to enhance tissue homeostasis and host-defense against pathogens. Importantly, DAMPs, or cell death products, establish a feedback loop that stimulates PRRs, causing inflammatory/immune response (Chaplin et al. 2018) as shown in Figure 1.4. This field has been thoroughly researched in the past. However, the proper annotation and development of more efficient computational resources and versatile methods continues to be a challenge.

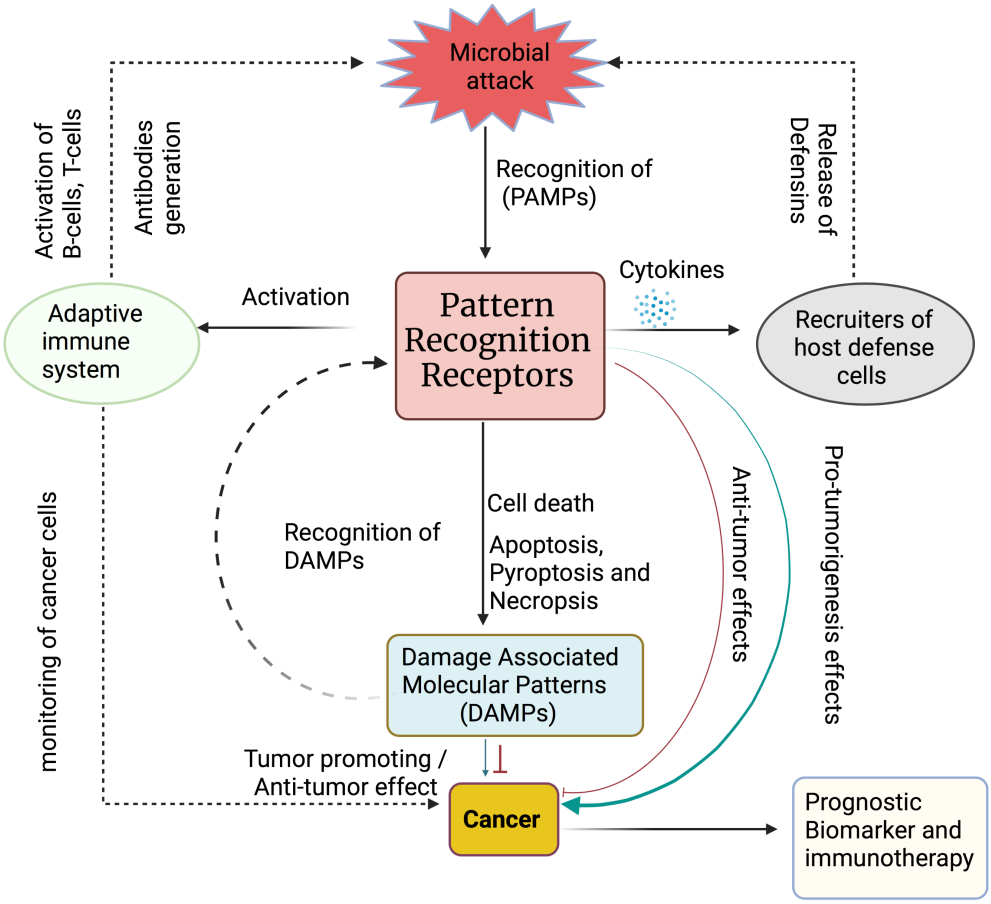


Figure 1.4 Mechanism and response of PRRs towards PAMPs and DAMPs and role in cancer.

1.5 Innate Immune System in Cancer

Besides having inflammatory roles and being involved in diseases like rheumatic disease (Mullen, Chamberlain, and Sacre 2015), autoimmune disorders atherosclerosis, sepsis, asthma (Lin, Verma, and Hodgkinson 2012), heart failure (Farrugia and Baron 2017), kidney diseases (Komada and Muruve 2019), bacterial meningitis, Parkinson's disease, stroke, Alzheimer's disease, viral encephalitis (V. Kumar 2019), immunodeficiency disorders like 'chronic granulomatous disease (CGD)', and 'X-linked agammaglobulinemia (XLA)'(Mortaz et al. 2017). Innate immune system does play a very vital role in cancer immunotherapy. The innate immune receptors like PRRs involved in cell death molecular mechanism that includes apoptosis, necroptosis and pyro-ptosis (Morimoto et al. 2021). PRRs, shows antitumoral activities in several cancers through activation in tumor cells. This

activation could trigger both pro- or antitumoral effects depending on the context (Shirota, Tross, and Klinman 2015) as shown in Figure 1.2. New therapies that promote anti-tumor immunity have been recently developed. Most of these immunomodulatory approaches have focused on enhancing T-cell responses, either by targeting inhibitory pathways with immune checkpoint inhibitors, or by targeting activating pathways, as with chimeric antigen receptor T cells or bispecific antibodies (Demaria et al. 2019). Many PRRs related genes have previously been associated with cancer development or progression. As a result, several drugs and biomarkers have been developed. However, the problem for identifying novel biomarkers and the creation of new prognostic methods remains open. In addition to the requirement of improved accuracy, novel methods are expected to overperform clinical features or compliment them. Given the crucial role of innate immune responses in immunity, harnessing these responses opens up new possibilities for long-lasting, multi layered tumor control.

1.6 Proposal's Origin

Several initiatives have been undertaken in the last decade to research adaptive immunity. Tremendous exposure like annotation, creating in-silico tool, making usage of biological insight to understand the mechanistic point has been explored well in this arm of immunity. Whereas, adaptive immunity get activated through first line of defense innate immunity and if any malignancy get resolved at first step there would not be any need to go further on another step. But, there is not as much work has been done in innate immunity. Although it has important role in fighting against infection and providing host defense, also it plays a vital role as pro and anti-tumoral molecules. The innate immune molecules requires a proper annotation so thus researcher use them for translational benefits in research and therapies. Several essential regulators have been identified, as well as their involvement in this complex system. In summary, it has been shown that some components and portions of the innate immune system are weakened in cancer cells, causing these injured cells to refuse to die and disseminate the harm to future generations. Because of our current understanding of the pathways, drugs that target these critical components and restore the survival/death balance have been developed. Furthermore, changes in the concentrations or status of innate immune molecule regulators are utilized to predict cancer prognosis and risk. The development of novel prognostic biomarkers/methods for cancer risk assessment, on the

other hand, remains a challenge. Likewise, given the importance of numerous clinical aspects in cancer genesis and progression, these prospective techniques should incorporate important elements in order to supplement or replace existing risk prediction systems. The innovative prognostic approaches can be used to provide more precise risk prediction and, as a result, more effective therapy planning.

1.7 Objective of the Thesis

To overcome these short-comings we have put effort to explore innate immune system in depth. Our present work mainly focuses on the innate immune system as the adaptive immune system is highly explored. The study is primarily divided into two broad categories (i) *in-silico* annotation of innate immune system (ii) identification of cancer biomarker and peptide therapeutics. For this, we have employed in-silico annotation first using the innate immune receptors (PRRs) and created a user-friendly webserver 'PRRpred', which can be used to predict whether a given protein is PRR or not. We have also created 'DefPred' for the classification of defensins and non-defensins. Besides this, we found the role of these receptors in cancer biomarker discovery in case of Endometrial cancer and related immune therapies. We have also created a universal prognostic biomarker for pan-cancer dataset. To check the interconnection and dependency of pathways in case of prognosis and their performance we have made the comparison between prognostic biomarker's performance and their biological insights using apoptotic pathway and PRR signaling pathway genes. Figure 1.5 outlines our overall work done in brief.

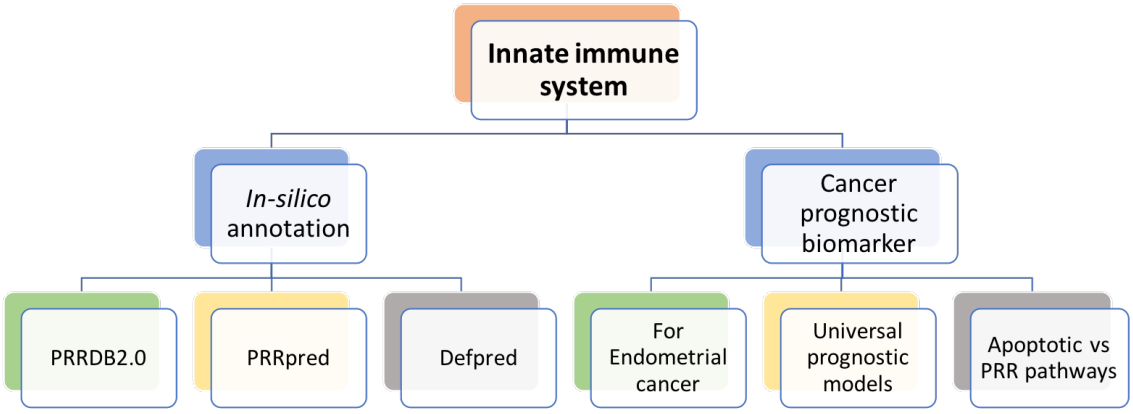


Figure 1.5 Outline of thesis

1.8 Organization of the Chapters

This thesis is divided into nine chapters, each of which contains the following information:

Chapter 1- In this part, the immune system is introduced and the underlying biological concept of the innate system. This is accompanied by a brief discussion of the various immune systems and innate system cells. Finally, the importance of innate immune cells, particularly PRR signaling genes, in cancer development and therapy is explored. The conclusion of this chapter emphasizes the importance of studying and annotating the innate immune system and its biomolecules, as well as the defense mechanism and use of PRR signaling genes for the identification of various new prognostic biomarkers and the construction of effective risk prediction models in the case of various cancers.

Chapter 2- This chapter provides a review of the literature on the innate immune system, its annotation work to date, and the use of immunotherapy for cancer. It also emphasizes the significance of various PRR-based combination treatments with conventional therapy in various cancers. In a nutshell, this chapter explains why the study was conducted.

Chapter 3- This chapter focuses on the thesis's first goal, which is the creation of a computational resource on Pattern Recognition Receptors (PRR). It is a ten-year update known as PRRDB2.0. The chapter goes into great detail about PRRs and their ligands/agonists. Details such as the name, source, function, and sequences of receptors and their agonists. It also contains derived information such as a Swiss-prot id, sequences in FASTA format, and a pub-chem assay. PRRDB2.0 contains information on more than 2700 PRRs and their ligands and is the largest informative collection known to date. The chapter also discusses the utility of the developed resource for improving and designing adjuvants that can aid in vaccination efficiency.

Chapter 4- This chapter is about the annotation of PRRs. We created an in-silico tool for the prediction of PRRs from a given protein sequence. It has also user friendly webserver utility named ‘PRRpred’ where a user can provide the input in FASTA format and it will predict whether the given protein is PRR or not. It has two prediction module the one sequence composition based and the other one is based on evolutionary information.

Chapter 5- This is a follow-up to a previous work on the annotation of innate immune system molecules. This research is primarily concerned with the prediction, design, and scanning of Defensins, which are host defence innate immune molecules. It also offers a user-friendly, all-devices compatible webserver called 'DefPred.' It features two *in-silico* models for determining if a protein sequence is defensin or not. Model-I can distinguish between any class of AMPs (Anti-Microbial Peptides) and defensins. The user can provide any AMPs sequence, and this model -I will predict whether or not the given input sequence is defensin. Model-II differentiates between defensin and any random protein sequence.

Chapter 6- - As it has been demonstrated that the PRR signalling pathway genes malfunction can occur at any of the multiple regulatory stages in various cancers, the genomic data corresponding to the entire PRR signalling pathway is used in this chapter. The prognostic significance of each of these genes in the context of endometrial cancer is furthermore investigated. Through their published functions in endometrial cancer, key genes are discovered and validated. Clinical features were also examined and taken into consideration and finally a hybrid prognostic biomarker has been made using 9 genes and clinical staging in case of endometrial cancer. Therapeutic possibilities have been proposed based on important biomarker genes and the downstream pathways they affect. In this situation, drug repurposing was also done, and a few FDA-approved drugs were also proposed in this chapter.

Chapter 7- The fundamental purpose of this chapter is to apply the concept of Chapter 6 to various tumors and utilize the data to build universal prognostic models. A universal prognostic biomarker applicable to a broad spectrum of cancers might have far-reaching consequences in the future.

Chapter 8- This chapter follows the previous study and demonstrates the interconnection and dependency of two key cancer pathways. According to the findings of Chapter 7, we examined apoptotic pathway and PRR signaling pathway biomarker genes in three cancer cases: THCA, MESO, and SKCM. A hypothesis was used to determine if changing the biological route may affect the performance of prognostic biomarkers in cancer, as well as whether these pathways are interconnected or act independently.

Chapter 9- In this chapter the thesis work finishes by providing a quick overview of the study and its contribution to the area of innate immune system research and its role and usage in therapy in cancer research.

2

Review of Literature

2.1 Overview of Innate Immune System

Organisms that are inhaled, ingested, or inhabit our skin and mucous membranes are constantly present. The pathogenicity of the organism (the virulence factors at its disposal) and the integrity of host defence mechanisms determine whether these organisms penetrate and cause illness. The immune system is a network of lymphoid organs, cells, humoral factors, and cytokines that interact with one another. The immune system's critical role in host defence is best seen when it fails; underactivity results in severe infections and tumours of immunodeficiency, overactivity in allergy and autoimmune illness (B. P. Kaur and Secord 2019). The detection systems (receptors and structures found on pathogens), the cells involved, and the nature of the processes differentiate innate (natural) immunity from acquired immunity. During an infection, innate immune responses emerge before acquired immune responses. Natural immunity includes cytokine, chemokine, and interleukin production; innate, cytokine-dependent nonspecific immunity of leukocytes; HLA-independent pathogen-killing cells; and phagocytosis (Sochocka and Blach-Olszewska 2005). While innate immunity is important for host defence against viral threats, it is also emerging as a key regulator of human inflammatory illness. Indeed, innate immune responses have been linked to the development of asthma and atopy, as well as a wide range of autoimmune diseases such as Type 1 diabetes, inflammatory bowel disease, and systemic lupus erythematosus. The new molecular explanation of how the innate immune system detects infection in order to activate protective immune responses has sparked a revival in the area of innate immunity. Innate immunity has abandoned its previous, derogatory label of 'non-specific immunity,' and is now a proud companion with the adaptive immune system in defending human hosts against pathogenic infections (Turvey and Broide 2010).

2.2 Innate Immune Molecules

Unlike the adaptive immune system, which is dependent on T and B lymphocytes, innate immune protection is accomplished by cells of both hematopoietic and non-hematopoietic origin. Macrophages, dendritic cells, mast cells, neutrophils, eosinophils, natural killer (NK) cells, and natural killer T cells are hematopoietic cells that participate in innate immune responses. In addition to hematopoietic cells, the skin and epithelial cells lining the respiratory, gastrointestinal, and genitourinary tracts have innate immune reactivity. Innate immune system

comprises and relies on Pattern Recognition Receptors (PRRs) proteins that recognise components that are often linked with infections (also known as Pathogen-Associated Molecular Patterns—PAMPs) (Chaplin et al. 2018). As shown in Figure 2.1 PRRs are specialised protein receptors aka innate immune receptors and can recognised PAMPs as well as Danger Associated Molecular Patterns (DAMPs). Different PRR families have been studied in the past, with transmembrane proteins like Toll-like receptors (TLRs) and C-type lectins receptors (CLRs) being the most studied, as well as cytoplasmic proteins like nucleotide-binding oligomerization domain (NOD)-like receptors (NLRs) and retinoic acid-inducible gene-I-like receptors (RLRs). TLRs are type-1 transmembrane proteins that detect PAMPs associated with invading pathogens both outside and within the cell, as well as in intracellular endosomes and lysosomes (Tartey and Takeuchi 2017; Kawai and Akira 2010; Hoving, Wilson, and Brown 2014b; Franchi et al. 2009; Loo and Gale 2011). CLRs are signaling transmembrane receptors that are important in antifungal immunity. The fundamental role of all PRRs is to detect PAMPs or DAMPs, which are important microbial components. The interaction of PRRs with PAMPs causes a variety of effects, including immune cell maturation, migration, and activation, as well as cytokine and chemokine production (Taghavi et al. 2017). The transcription of genes controlling proteins implicated in the inflammatory response, such as type I interferons (IFNs), proinflammatory cytokines, chemokines, antimicrobial proteins, and so on, is upregulated by most PRRs. PRRs are categorised into mainly three subtypes on the basis of their presence and activity (i) Intracellular PRRs (ii) Soluble PRRs (iii) Cell surface PRRs. As name suggest intracellular are those PRRs that resides and found inside the cells like TLR3, TLR7, TLR8 and TLR9. Soluble PRRs are soluble in nature like Mannose Binding Lectin (MBL) and Ficolin. The PRRs which present on the cell surface are known as cell surface PRRs and they include TLR2, TLR4 and TLR5.

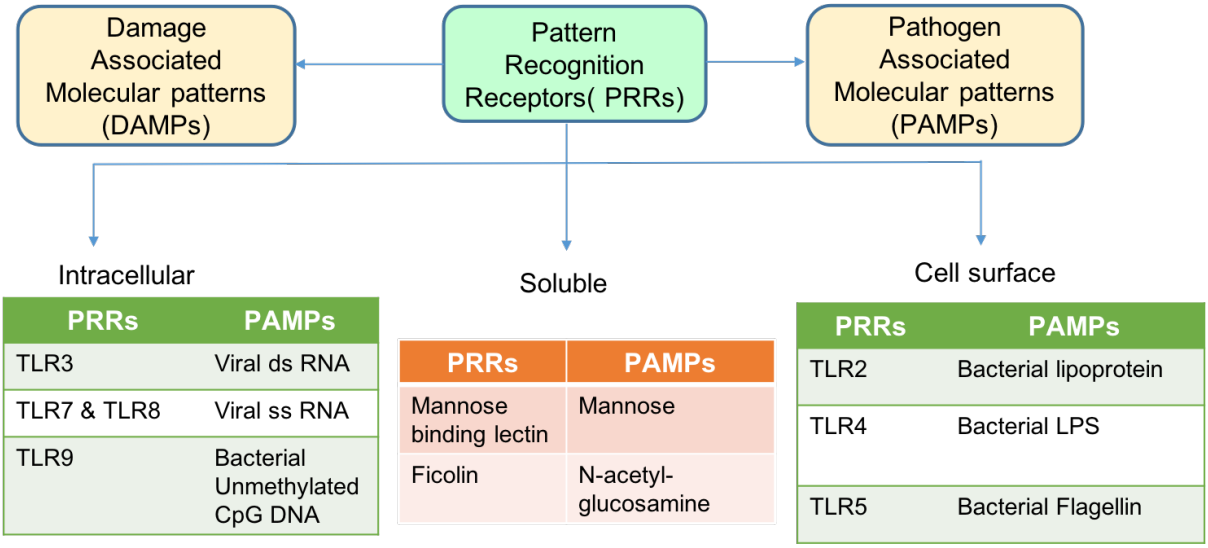


Figure 2.1 Representation of overview of PRRs

2.3 Role of PRR and its Ligand

Whenever there is an entry of antigen into the body, the immune system attempts to eliminate it. The adaptive immune system takes its time and develops long-term immunological responses, whereas the innate immune system is the body's first line of defense (Chaplin et al. 2018). Innate immune cells do have Pattern recognition receptors (PRRs) that assists in the detection of pathogens. They recognize pathogen associated molecular patterns (PAMPs) which are molecular patterns present in microbes (Haghparast, Zakeri, and Ramezani 2016). Microbes cause cytosolic buildup of inactive IL-1 precursor and caspase-1 activation during infection, the latter of which catalyzes the cleavage of the IL-1 precursor pro-IL-1 (Martinon, Burns, and Tschopp 2002; Mariathasan et al. 2006). Martinon et al. found a protein complex called the inflammasome that is responsible for this catalytic activity (231). The adaptor ASC (apoptosis-associated speck-like protein containing a CARD), pro-caspase-1, and a member of the NLR family, such as Ipaf (Ice protease-activating factor), NALP (NAcht LRR protein) 1, or NALP3/Cryopyrin, make up this inflammasome (Mariathasan et al. 2006; Martinon, Burns, and Tschopp 2002). There are various families of PRRs majority including TLRs, RLRs, NLRs and CLRs. Antigen presenting cells (APCs) recognizes PAMPs with the help of PRRs. After recognition antigens/ foreign particles are processed in APC and then get loaded onto major histocompatibility complex (MHC) molecules as shown in Figure 2.2. Adjuvants, in general, activate PRRs in immune cells to boost the innate immune system. The majority of immune

stimulatory adjuvants serve as PRR ligands, enhancing an activation pathway and inducing cytokine release (Coffman, Sher, and Seder 2010). Components of injured or dying host cells also contribute to adjuvant activity as a result of inflammasomes (Coffman, Sher, and Seder 2010). Aptamers are oligonucleotide molecules that have been chosen from a huge library to bind to a specific target. Aptamers can be employed in a range of medicinal, diagnostic, and target-binding applications as alternative to antibodies. It has been seen when delivered with vaccine, the CD28 aptamer dimer had a costimulatory effect, evoking a stronger cellular response. Affimer molecules are tiny proteins that bind to their targets with nanomolar affinity. These non-antibody binding proteins have been created to imitate the molecular recognition features of monoclonal antibodies in a variety of applications, including diagnostic tools and biotherapeutics. Nanomaterials have been shown to work as adjuvants by improving antigen transport to the immune system or potentiating innate and adaptive immune responses. Vaccine Adjuvants with Aluminium For almost eighty years, aluminium-based vaccine adjuvants have been used safely in human vaccination against illnesses such as DTaP (Diphtheria, Tetanus, acellular Pertussis), Human Papillomavirus, Pneumococcal, Hepatitis A, and Hepatitis B (Baylor, Egan, and Richman 2002)

TLR ligands are powerful immunomodulators that can affect a variety of immune responses. Once TLR recognises its ligands it stimulates the cells downstream and different patterns of gene expression are induced. The difference in TLR signalling is due to adopter molecules like MyD88 (myeloid differentiation primary response gene 88) and TRIF (toll/interleukin-1 receptor domain containing adaptor protein inducing interferon-b). MyD88 helps in the production of inflammatory cytokines by stimulating nuclear factor-kB (NF-kB), whereas TRIF assists in the production of type-I interferons (IFN) (Akira 2011). Adjuvants with structures similar to the different ligands of PRRs can activate innate immunity by stimulating their respective receptors. Cytokines released by innate immune system, can trigger adaptive immunological responses by boosting T cell responses, activating humoral immunity, or a combination of the two. Cell-mediated immunity is involved in the Th1 response, whereas humoral responses are used to eliminate external antigens by Th2 cells. The development of new adjuvants capable of eliciting protective CD8+ T cell responses is fraught with difficulties. Combination of a promising adjuvant with an antigen is required to promote functional CD8+ T cell development (Coffman, Sher, and Seder 2010).

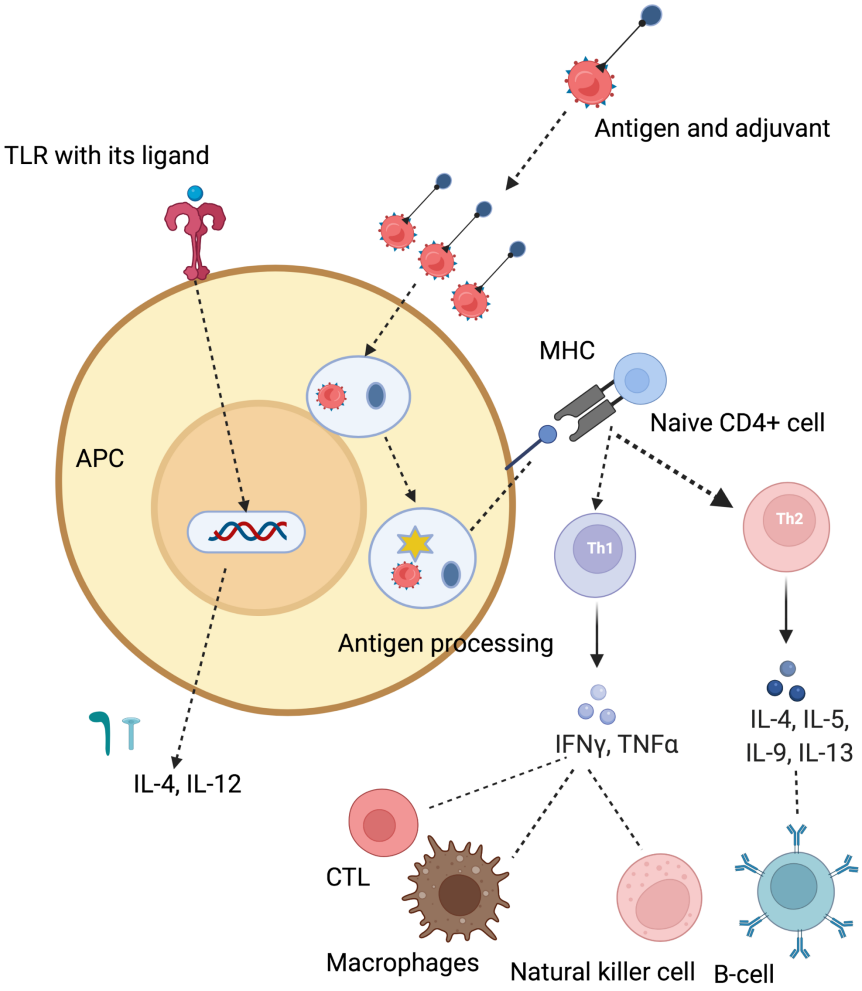


Figure 2.2 Representation of the mechanism of PRRs

2.3 Available Databases for Vaccine Adjuvants

Various databases have been created throughout the world to assist the scientific community during the last few decades. These databases provide a variety of information about the human immune system as well as biological molecules that can trigger immunological responses and also act as adjuvants (Table 3). PRRDB (Lata and Raghava 2008a) which was first developed in 2008, has detailed about the experimentally verified pattern recognition receptors and their agonist/ligands. It got updated recently as PRRDB2.0 (D. Kaur et al. 2019) possess about 5 times more information than the previous one. The information is very beneficial in designing vaccine adjuvants. Vaxjo (Sayers et al. 2012) is a web based vaccine adjuvant database developed in 2012. It is an analytic system for storing, curating, and analyzing vaccine adjuvants and their applications in vaccine development. Vaxjo currently has 103 vaccine

adjuvants in its database. 98 of these adjuvants have been utilized in 384 VIOLIN vaccines against 81 infections, malignancies, and allergies.

The most extensive web-based vaccination database and analysis system is the Vaccine Investigation and Online Information Network (VIOLIN) (Xiang et al. 2008). The vaccination information has been yielded in this database from approx. 1,600 peer-reviewed articles. VIOLIN has over 3,000 vaccinations or vaccine candidates for over 190 diseases. Over 3,000 vaccinations or vaccine candidates for over 190 diseases are presently available in VIOLIN. Manual curation of approximately 1,600 peer-reviewed articles yielded the vaccination information in the database. VIOLIN, unlike most other vaccine databases, concentrates on vaccination research data. Unlike most other vaccine databases, it concentrates on vaccination research data. AntigenDB (Ansari, Flower, and Raghava 2010) is a database that contains exhaustive information about experimentally verified antigens, including structural and functional annotation. PolysacDB (Aithal et al. 2012) is a maintained database of antigenic polysaccharides. It has extensive information regarding antigenic polysaccharides of microbial origin from literature and digital sites. It has around 1,554 total entries in which there is information on 149 different antigenic polysaccharides from 347 various bacteria. Each item regarding antigenic polysaccharide has details like its origin, role, respective antibodies, utilities, conjugation method, potential epitopes implicated. These database can be beneficial in the development of vaccines based on proteins or antigens.

Table 2.1. List of databases or repositories developed for maintaining adjuvants resources

Name	Description	Weblink	Year	Working Status
PRRDB	pattern recognition receptor database	https://webs.iiitd.edu.in/raghava/prrdb/	2008	Yes
PRRDB 2.0	updated pattern recognition receptor database	https://webs.iiitd.edu.in/raghava/prrdb2/	2020	Yes
Vaxjo	vaccine adjuvant database and its application for	http://www.violinet.org/vaxjo/	2012	Yes

	analysis of vaccine adjuvants and their uses in vaccine development			
VIOLIN	a comprehensive vaccine database and analysis system	http://www.violinet.org	2014	Yes
Antigen db	a database of pathogenic antigens	https://webs.iiitd.edu.in/raghava/antigendb/index.html	2010	Yes
Innate db	comprehensive information on innate immunity	http://www.innatedb.com	2013	Yes
Polysac DB	Repository of microbial polysaccharide antigens and their antibodies	https://webs.iiitd.edu.in/raghava/polysacdb/	2012	Yes

2.4 Tools for Designing Vaccine Adjuvants

Efforts have been made in the last decade to create data-driven techniques for predicting biomolecules that has immunomodulatory response and can act as adjuvants. Few of them are listed in Table 2.2. The foreign RNA sequence of a disease is detected by our innate immunity system, which then activates the immune system to clear the body of the infection. RNA-based immunotherapy and vaccination adjuvants can take use of RNA's immunomodulatory properties. The immunomodulatory impact of an RNA sequence is undesirable in siRNA-based treatment because it may cause immunotoxicity. 'imRNA' (Chaudhary, Nagpal, et al. 2016) is a method that provides the facility to create RNA-based medicines, vaccine adjuvants by constructing a single-stranded RNA (ssRNA) sequence with desirable immunomodulatory properties. The 'VaxinPAD', (Nagpal et al. 2018) predicts immunomodulatory peptides, paves the way for the development of rational peptide-based vaccination adjuvant design. The research is the first attempt to create models for predicting immunomodulatory peptides for vaccine adjuvant development. 'VaccineDA' (Nagpal et al. 2015) is the first of its kind to attempt to create an in silico platform for designing oligodeoxynucleotide (ODN) based vaccination adjuvants. The majority of these ODNs contain CpG sequences that can activate the innate immune system. The current work is the first of its kind to attempt to create an in silico platform for designing ODN-based vaccination adjuvants. VaccineDA offers a number of in silico modules that give users with the tools they

need to create ODN-based vaccine adjuvants. Various cytokines also act as immunomodulators and play a very significant role in innate immunity. Tools like IL6pred, IL2pred, IL4pred and IL10pred (Dhall et al. 2021; Anjali Lathwal et al. 2021; Dhanda et al. 2013; Nagpal et al. 2017) have been provided in the past for the prediction and designing of their respective interleukin inducing peptides. IFN epitope (Dhanda, Vir, and Raghava 2013) is a web tool for the prediction of IFN-γ inducing peptides. It also provides facility of virtual screening of peptide libraries and the identification of IFN-γ inducing regions in antigen.

Table 2.2. *In-silico* tools for prediction of adjuvants for vaccine.

Prediction server	Biomolecule	Description	Weblink	Year	Working Status
imRNA	SiRNA, RNA	Prediction of Immunomodulatory potential of an RNA	https://pubmed.ncbi.nlm.nih.gov/26861761/	2016	Yes
VaxinPad	Peptide	Prediction of antigen presenting cell modulators	https://webs.iiitd.edu.in/raghava/vaxinpad/	2018	Yes
VaccinDA	Nucleic Acid	oligodeoxynucleotide-based vaccination adjuvants: prediction, design, and genome-wide screening	https://webs.iiitd.edu.in/raghava/vaccineda/	2015	Yes
IL6pred	Cytokine	Prediction and design of IL-6 inducing peptides	https://webs.iiitd.edu.in/raghava/il6pred/	2021	Yes
IL10pred	Cytokine	Immunosuppressive peptides 's prediction and designing based on IL-10 inducing potential	https://webs.iiitd.edu.in/raghava/il10pred/	2017	Yes
IL2pred	Cytokine	Prediction of Interleukin 2 Inducing Peptides	https://webs.iiitd.edu.in/raghava/il2pred/	NA	Yes
IL4pred	Cytokine	Designing and discovering of interleukin -4 inducing peptides	https://webs.iiitd.edu.in/raghava/il4pred/	2013	Yes
IFN epitope	Cytokine	Predicting and designing interferon gamma inducing epitopes	https://webs.iiitd.edu.in/raghava/ifnepitope/	2013	Yes

2.5 PRRs in Cancer

Several PRR molecules have been found in/on cancer cells from many organs, including the lung, head and neck, colon, breast, stomach, ovary, and others (Damasdi et al. 2017; Fukata et

al. 2007; Gowing et al. 2017; Ikehata et al. 2018; N. Jiang et al. 2017; Park, Chung, and Kim 2017; Royse et al. 2017; Yue et al. 2017). The interactions between tumour cells and TLRs are intricate. They include not only the detection of PAMPs of microbial origin, but also interactions with tumor-infiltrating cells (TIC) such as NK cells, dendritic cells (DCs), CD8+ T cells, innate lymphoid cells, and others (Matsumoto et al. 2017). TLRs expressed on TIC are activated by DAMPs (tumour debris), resulting in antigen presentation to CD8+ T cells and an anti-tumor impact. In general, however, TLR expression appears to be tumor-promoting in the majority of cancers. TLR2, TLR4, and TLR9 are expressed in pancreatic cancer cells, although their prevalence of risk factors differ. TLR4 activation enhances angiogenesis (Sun et al. 2016), but TLR9 cytoplasmic expression has been linked to improved patient survival (Leppanen et al. 2017). TLR signalling increases autoregulatory tumour cell proliferation and anti-apoptotic Bcl-xL expression (Grimmig et al. 2016; Won et al. 2017). TLR4 expression has been widely acknowledged in the development of hepatocellular carcinoma by multiple pathways, including an increase in Treg cells, liver resident follicular helper-like T cells, and enhanced synthesis of pro-inflammatory and malignancy-related chemicals (I. J. Song et al. 2018). Apart from TLR4, additional TLRs such as TLR2, TLR3, and TLR9 have already been identified in cancer and hepatic cirrhosis (Yin et al. 2016). TLR4 expression was designated as a probable carcinogenic agent in hepatocellular carcinoma due to its ability to enhance the amount of many pro-inflammatory and malignancy-related molecules such as NANOG, Caspase-1, and others (Sepehri et al. 2017). TLR5 and TLR7 expression was associated with tumour recurrence in HPV-positive oropharyngeal carcinoma. TLR5 and TLR7 expression were both associated with poor disease-specific survival (Jouhi et al. 2017).

Table 2.3 Effect of various TLRs on different cancer.

Types of TLR	Types of Cancer	Effect of TLR on Tumor
TLR2	Oral Squamous Cell Carcinoma	Progression
TLR4	Head and Neck	Progression
TLR5, TLR7	Squamous Cell Carcinoma	Recurrence
TLR2, TLR4, TLR9	Pancreatic	Progression
TLR1/TLR2, TLR6	Chondrosarcoma	Suppression

Endosomal TLR7 and TLR8 agonists that recognise ssRNA are the most well-known TLR agonists. Imidazoquinoline, subsequently known as imiquimod, a TLR7 agonist, was shown to have antiviral action, first in animal models and then in people (Y. C. Chang et al. 2005; Mauldin et al. 2016). 852A, another TLR7 agonist, was discovered to induce plasmacytoid DCs to create IFN type I and to activate both CD8+ T cells and NK cells, resulting in an anti-tumor response (Inglefield et al. 2008; Weigel et al. 2012) TLR9 agonists detect unmetylated CpG dinucleotides (CpG ODN). The latter, when artificially created, induces a variety of desired immunological responses, including improved innate immunity and adaptive Th1 response (Krieg 2007). Polyinosinic–polycytydylic acid (poly:C), a double-stranded RNA that can function as a TLR3 ligand, has been found to decrease the development of radioresistant Lewis lung cancer in mice when used in conjunction with radiation (Yoshida et al. 2018).

Table 2.4 Types of TLR and its ligand in cancer immunotherapy

Type of TLR	Type of Cancer	Ligand/Agonist used
TLR7	Hematologic tumors	852A
TLR9	Myeloma	C792
TLR3	Advanced solid tumors	Poly(I:C)
TLR9	Myeloma	C792
TLR7/TLR8	Basal cell, other skin cancers	Imiquimod
TLR7	Melanoma	852A

2.6 PRRs as Targeted Therapy in Cancer

Early research on PRR-related drugs mostly focused on monotherapy with TLR-related treatments, however most monotherapies failed to provide positive outcomes (Y. H. Kim et al. 2010). Furthermore, patients had a higher prevalence of mild to severe systemic influenza-like symptoms. The realisation that cell death generated by cell suppressive treatment may be the consequence of immunogenic death has prompted researchers to focus their efforts on DAMP synthesis in tumour cells using radiation and chemotherapy as a means of activating the immune system. PRR agonists have the potential to be employed as vaccine adjuvants as well as to enhance systemic therapies such as chemotherapy, targeted therapy, and immunotherapy

(Table 2.5). Indeed, some have demonstrated effectiveness against immunotherapy resistance (Shekarian et al. 2017). Due to the fact that tumour development leads PRR agonists to have a negative regulatory impact on cancer immunity, PRR agonists paired with immune checkpoint inhibitors may be more appealing therapy choices. Furthermore, when used in combination with immune checkpoint inhibitors, PRRs can remodel the immune milieu and change a "cold" tumour into a "hot" one, improving therapeutic effectiveness even further.

Table 2.5 Different type of combinatorial therapies in case of various cancer.

Therapy	Cancer Type	Targeted PRR	Ligand/Agonist
PRR+ chemotherapy	Squamous cell NSCLC	TLR2	CADI-05
PRR+ chemotherapy	melanoma, NSCLC	TLR9	CpG 7909
PRR+ chemotherapy	breast cancer cutaneous metastases	TLR7/8	imiquimod
PRR+ radiotherapy	B-cell and T-cell lymphomas	TLR9	CpG 7909
PRR+ radiotherapy	Hepatocellular Carcinoma	TLR3	poly-ICLC
PRR+ radiotherapy+ chemotherapy	Glioblastoma	TLR3	poly-ICLC

2.7 Conclusion

Pattern recognition receptors have long been thought to be a minor biological phenomena. It began to change as their role in the infection became obvious, as did the relationship between innate and acquired immunity. Multiple past studies have revealed the detailed mechanism of PRRs that are requisite part of innate immune system. Yet specific proper annotation is lacking for the ligands/ agonist corresponding to PRRs. *In-silico* web resources and updated knowledgebase for PRRs for better understanding and designing vaccine adjuvants is required. Also, due to its dual role in cancer PRR can be use as targeted therapy but, utilization of these PRR and their agonist in prognosis of cancers is not explored yet. Furthermore, computational tools and databases that provide updated information and insight of PRR mechanism and use as biomarkers in cancer are not available.

3

Archive on Pattern Recognition Receptors

3.1 Introduction

Innate immunity which is also known as first line of defense is found in almost all kind of plants and animals. It is originated from the latin word 'Innatus' which mean 'Inborn'. It is non specific in nature ad comprises of cells and mechanism for providing defense. Pathogen recognition being the initial and significant part of the defense mechanism . Besides recognition of pathogens and contribution in acute inflammation it also activates the adaptive immunity. Innate immune cells comprises of germline specialized receptor known as Pattern Recognition Receptors (PRRs). They recognize pathogen- and damage-associated molecular patterns (PAMPs and DAMPs) on invading microorganisms. (Akira, Uematsu, and Takeuchi 2006; Takeuchi and Akira 2010). Different PRR families have been studied in the past, with transmembrane proteins like Toll-like receptors (TLRs) and C-type lectins receptors (CLRs) being the most studied, as well as cytoplasmic proteins like nucleotide-binding oligomerization domain (NOD)-like receptors (NLRs) and retinoic acid-inducible gene-I-like receptors (RLRs). TLRs are type-1 transmembrane proteins that detect PAMPs associated with invading pathogens both outside and within the cell, as well as in intracellular endosomes and lysosomes. Innate immune cells' pathogen identification and Toll-like receptor-targeted therapies (Tartey and Takeuchi 2017; Kawai and Akira 2010; Hoving, Wilson, and Brown 2014a; Franchi et al. 2009; Zhu et al. 2018). CLRs are transmembrane signalling receptors that play an important role in antifungal immunity. They are lectin-like receptors that recognize both exogenous and endogenous ligands and have at least one C-type lectin-like domain. (Hoving, Wilson, and Brown 2014a; Franchi et al. 2009; Zhu et al. 2018; J. Tang et al. 2018). NLRs and RLRs are intracellular cytosolic sensors. NLRs, generally associated with bacterial recognition, are composed of a central nucleotide binding domain and C-terminal leucine-rich repeats, whereas RLRs are the helicases that sense PAMPs with viral RNA (Y. K. Kim, Shin, and Nahm 2016; Kawai and Akira 2009).

Additionally many other receptors like scavenger, mannose and β-glucan receptors are involved in phagocytosis. Complement receptors, collectins, ficolins and pentraxins are some of the secreted PRRs (Paveley et al. 2011). Identification of PAMPs/ DAMPs which are important microbial component is an elemental role of PRRs. The interaction between PRRs and their corresponding ligands downstream various effects like maturation, migration and activation of immune cells. It also includes production of cytokines and chemokines (Taghavi et al. 2017).

Usually PRRs upregulates the transcription of genes controlling proteins implicated in the inflammatory response, such as type I interferons (IFNs), proinflammatory cytokines, chemokines, antimicrobial proteins, and so on. They also intensify the transcription and translation of proteins involved in PRR signalling regulation, which might lead to an adaptive immune response (Akira, Uematsu, and Takeuchi 2006; Mogensen 2009; Fearon and Locksley 1996) (Figure 2.1). The adaptive immune system, also known as the "specific immune system," is made up of particular cells and comprises humoral and cell-mediated immunity. It assists in the elimination of pathogens in the late stages of infection. The innate immune system can readily distinguish between different types of infections thanks to PRR, and consequently generates the most efficient adaptive immune response to eliminate the pathogens and their toxic molecules. (Iwasaki and Medzhitov 2015; Jain and Pasare 2017; Palm and Medzhitov 2009). This field has been extensively explored in the past, and researchers have established a lot of computational tools such as MHCBN, IEDB, and Bcipep etc (Bhasin, Singh, and Raghava 2003; Vita et al. 2019; Saha, Bhasin, and Raghava 2005). BepiPred 2.0, Bcepred, Lbtope, IgPred, PEASE, etc. aid in predicting epitopes in humoral-mediated immunity (Jespersen et al. 2017; Saha and Raghava 2004; Lian et al. 2015; Gupta et al. 2013; Sela-Culang, Ofran, and Peters 2015). ProPred 1 and NetMHCstabpan help in predicting MHC-I binder, whereas ProPred, MHC2Pred and EpiDOCK predict MHC-II binders (Harpreet Singh and Raghava 2003; Rasmussen et al. 2016; Bhasin and Raghava 2007; H Singh and Raghava 2001; Atanasova et al. 2013). We've released review papers that go into great depth regarding immunology resources and *in-silico* tools (Dhanda et al. 2017; Usmani, Kumar, Bhalla, et al. 2018). We suggest that computational resources in the PRRs–PAMPs field need to be revived. How specific PRRs sense invading pathogens, mechanisms involved in immune response against PAMPs, downstream signalling cascades involved in eliciting immune response, and other questions/challenges must be clearly inferred to retain better therapeutic strategies against a variety of infectious diseases caused by invading pathogens, are some of the questions/challenges that must be clearly inferred to retain better therapeutic strategies against a variety of infectious diseases caused by invading pathogens. In 2008, the first version of PRRDB, a database of PRRs and their ligands, was released (Lata and Raghava 2008b). PRRDB was helpful in the development of additional resources such as AntigenDB (Ansari, Flower, and Raghava 2010) and PolysacDB (Aithal et al. 2012), as well as the prediction of pattern receptor recognition families (Gao et al. 2012b). Innate immunity has been better

understood since 2008, and numerous more pathogen-associated molecules have been investigated and identified. As a result, there is a strong need to improve and update as much information as possible regarding PRRs and their ligands. PRRDB 2.0 is an updated and comprehensive database of PRRs and their ligands. The updated version includes detailed information on receptors, such as their domain and localization, as well as elaborative functions such as the role, occurrence, and sequence of their ligands. Furthermore, PubChem assays and experimental procedures that elucidate PRRs and their ligands as well as their available structures have been included to the updated version, which were previously missing.

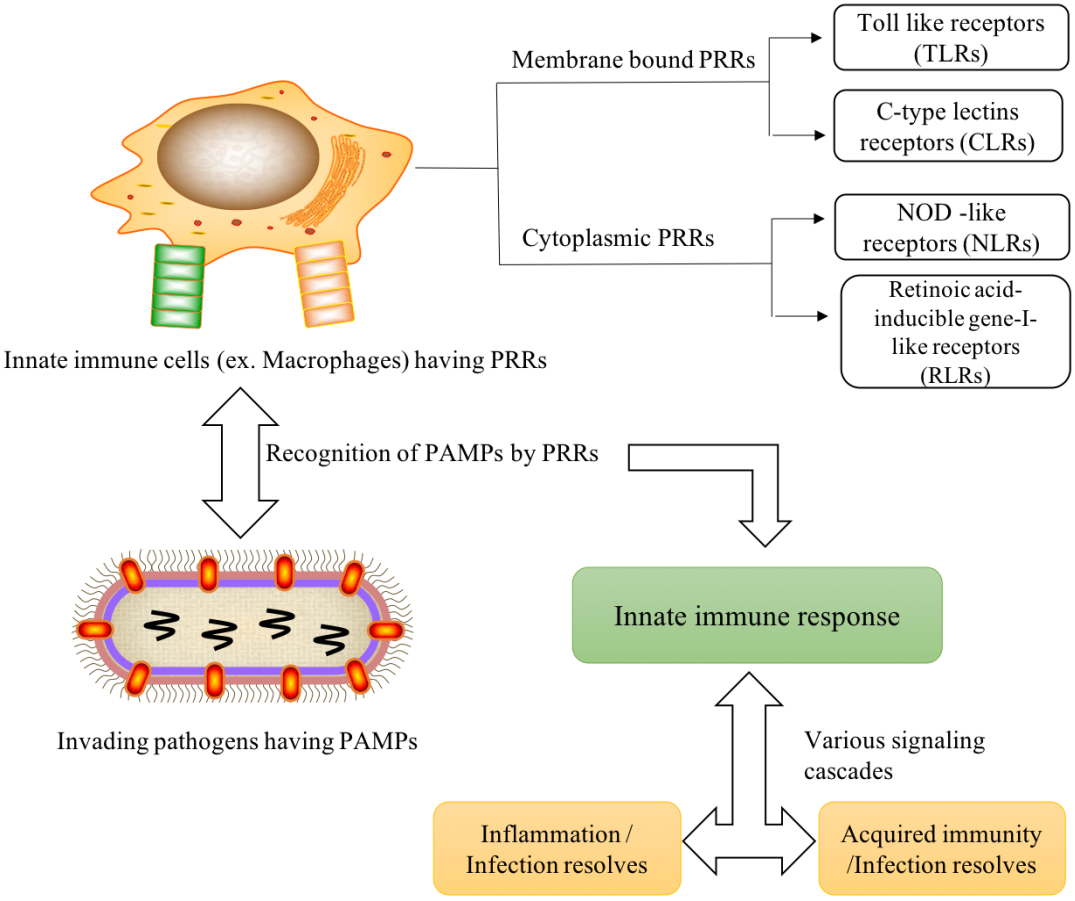


Figure 3.1 Graphical representation of immune mechanism through PRRs and PAMPs association after the microbial invasion (source ~ Kaur et al. 2019).

3.2 Material and Methods

3.2.1 Data Collection

PubMed was examined using keywords like ‘Pattern recognition receptors’ and ‘Pathogen-associated molecular patterns’, specifically published for 10 years from 2008 to 2018. The cumulative hits obtained were ~30 000. We screened all the abstracts manually and selected ~ 3000 abstracts for further investigation.

3.2.2 Data Curation

PRRDB2.0 has extensive information in the form of primary as well as secondary. Primary data comprises of sole information. Extracted from the research articles published and are linked under ‘PMID’. While secondary information are the derived one. Information have been provided in the form of tabular format for both receptors and their ligands (PAMPs/DAMPs). The primary information fields regarding ligands are (i) Ligand name: represents the name of particular ligand (PAMPs/DAMPs) (ii) Ligand source: describes the actual source or origin of that ligand; (iii) Ligand type: represents the category of ligands such as lipid, peptidoglycan, lipopolysaccharide, protein etc.; (iv) Occurrence of ligand: represents either natural or synthetic occurrence of ligands; (v) Role of ligands: provides extensive information about corresponding ligands’ role in activating the immune system.

Apart from this, PRRs have been organized under headings such as (vi) Receptor's name: represents the name of PRRs used in the literature; (vii) Receptor source: describes the receptor's true source or origin. (viii) Receptor type: refers to the many types of PRRs, such as TLRs, CLRs, RLRs, NLRs, and so on; (ix) Receptor localization: transmits the receptor's location or the cell type from where it was discovered; (x) Domain: A domain inside a PRR, such as the Leucine-rich domain in TLR9 or the lectin domain in CLRs, represents a specific domain within that PRR and (xi) Function: when linked with their unique ligands, indicates the role or function of PRRs in triggering the innate immune system via signal cascades. Under the heading 'Assay utilized,' the experimental technique or particular assays used in the associated literature are also curated. In addition to the aforementioned information, a hyperlink has been provided to all of the PubChem assays known to date for that particular PRR (S. Kim et al. 2019). PubChem and Swiss-Prot were used to compile various key pieces of information that were not included in the original study paper, such as PRR sequences and ligands (Prasad et al. 2020). The Protein Data Bank (PDB) was also used to get experimentally known PRR structures (Burley et al. 2019). We attempted to give predicted PRR structures

using structure prediction algorithms, namely PHYRE2, in circumstances where the structures had not been experimentally characterized (Kelley et al. 2015).

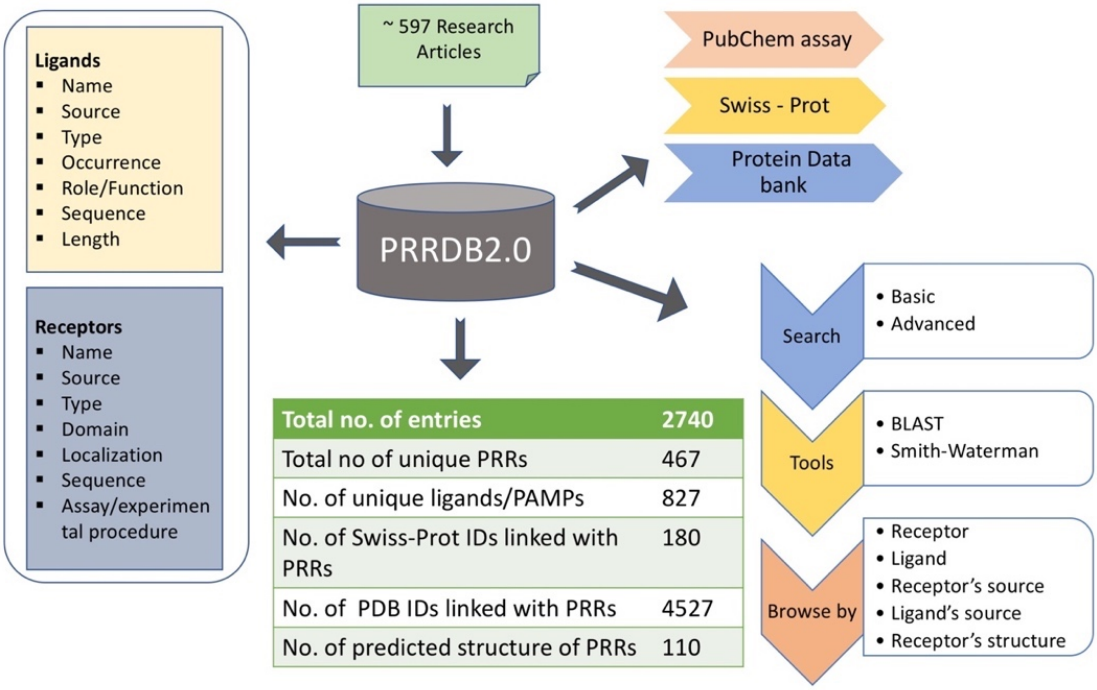


Figure 3.2 Digramatic representation of database architecture its organization and its facilities.

3.2.3 Database architecture and Web Interface

All the information obtained from the literature studies on PRRs was stored in SQL table and provided as a user-friendly interface in PRRDB2.0 (webs.iiitd.edu.in/raghava/prrdb2) which is based on Linux based Apache Server (LAMP). The Front-end web interface was made using bootstrap, a responsive development framework that includes HTML, CSS, and java script. MySQL client program was used to create the back-end database, and all the data handling/manipulation was done using the structured query language (SQL). The overall architecture of the PRRDB2.0 is in Figure 3.2. The information obtained from the research articles and the patents is summarized in tabular form under 25 fields in the database. We carefully searched the papers for every experimental detail, and relevant information was then included in the database.

3.3 Results

3.3.1 Data Statistics

PRRDB 2.0, the latest version of the PRR database, has 2740 entries derived from 597 research publications. A comprehensive update on PRR and its ligands has been published. In addition to the 353 receptors from the first edition of PRRDB, we have included 2374 more receptors, bringing the total number of PRRs to 2727. Similarly, PRRDB 2.0 includes 2197 total ligands researched in the last ten years, as well as 353 ligands from the previous database, for a total of 2550 ligands. In all, 2740 entries in PRRDB 2.0 include information on 2727 total, 467 distinct PRRs and 2550 total, 827 unique ligands.

The primary types of receptors accessible in PRRDB 2.0 are shown in Figure 3.3A. Because TLRs are the most well-known and researched of the PRRs, they are mentioned in 62% of entries. NLRs—241, CLRs—135, Scavenger—88, Syk-coupled CLRs—63, RLRs—40, Mannose receptor—33, PGRPs—25, and RAGE—22 are among the other entries. TLRs and CLRs are well known membrane-bound pathogen receptors. They cover, around 72% of the PRRs that are curated in PRRDB 2.0. Whereas, cytoplasmic PRRs covered in PRRDB2.0 are 10% only. Figure 3.3B depicts the graphical distribution of entries for various ligand types, including 496 entries for nucleic acids, 353 entries for protein-type ligands, lipopolysaccharides—207, peptidoglycan—111, carbohydrates—88, lipoproteins—85, glycoprotein—41, lipopeptide—37, glucan—31, lipid—25, polysaccharide—16, amphiphile—53 and a few others. The majority of the ligands in PRRDB 2.0 (79%) have natural sources. The majority of PRRs 48 % come from humans 315 from and mice. These PRRs bind to ligands that are predominantly found in bacteria (52%) and viruses (15%), as well as fungi (6%). These are represented in Figure 3.4.

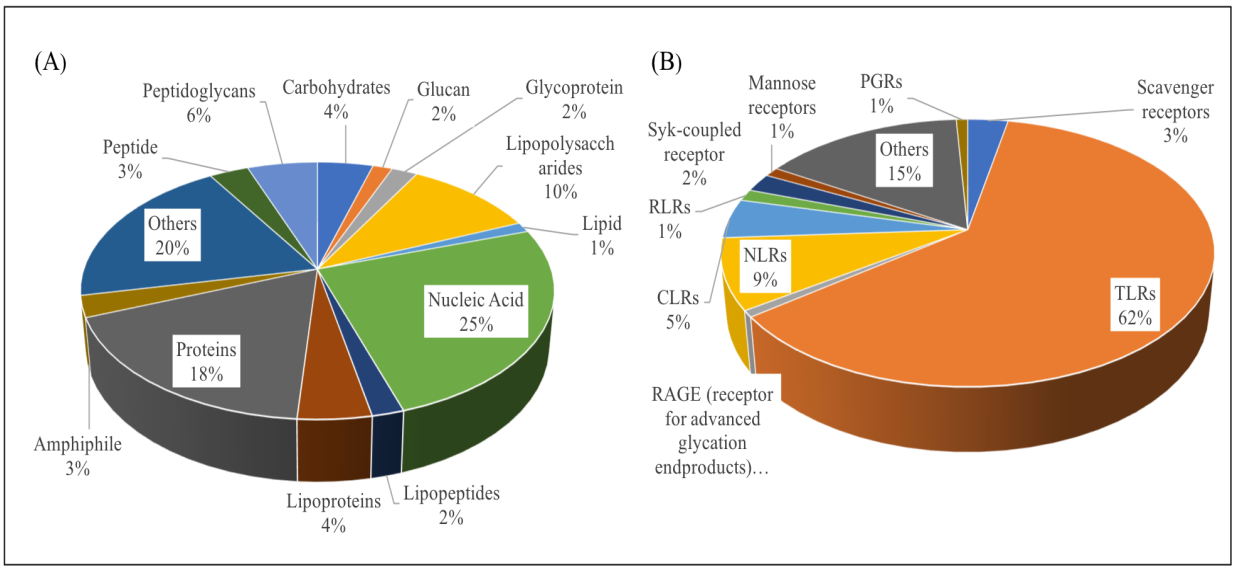


Figure 3.3 Representation of percentage distribution of different type of (A) Ligands (B) PRRs available in PRRDB2.0

* TLR: Toll like receptors, * NLRs: Nucleotide- binding oligomerization domain (NOD) like receptors, * PGRs : Peptidoglycan recognition proteins,* RLRs: Retinoic acid inducible gene-I like receptors, * CLRs: C-type lectin receptors, * RAGE: Receptors for advanced glycation end products.

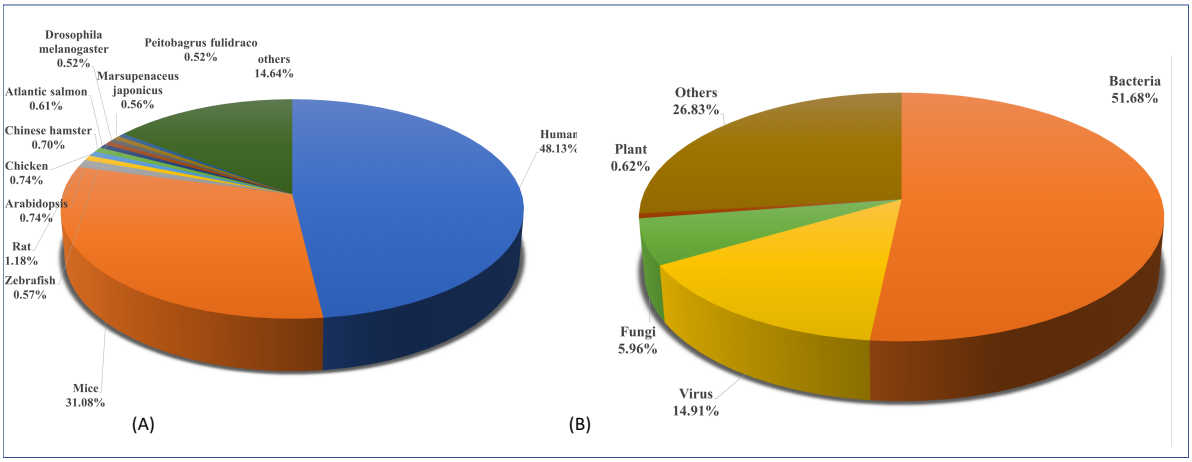


Figure 3.4 Representation of percentage distribution of sources of (A) PRRs (B) Ligands.

3.3.2 PRRDB and PRRDB2.0 Comparison

PRRDB was created in 2008 and consists of two tables; one of the PRRs, which contains 491 entries, contains information such as the receptor's name, source organism, sequence and

length, family, and type. Another table contains 266 items for ligands, including information such as name, source, ligand class, origin, and receptor (Lata and Raghava 2008a). We have added more information about each PRR and ligand in the updated version, resulting in a total of 2740 entries. We have provided more information about each PRR and ligand in the updated version, resulting in a total of 2740 entries. In addition to its name, source, type, and origin, we attempted to describe the role of ligand in immune system activation. Similarly, the new version includes detailed information about each PRR, including its name, source, type, sequence and length, localization and domain, and function. Table 3.1 shows a comparison of statistics. It just displays the most recent database change. In the new version, the experimental protocol or assay is also curated. Furthermore, for maximum information, data has been connected with Swiss-Prot, PubChem, and PDB.

Table 3.1 Represents the overall comparison between data statistics of PRRDB and PRRDB2.0

Field/Information	PRRDB	PRRDB 2.0
Total no. receptors	353	2727
Total no. ligands	354	2550
Total no. of sequence of receptors	221	1784
Total no. of Sequence of ligands	241	1583

Total no. of receptors has been increased to 2727 from 353 and total no. of ligands has been increased to 2550 from 354. Their no. of sequences has also been updated from 221 to 1784 and 241 to 1584 for receptors and ligands respectively. Majority of the PRRs available in PRRDB2.0 are TLR, CLR, NLR and Mannose based on their numbers as shown in Table 3.2. Information about TLR has been improved to 1737 from 185, for CLR its 135 from 27. NLR and mannose has been updated from 15 to 241 and 26 to 33 respectively.

Table 3.2 Represents the comparison of major types of PRRs available in PRRDB and PRRDB2.0

Receptors	PRRDB	PRRDB 2.0
TLR*	185	1737
CLR*	27	135
NLR*	15	241
Mannose	26	33

Table 3.3 Represents the comparison of entries for major types of ligands and entries for major sources of receptors available in PRRDB and PRRDB2.0

Entries for major types of ligands			**Entries for major sources of receptors**		
Field/Information	PRRDB	PRRDB 2.0	Field/Information	PRRDB	PRRDB 2.0
Peptide	15	62	Human	146	1092
Nucleic acid	68	496	Mice	102	717
PAMP	54	376	Chicken	0	17
DAMP	0	247	Hamster	15	16
Protein	60	353	Rat	3	27
LPS	16	207	Zebrafish	0	13
Peptidoglycan	8	111	*Arabidopsis*	1	17
Carbohydrates	37	88			

3.4 Implementation of web-resource

On a single platform, PRRDB 2.0 may be used to obtain comprehensive information on any PRR. For example, if a user wants to learn more about TLR 3, which identifies lipopolysaccharides associated with gram-negative bacteria, they should put TLR3 into the search box on the basic search page and look up the receptor's name, as shown in Figure 3.5

A. As illustrated in Figure 3.5 B, a single click on the search button will lead to a list of 400 items saved in PRRDB 2.0, each of which is distinguished by a unique ID. Each ID will take you to a thorough display page with all of the information you need, as well as connections to PubChem, PubMed, and Swiss-Prot. In addition, as shown in Figure 3.5 C, the TLR3 sequence is also accessible in FASTA format. In addition, PRRDB 2.0 allows users to explore all 27 experimentally verified structures for TLR3 and its complexes that are kept in the PDB.

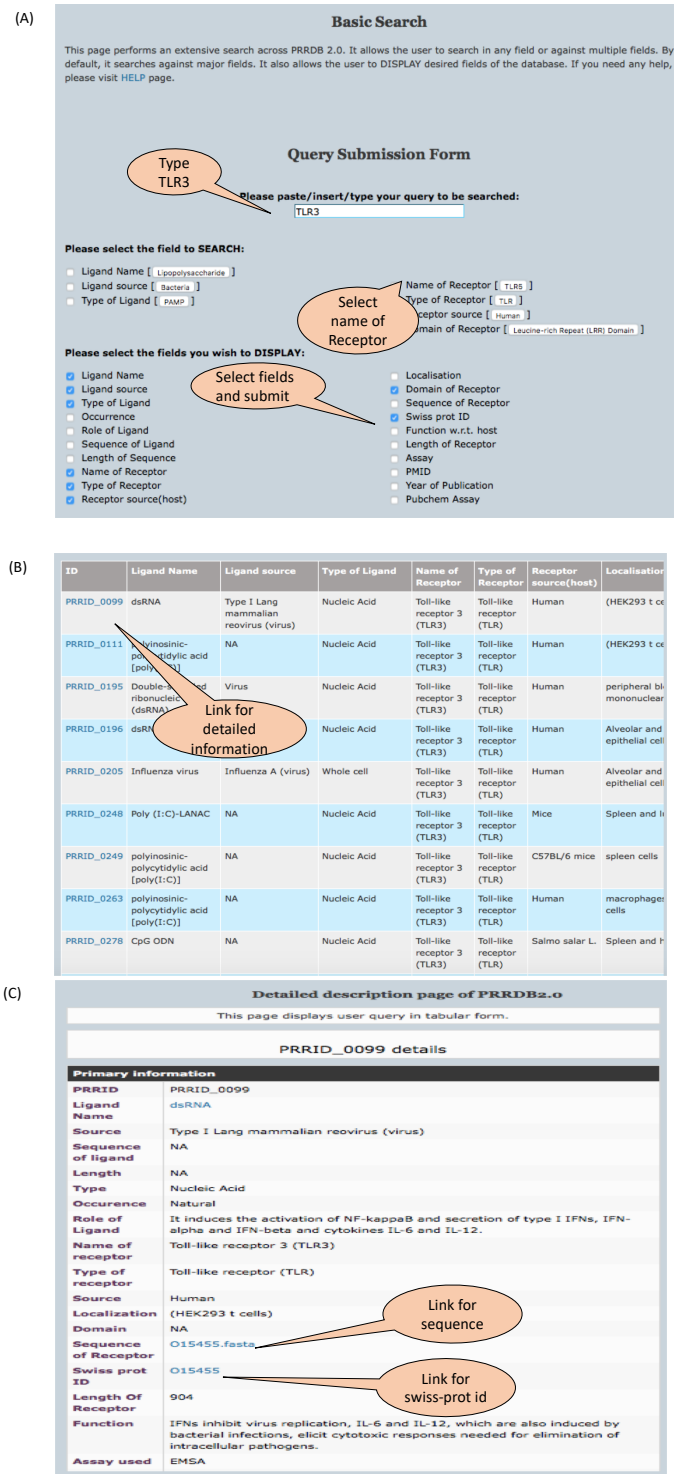


Figure 3.5 Representation of the screenshots of PRRDB2.0 demonstrating the (A) submission of query in basic search page (B) result page after submission of the query in result page (C) detailed information of the result page.

3.5 Conclusion and Summary

In the early stages of infection, the innate immune system is responsible for pathogen identification and elicitation of proinflammatory responses against invading pathogens, whereas the adaptive immune system kills the pathogen and builds immunological memory in the late stages. As stated in the introduction, PRRs have a large repertoire that detects a variety of pathogens. The fact that host PRRs recognize a wide spectrum of microorganisms in different life cycles and with varied metabolic compositions is remarkable. Another astonishing fact is that all classes of pathogens are sensed by more than one type of PRRs through various ligands and lead to a rapid proinflammatory response through various intracellular signal cascades (Mogensen 2009). Despite huge advances in innate immune-related research over the previous few decades, there is still a lot of ambiguity. TLRs are the most well-studied PRRs, although cytoplasmic PRRs also play an important role in the accumulation of diverse immunological responses, which requires further research. Other PRRs, such as mannose receptors, scavenger receptors, and a few secreted PRRs, also require additional investigation. A greater understanding of cross-talk between various PRRs is required. PRRDB 2.0, we hope, will help in the retrieval of all previously found data and queries. PRRDB 2.0, with over 2700 entries, provides improved coverage of all PRRs and their ligands ever investigated. We expanded the data set by adding new areas and emphasizing the importance and specificity of PRRs and ligands in activating the immune response. The hyperlinking of Swiss-Prot, PDB, and PubChem will deliver the most information in one location. The new edition, we feel, will be extremely beneficial to the scientific community.

4

Annotation of Pattern Recognition Receptors

4.1 Background

"Pattern recognition receptors" (PRRs) are the proteins that are germline encoded and bind to pathogen associated molecular patterns (PAMPs) and damage associated molecular patterns (DAMPs) to detect invading infections. PRRs recognize DAMPs that are chemical produced by injured cells only. The recognition triggers a cascade of signaling events which leads to microbicidal and pro- inflammatory response downstream. This act as a important link between innate and adaptive immune response (Mogensen 2009). The major families of PRRs include Toll-like receptors (TLRs), nucleotide-binding oligomerization domain (NOD)-Leucine Rich Repeats (LRR)-containing receptors (NLR), retinoic acid-inducible gene 1 (RIG-1)-like receptors (RLR), and C-type lectin receptors (CLRs). TLRs and CLRs are transmembrane proteins whereas, NLRs and RLRs are cytoplasmic protein receptors. The PRRs are essential for the identification of virus, bacteria and fungus (Kawai and Akira 2009). Phagocytosis is triggered by some of the specialized PRRs like scavenger receptors, mannose receptors, and glucan receptors. The other category is secreted PRRs that consist of colectins, ficolins serum amyloid, lipid transferases, pentraxins and peptidoglycan recognition proteins (PGRs) (Dilraj Kaur et al. 2019a).

Multiple previous studies show the importance of PRRs in various diseases like heart failure (Farrugia and Baron 2017) cancer (O' Donovan, Mao, and Mele 2019; do Prado et al. 2019; S. Qin et al. 2019; Haider et al. 2019), autoimmune disorders (Farrugia and Baron 2017; V. Kumar 2019), kidney disease (Komada and Muruve 2019), asthma, atherosclerosis, sepsis (Lin, Verma, and Hodgkinson 2012), Parkinson's disease , immunodeficiency disorders like chronic granulomatous disease (CGD), and "X-linked agammaglobulinemia (XLA)" (Mortaz et al. 2017). Thus, PRRs have seems to have a vital role in the therapeutic research mainly in adjuvant designing (Olive 2012; Shirota, Tross, and Klinman 2015; Dowling and Mansell 2016; Garlapati et al. 2009). Therefore, it is indispensable to have a profound understanding of biological machinery and functional role of PRRs in our immune system. Usually, PAMPs and DAMPs recognised by their PRRs and this begins the recruitment of leukocyte (Mogensen 2009). Innate immune cells like macrophages, dendritic cells, monocytes and mast cells, whereas epithelial cells and fibroblasts are non-immune cells that express PRR (D. Tang et al. 2012). A cascade of downstream signalling is triggered by pattern recognition receptor-ligand interaction and their combined conformational modifications. As a result of this cascade, transcriptional and post-translational alterations occur (Mogensen 2009). The conventional

approaches to identify PRRs consists of various experimental techniques like Quantitative real-time PCR (Kaiser et al. 2013), immunofluorescence (D'Souza et al. 2013), Cell viability assay, Immunoblot and Immuno-precipitation (Kennedy et al. 2004), PAMP binding assay (S. Jiang et al. 2017; C. Yang et al. 2017) , ELISA (P. Yang et al. 2010; Miao et al. 2010; Pohlmann et al. 2003), Growth-inhibition assay (Krol et al. 2010) and Microbial Binding and Agglutination Assay (S. Jiang et al. 2017).

These experimental approaches are extremely precise, but they are also expensive and time-consuming. Recent technological advancements have resulted in the creation of several in-silico methodologies for predicting a protein's function. These approaches are not only speedier and less costly, but they are also repeatable. Data for such prediction systems may be found in a variety of web-based sites, databases, and repositories such as IEDB (Dhanda et al. 2019), VAXJO (Sayers et al. 2012), IIDB (Korb et al. 2008), InnateDB (Breuer et al. 2013) and VIOLIN (Xiang et al. 2008). PRR prediction is necessary to facilitate research and efficient therapeutic design because of its key role in innate immunity. Only one prediction technique (Gao et al. 2012c) for PRR sub-family classification has been established previously, based on data acquired from the PRRDB (Lata and Raghava 2008a). Due to a lack of data, this approach adopted a more liberal dataset preparation requirement (CD hit at 90% threshold). Following that, this dataset was utilised to train and evaluate machine learning models. The model prediction results might be biased since their final dataset comprises sequences which are homologous.

We devised a technique employing the biggest available dataset, obtained from the PRRDB 2.0 (Dilraj Kaur et al. 2019a) database, with standard procedures, to complement and overcome the constraints of the existing approach. Without lowering the number of sequences in the dataset, we employed techniques that divided the data into five data sets in such a manner that no two proteins in two separate subsets had more than 40% sequence similarity (Bendtsen et al. 2004; Garg and Raghava 2008a). We examine the performance of BLAST on our dataset to better perceive the strengths and limitations of the typical similarity-based approach. We constructed conventional machine-learning-based classification models for predicting PRRs utilising a variety of descriptors such as residue composition and dipeptide composition in the second step (M. Kumar, Gromiha, and Raghava 2007). It has already been demonstrated that evolutionary data delivers more information than a single sequence (H. Kaur and Raghava 2004; Garg and Raghava 2008b). Thus, we created models based on evolutionary data such as

the composition of the position-specific scoring matrix (PSSM) profile (M. Kumar, Gromiha, and Raghava 2007). Finally, we created hybrid models that incorporate the strengths of the various methods employed in this research (Garg and Raghava 2008b; Bhasin and Raghava 2004). We demonstrate that the hybrid model, which combines BLAST-based similarity search with a PSSM profile-based Random Forest (RF) classifier, is the most effective *in-silico* PRR prediction approach. To support and encourage further study on PRRs, this model is freely accessible for public usage in the form of the web-server "PRRpred" (http://webs.iiitd.edu.in/raghava/prrpred/).

4.2 Methodology

4.2.1 Data Extraction and Pre-processing

The sequences of PRRs (positive data) were retrieved from the PRRDB2.0 database (Lata and Raghava 2008a). The total number of PRRs captured was 2,727 at first, but after removing identical sequences, the number of unique PRRs was decreased to 179. The negative dataset was produced by gathering random sequences that were not PRRs from Swiss-Prot (The UniProt Consortium 2017). There were 274 Non-PRR sequences in the negative sample. We utilised a technique previously published by (Bendtsen et al. 2004) and (Garg and Raghava 2008a) to produce subsets that are non-redundant without lowering the amount of sequences. We have employed a threshold of 40 % sequence similarity using “CD-HIT” on both positive (PRRs) and negative (Non-PRRs) datasets to obtain clusters. Based on the cut-off, each cluster is a group of identical sequences. There were 106 clusters from positive data and 210 clusters from negative datasets found in total. Figure 4.1 depicts the distribution of sequences in the clusters. There were 100 clusters from the positive or PRRs dataset having less than three protein sequences, whereas there were 200 clusters from the negative/ Non-PRRs dataset having less than three protein sequences. Likewise, there were 5 clusters from both positive and negative datasets having no. of protein sequences from 4-6. Five subsets were constructed from the CD-HIT clusters for the positive dataset. The first cluster's sequences were assigned to the first subset, the second cluster's sequences to the second subset, and so on. This method was repeated until all sequences (included in CD-HIT produced clusters) were dispersed evenly among the five subsets as shown in Figure 4.2. A similar procedure was used to produce five negative subsets from the negative dataset. This technique ensures that the subsets are distinct to one another (no more than 40% similarity between sequences in two subsets), which is

advantageous for unbiased machine learning model training and testing, as well as the selection of a superior classification model. The goal of this procedure is to construct a non-redundant dataset while retaining the same number of proteins (Bendtsen et al. 2004; Garg and Raghava 2008a).

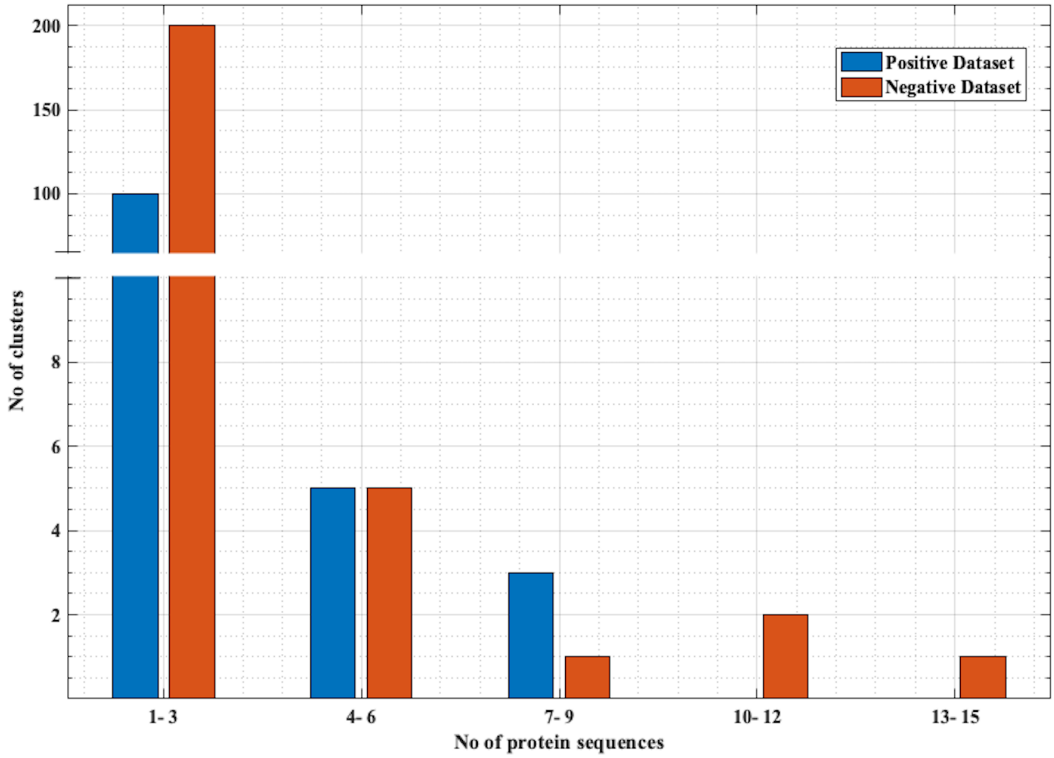


Figure 4.1 Sequence distribution in 'CD-HIT' clusters generated from positive/PRRs and negative/Non-PRRs dataset. The x-axis shows the number of sequences, while the y-axis reflects the number of clusters that include those sequences.

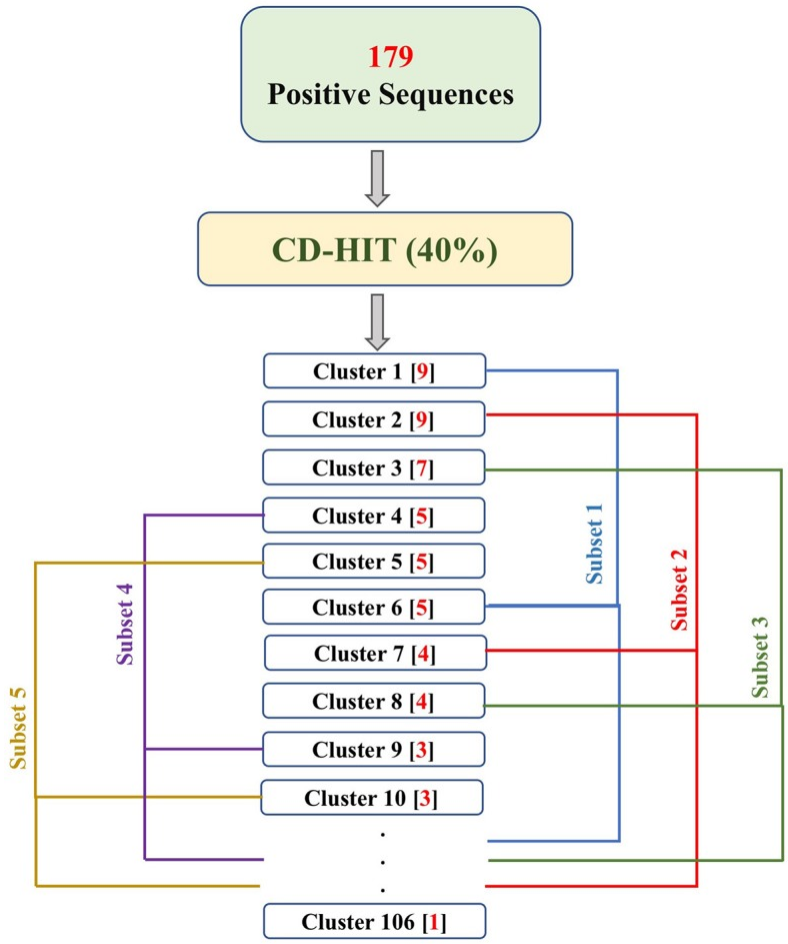


Figure 4.2 The flowchart explains the process of fractioning positive clusters obtained from CD-HIT into five subsets. The numbers in the parentheses, following the cluster names, represent the number of sequences in that cluster. As a result, Subset 1 contains sequences of clusters 1, 6, 11, …, 106; Subset 2 contains sequences of cluster 2, 7, 12, …, 102; Subset 3 contains sequences of cluster 3, 8, 13, …, 103; Subset 4 contains sequences of cluster 4, 9, 14, …, 104; and Subset 5 contains sequences of cluster 5, 10, 15, …, 105.

4.2.2 Similarity Search (BLAST)

Based on "pBLAST (BLAST+ 2.7.1)", a similarity search module was created (Camacho et al. 2009). Five-fold cross-validation was used to test this module's performance. For this, a train set was used to make a local database against which the query sequences (sequences in the test set) were searched at an *e*-value of 0.001. The approach is done five times (once for each training and test set), with the evaluation metrics recorded each time (Results). Finally, the whole positive (179 PRRs) and negative dataset (274 Non-PRRs) have been integrated in the web-server implementation to provide a repository/database of 453 proteins against which the user's unseen query protein may be searched.

4.2.3 Features Extraction

4.2.3.1 Composition Based

“Pfeature” was used to extract amino acid composition (AAC) and di-peptide composition (DPC), which were employed as features to offer residue information for a protein. AAC is a 20-length vector for a protein sequence, with each element representing the proportion of a given kind of residue in the sequence. DPC, on the other hand, is a 400-length vector that specifies the amino-acid composition of pairings of amino acids in the protein sequence (e.g., ‘L-M’, ‘G-L’, and so on). “Pfeature” can provide you with further details (Pande et al. 2019).

4.2.3.2 Evolutionary Information Based

Using PSI-BLAST, we were able to collect evolutionary information for a protein in this study. Similar to previous research, we included evolutionary information in the form of a ‘PSSM-400’ composition profile as a feature (M. Kumar, Gromiha, and Raghava 2007; Kaundal and Raghava 2009; Zhang, Liu, and Tramontano 2017; Verma, Varshney, and Raghava 2010; M. Kumar, Gromiha, and Raghava 2011). PSSM-400 is a 20 x 20 dimensional vector that represents the composition of occurrences of each of the 20 amino acids that correspond to each amino acid type in the protein sequence. “Pfeature's” (Pande et al. 2019) "Evolutionary Info" module was used to construct a PSSM matrix for each protein sequence, which was then normalised and transformed to a 20 × 20 ‘PSSM’ composition vector.

4.2.4 Machine Learning Based Models

To create prediction models, we utilised the sklearn package from Sci-Kit, which has a variety of classifiers. Each of these strategies necessitates the use of feature vectors with a predetermined length. The most important information regarding variable-length proteins was transformed into constant vectors of similar dimensions (‘AAC, DPC, PSSM-400’), which were then employed as input characteristics. To acquire the greatest performance on the training set, we utilised Sci-GridSearch Kit's module to tune hyper-parameters. Eventually, the best model was used for the test dataset. Five-fold cross-validation was used to accomplish this process, and the average performance of five-folds was evaluated. The prediction models were then developed using a variety of Machine Learning-based classifiers. To handle linear data, the most basic classifier, Logistic Regression (LR), was employed, while for non-linear data, sophisticated classifiers like Random Forest (RF), Support Vector Machine (SVM), Extra

Trees (ET), K-Nearest Neighbor (KNN), and Multi-Layer Perceptron (MLP) were utilised. Many bioinformatics research have effectively used all of these machine learning approaches (Nagpal et al. 2017; Chauhan, Mishra, and Raghava 2010; Chaudhary, Kumar, et al. 2016; Piyush Agrawal et al. 2019; Laurie and Goss 2013).

4.2.5 Cross Validation Techniques

Using the five-fold cross-validation approach, the performance of the modules built in this study was assessed. Positive and negative subsets were used to create training and test sets. The training set was created by combining four positive and four negative subgroups. The test set was created by combining the remaining one positive and one negative subsets. This method is done five times, with the result that the combination of a positive subset and its matching negative subset is only utilised as a test set once. As stated in the following sections, we used these five training and test sets to perform five-fold cross-validation to pick the best machine learning models and to construct a BLAST similarity search-based module. Five-fold cross-validation is a typical procedure that has previously been used effectively in a number of machine learning studies (Nagpal et al. 2017; Chauhan, Mishra, and Raghava 2010; Harinder Singh et al. 2016, 2015; Chaudhary, Kumar, et al. 2016; Piyush Agrawal et al. 2019).

4.2.6 Evaluation Parameters and Hybrid Models

Threshold independent and dependent score factors were utilised to assess each model employed in the study. Sensitivity (Sens), Specificity (Spec), Accuracy (Acc), and Matthew's correlation coefficient are the threshold dependent parameters employed here (MCC). “Sens” is defined as true positive rate (TPR) i.e., correctly predicted positives with respect to actual total positives, whereas true negative rate (TNR) is defined by “Spec.” “Acc” is the ability of the model to differentiate between true positives and true negatives, while MCC is the correlation coefficient between predicted and actual classes. Following relations were used to calculate these:

$$Sens = \frac{TP}{P} X100$$

$$Spec = \frac{TN}{N} X100$$

$$ACC = \frac{TP + TN}{P + N} X100$$

$$MCC = \frac{dTPXTN - FPXFN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

Where TP stands for correctly predicted positives and TN stands for correctly predicted negatives. The total sequences in the positive set are denoted by P, whereas the total sequences in the negative set are denoted by N. FP stands for real negative sequences that were incorrectly predicted as positive, whereas FN stands for incorrectly predicted positive sequences. These score parameters are well-known and have been used in several research to assess model performance. The value of the "Area Under Receiver Operating Characteristic Curve" (AUROC) is a threshold independent parameter derived by plotting the True Positive Rate (TPR or Sens) against the False Positive Rate (FPR) (FPR or 1-Spec) (Vinod Kumar et al. 2018). Hybrid models were created that integrated the BLAST prediction score with the ML-based scores, as done in Algpred (Saha and Raghava 2006), to increase the accuracy of the machine learning-based models even further. Positive prediction (PRRs) received a score of "+0.5," negative prediction (Non-PRRs) received a score of "0.5," and no hits received a score of "0." (NH). This score was added to the machine learning model's score (i.e., prediction probability of positive class). In a method called five-fold cross-validation method, this is done for each of the sequences in the test set. The scoring metrics for each ML model were then assessed at various probability cut-offs based on this combined score. Figure 4.3 depicts the workflow of the study.

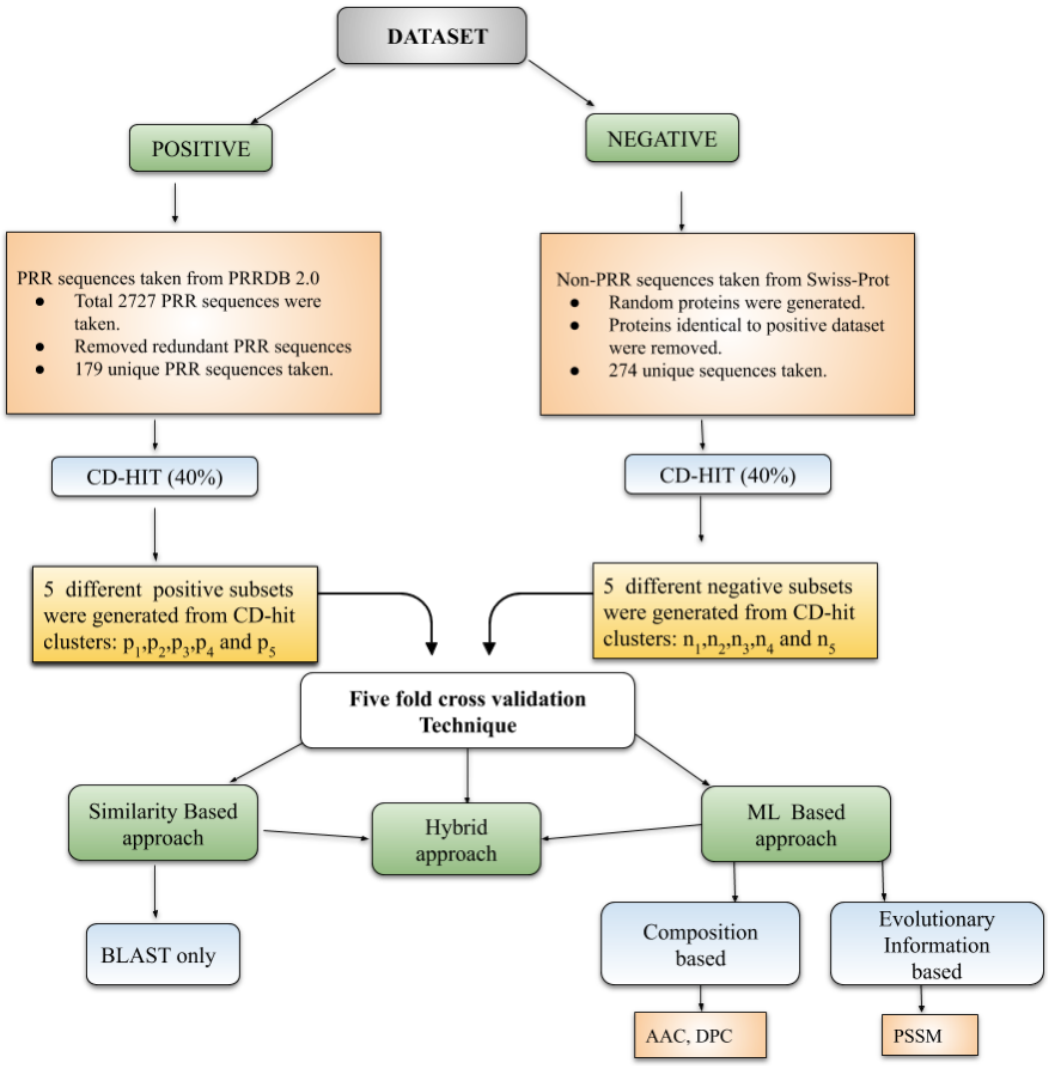


Figure 4.3 The pipeline of the study

4.3 Results

4.3.1 Similarity Search Based Prediction

BLAST is a popular software application that is frequently used for similarity searches. As a result, we employed BLAST to distinguish between PRRs and Non-PRRs. We utilized five-fold cross-validation to avoid bias, which involved searching proteins in the test set against the training set using BLAST at varying e-value cut-offs (Table 4.1). To cover all of the proteins in our training sets, this method is performed five times. There are 179 PRRs in the positive dataset and 274 Non-PRRs in the negative dataset. Table 4.1 shows that the number of correctly predicted PRRs jumped from 74.30 to 82.12 percent when the "e-value" was reduced from 10^{-}

^{9} to 10^{-0} or 1. While sensitivity increases with an increase in e- value, the rate of error i,e percentage of Non-PRRs also grew up. Specificity improved from 32.48 to 49.68 percent in Non-PRRs, while the error rate went from 1.67 to 10.05 percent, with an e-value of 10^{-9} to 10^{-0}. Due to a huge number of no-hits, the overall accuracy of BLAST was only approximately 51% at e-value 10^{-3}. This low result demonstrates that BLAST is ineffective in distinguishing PRRs from non-PRRs with high accuracy.

Table 4.1 shows performance of BLAST on both datasets using five-fold cross validation on different e-values.

Blast (e-value)	Positive hits (PRRs)		Negative hits (Non-PRRs)	
	PRRs (Sens)	**Non-PRRs (Error)**	**Non-PRRs (Sens)**	**PRRs (Error)**
10^{-9}	133 (74.30)	4 (1.45)	89 (32.48)	3 (1.67)
10^{-8}	134 (74.86)	4 (1.45)	90 (32.84)	4 (2.23)
10^{-7}	134 (74.86)	5 (1.82)	90 (32.84)	4 (2.23)
10^{-6}	135 (75.41)	5 (1.82)	93 (33.94)	4 (2.23)
10^{-5}	136 (75.97)	7 (2.55)	98 (35.76)	5 (2.79)
10^{-4}	136 (75.97)	7 (2.55)	99 (36.13)	6 (3.35)
10^{-3}	**138 (77.09)**	**8 (2.92)**	**101 (36.86)**	**6 (3.35)**
10^{-2}	139 (77.65)	10 (3.64)	102 (37.22)	6 (3.35)
10^{-1}	140 (78.21)	20 (7.29)	107 (39.05)	7 (3.91)
1	147 (82.12)	65 (23.72)	135 (49.27)	18 (10.05)

4.3.2 Models developed using Machine learning Techniques

4.3.2.1 Sequence's Composition Based

We employed two important sequence composition-based criteria, namely (i) amino acid composition and (ii) dipeptide composition, to build a technique for classifying PRRs and Non-PRRs. Prediction models were developed using a variety of machine learning approaches (e.g., "SVM, KNN, and RF"). In both the positive and negative datasets, we investigated at the

frequency of the 20 amino acids. A study of the amino acid content of PRRs and Non-PRRs revealed that PRRs had more residues “L, N, S, and Q”, whereas Non-PRRs have more residues “A, D, E, K, and V”. (Figure 4.2). As seen in Figure 4.4, the composition of PRRs differs from the composition of Non-PRRs. As a result, models for distinguishing two classes may be developed using the amino acid composition (AAC) feature. To create binary classification models, the following machine learning approaches were used: Extra-trees (ET), Random forest (RF), Support vector machine (SVM), K nearest neighbour (KNN), Logistic regression (LR), and Multi-layer perceptron (MLP). On the training dataset, ET-based models had a maximum AUROC of 0.90 and an MCC value of 0.63, as shown in Table 4.2. On the test dataset, we got an AUROC of 0.88 and an MCC of 0.63. Models were also built utilising dipeptide composition and a variety of machine learning approaches. On the test set, LR had the best performance, with an average accuracy of 80.25 percent, MCC of 0.59, and AUROC of 0.87, while on the training dataset, it had an average accuracy of 82.57 percent, MCC of 0.64, and AUROC of 0.88. In the case of LR, overall test accuracy was 83 percent, with MCC of 0.64 and AUROC of 0.88.

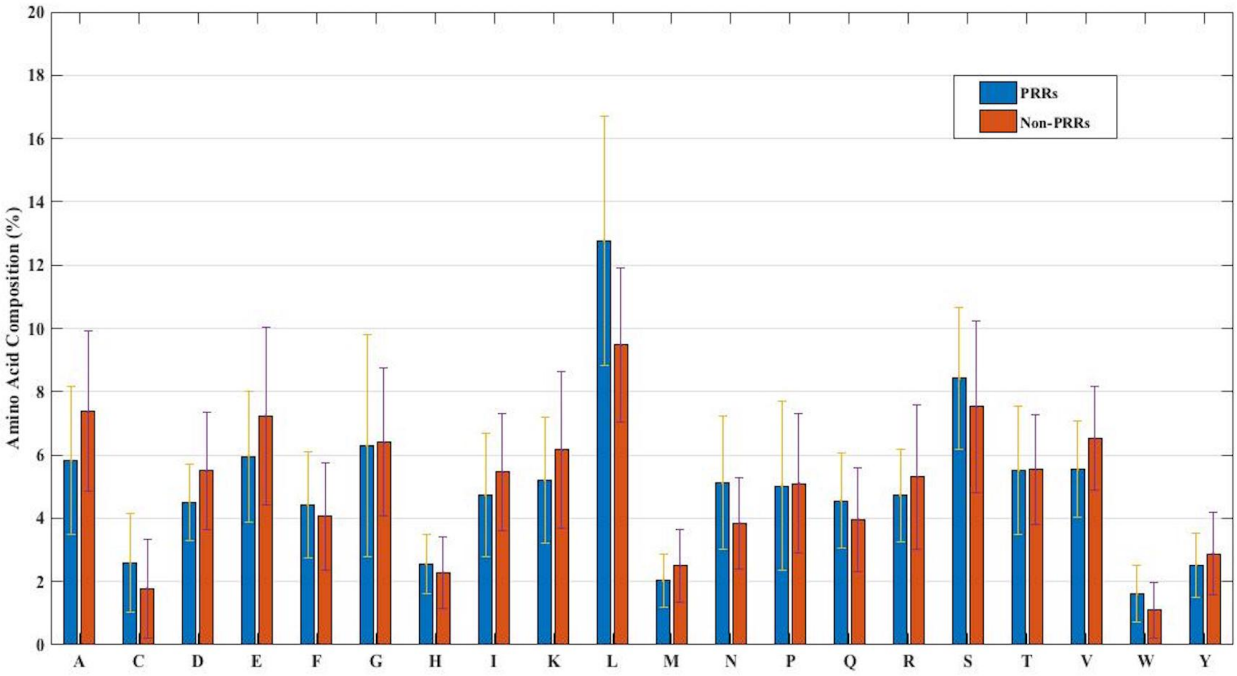


Figure 4.4 The figure represents the amino acid composition in percentage of PRRs and Non-PRRs

Table 4.2 The performance of models based on different machine learning techniques for positive dataset (PRRs) created using Amino acid composition of protein sequences.

Method		Train Dataset (Average)					Test Dataset (Average)				
Model	**Hyper-parameters**	**Sens**	**Spec**	**Acc**	**AUROC**	**MCC**	**Sens**	**Spec**	**Acc**	**AUROC**	**MCC**
ET	**ne=90**	80.71	82.56	81.73	0.90	0.63	77.06	84.08	82.46	0.88	0.63
SVM	**C=5, g=0.01, k=rbf**	78.07	83.83	81.62	0.87	0.62	77.95	82.31	81.06	0.88	0.60
RF	**ne=100**	77.82	81.46	80.08	0.88	0.59	77.42	80.85	79.97	0.87	0.58
LR	**C=1**	77.98	82.50	80.77	0.86	0.60	76.12	81.57	79.57	0.86	0.58
MLP	**a=tanh, HL=(19,), m=200, s=adam**	77.02	82.77	80.50	0.86	0.59	78.88	77.94	78.90	0.87	0.57
KNN	**al=ball_tree, nn=20, w=distance**	76.17	79.06	77.91	0.85	0.55	77.74	75.00	76.97	0.86	0.53

*g=gamma, ne=n_estimators, k=kernel, a=activation, HL=hidden layer size, s=solver, al=algorithm, w=weight, m=max_iter and nn=n_neighbours.

4.3.2.2 Evolutionary Information Based

The sequence profile has already been proven to give more information than a single sequence. As a result, in our work, we first use PSI-BLAST software to construct a sequence profile matching to a protein. We compute the composition of a sequence profile or PSSM profile in order to create a fixed number of features (see section Materials and Methods). PSSM-400, a fixed-length vector of 400 components, is used to express the PSSM profile's composition. 'PSSM-400', which contains a fixed-length vector of 400 elements, is used to represent the PSSM profile composition. For our dataset, we generated PSSM-400 composition profiles and utilised them as feature vectors to develop classification models. Similarly to the AAC and

DPC-based approaches, we employ numerous classifiers for training and testing, such as 'SVM, RF, ET, MLP', and others. On the training dataset, models based on evolutionary information had an AUROC of 0.87 and an MCC of 0.64, as shown in Table 4.3. Similarly, the maximum AUROC on the test dataset was 0.89, with MCC of 0.66. In terms of MCC, the 'PSSM'-based prediction model outperformed the composition-based prediction models. In terms of AUROC, both composition and PSSM-based techniques performed almost identical.

Table 4.3 The performance of models based on different machine learning techniques for positive dataset (PRRs) developed using evolutionary information ("PSSM-400") of protein sequences.

Method		**Train Dataset (average)**					**Test Dataset (average)**				
Model	**Hyper-parameters**	**Sens**	**Spec**	**Acc**	**AUROC**	**MCC**	**Sens**	**Spec**	**Acc**	**AUROC**	**MCC**
SVM	**C=10, g=0.5, k=rbf**	77.80	85.89	82.78	0.87	0.64	79.74	85.46	83.64	0.89	0.66
LR	**C=1000**	77.31	86.37	82.84	0.87	0.64	80.80	81.07	81.13	0.89	0.61
KNN	**al=ball_tree, nn =6, w=distance**	72.80	83.48	79.36	0.86	0.57	78.40	82.50	81.07	0.87	0.60
RF	**ne=80**	75.95	85.01	81.55	0.87	0.61	79.07	81.41	80.74	0.86	0.60
MLP	**a=logistic, HL=(14,), m=200, s=adam**	75.26	85.09	81.28	0.86	0.61	79.07	81.03	80.26	0.88	0.59
ET	**ne=70**	80.33	78.79	79.36	0.88	0.58	83.73	74.97	79.15	0.87	0.59

*g=gamma, ne=n_estimators, k=kernel, a=activation, HL=hidden layer size, s=solver, al=algorithm, w=weight, m=max_iter and nn=n_neighbours.

4.3.2.3 Combination of Composition and evolutionary information

The PSSM-400 and amino acid compositions were combined to create a 420-length feature vector. Several classifiers were trained and tested using five-fold cross-validation. Using LR on training sets, we obtained an AUROC value of 0.89 and an MCC value of 0.66, as shown in Table 4.4. Similarly, MLP gave the highest AUROC with an AUROC of 0.90 on the training dataset and 0.67 on the test dataset. As a result, as compared to employing evolutionary information-based features (PSSM) or composition-based features (AAC or DPC) alone, performance has improved. In Table 4.4 comparison with (Gao et al. 2012c) has been also done. The model accuracy in (Gao et al. 2012c) was claimed to be 97–98%; nevertheless, such a relaxed redundancy reduction technique utilises sequences that can be highly similar to a high degree. Figure 4.5 depicts the ROC curves for various classifiers such as AAC, PSSM, and the combination of AAC and PSSM.

Table 4.4 On the PRR dataset, the performance of multiple machine learning techniques-based models constructed by integrating composition (AAC) and evolutionary information (PSSM-400) based features for protein sequences. Comparison with Gao, et.al is shown in the last row.

Method		**Train Dataset (Average)**					**Test Dataset (Average)**				
Model	**Hyper-parameters**	**Sens**	**Spec**	**Acc**	**AUROC**	**MCC**	**Sens**	**Spec**	**Acc**	**AUROC**	**MCC**
MLP	**a=tanh, HL,=(70,), m=200, s=adam**	77.70	86.54	83.20	0.88	0.65	81.23	85.50	84.19	0.90	0.67
LR	**C=1000**	82.97	83.49	83.34	0.89	0.66	83.59	81.49	82.67	0.90	0.64
RF	**ne =60**	80.32	83.24	82.16	0.88	0.63	80.43	82.44	82.16	0.87	0.63
ET	**ne =100**	77.72	85.25	82.35	0.89	0.63	78.96	83.75	82.13	0.88	0.63
SVC	**C=5, g=0.01, k=rbf**	81.65	83.35	82.73	0.89	0.65	80.62	81.56	81.72	0.88	0.62

KNN	**al=ball_tree, nn =20, w=distance**	80.20	76.60	78.12	0.87	0.56	80.41	72.88	76.35	0.86	0.52
SVC	(Gao et al. 2012c)	-	-	-	-	-	-	-	97.9	-	-

*g=gamma, ne=n_estimators, k=kernel, a=activation, HL=hidden layer size, s=solver, al=algorithm, w=weight, m=max_iter and nn=n_neighbours.

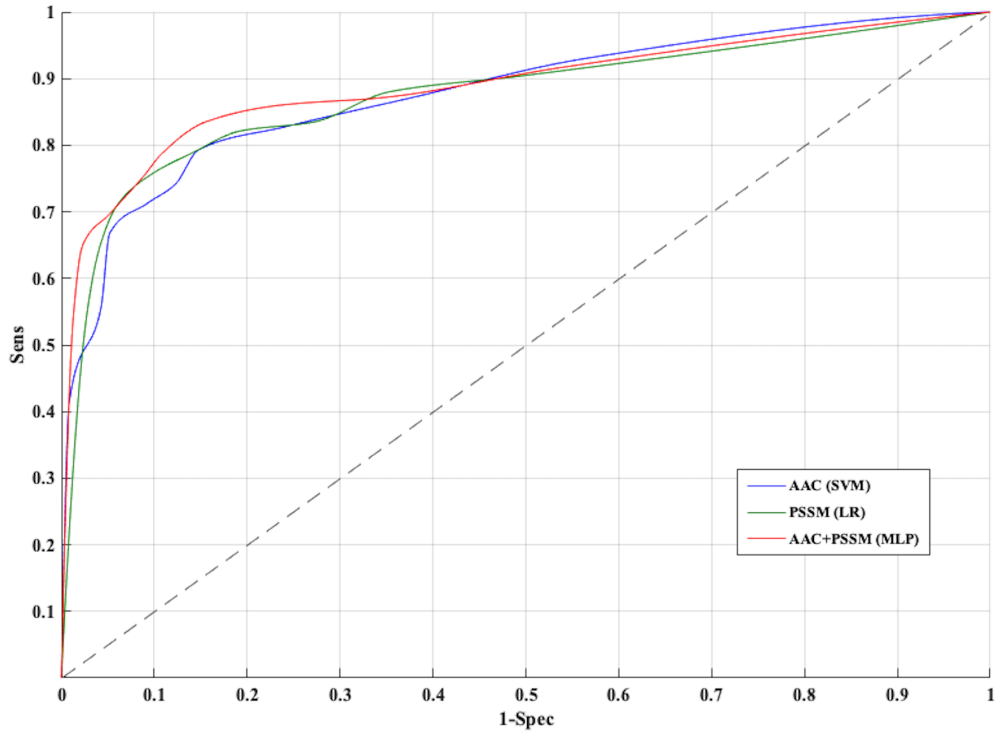


Figure 4.5 Receiver operating characteristic curves for five-fold cross-validation utilising AAC, PSSM, and AAC+PSSM employing Support vector machine (SVM), Logistic Regression (LR), and Multi-layer Perceptron (MLP), respectively.

4.3.2.4 Hybrid Models

The preceding results show that both the similarity-based approach and the machine learning-based approaches have pros and cons. As a result, we attempted to devise a strategy that combines the advantages of both approaches. Based on the hits against PRRs, the e-value of 10^{-3} was chosen for the BLAST-based similarity search approach. Since the probability of right prediction was found to be quite high (77.09 %) for this e-value, and the rate of error was very low (2.92 %). Despite the fact that the number of no-hits was too large at this cutoff (80%), it was offset by a good prediction accuracy. Proteins were initially categorised using machine learning models in order to merge the two approaches. In the second stage, the proteins were categorised again using BLAST, with the query proteins that showed similarity to PRRs at an e-value of 10^{-3} being designated as PRRs. Because of the high probability of successful prediction of the BLAST-based similarity search approach, we chose it over machine learning-based models in predicting PRRs. Simply stated, we employed machine learning techniques to categorise proteins as PRRs or Non-PRRs when there was no BLAST result for the query protein at a BLAST e-value of $10^{-3.}$ This hybrid method increased coverage that was previously lacking when using BLAST alone. As seen in Table 4.5, when BLAST was included, the performance of machine learning algorithms increased dramatically. Our best hybrid model 'RF' based on PSSM attained an accuracy of 91.39 percent, an AUROC of 0.95, and an MCC of 0.82. In general, the performance of all hybrid models was shown to be superior to that of BLAST-based similarity searches and machine learning models.

Table 4.5 Shows the performance of several machine learning techniques-based models on the test data when paired with BLAST hits at "e-value" 10^{-3}

Feature	Model	Hyper-parameters	Sens	Spec	Acc	AUROC	MCC
PSSM	**RF**	**C=80**	83.24	96.72	91.39	0.95	0.82
AAC	**RF**	**C=100**	82.12	94.53	89.62	0.92	0.78
AAC+PSSM	**ET**	**ne =100**	87.15	89.78	88.74	0.95	0.77
DPC	**SVC**	**C=2, g=0.01, k= rbf**	79.89	92.34	87.42	0.93	0.73

4.4 Implementation of Web Resource

One of the key purposes of this research is to serve the scientific community. We created a simple web server "(http://webs.iiitd.edu.in/raghava/prrpred/)" that allows users to predict whether or not a particular protein is a pattern recognition receptor. Under prediction, the server's web interface contains two sub-modules: (i) Composition Based and (ii) Evolutionary Information Based. The "Composition Based" module enables a user to find a protein sequence based on amino acid composition. This module also gives the user the choice of using the non-hybrid technique, which is solely AAC-based, or the hybrid method, which is AAC+BLAST-based. The "Evolutionary Information Based" module assists the user in predicting PRRs based on evolutionary information from a protein sequence. The PSSM-400 composition profile for the input protein sequence is generated and used as a feature vector for prediction in this step. This module, like the composition-based module, supports non-hybrid and hybrid models. The web server was built using a responsive HTML template to adjust to the browsing device. As a result, our web server is compatible with a broad range of devices, such as desktop computers, tablets, and smartphones.

4.5 Discussion and Conclusion

Understanding innate immunity has advanced rapidly in recent years, particularly in terms of the processes by which pathogens are detected and how signalling molecules respond to them. Because of its involvement in combating pathogens during the early phases of infection, innate immunity is earning more attention than adaptive immunity, whereas adaptive immunity enters the picture later. Adaptive immunity is made up of receptors that are extremely specific to antigens (Zhu et al. 2018). In contrast, innate immunity is made up of specialised receptors known as PRRs that sense infectious microorganisms and generate inflammatory responses to eliminate them (Pahari et al. 2017). In the past, some crucial implications of PRRs have been documented in the context of adjuvant design, therapeutic targets, immunomodulator design, cancer immunotherapy, and so on (Zhu et al. 2018; Vasou et al. 2017; Mullen, Chamberlain, and Sacre 2015). To understand innate immunity, a comprehensive database of pathogen-recognizing receptors, such as 'PRRDB' (Lata and Raghava 2008b) is required. These kind of knowledge-based tools can help researchers working on innate immunity and medication development. In addition to resources, methods for annotating freshly sequenced PRRs are required. Recently, SVM was used to construct a technique for predicting PRRs and

subfamilies (Gao et al. 2012a). This approach develops models based on amino acid and pseudo-amino acid composition ('PseAAC') utilising PRRDB datasets (Lata and Raghava 2008b). The prediction was based on 332 PRR sequences (from various families) retrieved from 473 sequences (with numerous similar 'UniProt-IDs') that were previously included in the database using "CD-HIT" at a 90% threshold. The model accuracy was claimed to be 97–98%; nevertheless, such a relaxed redundancy reduction technique utilises sequences that can be highly similar to a high degree. In this work, we developed classification models using a dataset collected from the newly upgraded version PRRDB2.0 (Dilraj Kaur et al. 2019b). In our case, the positive dataset comprises of PRR sequences with unique UniProt IDs, decreasing the duplicate data (1,784 sequences) to 179 sequences. Second, a CD hit threshold of 40% was used to separate both the negative (274 random Non-PRR sequences from swiss-prot) and positive datasets into five subsets each. This aided in lowering homology bias between the train and test datasets, resulting in more precise model training during five-fold cross-validation.

In this study, we investigated many methods for predicting PRRs. To identify PRRs from Non-PRRs, we employed several protein features such as composition-based features (AAC and DPC) and evolutionary information-based features ("PSSM") to create machine learning-based models. For the same purpose, we applied a combination of composition-based features and evolutionary information-based features. These techniques were adopted for the first time in the investigation of PRR prediction. To do this, we used a range of classifiers from Scikit-sklearn, including "SVM, RF, ET, and MLP". We started with BLAST alone categorization because it is simple and widely used. BLAST had a very high accuracy (e-value of 10^{-3}) if a hit was detected, but it was unable to predict about 80% of sequences (No-Hits) during five-fold cross-validation. As a result, we used a hybrid strategy for the problem at hand, combining ML-based approaches with BLAST. The main benefit of this technique is that proteins that could not be predicted using BLAST alone can now be predicted using ML. We used a wide range of classifiers to try this method with each of the protein-features and their combinations. The best results were obtained in the hybrid case of PSSM and BLAST. The best results were obtained in the hybrid case of PSSM and BLAST. This hybrid model's formulation was implemented in a free web-server. Using the web-server, this model will first predict the positive ("PRR") or negative ("Non-PRR") class for an unknown protein sequence, based on a BLAST search against the whole database (179 PRRs+274 Non-PRRs). If the outcome is a "No-Hit," the RF model trained on the entire set will make the prediction. The web-server is

open to the public and simple to use. We hope that the work done here will be useful for the annotation of PRRs and will help to further ongoing research in the field of innate immunity.

5

Classification of Defensins and Non-defensins

5.1 Introduction

Defensins are a class of antimicrobial peptides ("AMPs") that play an important role in the innate immune system. They are critical effector components in a host's defence against infections due to their broad-spectrum antibacterial activity (Raj and Dentino 2002; Mookherjee et al. 2020; Ting et al. 2020) Defensins are classified into two types based on their configuration: "α-defensins (α-helices)" and "β-defensins (β-sheets)". Defensins are small, cationic peptides that help phagocytes, skin, and mucosa combat germs. They are also antimicrobially active against viruses, mycoplasma, tumours, and fungus. They have an amphipathic property and use it to act on the membrane or envelope the wall (P. K. Singh et al. 1998; Semple and Dorin 2012; Prasad et al. 2019). Neutrophils and epithelial cells are important cellular secretors of these peptides, although defensins are also produced by monocytes, macrophages, dendritic cells, and lymphocytes (Solanki et al. 2021). Defensins are widely distributed throughout distinct body compartments in virtually all living species, according to earlier research; nevertheless, they appear to be enhanced in particular pathogenic body cells (Robert I Lehrer, Bevins, and Ganz 2005). Through the cells that create them, these host defence peptides contribute in the battle against bacterial, viral, and fungal infections (Robert I Lehrer, Bevins, and Ganz 2005).

Defensin peptides primarily disrupt the structure of bacterial cell membranes as part of their action mechanism, resulting in membrane permeabilization and the release of nutrients from the bacterial cell (D. Yang et al. 2009). They accomplish this by adhering to the membrane and developing damaging holes in it. Defensins are activated by a variety of stimuli (Sudheendra et al. 2015). Dendritic cells, monocytes, neutrophils, eosinophils, and epithelial cells produce and release the majority of them. Figure 5.1 shows that, in addition to antibacterial action, defensins are actively involved in a variety of immune-modulatory processes such as mitogenesis, cytokine production, and histamine release. Many developing strains of pathogens (i.e., bacteria, fungus, parasites) are resistant to current medications, notably antibiotics, in the era of drug resistance (Maryam, Usmani, and Raghava 2021; Seung, Keshavjee, and Rich 2015). This includes multidrug-resistant strains that are resistant to most of the existing drugs (Blasco, Leroy, and Fidock 2017; Bhardwaj et al. 2011; Boyanova, Markovska, and Mitov 2019). To control the treatment of drug-resistant disease strains, experts are looking for alternatives to antibiotics (Golkar, Bagasra, and Pace 2014; Othieno, Njagi, and Azegele 2020). Protein/peptide-based therapies are one of the promising antibiotic alternatives. The number of

peptide-based therapies authorised by the FDA has increased significantly during the previous two decades (Usmani, Kumar, Bhalla, et al. 2018; Usmani et al. 2017; D'Aloisio et al. 2021). Poly(2-oxazoline)s, which are synthetic analogues of host defence peptides (M. Zhou et al. 2021), as well as daptomycin, gramicidin, and colistin, are among the FDA-approved AMPs (M. Zhou et al. 2020).

AMPs are one of the most frequent groups of therapeutic peptides used to eliminate microbial infections, including drug-resistant strains (da Silva et al. 2020; Nuti et al. 2017). In the past, numerous computational resources and methods have been developed for predicting AMPs including chemically modified AMPs (Piyush Agrawal and Raghava 2018; Meher et al. 2017). In addition to AMPs, a variety of approaches for predicting peptides for killing a specific kind of bacterium have been developed, including the prediction of antibacterial, antituberculosis, antiviral, antifungal, and anti parasite peptides (Qureshi, Tandon, and Kumar 2015; Thakur, Qureshi, and Kumar 2012; Mehta et al. 2014; Lata, Sharma, and Raghava 2007; Lata, Mishra, and Raghava 2010; Usmani, Kumar, Kumar, et al. 2018; Piyush Agrawal et al. 2018). Despite the fact that these antimicrobial peptides represent a viable alternative to small-molecule-based medicines, their toxicity, half-life, and allergenicity pose significant hurdles (Sharma et al. 2020; Gupta et al. 2015; Mathur et al. 2018). As a result, there is a need to investigate a new class of AMPs, which are employed by hosts to fight themselves against infections. Because they are damage-associated molecular patterns (DAMPs) and are released in the host, these defensins offer various benefits over AMPs. As a result, they are less toxic and well tolerated by the body. Pattern recognition systems (PRRs) identify them when they occur naturally (Pouwels et al. 2014; Dilraj Kaur et al. 2019b). Several approaches for predicting defensins and their classes have been established in the past (Kumari et al. 2012; Y. Zuo et al. 2015; Y.-C. Zuo and Li 2009; Y. Zuo et al. 2019). In the section Comparison With Existing Tools, we explored the various tools.

In this study, we offer a dependable approach for predicting defensins with high accuracy. To build the biggest datasets feasible, we methodically gathered defensins, AMPs, and non-defensins from multiple sources. We attempted to investigate the differences and similarities between defensins and AMPs in this work. Defensins and AMPs showed substantial variations. As a result, we created models for distinguishing antimicrobial peptides and defensins. In addition, we created models for distinguishing between defensins and non-defensins. We created a standalone programme as well as a web server to assist the scientific community.

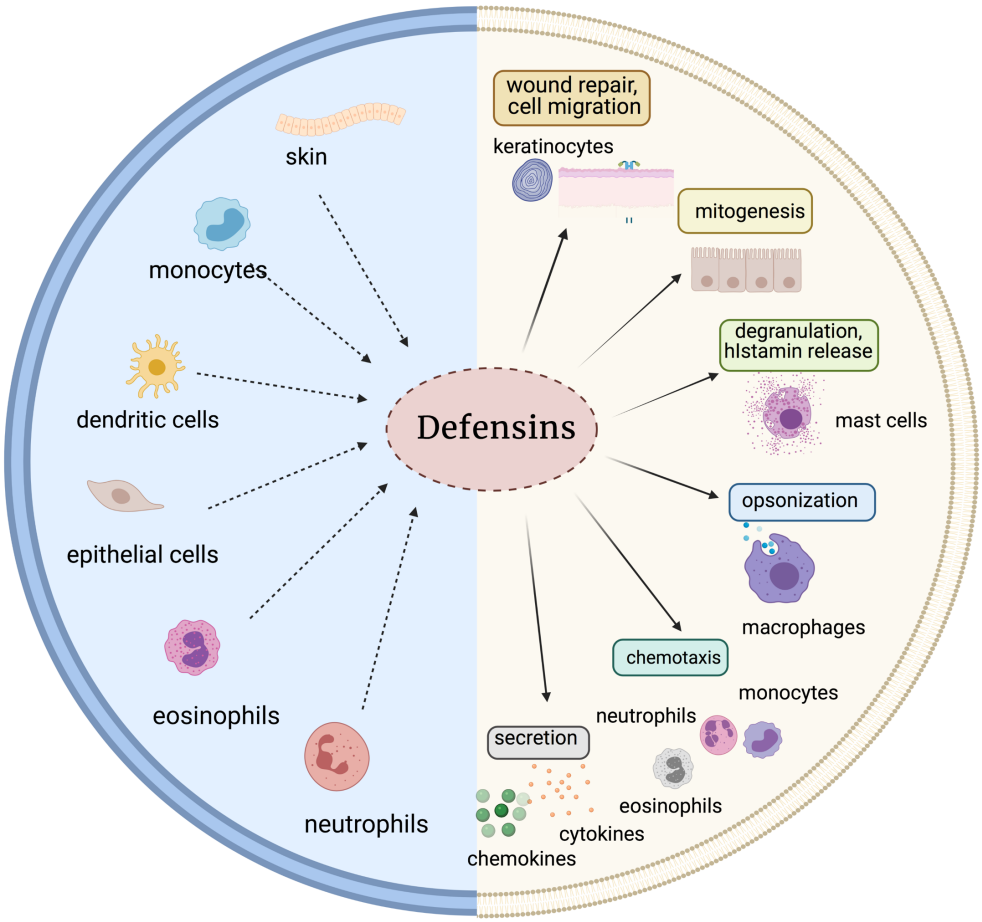


Figure 5.1 Visual representation of secretory cells and immunomodulatory function of defensins.

5.2 Methodology

5.2.1 Generation of Datasets

Defensins were collected from a variety of sources, including earlier research (Y. Zuo et al. 2015; Y.-C. Zuo and Li 2009; Y. Zuo et al. 2019), DRAMP2.0 (Kang et al. 2019), and CAMPR3 (Waghu et al. 2016). We only gathered antimicrobially active defensin sequences that have been experimentally validated. Defensins have a wide variety of lengths "(5–120 residues)", however the majority of them (77.59 percent of all sequences) contain "10–60" residues. As a result, in our study, we eliminated all defensins with less than 10 or more than

60 residues. We also filtered out sequences that had non-natural or non-standard amino acids ("B, J, O, U, X, and Z"). Finally, 1,036 distinct defensins were identified. These defensin sequences were used to generate two datasets, which are detailed further below.

(a) **Main Dataset:** Defensins are positive sequences in our primary dataset, while AMPs are negative sequences. As previously stated, we obtained 1,036 defensins from various sources. The 'CAMPR3' made accessible us with 2,297 experimentally validated AMPs. Basically, we took all peptides except those from the defensin family. The sequence lengths were limited to 10 to 60 residues, like with defensins. We also eliminated sequences that included amino acids that were not naturally occurring. In all, we have 1,036 experimentally confirmed defensins and 1,035 AMPs in our main dataset.

(b) Alternate Dataset: In our other dataset, we have both defensins and non-defensins. To find non-defensins, we used the following searches in "Swiss-Prot" (The UniProt Consortium 2017) "Non-AMPs" and "Non-Defensin" and "Not antibacterial" and "Not antifungal" and "Not antiviral" and "Not antiparasitic" and "Not antimicrobial" proteins. Initially, we acquired 42,357 protein sequences, from which we randomly picked 1,055 unique sequences with residue counts ranging from 10 to 60. As shown in Figure 5.2, our alternative dataset comprises 1,036 defensins and 1,054 non-defensin sequences.

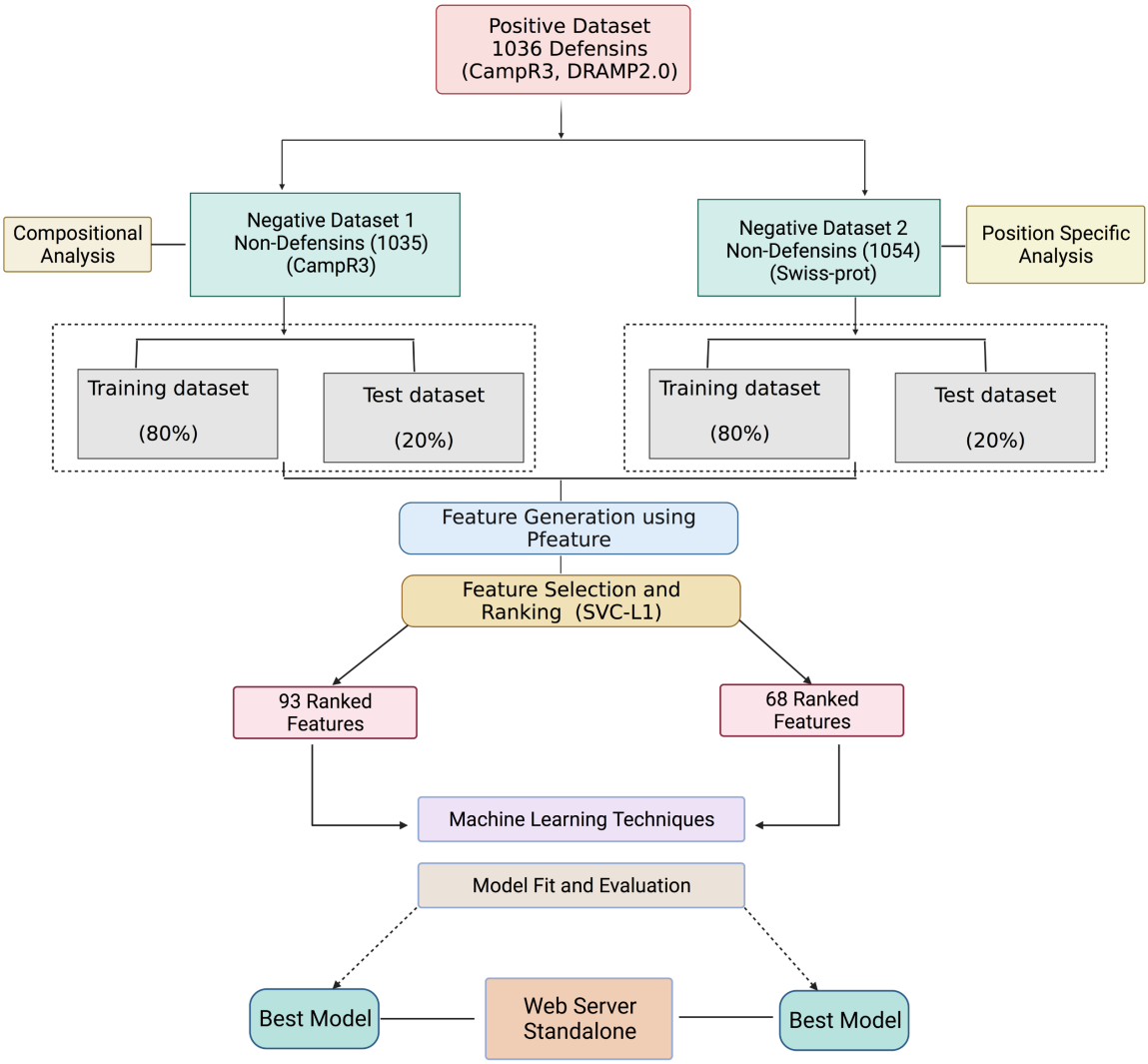


Figure 5.2 The workflow of the study

5.2.2 Feature's Generation, Selection and Ranking

In this study, the standalone version of "Pfeature" (Pande et al. 2019) was used to generate a variety of features from protein sequences. "Pfeature" can compute thousands of features/descriptors of protein or peptide sequences. We used "Pfeature's" composition-based function module to generate a vector of 8,968 features. Aside from this, we evaluated other composition features from "Pfeature" on both datasets independently. One of the study's primary issues is identifying an important group of features from the huge dimension of features. The 'SVC-L1-based' feature selection technique was applied, which combines the

support vector classifier (SVC) with a linear kernel and is penalised with L1 regularisation. “SVC-L1” was chosen because it employs several approaches to choose the appropriate features from a large number of feature vectors and is extremely fast in contrast to other methods (Aggarwal 2014). Its primary purpose is to minimise the objective function, which includes the loss function and regularisation. To reduce dimensions, the “SVC-L1” algorithm selects non-zero coefficients and then applies the L1 penalty to choose relevant features. The L1 regularisation builds sparse models throughout the optimization phase by eliminating a few features from the model and setting the coefficients to zero. The sparsity is controlled by the "C" parameter, which is proportional to the number of features chosen; the lower the "C" value, the fewer features are chosen. We utilised the default value of 0.01 for parameter "C” (Tandelilin 2010). Following that, the relevance of these features in protein classification was assessed using the software "feature selector." Using a “DT-based” technique called the Light Gradient Boosting Machine, the software "feature selector" ranks the features based on the number of time a feature is used to partition data across all trees.

5.2.3 Machine Learning Techniques

Several machine learning methods were employed in this work to create classification models using Python's “scikit-learn” module (“Scikit,” n.d.). It contains of extra tree (ET), random forest (RF), logistic regression (LR), support vector machine (SVM), k-nearest neighbours (KNNs), and multilayer perceptron (MLP). Using "GridSearch," several hyperparameters corresponding to these classifiers were tweaked, and only the best results were used.

5.2.4 Validation Techniques

To offer internal and external validation, we divide our datasets into training and validation sets in proportions of 80% and 20%, respectively. In the case of internal validation, we employed a five-fold cross-validation approach, in which training set sequences are arbitrarily split into five equivalent folds (Nagpal et al. 2018; Dilraj Kaur, Arora, and Raghava 2020). Thereafter, four of these folds are used for training and the remaining fold is used for testing. The procedure is replicated five times until each of the five folds has been used for testing at least once. Finally, the model's performance is computed by averaging the performance on the five folds. Internal validation is the process of optimising parameters on an 80 percent training

dataset to attain the optimal performance. To validate the performance of our models, we use a 20 percent validation dataset, which is referred to as external validation.

5.2.5 Evaluation Parameters

To test the performance of several machine learning classification models, we employed well-established assessment criteria. In this research, we employed both threshold-dependent and independent parameters such as sensitivity (Sens), specificity (Spec), and accuracy (Acc). A receiver operating characteristic (ROC) curve was produced between sensitivity and 1-specificity to evaluate the models' findings. Following that, we assessed using the standard threshold-independent parameter AUROC (area under the ROC curve) values. These properties were quantified using the following equations:

$$Sens = \frac{TP}{P} X100$$

$$Spec = \frac{TN}{N} X100$$

$$Acc = \frac{TP + TN}{P + N} X100$$

$$MCC = \frac{TPXTN - FPXFN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

where TP denotes true positive, FP denotes false positive, TN is true negative, and FN denotes false negative.

5.3 Results

We performed some preliminary analyses on the main and alternate dataset sequences to learn more about the preferences of various kinds of residues. Following that, the models were built using the "main" and "alternate" datasets. The sections that follow have to get into great detail on these studies as well as the performance of the models.

5.3.1 Compositional Preference Analysis

For defensins, AMPs, and non-defensin peptides, the amino acid composition (AAC) was computed. The usual amino acid composition of defensin, antimicrobial, and non-defensin peptides is depicted in Figure 5.3. In comparison to AMPs, defensins have a greater amino acid composition for particular types of residues (i.e., “C, D, E, N, R, T, Y”), as seen in Figure 5.3. Defensins have a greater amino acid composition for the following types of residues than non-defensins: “C, G, R, and Y”. Similarly, in comparison to non-defensins, AMPs have a higher composition for particular types of residues (e.g., “C, I, K, L”). These findings suggest that, despite the fact that both defensin and AMPs have antimicrobial activity, they have distinct residue preferences. These findings suggest that antimicrobial peptide prediction is not appropriate for predicting defensins because they prefer distinct types of residues. In addition, we used the "Mann–Whitney test" to establish the statistical significance of these three groups. We discovered that 54 of 60 pairs were statistically significant. A and W were non-significant amino acid residue pairings in AMPs and non-defensins. M residue is a non-significant pair in AMPs and defensins. At the same time, non-defensins with defensins have non-significant pairings of F, H, and T.

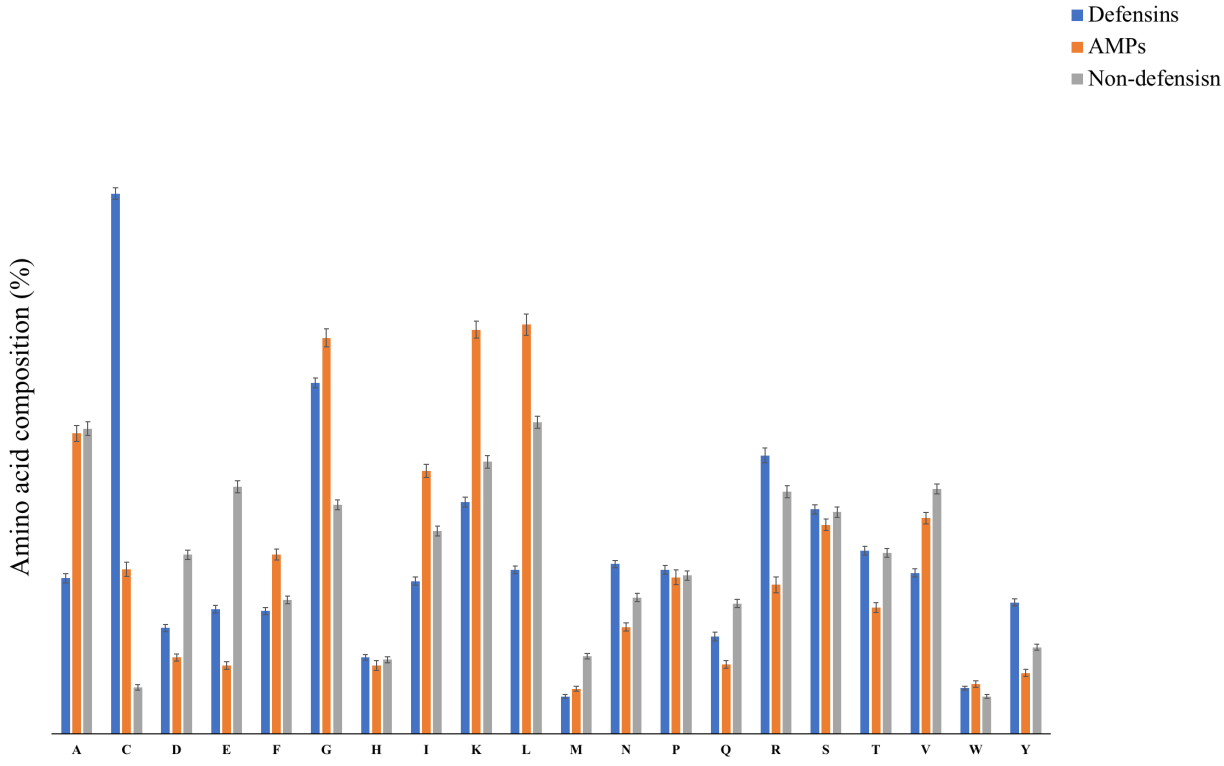


Figure 5.3 Visual representation of analysis of the average amino acid composition of defensins, AMPs, and non-defensins.

5.3.2 Positional Preference Analysis

The preference of a given amino acid at a specific position in the protein sequence was investigated in this study. Figure 5.4 depicts a two-sample logo (TSL) for the main and alternative datasets. The relative abundance of the sequence is represented by the most important amino acid residue. It is vital to remember that the first 10 positions represent peptide N-terminal residues, while the final 10 positions represent peptide C-terminus. The amino acid "C" was shown to be abundant at positions 1, 2, 3, 5, 6, 7, 8, and 9 of the C-terminus, as well as positions 3, 4, 5, 6, 8, and 9 of the N-terminus. Furthermore, the amino acid "N" was abundant at position 10 of the C-terminus, whereas "S" was enriched at position 7 of the N-terminus. Non-defensins, on the other hand, have an abundance of "K," "L," and "A" at various positions in both the C- and N-termini.

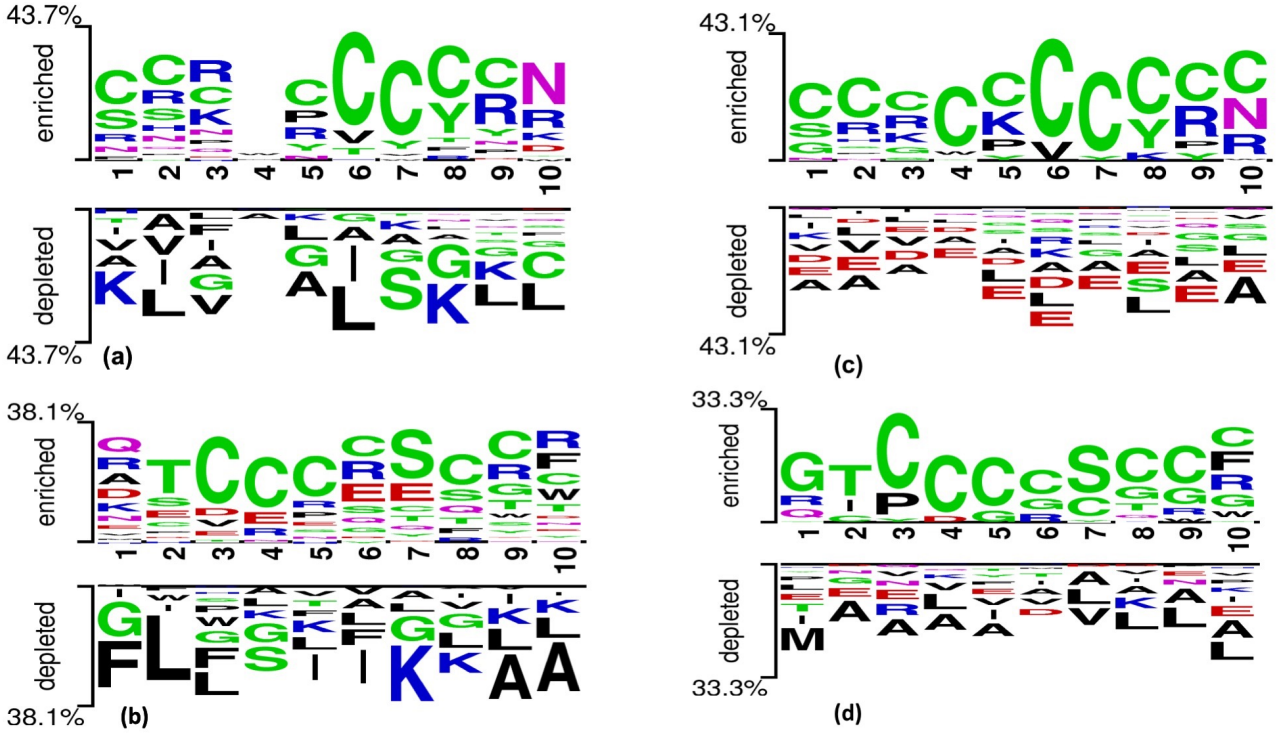


Figure 5.4 Two sample logos produced from the (a) C-terminus (last 10 residues) of the main dataset, C is enriched at position 1, 2,5,6,7,8 and 9, N at position 10 (b) the N-terminus (first 10 residues) of the main dataset, C is enriched at 3,4,5,6,8,9, S at position7 and R at position 10 (c) the C-terminus (last 10 residues) of the alternate dataset, C is enriched at position from

1-10 (d) and the N-terminus (first 10 residues) of the alternate dataset G is enriched at position 1, T at position 2, C is 3,4,5,6 8,9,10 and S at position at 7.

5.3.3 Models Development for Prediction

5.3.3.1 Selection of Features

First, we used the Pfeature software to compute a wide range of features. We have eliminated all the irrelevant features. Based on the SVC-L1 feature selection approach described in the Materials and Methods section, 93 significant features for the main dataset and 68 significant features for the alternate dataset were identified. For each of these datasets, all features were ranked based on their normalised an cumulative scores using the "feature selector" tool.

5.3.3.2 Models Based on Machine Learning on Selected Features

As previously stated, the SVC-L1-based selection technique reduced a total of 8,948 features retrieved from Pfeature's composition-based module to 93 (main dataset) and 68 (alternative dataset) features. On both datasets, a variety of machine learning classifiers such as SVM, LR, KNN, RF, MLP, and ET were used. Table 5.1 depicts the performance of various models. SVM clearly outperforms other methods for the main dataset, with AUROC and Matthews correlation coefficient (MCC) values of 0.98 and 0.88, respectively, at the training dataset. The corresponding validation dataset has an AUROC of 0.97 and an MCC of 0.87. With 0.97 AUROC and 0.84 MCC on the training dataset and 0.97 AUROC and 0.85 MCC on the validation dataset, LR was the second best model. Similarly, SVM was the best model for the alternate dataset, with 0.99 AUROC and 0.94 MCC on the training dataset and 0.99 AUROC and 0.96 MCC on the validation dataset.

Table 5.1 Machine learning model performance on SVC-L1 selected features for Main Dataset and Alternate Dataset.

MAIN DATASET											
		Training Dataset					Validation Dataset				
Model	Hyperparameters	Sens	Spec	ACC	AUROC	MCC	Sens	Spec	ACC	AUROC	MCC
SVM	C=2, g=1, k=rbf	93.24	94.81	94.03	0.98	0.88	93.72	93.24	93.48	0.97	0.87

LR	C=1	92.4	91.67	92.03	0.97	0.84	92.75	91.79	92.27	0.97	0.85
ET	ne=30	93.73	94.08	93.9	0.98	0.88	93.24	93.72	93.48	0.97	0.87
RF	ne=90	91.07	95.41	93.24	0.98	0.87	91.3	95.17	93.24	0.98	0.87
KNN	al=ball-tree, nn=10, w=distance	92.52	94.32	93.42	0.97	0.87	92.27	90.82	91.55	0.96	0.83
MLP	a=identity, HL=3, m=100, s=adam	92.4	89.73	91.07	0.95	0.82	93.72	87.92	90.82	0.96	0.82

ALTERNATE DATASET											
		Training Dataset					**Validation Dataset**				
Model	Hyperparameters	Sens	Spec	ACC	AUROC	MCC	Sens	Spec	ACC	AUROC	MCC
SVM	C=2, g=0.5, k=rbf	95.05	98.46	96.77	0.99	0.94	97.1	99.05	98.09	0.99	0.96
LR	C=10	94.93	97.86	96.41	0.99	0.93	94.69	98.58	96.65	0.99	0.93
ET	ne=50	94.09	98.93	96.53	0.99	0.93	94.69	99.53	97.13	0.99	0.94
KNN	al=brute, nn=10, w=distance	92.88	98.22	95.57	0.99	0.91	94.69	98.58	96.65	0.98	0.93
RF	ne=70	95.66	97.27	96.47	0.99	0.93	96.14	97.16	96.65	0.99	0.93
MLP	a=tanh, HL=10, m=100, s=adam	92.4	97.86	95.16	0.98	0.9	93.72	98.1	95.93	0.98	0.92

* g: gamma, ne: n_estimators, k: kernel, a: activation, HL: hidden layer size, s: solver, al : algorithm, w:weight, m : max_iter, nn: n_neighbour

5.3.3.3 Machine Learning-Based Models on Selected Top-Ranked Features

We evaluated the importance of various feature sets in addition to developing prediction models over entire selected features. The aim was to find a feature set with the fewest features that can consistently distinguish between defensins with AMPs and non-defensins with high

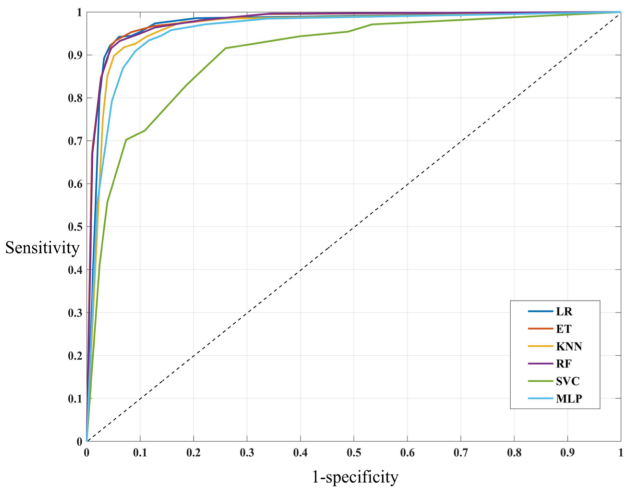
AUROC and accuracy. As a consequence, we developed multiple models based on the top (10, 20, 30,..., 93) features in the main dataset and the top (10, 20, 30,..., 68) features in the alternate dataset, and assessed them on the training and validation datasets. The best features, as seen in the results, were determined, namely the top 60 for the main dataset and the top 50 for the alternate dataset. For the main dataset, SVM (training: 0.98 AUROC, 0.88 MCC, and validation: 0.98 AUROC, 0.88 MCC) is the best model, followed by LR (training: 0.96 AUROC, 0.82 MCC and validation: 0.97 AUROC, 0.83 MCC). As demonstrated in Table 5.2 and Figure 5.5, the best model for the alternate dataset is SVM (training: 0.99 AUROC, 0.93 MCC and validation: 0.99 AUROC, 0.96 MCC), followed by LR (training: 0.99 AUROC, 0.91 MCC and validation: 0.98 AUROC, 0.90 MCC).

Table 5.2 Machine learning models' performance on the top 60 features of the main dataset and the top 50 features of the alternate dataset.

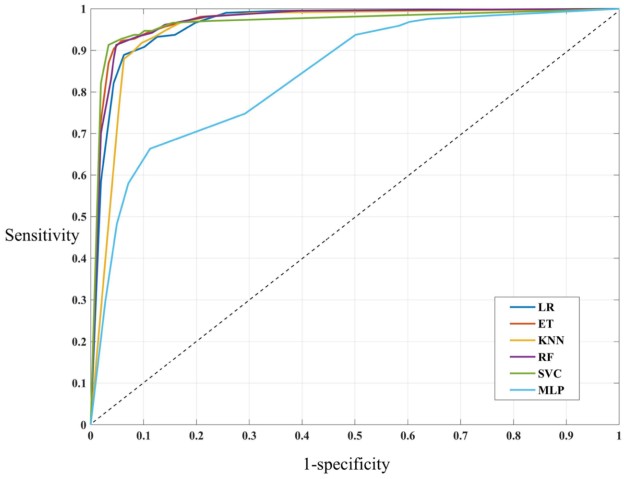
Main top 60		**Training dataset**					**Validation dataset**				
Model	Hyperparameters	Sens	Spec	ACC	AUROC	MCC	Sens	Spec	ACC	AUROC	MCC
SVM	C=2, g=1, k=rbf	89.26	96.74	93	0.98	0.86	90.82	97.1	93.96	0.98	0.88
LR	C=0.1	86.85	93.24	90.04	0.96	0.8	88.89	93.72	91.3	0.97	0.83
ET	ne=50	92.4	95.41	93.9	0.98	0.88	92.4	95.41	93.9	0.98	0.88
RF	ne=60	91.68	95.29	93.48	0.98	0.87	91.3	94.69	93	0.98	0.86
MLP	a=tanh, HL=17,m=100, s=adam	74.79	70.77	72.78	0.85	0.46	91.79	91.3	91.55	0.96	0.83
KNN	al=ball-tree, nn=10, w=distance	91.8	93	92.4	0.97	0.85	91.79	90.34	91.06	0.96	0.82

Alternate top 50		**Training dataset**					**Validation dataset**				
Model	Hyperparameters	Sens	Spec	ACC	AUROC	MCC	Sens	Spec	ACC	AUROC	MCC
SVM	C=2, g=1, k=rbf	95.17	97.98	96.59	0.99	0.93	97.1	99.05	98.09	0.99	0.96

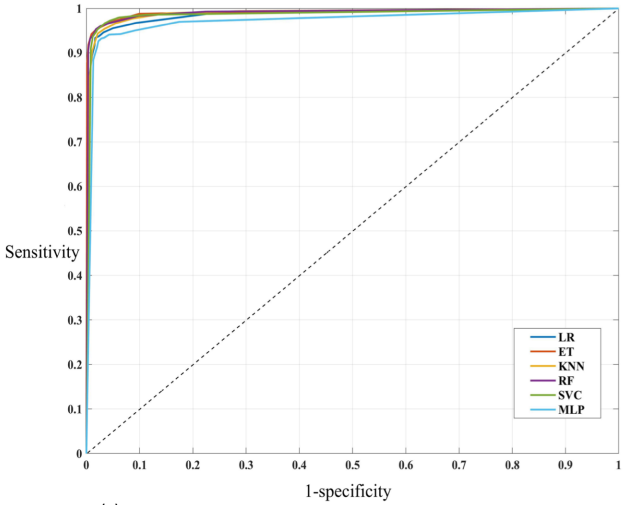
LR	C=1	95.54	95.02	95.28	0.99	0.91	95.65	95.73	95.69	0.98	0.91
ET	ne=40	95.17	98.22	96.71	0.99	0.93	95.65	98.58	97.13	0.99	0.94
KNN	al=ball-tree, nn=9, w=distance	94.33	97.86	96.11	0.99	0.92	95.65	98.1	96.89	0.98	0.94
RF	ne=50	95.3	98.22	96.77	0.99	0.94	96.65	97.63	96.65	0.99	0.93
MLP	a=tanh, HL=15, m=100, s=adam	92.64	97.75	95.22	0.98	0.91	92.27	97.63	94.98	0.98	0.9



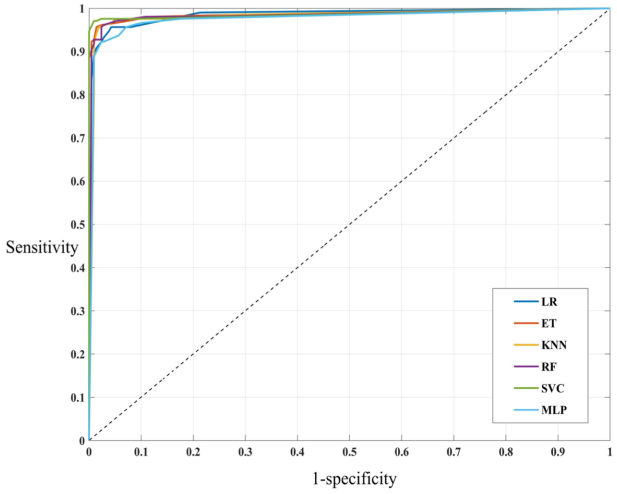
(a)



(b)



(c)



(d)

Figure 5.5 AUROC plots (a) main (top 60 selected features from the training datasets), (b) main (top 60 selected features from the validation datasets), (c) alternate (top 50 selected features from the training datasets), and (d) alternate (top 50 selected features from the validation datasets)

5.3.4 Existing Methods : Comparison

We also compared our models produced in this study to previous methods developed. As demonstrated in Table 5.3, these approaches have been refined over time on a variety of datasets of varying size and nature. As a result, direct comparisons of these approaches with other methods are not feasible. Previous research collected defensin peptides from the Defensins Knowledgebase, which was created in 2006 (Seebah et al. 2007), or from Swiss-Prot (The UniProt Consortium 2017). The size of the dataset is one of the constraints of prior studies. In this work, we used the largest dataset available to create trustworthy models using data from several sources. In addition, to distinguish defensin from antimicrobial peptides and non-defensins, we constructed two datasets termed the main and alternate datasets. Our web service not only predicts defensin but also scans for defensin peptides in proteins and designs highly effective defensins. In contrast, the majority of previous web services are no longer in use. This warrants the creation of a new approach to supplement current ones.

Table 5.3 Describing main components of existing techniques and DefPred, such as dataset source, data size, significant characteristics, type, performance, and so on.

Study	Source of Dataset	Size of data	Major Features	Classifier used	Type	Accuracy	Webserver availability, status	PMID
Zuo, YC et al. 2009	Defensin Knowledgebase	286P	ID_RAAA	Jack-knife Test	Prediction	91.36%	No	19591890
Shreyas, K et al. 2009	Pubmed, iHop, Uniport, HubMed	238P, 238 N	RQA Descriptors	RF	classification	78.12%	No	NA
Kumari, SR et al. 2012	NCBI, UNIPROT	383P, 383 N	AAC, DPC, PSAAC	SVM	classification	99%	Yes, Inactive	22670676
Zuo, YC et al. 2015	Defensin Knowledge base	333P	iDEF-PseRAAAC	SVM	Prediction	85.59%	Yes, Inactive	26713618

Zuo, YC et al. 2019	Defensin Knowledge base	328P	iDEF-PseRAAC	SVM	Prediction	91.16%	Yes, Active	31391777
DefPred	CAMPR3, DRAMP2.0, Defensin Knowledgebase, Swiss PROT	1036P, 1035N, (main), 1036 P , 1054N (alternate)	Selected features	SVM	Prediction	93.96%(main) 98.09% (alternate)	Yes, Active	NA

5.3.5 Implementation of the Web Resource

To assist the scientific community, we created a user-friendly prediction web server that integrates multiple modules to predict defensin proteins. The study's prediction models are used in the web server. Users will predict if a query peptide is defensin or non-defensin based on the score of the prediction models at different thresholds. The five major modules of the web server are Predict, Protein-scan, Design, Downloads, and Algorithm. Using the "Predict" module, the user may distinguish between defensin and non-defensin peptides. Both the positive and negative datasets used in this study are available in FASTA format for download. The web server "DefPred" was built using HTML, Java, and PHP scripts. Below is a thorough explanation of these modules:

The Predict module predicts whether or not the provided protein sequence is defensin. Users can enter several peptides in FASTA format into the box or upload a file that contains the same. This module allows the user to predict defensins from AMPs using model-1 established on the main dataset. Model-2 was created to predict both defensins and non-defensins. The Design module enables the user to produce all potential analogues for a sequence and then rank these peptide sequences according to their ratings. This helps the user to find the best defensin analogue. The Scan module was created to discover protein areas with defensin-like properties. In order to better serve the community, we created a stand-alone Python application. In addition, we have given a stand-alone facility in the form of Docker technology. This standalone programme is incorporated into our "GPSRdocker" package, which may be obtained from the website https://webs.iiitd.edu.in/gpsrdocker/ (P. Agrawal et al. 2019).

5.4 Discussion and Conclusion

Antibiotic resistance is spreading among microorganisms all across the world, and conventional therapies for drug-resistant pathogens are inadequate. With increased pathogen drug resistance, the concern of a post-antibiotic era needs the development of alternatives to standard antibiotics or small molecule-based medicines. Because of their various therapeutic effects, AMPs represent a class of prospective therapeutics with curative promise. Several organisms' innate immune systems rely significantly on these evolutionarily conserved molecules. Defensins are a subclass of AMPs with several roles and modes of action, making them less likely to be drug resistant (Robert I Lehrer, Bevins, and Ganz 2005; R I Lehrer, Lichtenstein, and Ganz 1993). Furthermore, defensins' peculiarities in the mechanism of microbicidal activity from other antibiotics make them useful in combating infections when taken in conjunction with traditional antibiotic therapies (Tai et al. 2015). Natural defensins are efficient, non-toxic microbicides that may be useful in treating infections caused by antibiotic-resistant pathogens. According to recent research, they do this by disrupting bacterial cell membranes but not mammalian cell membranes. With this knowledge, producing next-generation defensins with improved biological activity profiles is a realistic goal that will allow defensins to be used to boost human health in the near future. New antimicrobials with defensin-based bactericidal and immunomodulatory properties may be effective against drug-resistant bacteria while also boosting survival from common illnesses when used in combination with standard antibiotic therapy (Tai et al. 2015). Furthermore, prior research has shown that defensin and antibiotic combinations may be used synergistically to combat infections, including biofilms, allowing for lower doses of both medications while still boosting treatment efficacy (Koo et al. 2017; Dostert, Belanger, and Hancock 2019; Y. Jiang, Geng, and Bai 2020). The growth of in-silico research, notably in the field of bioinformatics, has resulted in the identification and definition of defensin features that allow them to carry out their varied spectrum of biological functions. However, because defensins and AMPs are so similar in nature, distinguishing defensins is difficult, making it difficult to develop entirely defensin-based treatments.

Our research tackles this issue by offering cutting-edge machine learning models that can be used to distinguish and predict defensins from other AMPs and defensins from other proteins (non-defensins). Furthermore, because the dataset is critical in machine learning as well as for a strong in-silico prediction model, we used updated repositories to construct a very thorough

and up-to-date dataset. TSL and compositional analysis studies were carried out to better understand the structure and positional preference of defensins. Previous research suggests that defensins are abundant in cysteine (C) amino acid (Solanki et al. 2021), which is consistent with our findings. The features of experimentally validated defensins found in the literature were used to create a variety of prediction models. From sequence data, the tool "Pfeature" was used to generate 8,968 features. The scikit package's SVC-L1 was used to choose features, which were subsequently ranked using feature selector algorithms. The compositional study revealed that some types of residues, such as C, R, N, L, and Y, are favoured in defensins whereas others, such as M, are not. . This was also supported by one of the highest-ranking selected features, AAC_C, which represents the amino acid composition of cysteine in a protein sequence. AAC_C came in first place in the main dataset and second place in the alternate dataset. CeTD_ SA1 is a composition-enhanced transition and distribution of group 1 (A, L, F, C, G, I, V, W) for solvent accessibility attribute, and PAAC1_E is the pseudo-amino acid composition of glutamic acid in the main dataset. In the case of the alternate dataset, CeTD_SS1, which is a composition of group 1 (A, L, F, C, G, I, V, W) residue for the secondary structure attribute, and BTC_T, which is the total bond composition present in the sequence, were two of the top-ranked features. The amino acid composition of cysteine is similar in both the main and alternate datasets, demonstrating that defensins stand out due to their higher "C" concentration (Figure 5.3). It's worth mentioning that the new feature selection procedures chose 93 features for the main dataset and 68 features for the alternate dataset, which contain the aforementioned features. We utilised these 93 and 68 features to develop the two classification models in our study. Furthermore, the performance of several models based on the top-ranked features was validated using a five-fold cross-validation approach. . To avoid over-optimization of the models, we desired a basic collection of features with the least amount of performance loss. For the final classification models, we picked the top 60 and top 50 features from the main and alternate datasets, respectively. Model-1, which used the main dataset, is an SVM classifier that achieved a maximum performance of 0.98 AUROC and 0.88 MCC in the training dataset and 0.98 AUROC and 0.88 MCC in the validation dataset for distinguishing defensins from AMPs, whereas model-2, which used the alternate dataset, distinguished defensins from non-defensins. Model-2 is another SVM classifier that outperformed the others on the training and validation datasets, with AUROC of 0.99 and MCC of 0.93 and AUROC of 0.99 and MCC of 0.96, respectively.

Despite several advancements, this study has a few drawbacks. The current work intended to create a method for predicting defensins/AMPs and defensins/non-defensins. Due to the small number of experimentally validated defensins, we used sequence data from all available species, including mammals, plants, and insects, to achieve this; however, the ideal process for developing a host-specific method for predicting defensins should only include data from the concerned host. Furthermore, structural features like as secondary structure data, surface accessibility rating, and disulfide bond information are not taken into consideration in our models. Furthermore, when it comes to prediction, our models exclude information about post-translational modifications (e.g., terminus modification, incorporation of chemical moieties, glycosylation, and phosphorylation). Although a systematic attempt was made in this study to develop the best models feasible given the existing conditions, it is anticipated that future research will be able to overcome these concerns in order to enhance prediction.

Finally, in order to better serve the scientific community, we created a web server called "DefPred," as well as a standalone version that included our top models. The standalone version is Python-based and provides the user with a plethora of choices. The accompanying server, on the other hand, is user-friendly and compatible with a variety of displays, including laptops, Android mobile phones, iPhones, and iPads. In addition, we have given a stand-alone facility in the form of Docker technology. This standalone programme is incorporated into our "GPSRdocker" package, which may be obtained from the website https://webs.iiitd.edu.in/gpsrdocker/. We expect that this work will aid vaccine designers as well as provide a better knowledge of immune defence response.

\

6

Prediction of Patients at High Risk and Possible Therapeutics in UCEC

6.1 Introduction

Endometrial cancer (EC) or uterine corpus endometrial carcinoma (UCEC) is the sixth most prevalent cancer in women. In 2020, 12,590 fatalities were expected from the 65,620 recorded cases (Siegel, Miller, and Jemal 2020). In contrast to the lowering trends for many prevalent malignancies, death rates for EC have stayed essentially constant (Cronin et al. 2018). Because of advancements in high-throughput technologies, it is becoming increasingly common to diagnose and prognostic EC at an early stage. Nevertheless, a significant number of individuals who develop metastasis or recurrent tumour have a poor prognosis. UCEC is divided into two primary subtypes: type I tumours, which account for about 75–80 percent of the pathologic subtypes and are endometrioid adenocarcinomas (Gottwald et al. 2010; Setiawan et al. 2013) and type II tumours, which are serous carcinomas and have a poorer diagnosis (Setiawan et al. 2013; Black et al. 2016). In individuals with UCEC, poor diagnosis and prognostic factors result in significant mortality and recurrence. To identify UCEC from other benign disorders, a multistep diagnostic method involving gynaecological examination, transvaginal ultrasonography, and endometrial biopsy is currently required. Figure 6.1 shows the anatomy of female reproductive system. Likewise, clinical features such as tumour grade, cervical involvement, lymph node status, histological subtype, depth of myometrial invasion, and lympho vascular space invasion (LVSI) are considered as prognostic factors for UCEC patients (Coll-de la Rubia et al. 2020). Biopsies provide information on tumour grade, histological subtype, and other clinical features. Although platinum-based chemotherapy and hormone therapy are first-line therapies for UCEC, primary hysterectomy and bilateral salpingo-oophorectomy are conventional treatments (Morice et al. 2016). However, the existence of distant metastases means that patients respond poorly to conventional treatment and have a relatively low 5-year survival rate of 17% (P. Chen et al. 2020). As a result, effective risk stratification strategies are necessary in UCEC for prognostic assessment and treatment decision making.

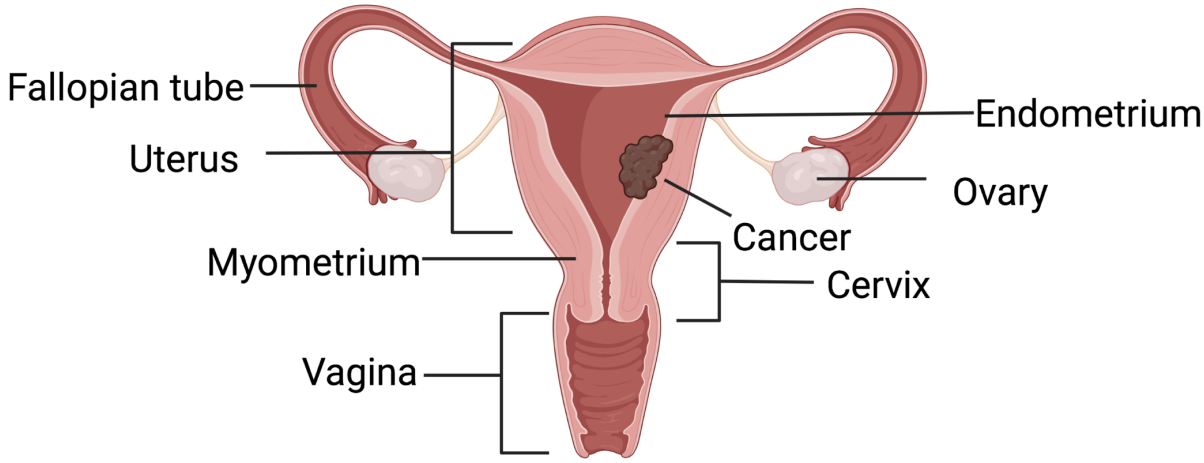


Figure 6.1 Anatomy of the female reproductive system, showing endometrial cancer.

Many biomarkers for UCEC diagnosis, classification, and prognosis have been found with the introduction of high-throughput sequencing technologies and public databases. In contrast to clinicopathological factors, these biomarkers are linked to underlying molecular pathways and provide compelling explanations for the pathogenesis of EC. Previous research has shown that a group of five genes (*BUB1B, CCNB1, CDC20, DLGAPS, and NCAPG*) can effectively predict the prognosis of endometrial I carcinoma. These genes' expression was higher, which helped predict a relatively large tumour grade and a poorer overall survival (OS) (Bian et al. 2020). Nine glycolysis-related genes (*CLDN9, AK4, PC, GPC1*, and *SRD5A3*) were linked to poorer survival, whereas *B4GALT1, GMPPB, B4GALT4*, and *CHST6* were linked to better survival. *GMPPB* had the greatest hazard ratio (HR; 1.544) and a p value of 0.0134 among these genes (Z.-H. Wang et al. 2019). In UCEC samples, higher *KLHL14* expression was linked with lower overall survival ($p = 0.0370$) and progression-free survival ($p = 0.081$) (Han, Yang, and Lin 2019). Six possible predictive tumour microenvironment (TME)-related genes (*CACNA2D2, CTSW, NOL4, SIGLEC1, TMEM150B, and TRPM5*) correlated with OS in UCEC patients (P. Chen et al. 2020).

Pattern recognition receptors (PRRs) have long been thought to have a role in recognizing microbial ligands and activating the immune system. Recent advances in bioinformatics have resulted in the creation of a database dedicated to PRRs (Lata and Raghava 2008a). The recently revised PRRDB2.0 (Dilraj Kaur et al. 2019a) contains the most recent information regarding receptors and their related ligands and suggests that ligands for toll-like receptors

(TLRs), a well-known family of PRRs, have anti-tumoral actions in numerous malignancies via activation in tumour cells. Depending on the circumstances, this activation might have both pro- and anti-tumoral effects (Goutagny et al. 2012). Some of the activities demonstrating that TLRs work as tumour promoters include increased angiogenesis and survival, acceleration of tumour invasion, and resistance to apoptosis (Ikebe et al. 2009). TLR pathways are important regulators of chemo-resistance, potentially via upregulating production of the anti-apoptotic protein B-cell lymphoma-2 (Bcl-2) via activated nuclear factor (NF)-B (Alvero et al. 2009). TLR1 overexpression was associated with better survival in pancreatic cancer (HR 0.68; 95 percent confidence interval [CI] 0.47–0.99; p = 0.044) (Lanki et al. 2019). TLR9 is enhanced in glioma development (C. Wang et al. 2010), and its lower expression is useful in predicting disease-free survival in triple negative breast cancer (Tuomela et al. 2012). TLR4 is also involved in the chemo resistance of ovarian cancer cells (Kelly et al. 2006). TLR3 and TLR4 expression were studied during the menstrual cycle, endometriosis, postmenopausal endometrium, and endometrial hyperplasia, and low TLR3 and 4 expression was related with a poor prognosis in UCEC (Allhorn et al. 2008). The association if various TLRs with cancer can be found in Table 6.1

Table 6.1 Shows the association of various TLRs with different cancer.

TLRs	Association	Reference
TLR3-4, 7 and 9	poor differentiation, high proliferation, and advanced stage in oesophageal cancer	(Sheyhidin et al. 2011)
TLR5	good prognosis in lung cancer	(Brackett et al. 2016)
TLR7	poor diagnosis in lung cancer	(Grimmig et al. 2015)
TLR1	better survival in pancreatic cancer	(Lanki et al. 2019)
TLR9	poorer survival in renal cell carcinoma	(Ronkainen et al. 2011)

The purpose of this study was to look at the changed expression profile of PRR genes in the context of survival prediction in patients with UCEC. The discovery of key biomarker genes

having a strong connection with survival can aid in risk group stratification and prognosis. As a result, the discovered biomarker genes can serve as a strong foundation for the investigation of novel therapeutic techniques in the treatment of UCEC. Our study used a variety of bioinformatics methodologies, including network-based approaches, Cox-proportional hazard (PH) survival studies, and clustering-based approaches, to identify important genes and construct highly accurate risk-prediction models. We also assessed the prognostic significance of different clinicopathological features and investigated the molecular mechanisms linked to the discovered genes in order to uncover relevant therapeutic molecules that could improve the survival of patients with UCEC.

6.2 Methods and Materials

6.2.1 Dataset Preparation

The dataset was first extracted from 'The Cancer Genome Atlas' (TCGA) using TCGA Assembler 2, and it comprised of quantile normalised RNA seq expression levels for 581 individuals with UCEC. Only 541 patients in the study have information regarding their OS and censoring. Gene Set Enrichment Analysis and the HUGO Gene Nomenclature Committee were used to generate the list of 331 PRR signalling pathway genes. Only 308 PRR genes have gene expression data available. Using custom Python and R programmes, the final dataset was reduced to 541 samples including RNA seq values for 308 PRR-related genes. Overall workflow is shown in Figure 6.2.

6.2.2 Survival Analysis

Based on the length of OS, HRs and 95% CI were calculated to predict the risks of mortality associated with high-risk and low-risk categories. Using the univariate unadjusted Cox-PH regression models, these were stratified based on appropriate cut-offs for various factors. The survival curves of high-risk and low-risk groups were compared using Kaplan–Meier (KM) plots. These datasets' survival analyses were carried out in R using the'survival' and'survminer' packages (V.2.42-6) (V.3.4.4, The R Foundation). Log-rank tests were used to determine the statistical significance of the survival curves. Wald tests were employed to assess the significance of the explanatory factors utilised in HR estimations. The model's prediction ability was measured using the Concordance Index (van der Net et al. 2009). p values of less

than 0.05 were considered significant. To assess the association between various factors, multivariate survival analysis based on Cox regression was used.

6.2.3 Network of gene co-expression

To find relationships between the genes, a co-expression network of PRR genes was built. We employed Pearson's correlation coefficient (PCC) for each PRR gene pair utilising gene expression value to determine statistically significant important genes and therefore design highly accurate risk-prediction models to create the gene co-expression network used here. We also assessed the predictive relevance of different clinicopathological features and investigated the molecular mechanisms linked to the discovered genes in order to uncover relevant therapeutic molecules that could improve the survival of patients with UCEC.

6.2.4 Clustering using kMeans and kMedoids

Unsupervised clustering using k-Means was utilised to find a predetermined number (k) of representatives/centroids. After forming k clusters, a representative gene from each cluster was selected. This decision was based on the lowest p value obtained from univariate survival analysis. Medoids from each cluster were chosen as representatives in the same way. This is due to the fact that, unlike centroids, medoids are always confined to cluster genes. We then utilised the obtained representative genes to construct a model.

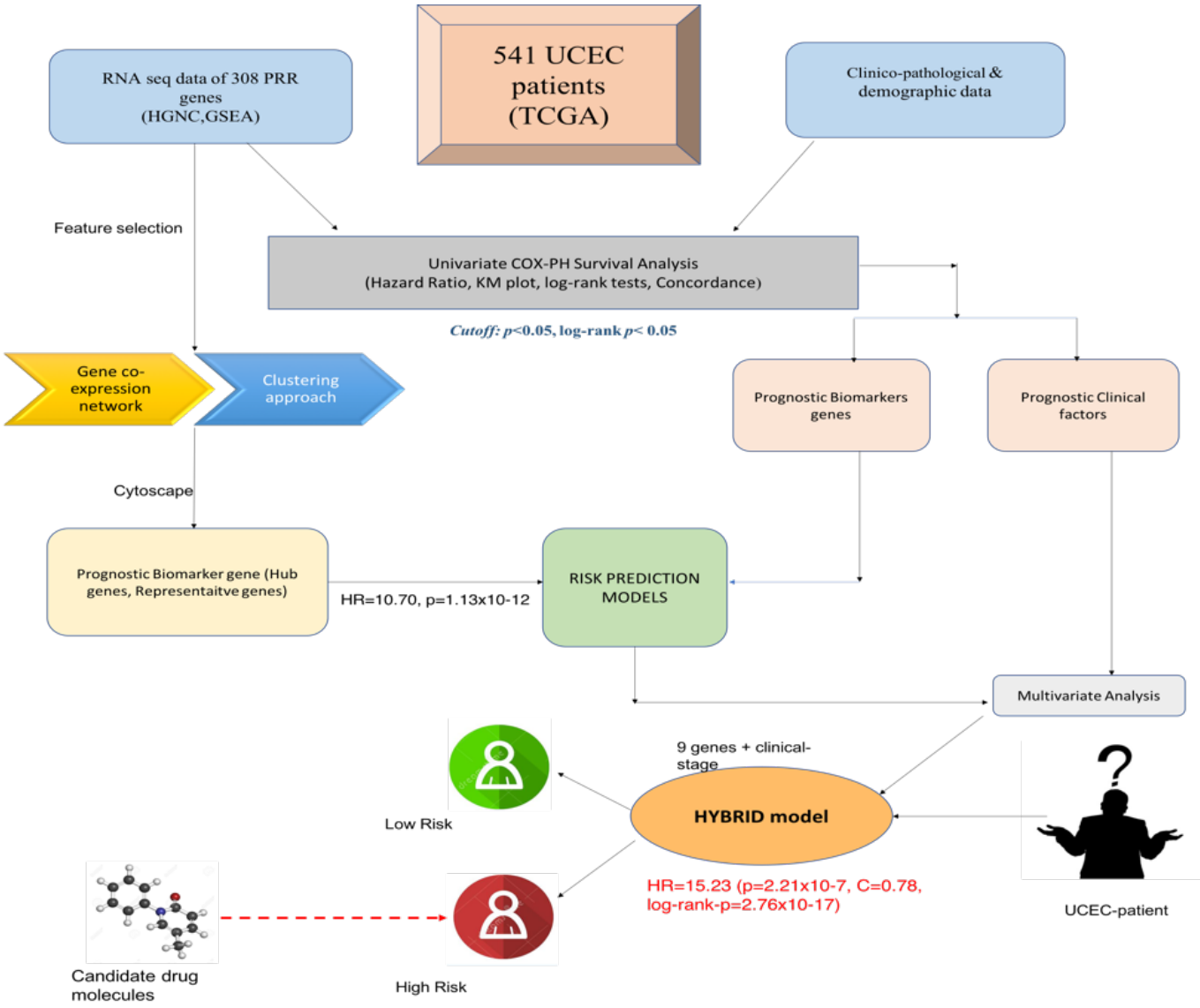


Figure 6.2 Overall workflow of the study

6.2.5 Models Based on Multiple Genes

6.2.5.1 Regression Models Based on Machine Learning

Regression models from the Python 'sklearn' module were used to fit gene expression data (independent variables) against OS time (target variable). Linear, Random forest, K-nearest neighbours, Ridge, Lasso, Lasso Lars, and Elastic Net were among the regressors utilised. A fivefold cross-validation approach, as used in earlier works (D. Kaur, Arora, and Raghava 2019), was used for fitting and test evaluations. To estimate HRs, CIs, and p values, a combination of all five analysed test datasets (predicted OS) was utilised to categorise the

actual patient OS at the median cut-off. Patients with a predicted OS greater than the median were labelled 'low risk,' while those with a predicted OS less than the median were labelled 'high risk.' The in-built function 'Grid search CV' was used to optimise and regularise the hyperparameters. The standard parameters, root mean squared error and mean absolute error, are used to describe model performance.

6.2.5.2 Prognostic Index

The prognostic index (PI) for a set of k genes was calculated as indicated in Eq. (1), as implemented by Li et al. (P. Li et al. 2018) and Wang et al. (Y. Wang et al. 2018).

$$\boldsymbol{PI = S_k \beta_k g_k} \qquad \textbf{(1)}$$

Where β indicates the regression coefficient derived from a univariate Cox regression for a gene g, the PI for a separate set of genes was utilised for stratifying risk groups, and standard metrics such as HR, p value, and so on were evaluated. Patients with a PI more than the PI's median were classed as high risk, while patients with a PI less than the PI's median were classified as low risk.

6.2.5.3 Model Based on Gene Voting

Each patient was assigned a risk label of 'high risk' or 'low risk' based on their unique gene expression (median cut-off). As a result, for n survival-associated genes, each patient was represented by a vector of n risk labels. The patient was finally categorised into one of the high-/low-risk groups using the gene voting approach based on the dominant 'label' (i.e. occuring more than n/2 times) in this vector. This is seen in Figure 6.3 for n = 7. Following that, conventional metrics were evaluated.

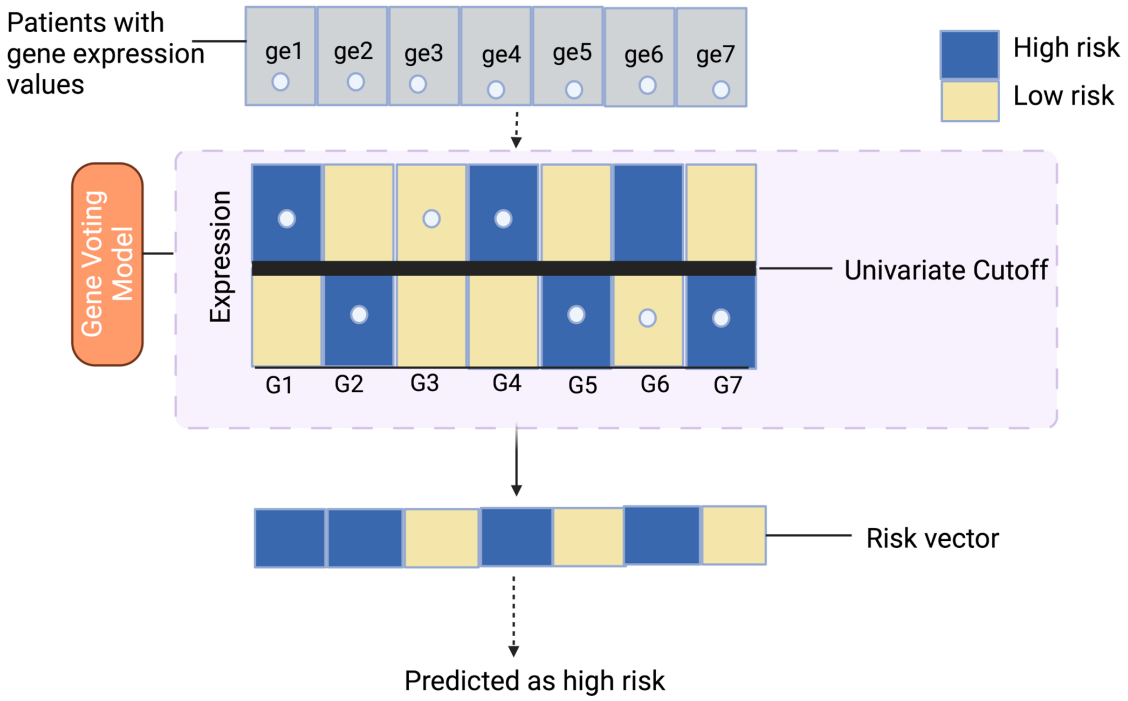


Figure 6.3 Using a seven-gene voting model ("G1, G2,... G7"), patients are classified as high- or low-risk. Based on dominant risk labels, the expression levels of these genes ("ge1, ge2,... ge7") are utilised to determine the risk vector.

6.3 Results

6.3.1 Pattern Recognition Receptor Genes Associated With Survival

All 308 PRR genes were subjected to a univariate Cox-PH analysis with median expression cut-offs. Patients having an expression value greater or lower than the gene's median expression value were classed as high or low risk, accordingly. Only 15 of the 308 genes were significant ($p < 0.05$). The survival study identified nine good prognosis markers (GPM; that is, genes that are positively correlated with patient OS time) and six bad prognostic markers (BPM; that is, genes that are negatively correlated with patient OS time). The findings for these genes are shown in Table 6.2, together with the metrics linked with the stratification of high-/low-risk patients.

Table 6.2 Shows the results of a univariate Cox regression with a cut-off of greater than the median.

Gene	HR	1/HR	p Value	C-index	95% CI	Log-rank (p)	q Value
CLEC1B	6.48	0.15	2.11E−06	0.58	2.99–14.04	1.01E−04	4.78E−08
CLEC3A	2.71	0.37	3.47E−03	0.58	1.39–5.28	7.16E−03	2.34E−03
MRC1	2.17	0.46	1.60E−02	0.6	1.16–4.09	1.23E−02	1.35E−02
IRF7	0.47	2.14	1.77E−02	0.59	0.25–0.88	1.43E−02	2.56E−02
CTSB	0.5	2	2.65E−02	0.61	0.27–0.92	2.31E−02	3.64E−02
FCN1	2	0.5	2.85E−02	0.56	1.08–3.72	2.47E−02	3.53E−02
RIPK2	0.5	2	2.86E−02	0.57	0.27–0.93	2.42E−02	4.62E−02
CLEC3B	0.51	1.94	3.35E−02	0.59	0.28–0.95	2.96E−02	4.05E−02
CLEC12B	1.86	0.54	3.82E−02	0.57	1.03–3.35	4.11E−02	4.02E−02
TLR4	0.53	1.89	3.88E−02	0.55	0.29–0.97	3.55E−02	3.82E−02
NLRP10	1.94	0.52	3.99E−02	0.57	1.03–3.65	5.00E−02	3.57E−02
NLRP9	0.53	1.89	4.15E−02	0.56	0.29–0.98	3.70E−02	3.04E−02
MAPKAPK2	0.53	1.88	4.35E−02	0.55	0.29–0.98	3.89E−02	2.36E−02
TNIP1	0.54	1.86	4.38E−02	0.56	0.29–0.98	4.02E−02	2.56E−02
SARM1	0.54	1.85	4.95E−02	0.55	0.29–1.00	4.54E−02	1.53E−02

Genes with HR more than one are considered bad BPM, while those with HR less than one are considered GPM. BPM is a bad prognostic marker, CI is a confidence interval, C-index is a concordance index, GPM is a good prognostic marker, HR hazard ratio, and q value is a false discovery rate-corrected p value.

6.3.2 Using Network Based Features for Risk Prediction

In this part, we attempted to pick features from the PRR gene network. The network was used to identify features/representatives in order to better understand connections in PRR genes. For feature selection, we employed the following methods: (1) network hub genes, (2) cluster medoids, and (3) cluster representatives. These chosen features or PRR genes were utilised to construct models to predict the survival of cancer patients**.**

6.3.2.1 Network's Hub Genes

A correlation matrix was generated from 308 PRR genes, with correlation calculated between all possible pairings of genes based on the expression data. The correlation matrix used to generate network edges using 'Igraph' for strongly correlated pairs of genes (|PCC| > 0.5), as explained in Sect. 6.2. To visualise and analyse the gene network, we utilised the 'Cytoscape' software. When the effective correlation was set at larger than 0.5, there were 116 nodes and 804 edges. Based on their degree, we chose the top 15 hub genes (*BTK, ITGB2, HAVCR2, FCRL3, CD163, CD300LF, CD68, CTSS, CLEC10A, CLEC12A, NR1H3, CLEC4E, CD209, ITGAM, and TLR8)*. These hub genes were utilised to construct a prognostic model for predicting the survival risk of UCEC patients. Our voting-based model, which used these 15 hub genes, produced an HR of 1.37 with a p-value of 0.294. Figure 6.4 depicts the network.

6.3.2.2 Cluster Medoids

We used k-medoids to cluster genes based on pairwise dissimilarity. The medoids were chosen for clusters with k = 5, 10, 15, 20, and 25. The gene voting model produced the best results at k = 5 (HR 1.85; p = 0.045).

6.3.2.3 Clusters Representatives

The representative gene was chosen based on the lowest p value obtained from each cluster's univariate survival study. Representative genes from each cluster were obtained and utilised to create a risk classification model. This procedure was carried out for k = 5, 10, 15, 20, and 25. As indicated in Table 6.3, the best result was obtained for k = 10 (HR 4.11; p = 3.7 10 5). We next filtered the representative genes within each cluster using a p< 0.05 cut-off. Table 6.3 shows that gene voting models produced the best results. Gene voting models produced the best outcomes, as seen in Table 6.4. We discovered that when k = 15, optimum risk segregation

was obtained, resulting in seven representative genes. The gene voting model created for the genes *CLEC1B, CLEC3A, CTSB, NLRP10, NLRP9, TNIP1, and SARM1* obtained HR 9.14 and p = 1.49X10^{-12}. The network was built with 'Cytoscape 3.7.1' and depicts seven separate clusters with their corresponding genes in Figure 6.5.

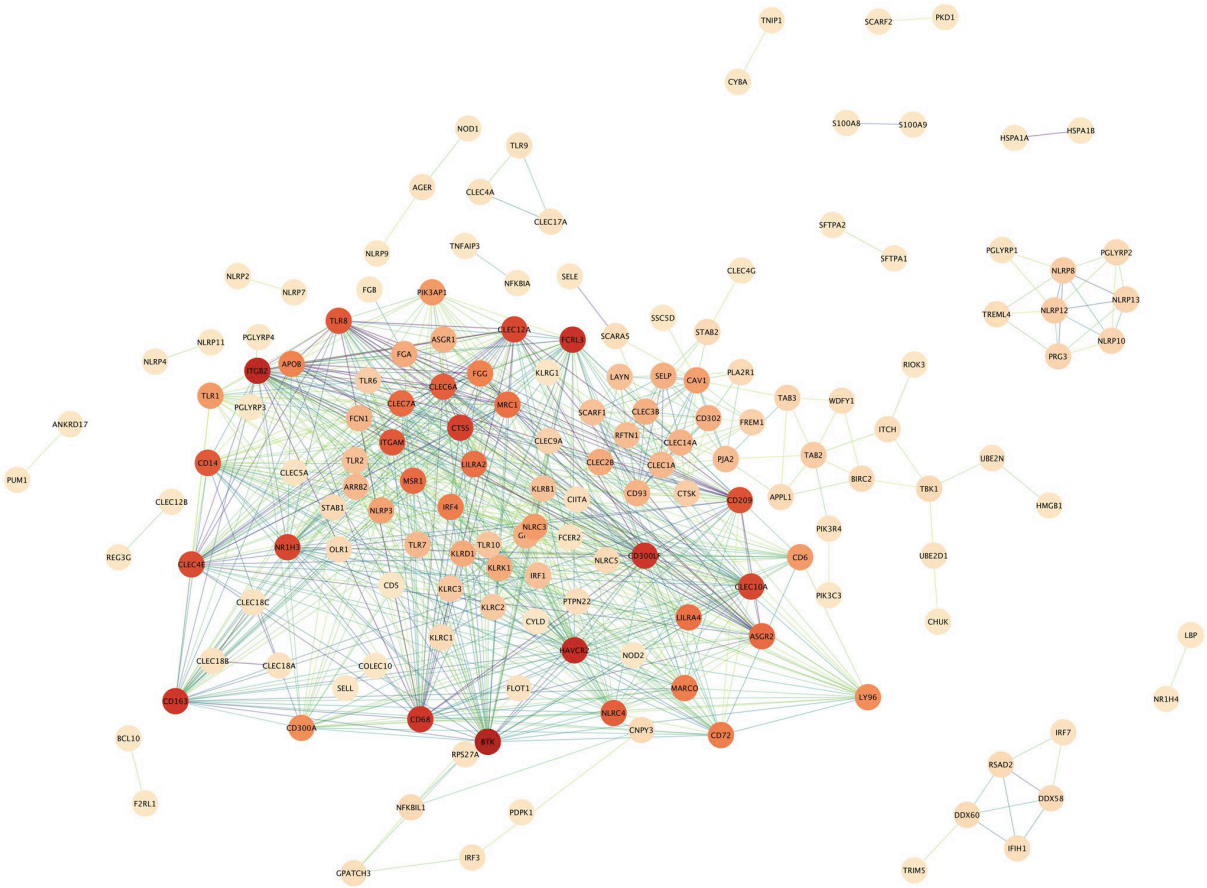


Figure 6.4 Pattern recognition receptor gene co-expression network analysis Interconnection of 15 hub genes; darker colour indicates a higher degree score, and darker edge indicates a greater clustering coefficient value.

6.3.3 Multiple Gene-Based Models for Risk Estimation

Using the expression profile of survival-associated PRR genes, several risk classification models based on machine-learning-based regression (MLR), PI, and gene voting were developed (based on p value). We examined 15 major genes in various combinations and discovered that a combination of nine genes performed best: *CLEC1B, CLEC3A, IRF7, CTSB,*

FCN1, RIPK2, NLRP10, NLRP9, and *SARM1*. Table 6.5 displays the findings of the various risk prediction models for these nine genes. The gene voting-based model outperformed the others: HR = 10.70 and $p = 10^{-12}$. This model also had the highest concordance index value of 0.76, and the high-/low-risk group survival curves were significantly separated with a log-rank $p=10^{-14}$. Figure 6.6 depicts the KM plot displaying the survival curves for the two risk categories. While the 5-year survival rate for low-risk individuals was nearing 85%, it dropped as low as 15% for high-risk patients. The PI-based model came in second, with an HR of 3.41 and $p\sim10^{-3}$, while the regression-based linear model came in third (and first among the MLR models), with an HR of 1.66 but a p value that was not statistically significant.

Table 6.3 Gene voting model results for selected representative genes from each cluster. The set of representative genes for the best model is shown by bold formatting.

S.no	Genes	Clusters (n)	HR	p Value	C-index	95% CI	log-rank (p)
1	*CLEC1B, CTSB, NLRP10, UBC, LTF*	5	3.53	3.74E−05	0.64	1.94–6.43	6.84E−05
2	***CLEC1B, CTSB, NLRP10, CLEC3A, UBC, ESR1, LTF, UBB, LGALS3BP, MAPKAPK2***	**10**	**4.11**	**4.38E−06**	**0.64**	**2.25–7.51**	**1.57E−05**
3	*CLEC1B, UBC, LTF, CTSB, NLRP10, UBB, LGALS3BP, S100A9, CLEC3A, RPS27A, APPL1, HSPA1A, SARM1, NLRP9, TNIP1, S100A9*	15	2.93	8.00E−04	0.62	1.56–5.49	5.41E−04
4	*CLEC1B, UBC, LTF, CTSB, NLRP10, DMBT1, LGALS3BP, S100A9, CLEC3A, RPS27A, APPL1, HSPA1A, SARM1, NLRP9, TNIP1, HSPD1, UBB, CYBA, FLOT1, HMGB1*	20	3.64	2.06E−05	0.66	2.01–6.59	2.06E−05
5	*CLEC1B, UBC, LTF, CTSB, NLRP10, HSPA1A, LGALS3BP, S100A9, CLEC3A, RPS27A, BIRC3, MRC2, CNPY3, NLRP9, TNIP1, HSPD1, UBB, CYBA, MRC1, HMGB1, VCAN, CFI, S100A8, FCN1, SARM1*	25	3.58	4.87E−05	0.64	1.93–6.61	2.86E−05

Table 6.4 Results of gene voting model for chosen significant representative genes only from each cluster

S.no	Representative genes	Clusters (n)	HR	p Value	C-index	95% CI	Log-rank (p)
1	*CLEC1B, CTSB, NLRP10*	5	3.62	1.12E−04	0.6	1.88–6.96	4.87E−04
2	*CLEC1B, CTSB, NLRP10, CLEC3A*	10	4.25	1.63E−06	0.53	1.30–13.84	4.70E−02
3	***CLEC1B, CLEC3A, CTSB, NLRP10, NLRP9, TNIP1, SARM1***	**15**	**9.14**	**1.49E−12**	**0.73**	**4.95–16.87**	**6.64E−12**
4	*CLEC1B, CLEC3A, CTSB, NLRP10, NLRP9, TNIP1, SARM1*	20	9.14	1.49E−12	0.73	4.95–16.87	6.64E−12
5	*CLEC1B, CLEC3A, CTSB, NLRP10, NLRP9, TNIP1, SARM1, MRC1, FCN1*	25	5.46	3.05E−08	0.67	2.99–9.95	9.00E−08

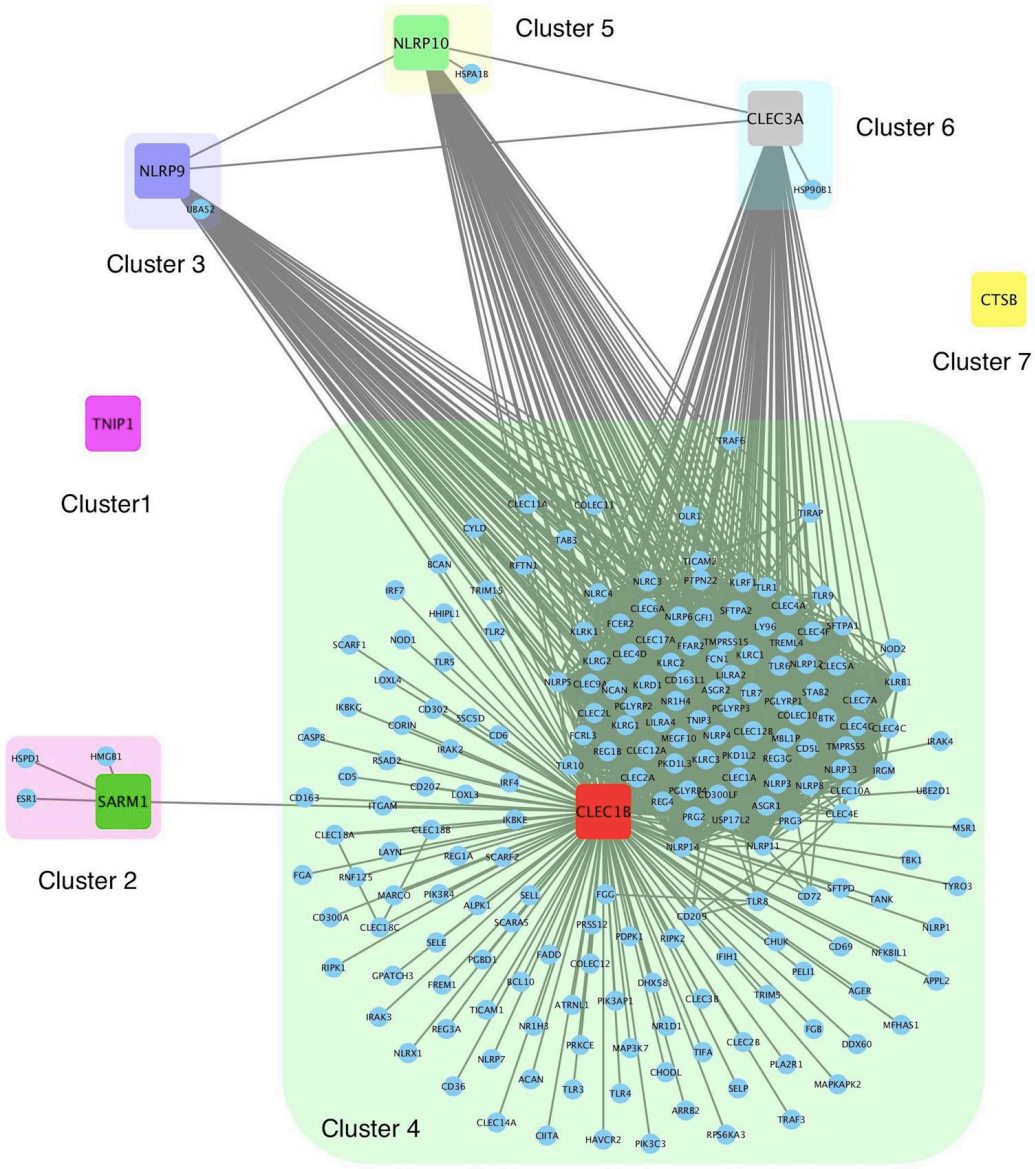


Figure 6.5 Clustered network: seven distinct clusters are depicted in various colours. Large squares emphasise and represent representative pattern recognition receptor genes.

Table 6.5 Different models' performance for using multiple gene expression profiles

Model	HR	p Value	C-Index	95% CI	Log-rank (p)
Voting	**10.7**	**1.13E−12**	**0.76**	**5.57–20.55**	**8.15E−14**
PI	3.41	9.70E−03	0.6	1.35–8.66	2.59E−03
Linear	1.66	1.00E−01	0.56	0.91–3.03	9.81E−02

Ridge	0.99	9.86E−01	0.52	0.55–1.79	9.86E−01
KNN	0.83	5.24E−01	0.52	0.46–1.49	5.24E−01
Elastic net	0.79	4.45E−01	0.55	0.44–1.43	4.47E−01
Random	0.77	3.85E−01	0.55	0.43–1.39	3.86E−01
Lasso	0.65	1.57E−01	0.56	0.36–1.18	1.53E−01
SVR	0.65	1.57E−01	0.56	0.36–1.18	1.53E−01
Lasso Lars	0.65	1.57E−01	0.56	0.36–1.18	1.53E−01

The use of bold font indicates statistically significant results (p value, log-rank $p<0.05$). CI confidence interval, C-index concordance index, HR hazard ratio, KNN K-nearest neighbours, PI prognostic index, and SVR support vector regression.

6.3.4 Sub-stratification of patients in clinico-pathological high-risk

Previous research has suggested that clinicopathological factors such as histologic diagnosis, ethnicity, clinical stage (Coll-de la Rubia et al. 2020), menopausal status, peritoneal washing, and so on] have a role in UCEC prognosis (H. Zhou et al. 2020). As a result, we conducted a univariate analysis to examine the relationship between these factors and OS in our dataset. The findings of the univariate analysis are shown in Table 6.6. Clinical factors influencing UCEC prognosis were clinical stage, residual tumour, peritoneal washing, grade, histologic grade, and menopausal status. As demonstrated in Figure 6.7, the gene voting model was able to stratify high-risk UCEC patients based on clinicopathological factors such as histologic diagnosis, peritoneal washing, menopausal status, neoplasmic grade, residual tumour, and clinical stage. The KM plots, together with low log-rank p values, show a substantial difference between high- and low-risk patients.

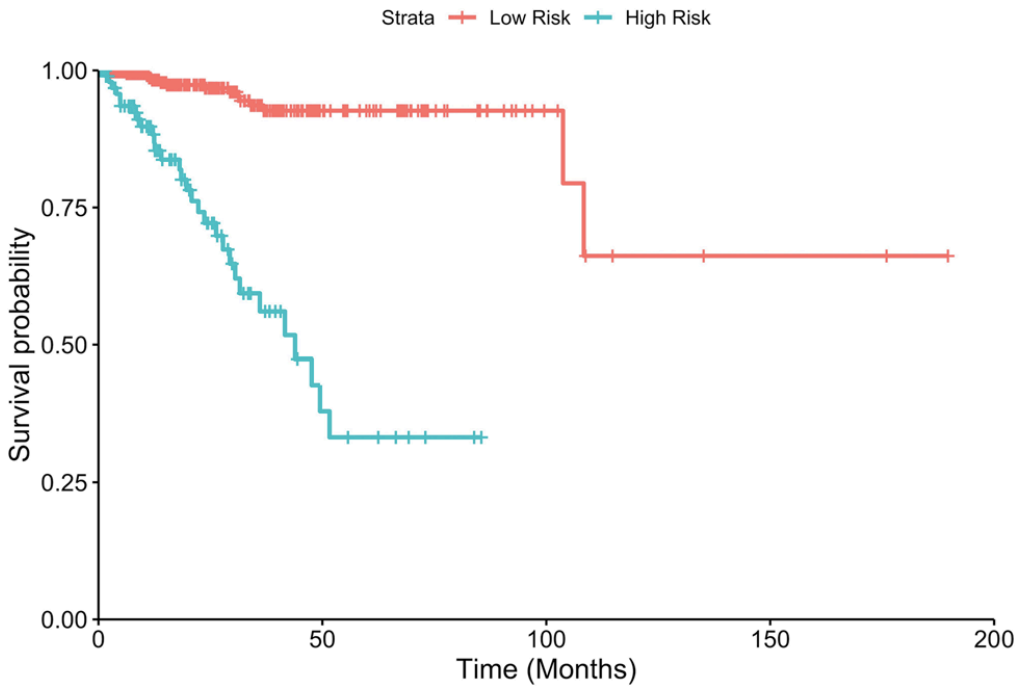


Figure 6.6 shows a Kaplan–Meier plot based on gene voting to highlight the risk stratification of patients with uterine corpus endometrial cancer. Patients with more than four 'high-risk' labels in the ten-bit risk vector are classified as high risk (hazard ratio 10.70, $p = 1.13X10^{-12}$, C = 0.76, log-rank-$p = 8.15X10^{-14}$), whilst others are classified as low risk (red).

Table 6.6 Univariate analysis using clinico-pathological features. Clinical stage, residual tumor, peritoneal washing, grade, histologic washing, menopause status are seen to be the significant factor.

Clinical Factors	Strata	n	HR	p-value	C-INDEX	95%CI(L)	95%CI(U)	Log-rank (p)
Clinical stage	III, IV vs I,II	541.00	4.44	1.31E-06	0.71	2.43	8.11	8.81E-07
Residual tumor	R0 vs R1,R2	411.00	0.29	8.06E-04	0.59	0.14	0.60	2.27E-03
Peritoneal washing	negative vs positive	407.00	0.31	2.42E-03	0.58	0.15	0.66	5.38E-03
Grade	G3, High grade vs G1,G2	541.00	3.27	2.46E-03	0.61	1.52	7.03	7.24E-04
Histologic_diagnosis	MSE, SEA VS EEC	541.00	2.29	6.61E-03	0.56	1.26	4.16	8.65E-03
Menopause status	Pre vs post	512.00	2.54	3.51E-02	0.53	1.07	6.02	5.92E-02
Age	>64 vs <=64	541.00	1.30	3.79E-01	0.54	0.72	2.33	3.79E-01

Race	White vs others	511.00	1.15	6.83E-01	0.50	0.58	2.29	6.79E-01
Ethnicity	Hispanic or latino vs Not hispanic not latino	388.00	1.44	7.23E-01	0.50	0.19	10.61	7.37E-01
History other malignancy	No vs Yes	541.00	1.18	7.84E-01	0.49	0.36	3.81	7.79E-01
Surgical approach	open vs minimally invasive	519.00	1.06	8.71E-01	0.51	0.54	2.07	8.71E-01

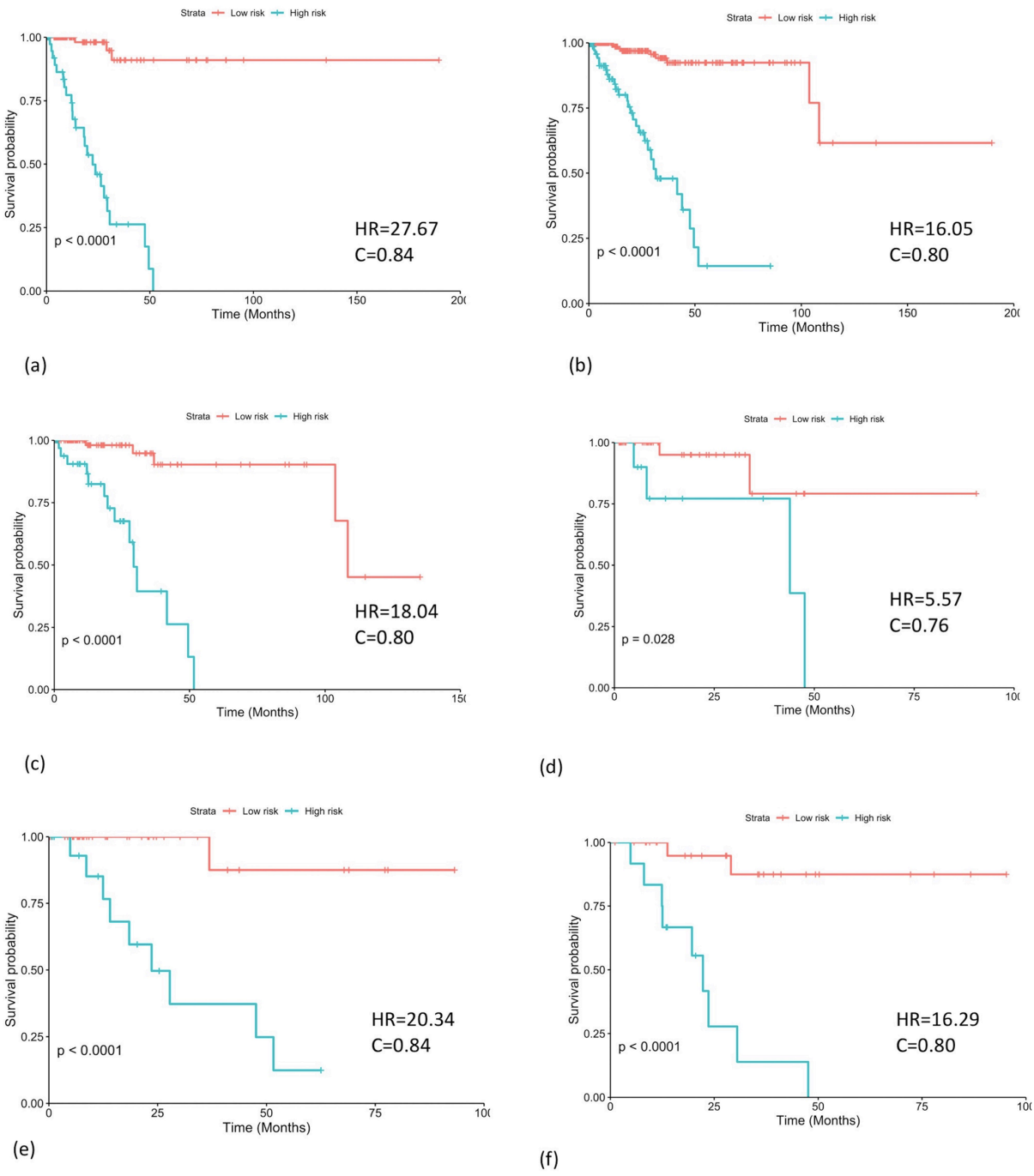


Figure 6.7 Sub-stratification of high risk groups using gene voting model (a) clinical stage (b) neoplasmic grade (c) histologic diagnosis (d) menopause status (e) peritoneal washing (f) residual tumour

6.3.5 Multivariate Analysis

We conducted a multivariate Cox regression survival analysis with seven major prognostic variables, including multiple gene voting model, clinical staging, residual tumour, peritoneal washing, histological subtype, menopausal status, and grade. The p-value for the gene voting model (HR 8.17; p<0.001) and the clinical stage (HR 3.11; p = 0.03) was significant, but not for the others, as shown by the forest plot in Figure 6.8. As a result, a hybrid model may be created utilising the gene voting model and the clinical stage to increase risk stratification even more.

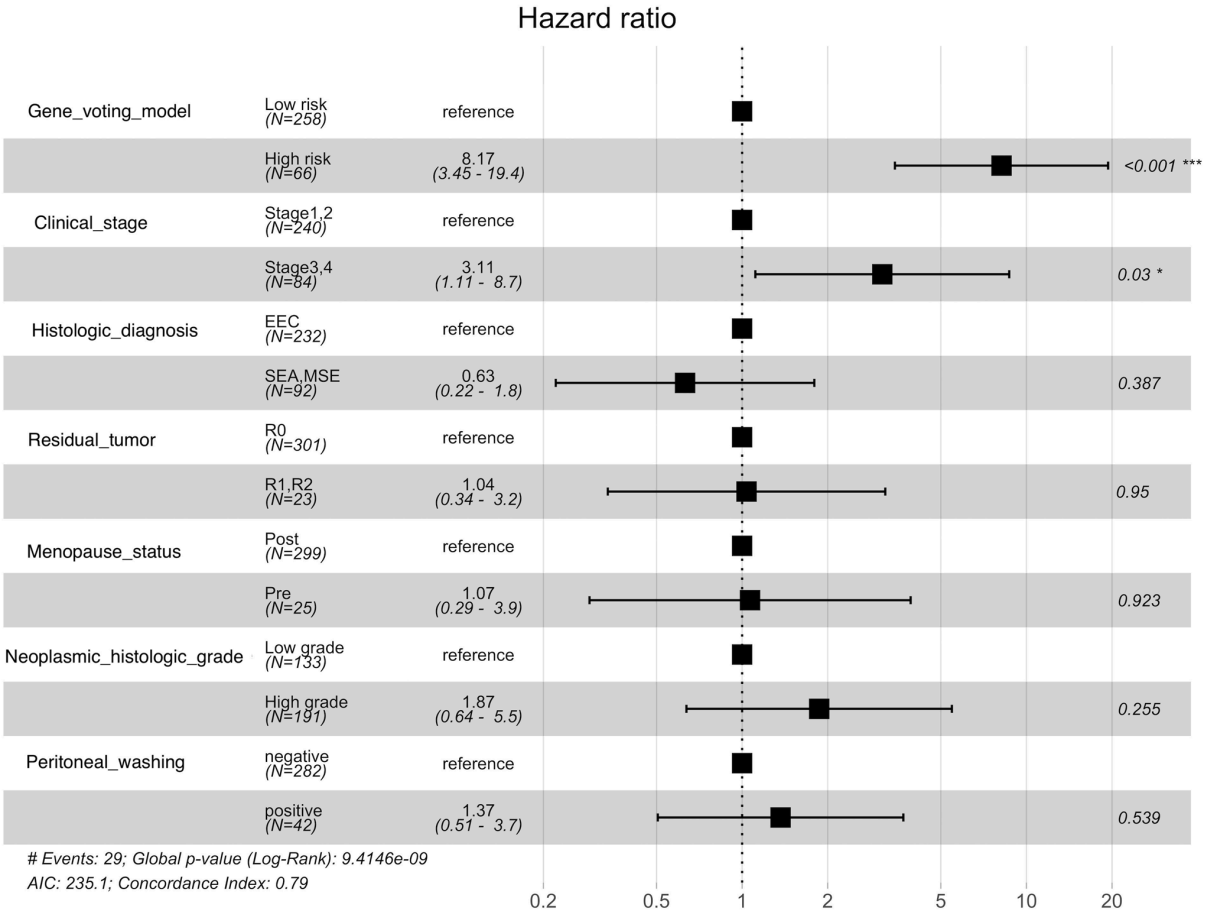


Figure 6.8 Multivariate analysis identifies gene voting model (hazard ratio 8.17; p 0.001) and clinical stage (hazard ratio 3.11; p = 0.03) as independent factors.

6.3.6 Hybrid Voting Model

Based on a multivariate Cox regression survival analysis, we built a hybrid voting model after obtaining the independent variables, i.e. multiple gene voting model and clinical stage. For risk stratification, this model integrated clinical stage with the nine gene voting models. As a result,

the risk vector for each patient was a 10-bit vector, with 1 bit given to the risk label according to the clinical stage. The model outperformed the nine-gene voting model (HR 15.23; p = $2.21X10^{-7}$, concordance index = 0.78, log-rank-p = $2.76X10^{-17}$). Figure 6.8 depicts the KM plot for the hybrid model.

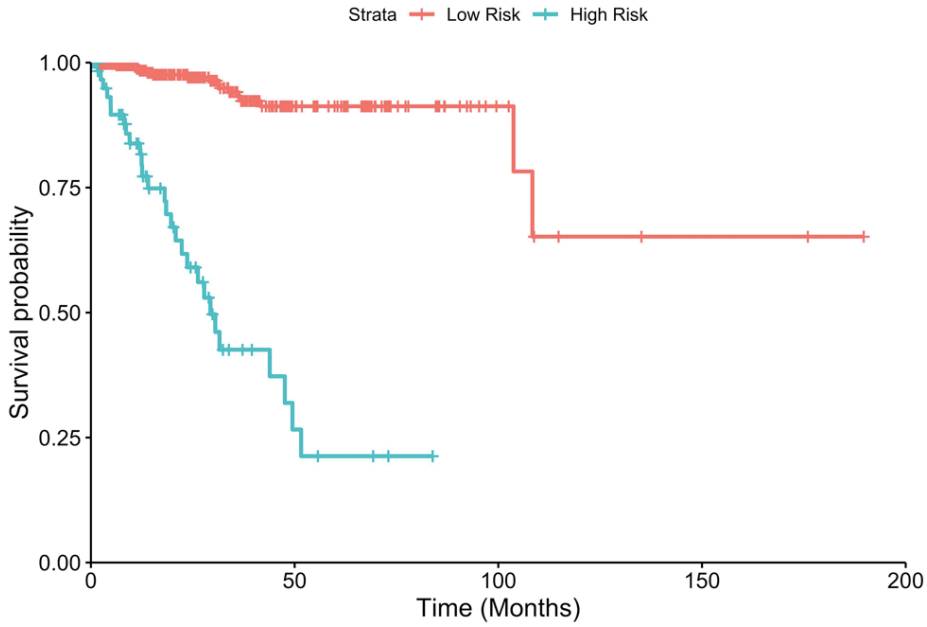


Figure 6.9 Using a nine-gene voting model and clinical stage, a hybrid model for risk stratification was developed (hazard ratio 15.23; p = 2.21 10 7, C = 0.78, log-rank-p = 2.76 10^{-17}).

6.3.7 Predictive Validation

We used subset of the samples of the complete dataset to undertake a predictive assessment of our gene voting model, as applied in Zhao et al. (Zhao et al. 2020) With 100 iterations, sampling sizes of 50%, 70%, and 90% were chosen. For each iteration of the gene voting model and the hybrid model, the HR and concordance index were calculated. Figure 6.10 depicts the boxplots that correlate to these results. Despite the small sample size, the median HR (15.34, 15.32, 15.02) and concordance index were higher. This strategy ensured that the risk stratification models were resilient and functioned well with randomly generated datasets of varying sizes.

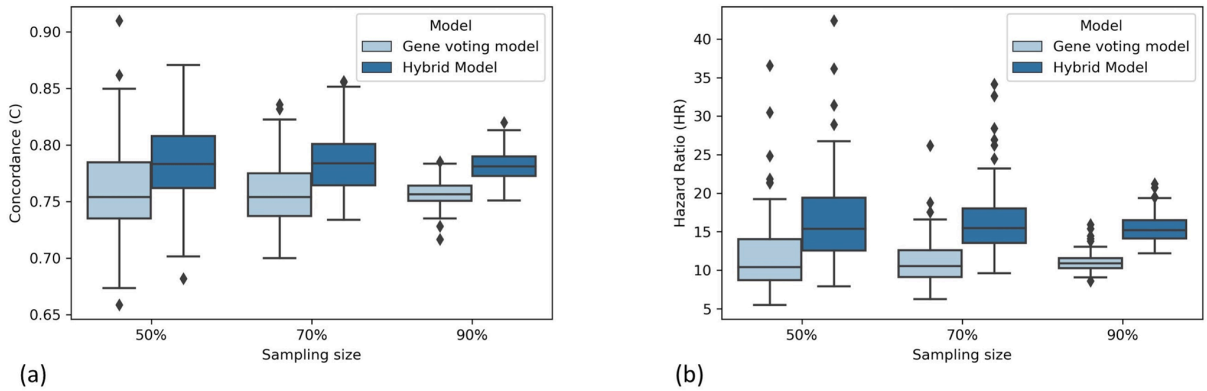


Figure 6.10 Validation of a voting-based model using predictive analytics. (a) For 100 rounds of data sampling, grouped boxplots matching to the estimated Concordance index (y-axis) (x-axis). (b) Similarly, using random sampling, the hazard ratio (y-axis) for several models is estimated (x-axis)

6.3.8 Hybrid Voting Model for Classification

The area under the receiver operating characteristic curve (AUROC) value was used to assess the hybrid model's performance. The true-positive and false-positive rates were calculated using the 'survivalROC' package as shown in Table 6.7. In this case, a prediction was termed a true positive if the OS was larger than the cut-off time and the patient was in the model's low risk category; the opposite was true for a true negative prediction, as illustrated in Figure 11a. The average OS time was 1.1 years. We discovered that, among various OS cut-offs, the model performed best at the 4.3-year cut-off. This cut-off is an excellent predictor of high- and low-risk patients. The classification based on the hybrid voting model produced an AUROC score of 0.86 using this cut-off. The ROC curve for this is depicted in Figure 11b.

Table 6.7 True positive and False positive at different threshold.

Threshold	TP (True Positive)/ Precision	FPR (False Positive)/ Recall
0	1.000	1.000
0.1	0.997	0.996
0.2	0.982	0.935
0.3	0.962	0.719
0.4	0.959	0.395
0.5	0.683	0.150
0.6	0.470	0.033

0.7	0.122	0.012
0.8	0.071	0.000
0.9	0.000	0.000

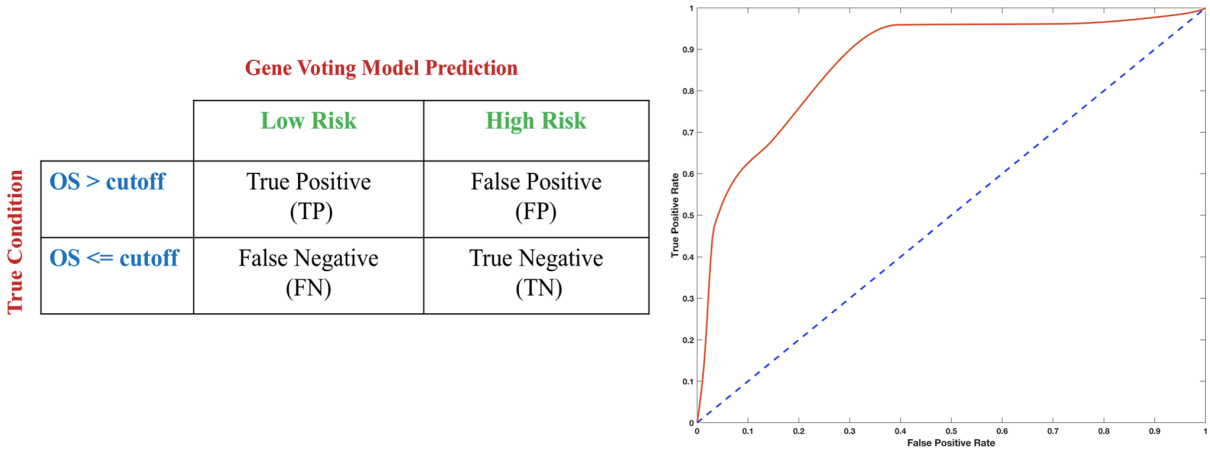


Figure 6. 11 (a) Terminology for assessing the confusion matrix. The first risk labelling was done using an overall survival (OS) cut-off, with patients with OS > cut-off labelled as positive or low risk, and patients with OS cut-off labelled as negative or high risk. (b) A receiver operating characteristic (ROC) curve area under the ROC curve (AUROC) of 0.86 was achieved for the gene voting model.

6.3.9 Therapeutic Agent Screening

Following the identification of genes that play a significant role, the selection of therapy is a critical step. Using the "Cmap2" database (Musa et al. 2018), we retrieved drug molecules that might re-modulate the overexpressed and under expressed genes, as Shen et al. (Shen et al. 2020) did. We used Cmap2 to query a list of probe identifiers matching to upregulated genes (*CLEC1B, CLEC3A, FCN1, NLRP10)* and downregulated genes *(RIPK2, SARM1, IRF7, CTSB, NLRP9*). Hexamethonium bromide (enrichment 0.834; p = $2.6X10^{-4}$) and isoflupredone (enrichment 0.955; p = $2.2X10^{-4}$) were the top positive and negative enriched compounds, respectively. It is known to be poorly absorbed from the GI tract and to not penetrate the blood–brain barrier. Isoflupredone is a synthetic glucocorticoid corticosteroid also known as delta-fludrocortisone and 9-fluoroprednisolone.

6.3.9.1 Drug Repurposing

We also looked at how approved drugs may aid prolong UCEC survival. We were able to extract five GPM and four BPM genes. The molecular processes of GPM genes (*RIPK2, SARM1, IRF7, CTSB, NLRP9*) that have a favorable prognosis in UCEC can be explored to identify positive regulators/inducers that might be employed as possible therapies. Inducing IRF7 can result in the generation of interferon (IFN)- and IFN-like proteins These play an important role in anti-tumor immunity and homeostasis and have been approved for use in the treatment of cancer by the US Food and Drug Administration (Abdolvahab et al. 2020). IFN is known to cause apoptosis in tumor cells in response to diverse stimuli and to stimulate dendritic cell differentiation (Gogas et al. 2006). IFN, on the other hand, activates immune cells such as macrophages and natural killer (NK) cells, which may enhance the anti-tumor impact (X. Q. Qin et al. 1998). Hairy cell leukaemia (Gogas et al. 2006), follicular lymphoma (Rohatiner et al. 2005), and renal cell carcinoma (Locatelli et al. 1999) have all been treated with it. Paclitaxel and imiquimod are *TLR4* and *TLR7* positive regulators, respectively (Solinhac 1985; Novak et al. 2008), and can be employed as therapeutic medicines for the indirect stimulation of *IRF7* and hence IFNs. Agonists such as lipopolysaccharide and imidazoquinoline anti-viral chemical single strand RNA may also aid in the activation of the signalling pathway and the generation of IFN. *CTSB,* which is important in the NLRP signalling pathway, can be favorably regulated by a variety of agonists, including pathogen-associated molecular patterns/damage-associated molecular patterns. Different interleukin (IL) therapies may potentially be a viable option in UCEC instances. *RIPK2/RIP2* is involved in the nucleosome-binding oligomerization domain-signalling pathway and induces the production of various pro-inflammatory cytokines (IL-1, IL-6, tumour necrosis factor-, IL-18), chemokines (monocyte chemoattractant protein-1, chemokine [C-X-C motif] ligand), and antimicrobial peptides (cathelicidin, defensins).

MESO-DAP1 and mifamurtide (licenced by the FDA) positively control *NOD1 and NOD2*, and so potentially act as indirect regulators of *RIPK2* (Hasuo and Akasu 2001). Bacterial peptidoglycan can potentially be used as an agonist to stimulate the pathway. The BPM genes (*CLEC1B, CLEC3A, FCN1, and NLRP10*) that exhibited poor prognosis in our analysis must be suppressed. Protein kinase B (AKT) abnormal overexpression has been seen in ovarian, lung, and pancreatic malignancies and is related with enhanced cancer cell proliferation and survival. As a result, targeting AKT might be a critical technique for cancer prevention and treatment. In osteosarcoma, the inhibition of *CLEC-3A* decreases cell proliferation and

promotes chemo sensitivity via the AKT/mTOR/hypoxia-inducible factor (HIF)-1 pathway (Ren et al. 2020). MK-2206 (NCT 01333475) is a synthetic AKT inhibitor in clinical phase II, whereas resveratrol is in clinical phase I (M. Song et al. 2019). If the studies are successful, they might be utilised as AKT inhibitors. These findings supported the biological feasibility of repositioning cancer medicines for EC treatment.

6.4 Discussion

Although UCEC has a fair prognosis if detected early, individuals with late stages have a dismal prognosis and a significant mortality rate. As a result, effective risk assessment methodologies are necessary for clinical decision making and therapeutic intervention. Clinical features such as tumour grade, cervical involvement, lymph node status, histological subtype, depth of myometrial invasion, and LVSI are important in UCEC risk categorization but inefficient due to their limitations. Various molecular prognostic indicators have therefore been proposed, assisted by the advent of high-throughput sequencing tools and the availability of a large quantity of experimental data. Previous research revealed many molecular pathways that lead to sophisticated molecular processes that are critical for cancer growth and development, such as the explanation of diverse signalling processes regulated by PRRs. This elucidation of PRRs' regulatory role has aided treatment decision-making in a variety of malignancies. PRR agonists are now employed as vaccine adjuvants in possible systemic therapies such as chemotherapy, targeted therapy, and immunotherapy (Bai et al. 2020) , such as AS15, which is a TLR4/9 agonist and is used as an adjuvant to vaccines dHER2 (truncated version of HER2) and lapatinib in breast cancer (Milani et al. 2013). Monophosphoryl lipid A, which is related with TLR4, has been found to be a strong vaccination adjuvant and to stimulate a type 1 T helper (Th1)-based immune response in human papillomavirus-induced cervical cancer (Gregg et al. 2017). Imiquimod in combination with TLR7 has been shown to promote apoptosis and trigger a cell-mediated immune response in basal cell carcinoma (Bubna 2015). When combined with cyclophosphamide, OM-174, a synthetic derivative of lipid A, is said to inhibit tumour development and prolong life in melanoma (D'Agostini et al. 2005). AS04, a TLR4 agonist, is licenced by the FDA for cervical cancer; imiquimod, a TLR7/8 agonist, is approved for different skin malignancies; and mifamurtide, a NOD2 agonist, is approved for osteosarcoma. The combination of PRR-based agonist treatment with immune checkpoint-targeted antibodies (such as anti-cytotoxic T-lymphocyteassociated protein-4 or anti-

programmed cell death ligand 1) may represent the future of cancer therapy (Bai et al. 2020). However, the significance of PRR signalling genes and their efficacy in UCEC treatment remains unknown. We employed messenger RNA (mRNA) expression data from the TCGA-UCEC cohort in this investigation. First, we used survival analysis to quantify each PRR gene's predictive performance. To identify essential PRR genes, we employed gene co-expression network-based feature selection and a clustering-based technique. The genes collected from this technique were then used to develop risk classification models. To improve the model's performance, we discovered 15 PRR-related biomarker genes linked with UCEC prognosis using Cox-regression survival analysis: *CLEC1B, CLEC3A, MRC1, IRF7, CTSB, FCN1, RIPK2, CLEC3B, CLEC12B, TLR4, NLRP10, NLRP9, TNIP1, SARM1, and MAPKAPK2.* A nine-gene *(CLEC1B, CLEC3A, IRF7, CTSB, FCN1, RIPK2, NLRP10, NLRP9, and SARM1)* voting-based model performed best and strongly classified high-risk clinical groups. Finally, following a thorough predictive comparison with other clinicopathological markers, we established a hybrid model that combines the expression patterns of nine genes with 'clinical stage' to accurately identify high- and low-risk individuals with UCEC. We also predicted candidate biomolecules that may alter gene expression and potentially act as medications in the therapy of UCEC.

The nine key biomarker genes identified (*CLEC1B, CLEC3A, IRF7, CTSB, FCN1, RIPK2, NLRP10, NLRP9, and SARM1)* have been found to have important regulatory roles in a variety of illnesses, including cancer. *RIPK2, SARM1, IRF7, CTSB, and NLRP9* were associated with a favourable prognosis, whereas *CLEC1B, CLEC3A, FCN1*, *and NLRP10* were associated with a bad prognosis. *CLEC1B*, also known as C-type lectin domain family 1 member B, has been proven to be effective in the treatment of atherosclerosis. *CLEC1B* plasma concentrations were shown to be directly related to an increased risk of carotid plaque development. The odds ratios (ORs) for platelet-derived growth factor receptor- were 0.79 (95 percent CI 0.66–0.94; p = 0.008) for 1-standard deviation increase (Mosley et al. 2018). The -chain C-terminus of aggretin is a possible target for the therapy of tumour metastasis via *CLEC-2* inhibition (C.-H. Chang et al. 2014). *CLEC2* appears to reduce AKT signalling and gastric cancer cell invasion by inhibiting the production of phosphoinositide 3-kinase (PI3K) subunits (L. Wang et al. 2016). *CLEC3A* overexpression enhances tumour development and a poor prognosis in invasive ductal breast cancer. *CLEC3A* knockdown using RNA interference effectively decreased the proliferation, migration, and invasion of breast cancer cells, which may be

mediated by the PI3K/AKT signalling pathway. As a result, it has an anti-cancer impact and has the potential to be a crucial therapy for breast cancer (Ni et al. 2018). *FCN1* can be used as an additional biomarker in individuals with acute myeloid leukaemia (p = 0.004; OR 2.95; 95 percent CI=1.41–6.16). Polymorphisms in the *FCN1* gene are linked to increased expression of specific mRNA in monocytes and granulocytes, as well as greater FCN1 serum levels. *NLRP10* binds to an apoptosis-associated speck-like protein, inhibiting NF-B activation and apoptosis as well as caspase-1-mediated IL-1 maturation, and so aids in the control of apoptosis and inflammation (Zambetti et al. 2012). Through Bcl-2 family members, SARM mediates intrinsic apoptosis. It inhibits Bcl extra-large (Bcl-xL) and decreases phosphorylation of extracellular signal-regulated kinase (Panneerselvam et al. 2013). *SARM1* can function as an epigenetic biomarker in colorectal cancer, aiding in the detection of cancer through underexpression and/or CpG methylation (Quyun et al. 2010). In cancer, *IRF7* promotes the formation of granulocytic myeloid-derived suppressor cells via S100A9 trans-repression (Q. Yang et al. 2017). The functional polymorphism rs12898 in cathepsin B (*CTSB*) may contribute to the susceptibility to primary hepatic cancer, and the variation A allele may enhance the risk of the cancer (Cui et al. 2019). *NLRP9* functions as an inflammasome-related molecule, making it a reliable non-invasive method for diagnosing breast cancer (Mearini et al. 2017). The measurement of *NLRP4 and NLRP9* expression may be useful in predicting Bacillus Calmette-Guerin failure as well as in making decisions about early radical surgery (Poli et al. 2017). One research (Manjang et al. 2021) discovered that by removing biomarker genes (and all genes involved in the same biological process), other signatures with the same prognostic prediction capabilities but opposing biological meaning could always be discovered. We conducted a similar analysis on our first list of 15 major biomarker genes to see whether the alternate collection of biomarker genes was similarly successful in our study. We picked the remaining set of six alternate biomarker genes for model development instead of the nine major biomarker genes (included in our final risk-prediction model). The highest performing model of the possible combinations of these six genes, with HR 4.65 and p value $1.09X10^{-5}$, consisted of five genes. These findings revealed that our final model, which included nine major biomarker genes, outperformed the competition. To answer our next question about the (co-expression-based or biological) correlation between main biomarker genes and alternate biomarker genes, we discovered that the majority of the alternate biomarker genes correlated with one of the nine main biomarker genes, *with MRC1* having a 0.45 correlation with *FCN1*. This explains why

the alternate collection of biomarker genes, rather not only the primary set of biomarker genes, linked with survival. We also attempted to discover the biological links between these nine and five biomarker genes. We observed that most of these genes have overlapping biological pathways. Thus, we concluded that the alternate genes are correlated with the main biomarker genes both in expression and in shared biological mechanisms. Therefore, the model with the main biomarker genes is the optimal choice in our case. In an additional analysis, we clustered the samples into three groups (high risk, mid risk, and low risk) according to their OS time. Thereafter, the expression-based correlation (PCC) amongst gene pairs were evaluated for these three groups. We then took the top ten highly correlated gene pairs from the low-risk group and examined whether the high correlation in these gene pairs was maintained in the mid- and high-risk groups. We observed that PCC values in all gene pairs decreased in the mid-risk group. It further decreased in the high-risk group. These findings suggest assumption there is a strong link between strongly correlated gene pairs agitated in high-risk patients. In other words, it's either low or none. PCC in these gene pairs may be linked to worse health. Patients with UCEC have a higher chance of survival. These observations indicate that the correlation between highly correlated gene pairs is disturbed in high-risk patients. In other words, low or no PCC in these gene pairs is potentially associated with poorer survival of patients with UCEC. These nine biomarker genes have the potential to be used as therapeutic targets for therapy. This can be accomplished by modifying the expression of these genes to produce the desired low-risk profile. In addition to in silico prediction of small compounds that achieve this aim, we investigated the molecular involvement of many of these candidate genes and therefore proposed novel agonists for medicinal repurposing. This study is acceptable, provides a realistic foundation for future research, and may be used as a reference for additional experimental attempts. To avoid the risk of data bias and establish the robustness of our model, we used a predictive validation approach in which the model's performance was evaluated by repeated sampling of the dataset across different sample sizes. However, in order to define their functions and offer a foundation for possible therapeutic use, the biomarkers identified in this study must be verified in external cohorts and may be combined with other known biomarkers. Furthermore, because to a lack of data on metastasis and recurrence in the TCGA database, we were only able to utilise OS to analyse patient prognosis, which is another limitation of our analysis.

6.5 Conclusion

A risk classification model based on nine PRR-related genes was devised to evaluate survival outcomes and provide personalised anticancer treatment in patients with UCEC. Our findings also imply that combining the suggested gene signature with clinical staging gives better prognosis than staging information alone. As a result, the findings of our study may serve as a potential change to UCEC's existing risk evaluation method. Overall, this work contributes to a better understanding of the oncogenic function of pattern recognition receptors in UCEC.

7

PRRs Based Universal Prognostic Biomarkers

7.1 Introduction

"Cancer" is the main cause of mortality globally, and its progression has been linked to a variety of regulatory variables (Sever and Brugge 2015). In recent years, there has been a lot of interest in researching the regulatory systems that contribute to cancer. A variety of biomarkers and risk prediction methods have emerged as a result of the exploration of these processes. The biomarkers/ methods developed in the past are apt for single specific cancer but they fail when utilized for another cancers. Nevertheless, some pan-cancer prognostic biomarkers have appeared as omics data has expanded. Some examples list an extensive analysis of multi-omics from 13 cancers and 7 genes were found to be associated with them (Zhao et al. 2020). In case of 8 cancers using mRNA expression of "*Siglec-15*" a risk stratification (HR=3.03, p=0.044) was achieved in THCA patients (B. Li et al. 2020) and further study that found that the mRNA expression levels of the gene, "Long intergenic non-coding" (D. Wang et al. 2020). Studies from the past also elaborated that the prognostic significance of genes such as "*WISP1*" which shows different expressions in cancer and the adjacent normal tissue (Liao et al. 2020), "*FUNDC1*", which was proved to be associated significantly in 8 cancers with maximum prognosis in LIHC (Yuan et al. 2019), and "HSP90AA1", which act as prognostic biomarker in 8 cancers (W. Chen et al. 2020). Besides this, recently, research has been added that tumor mutational load and indel burden is linked with the prognosis in 14 cancers with maximum performance in CHOL (Wu et al. 2019). Although these researches are intriguing, developing a more reliable and accurate biomarker across various types of cancer remains an unsolved problem.

Pattern recognition receptors (PRRs) have long been known for their involvement in identifying microbial ligands and activating the immune system, as well as their pro and anti-tumor roles in cancer. TLR pathways are important regulators of chemo-resistance, potentially through activating NF-kB and increasing the production of the antiapoptotic protein Bcl-2 (Alvero et al. 2009). TLR receptors such as TLR3, 4, 7, and 9 expression corresponds with poor differentiation, high proliferation, and advanced stage cancer in oesophageal cancer (Sheyhidin et al. 2011). TLR5 expression is associated with a good prognosis in lung cancer, but TLR7 expression is associated with a bad diagnosis. TLR9 expression is related with poorer survival in renal cell carcinoma (Ronkainen et al. 2011) and is enhanced in the progression of glioma (C. Wang et al. 2010).

However, the horizon of these investigations was confined to specific tumors and a limited range of genes/proteins. Because PRR genes contain many regulatory genes/proteins, determining the prognostic importance of the most significant number of genes/proteins participating in immunomodulatory signaling across several cancers might provide a better understanding. It may potentially uncover new targets and aid in creating more effective biomarkers for cancer prognosis.

7.2 Methodology

7.2.1 Dataset Preparation

TCGA-Assembler-2 was used to extract 'The Cancer Genome Atlas' (TCGA) normalized gene expression datasets and the raw counts for pan cancer i,e 33 types of cancers (Wei et al. 2018). A “pan-cancer” dataset was created by integrating all of the samples with raw gene expression levels from 33 different cancers. The list of 331 pattern recognition receptor signaling pathway genes was taken from Gene Set Enrichment Analysis (GSEA) and HUGO Gene Nomenclature Committee (HGNC). The gene expression data for these 331 genes were taken from the TCGA datasets that were downloaded and from the pan-cancer datasets. We have taken only those patients' datasets for whom the overall survival and censoring information were provided. The accumulative total samples number in the dataset was 9569, and the total number of samples, which is denoted by N, is shown in Table 7.1.

7.2.2 Models for predicting survival

To test for survival-associated genes based on expression data, unadjusted “Cox proportional hazards” (Cox-PH) regression models were applied. The Cox-PH models were implemented using the R packages “survival” and “survminer”. “Hazard ratios” (HR) were calculated using this, as well as “confidence intervals” (% 95 CI) and “p-values”. HR is a hazard rate ratio that represents the death risk associated with one group compared to another by employing an appropriate gene-expression cutoff. We employed “Kaplan-Meier” (KM) plots and “log-rank tests” to compare survival curves between two risk categories. With HR more than or less than 1 and $p<0.05$, survival related genes were identified. The model's prediction performance was assessed using the Concordance (C) measure. According to (A Lathwal, Arora, and Raghava 2019), the “Prognostic Index” (PI) for n genes, “g1, g2,... gn”, with cox coefficients “β1, β2,

β3 … βn" derived from univariate Cox-PH analyses using median cut-offs was defined as PI=B.g, where "g=[g1 g2 g3.... gn]" and "B==[β1 β2 β3 … βn]". Following that, risk groups were separated using a univariate Cox-PH regression model. The cut-off value for PI was calculated in R using "cutp" from the "survMisc" package. HR, p, % 95 CI, and C values are used to estimate the model's performance. In addition, for an n-gene voting model, each patient sample is allocated an n-bit vector. Further, each bit is labeled as high or low risk using Cox-PH univariate models. This labeling is based on appropriate categorization by each gene. At last, the sample is assigned an overall risk label which is totally based on most of the labeled bits (i,e larger than n/2)

7.3 Results

7.3.1 Prognostic Biomarker Gene's Identification

Using each cancer's dataset, a "univariate Cox-PH" survival analysis was done on 331 genes. Genes were classified as either good ("GPM") or bad prognostic markers ("BPM"). Table 7.1 displays the number of survival-associated genes for each tumour, as well as additional information. It is shown that BPM genes outnumber GPM genes in the majority of cancers, demonstrating the negative significance of elevated expression of certain PRR genes in cancer. Table 7.1 also lists the top 10 genes (at most) for each tumour based on "p-values" from univariate survival analysis. None of the 331 genes were shown to be substantially related to survival in three types of cancers namely DLBC, TCGT, and PCPG.

Table 7.1 The table displays the total number of patient samples ("N"), the number of "BPM" and "GPM" genes, and the top 10 survival linked genes for 33 cancers.

Cancer	N	GPM	BPM	Total	Top Genes
ACC	79	27	15	42	*LGALS3BP,PKD1L2,NLRP4,ACAN,MAPKAPK3,PJA2,TNFAIP3,RNF125,CYLD,NLRP1*
BLCA	404	15	8	23	*IRF3,IKBKB,AGER,HHIPL1,CLEC2D,ATRNL1,ALPK1,UBQLN1,CLEC12A,UNC93B1*
BRCA	1091	38	14	52	*NFKBIA,TNIP1,TNIP2,CLEC2D,NLRP6,REG4,ZCCHC3,XIAP,ANKRD17,USP17L2*
CESC	304	26	12	38	*LOXL2,CD6,NR1H4,NLRC5,LILRA4,CD5,NLRC3,HHIPL1,SELL,CLEC10A*
CHOL	36	4	3	7	*C1QBP,TAB3,CD300LF,SSC4D,DGCR2,OLR1,TMPRSS4,*

COAD	297	5	8	13	*MBL1P,COLEC11,TMPRSS15,DGCR2,BCL10,SCARB2,MAP3K7,TAB1,IRF7,CLEC2A*
DLBC	47	0	0	0	-
ESCA	183	8	4	12	*UNC93B1,DHX58,PRSS12,LGR4,FGA,TAB3,ATRN,SFTPA2,HMGB1,CD248*
GBM	160	4	18	22	*CLEC4C,PIK3C3,SCARA3,RTN4,HSP90B1,CTSB,CTSL,NLRP12,CD248,NOD1*
HNSC	519	32	16	48	*PKD1,PRSS12,KLRB1,CD5,AGER,CACTIN,NLRC3,MAPKAPK2,FCRL3,NLRP1*
KICH	65	3	7	10	*TIFA,PGLYRP2,REG3A,LBP,SFTPD,NR1D1,APPL2,APPL1,IKBKB,SELE*
KIRC	532	52	89	141	*NPLOC4,TLR9,TRAF6,APPL1,LGALS3BP,CNPY3,IKBKG,PRKCE,PJA2,PKD1L2*
KIRP	287	8	26	34	*SFTPD,LBP,RIPK2,CLEC2L,TANK,LYN,UBE2D1,NLRP5,IRF1,STAB2*
LAML	173	18	27	45	*MAP3K1,CLEC11A,RFTN1,CD300A,TLR9,ESR1,IRAK1,CLEC18A,CLEC5A,MRC1*
LGG	511	31	112	143	*MYD88,DGCR2,CD302,MAP3K1,CLEC18B,CLEC18A,CIITA,IRGM,TAB1,CD69*
LIHC	369	13	15	28	*NLRC3,NLRP8,CD5,CLEC3B,TREML4,KLRK1,CLEC2L,FCRL3,NLRP9,ITCH*
LUAD	497	20	18	38	*NLRP10 ,TRIM15,UBE2N,LOXL2,NLRP2,AGER,FADD,NLRC3,CD302,CTSL*
LUSC	488	6	21	27	*CD14,RPS27A,FGA,CLEC18C,NLRP12,LOXL2,ESR1,DAB2IP,MAP2K6,S100A1*
MESO	86	18	25	43	*LOXL2,LAYN,OTULIN,RTN4,UBE2N,INAVA,CACTIN,KLRB1,PKD1L3,TRIL*
OV	305	22	0	22	*CASP8,HMGB1,NLRP4,CD300A,BIRC3,STAB1,HSPD1,LGR4,NLRP12,CACTIN*
PAAD	178	26	27	53	*ITCH,WDFY1,IRAK4,FADD,PTPRS,SARM1,BIRC2,ERBIN,IRAK2,TMPRSS4*
PCPG	179	0	0	0	-
PRAD	497	3	0	3	*RIOK3,PGLYRP1,TREML4*
READ	96	1	2	3	*IRAK1,PGBD1,ASGR2*
SARC	257	37	10	47	*RNF125,SELE,TICAM1,CLEC10A,SELP,IRF1,HPN,NLRP11,DHX58,LILRA4*
SKCM	449	78	11	89	*TLR4,KLRD1,CLEC6A,RSAD2,IFIH1,BIRC3,KLRC1,TLR8,KLRC2,TNIP3*
STAD	413	5	18	23	*NLRP14,CAV1,PJA2,NOD2,VCAN,TICAM1,TREML4,S100A9,SELE,REG1A*
TGCT	133	0	0	0	-
THCA	505	10	9	19	*CLEC4C,PKD1L3,TREML4,KLRB1,LAYN,FADD,CHODL,NOD1,CLEC4D,APPL1*
THYM	119	1	3	4	*PGLYRP1,KLRC1,CLEC4C,CD6*
UCEC	541	3	10	13	*CD68,NPLOC4,FLOT1,TYRO3,KLRG2,ESR1,KLRC3,CD163L1,HMGB1,PKD1L3*
UCS	57	5	5	10	*NLRP2,KLRF1,TICAM1,MAPKAPK2,NLRP14,LGALS3BP,HSP90B1,TREML4,MAP3K1, SMPDL3B*
UVM	80	0	43	43	*REG3G ,LBP,IRAK1,COLEC11,CD5L,RNF125,SCARB1,CLEC11A,IKBKG,CLEC12A*

7.3.2 Models for Predicting Specific Cancer

The best genes listed in Table 7.1 were used to build risk stratification models in 30 different cancers, except the cancers with no substantial genes in prognosis like TCGT, PCPG, and DLBC. To categorize patients into risk categories, both "gene voting-based" models and PI models were employed. The "HR, p-values, and C index" were then estimated. The voting models delivered the best outcomes, as shown in Table 7.2. We haven't shown the models for UVM, UCEC, PRAD and READ as the p-value wasn't significant. As it is clear from the table that THCA has the best prognostic model with HR=53.57, p=1.41E-04, whereas SKCM has the least performing prognostic model with HR=1.99, p= 2.33E-05.

Table 7.2 The table shows the performance of prognostic models for each specific cancer type.

Cancer	HR	p-value	logrank-p	C	%95 CI lower	%95 CI upper
THCA	53.57	1.41E-04	5.46E-09	0.84	6.89	416.23
KICH	37.79	6.34E-04	3.02E-06	0.86	4.70	303.63
ACC	15.11	1.58E-06	7.39E-09	0.78	4.99	45.80
THYM	12.73	2.27E-02	6.15E-03	0.80	1.43	113.60
CHOL	11.79	1.35E-03	6.08E-05	0.77	2.61	53.31
KIRP	7.56	8.65E-07	6.25E-08	0.75	3.38	16.94
LGG	4.83	1.04E-11	6.28E-13	0.74	3.07	7.61
COAD	4.76	1.81E-06	3.10E-06	0.71	2.51	9.02
UCS	4.39	3.12E-04	2.68E-04	0.65	1.96	9.82
MESO	4.24	9.53E-07	4.90E-07	0.68	2.38	7.54
BRCA	3.97	1.81E-10	1.83E-11	0.64	2.60	6.07
PAAD	3.88	1.82E-06	7.87E-07	0.68	2.22	6.78
KIRC	3.59	3.15E-13	9.69E-15	0.66	2.55	5.06
ESCA	3.55	1.34E-05	2.10E-05	0.65	2.01	6.29
BLCA	3.22	1.34E-08	3.63E-09	0.64	2.15	4.81

LUAD	3.15	1.14E-09	1.39E-09	0.66	2.18	4.55
LIHC	3.07	3.37E-07	8.17E-07	0.66	2.00	4.73
STAD	3.04	1.48E-06	8.76E-07	0.64	1.93	4.79
SARC	2.92	7.16E-06	4.97E-06	0.66	1.83	4.65
LAML	2.89	9.94E-08	6.35E-08	0.61	1.95	4.26
CESC	2.76	1.50E-04	1.12E-04	0.67	1.63	4.68
LUSC	2.49	2.61E-08	2.75E-08	0.61	1.81	3.44
GBM	2.48	8.38E-06	1.42E-05	0.62	1.66	3.70
HNSC	2.48	9.10E-08	2.48E-08	0.61	1.78	3.45
OV	2.13	2.31E-06	2.98E-06	0.60	1.56	2.92
SKCM	1.99	2.33E-05	2.39E-05	0.59	1.45	2.74

7.3.3 Universal Prognostic Biomarkers and Models

We obtained that among 30 types of cancers, 12 genes play a prognostic role in almost 6 cancers, at least 25 % of the total cancer data. The 12 genes biomarker signature composed of genes named *UNC93B1, ALPK1, APPL1, CASP8, CD5, CLEC2D, HMGB1, HSP90B1, IKBKG, IRF1, KLRB1, NCAN*. The role of these 12 genes as GPM or BPM in different types of cancer is shown through Figure 7.1. There is almost equal distribution of GPM and BPM roles of these genes among the cancers. *CD5* and *KLRB1* are GPM genes which means that their high expression is associated with Low-risk patients. At the same time, *HSP90B1* is a BPM gene which indicates that its high expression is directly linked to High-risk patients. Different models like the PI and the gene voting models were created using these 12 genes as biomarker panels in all cancers. The result for the gene voting model is displayed in Table 7.3. The model performed best in the case of THYM, KICH, and UVM. In the cancers like KIRP, STAD, and CHOL, the model's performance is moderate to poor. Thus, for these cancers, the prognostic biomarker, which is cancer-specific, should be relied on for better performance and risk stratification.

Table 7.3 The performance of the universal prognostic model for risk prediction in all cancers.

Cancer	N	HR	p-val	C	%95 CI L	%95 CI U	Logrank(p)
THYM	119	14.03	1.74E-02	0.68	1.59	123.79	4.07E-03
KICH	65	8.00	9.65E-03	0.73	1.66	38.59	3.50E-03
UVM	80	7.94	7.14E-03	0.74	1.75	35.96	1.18E-03
KIRP	287	4.06	1.43E-04	0.68	1.97	8.37	1.06E-04
STAD	47	3.62	2.01E-01	0.63	0.50	25.93	2.15E-01
CHOL	36	3.38	2.73E-02	0.68	1.15	9.97	2.18E-02
PAAD	178	2.91	7.68E-05	0.65	1.72	4.95	6.10E-05
OV	305	2.91	7.68E-05	0.65	1.72	4.95	6.10E-05
KIRC	532	2.91	7.68E-05	0.65	1.72	4.95	6.10E-05
PRAD	497	2.79	1.62E-01	0.71	0.66	11.79	1.53E-01
TGCT	497	2.79	1.62E-01	0.71	0.66	11.79	1.53E-01
PCPG	179	2.54	3.07E-01	0.54	0.42	15.25	3.00E-01
LGG	511	2.37	4.76E-05	0.64	1.56	3.59	4.86E-05
THCA	505	2.30	1.23E-01	0.47	0.80	6.65	1.20E-01
CESC	304	2.26	1.75E-03	0.63	1.36	3.76	1.91E-03
MESO	86	2.10	9.73E-03	0.55	1.20	3.70	9.81E-03
DLBC	47	2.04	4.75E-01	0.58	0.29	14.55	4.80E-01
SKCM	96	1.99	4.25E-01	0.59	0.37	10.87	4.08E-01
READ	96	1.99	4.25E-01	0.59	0.37	10.87	4.08E-01
ESCA	183	1.98	1.19E-02	0.60	1.16	3.38	1.23E-02
HNSC	519	1.94	3.23E-05	0.58	1.42	2.65	4.09E-05
BLCA	404	1.89	9.23E-04	0.60	1.30	2.76	9.95E-04
ACC	79	1.85	1.24E-01	0.59	0.84	4.06	1.26E-01
SARC	257	1.79	1.25E-02	0.59	1.13	2.82	1.23E-02
BRCA	1091	1.71	6.59E-03	0.59	1.16	2.52	6.45E-03
GBM	160	1.71	6.59E-03	0.59	1.16	2.52	6.45E-03
UCS	57	1.58	2.08E-01	0.54	0.77	3.24	2.13E-01
LUAD	497	1.48	3.00E-02	0.56	1.04	2.12	3.07E-02
COAD	297	1.48	3.00E-02	0.56	1.04	2.12	3.07E-02
LUSC	488	1.40	4.34E-02	0.54	1.01	1.95	4.64E-02
LAML	173	1.37	1.01E-01	0.55	0.94	2.00	1.05E-01

LIHC	257	1.02	9.27E-01	0.50	0.63	1.65	9.27E-01
UCEC	541	0.88	6.85E-01	0.51	0.48	1.62	6.84E-01

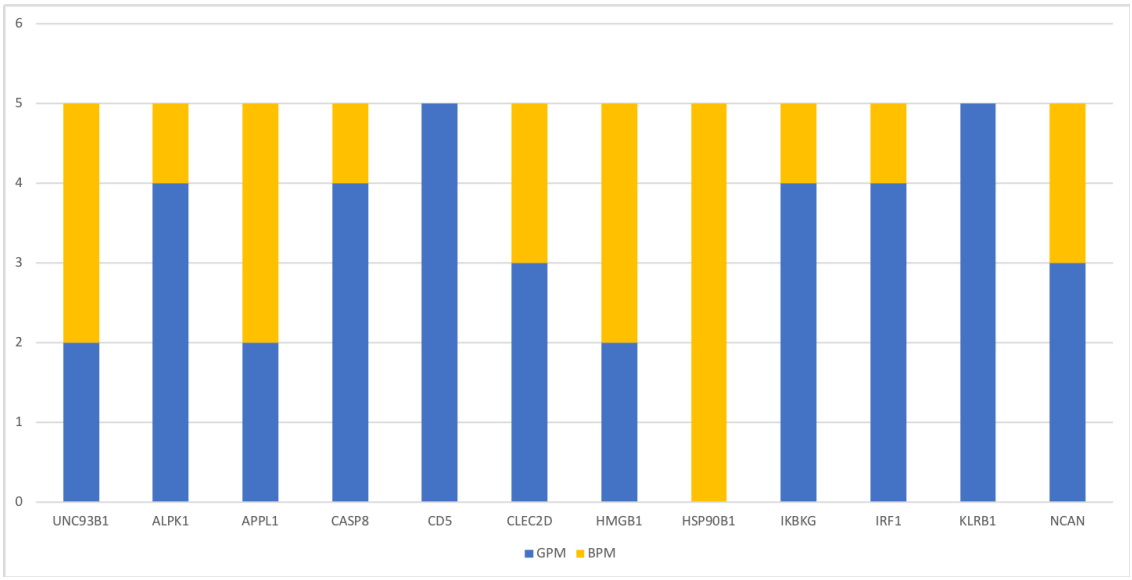


Figure 7.1 Multiple cancer survival linked genes. The distribution of genes from 12 genes panel across the cancers is shown. Y-axis denote the no. of cancers in which these gene play role as prognostic biomarker.

7.4 Discussion & Conclusion

In the past, various attempts have been made, and several cancer-specific prognostic models have been established, which are only relevant to single cancer. This study aimed to find "universal or multi-cancer" prognostic biomarkers and build models for predicting survival risk in different types of cancer patients. To do this, we assessed the prognostic significance of 331 PRR gene expression in 33 cancers in terms of patient overall survival. We discovered prognostic biomarker genes for only 30 different types of cancers initially. HR_{SKCM}=1.99 and HR_{THCA}=53.57 were attained by the cancer-specific prognostic models. Furthermore, a comprehensive study was carried out in order to uncover universal biomarker genes across a wide range of cancers. Our best prognostic model included 12 genes (*UNC93B1, ALPK1, APPL1, CASP8, CD5, CLEC2D, HMGB1, HSP90B1, IKBKG, IRF1, KLRB1, NCAN)* and classified risk groups across 6 cancers (HR_{THYM}=14.03, HR_{UCEC}=0.88). Besides offering a

complete assessment of the prognostic potential of PRR signaling genes in diverse cancer types, our research might aid in the development of adaptable risk management and treatment options for cancer patients.

8

Apoptosis Vs PRR Signaling Pathway in Multiple Cancer

8.1 Introduction

Cancer is a complicated and multidimensional illness defined by the unrestricted growth of aberrant cells capable of attacking or spreading throughout the body ("Pan-Cancer Analysis of Whole Genomes." 2020). This illness develops as a result of a breakdown of balance in homeostasis between cell survival and cell death (Y. Chen et al. 2016). Multiple signalling pathways implicated in cancer development emphasize the need for more investigation (Fakhri et al. 2021; Ochwang'i et al. 2014). Despite advancements in cancer research, revealing more involved pathways and molecular targets is crucial. Alteration in the expression of tumour suppressor genes, oncogenes, and apoptotic genes is important in the pathophysiology of cancer (Slattery et al. 2017; Debatin 2004). Furthermore, many inflammatory, oxidative stress, autophagy, and apoptotic dysregulated pathways are implicated in cancer start and progression (Monkkonen and Debnath 2018; Postovit et al. 2018; Mileo and Miccadei 2016). Among them pattern recognition receptor signalling pathway and apoptotic pathway are the two major pathway to be involved in cancer. A wide range of intracellular chemicals have been discovered as causing cancer cells to proliferate uncontrollably. Chronic inflammation is thought to be a critical driver of tumorigenesis's initiation and progression (Marelli et al. 2017). The innate immune system plays an important role in the inflammatory response to pathophysiological stimuli. The key sensors and pattern recognition receptors (PRRs) of the innate immune system that trigger stimuli (signal)-specific pro-inflammatory responses are toll-like receptors (TLRs) and inflammasomes. Chronic activation of PRRs has been linked to the aggressiveness of several malignancies as well as a bad prognosis. PRR involvement in carcinogenesis was previously thought to be restricted to infection- and injury-driven carcinogenesis, where they are activated by pathogenic ligands. With the identification of damage-associated molecular patterns (DAMPs) as PRR ligands, the function of PRRs in carcinogenesis has been expanded to include various non-pathogen-driven neoplasms. Dying (apoptotic or necrotic) cells release plenty of DAMPs, resulting in prolonged activation of PRRs and chronic inflammation and carcinogenesis. By regulating pro-inflammatory cytokines, metalloproteinases, and integrins, prolonged TLR activation increases tumour cell proliferation and accelerates tumour cell invasion and metastasis. Because PRRs play such an important part in carcinogenesis, targeting PRRs looks to be an effective anticancer technique. As a result, addressing the major aberrant proteins and pathways is a promising strategy to cancer treatment.

Cell death, particularly apoptosis, is perhaps one of the most extensively researched topics among cell biologists. Understanding apoptosis in disease circumstances is critical since it not only provides insights into disease pathogenesis but may also provide hints as to how the disease might be treated. There is an imbalance between cell division and cell death in cancer, and cells that should have perished did not get the instructions to do so. The issue might develop at any point during the apoptotic process. Downregulation of p53, a tumour suppressor gene, resulting in decreased apoptosis and increased tumour growth and progression (Bauer and Helfand 2006), and p53 inactivation, regardless of the method, as been related to numerous human malignancies (Morton et al. 2010; Bauer et al. 2002). As a result, apoptosis is vital in both carcinogenesis and cancer therapy.

The discovery of prognostic biomarkers for predicting cancer development is a critical issue for two reasons. For instance, such biomarkers can be used to treat patients in a medical context. Second, it is expected that investigating the biomarkers themselves will yield unique insights into disease mechanisms and the underlying biological processes that produce aberrant behaviour. Tons of signature have been developed for pan-cancer using gene expression which are associated with overall survival of the patients. As a result, such signatures have been proposed as biological explanations for breast cancer and therapeutic actions. In this study, we will compare two major dysregulated pathway in case of cancer (PRR signalling and apoptosis) and will find out whether the prognostic biomarkers developed using these two pathway genes perform same or different provided sensible biological interpretation. We will have look on specific cancer prognostic biomarker using these two pathway and see whether they are correlated or independent of each other.

8.2 Methodology

8.2.1 Dataset Generation and Pre-processing

TCGA-Assembler-2 was used to extract 'The Cancer Genome Atlas' (TCGA) normalized gene expression datasets and the raw counts for pan cancer i,e 33 types of cancers (Wei et al. 2018). A “pan-cancer” dataset was created by integrating all of the samples with raw gene expression levels from 33 different cancers. The list of 331 pattern recognition receptor signaling pathway genes was taken from Gene Set Enrichment Analysis (GSEA) and HUGO Gene Nomenclature Committee (HGNC). The gene expression data for these 331 genes were taken from the TCGA datasets that were downloaded and from the pan-cancer datasets in case of PRR genes only. In

case of apoptotic genes a list of 165 apoptotic genes was constructed (Sanchez-Vega et al., 2018). These 165 gene expression data were retrieved from the downloaded "TCGA" cancer datasets and pan-cancer datasets. Only those patient samples were retained for all the datasets for whom overall survival and censoring information were available. The total number of samples in the pan-cancer dataset was 9569 in both the case, and the total number of samples in each cancer cohort, "N", is shown in Table 8.1 for PRR and Table 8.2 for apoptotic genes.

8.2.2 Models for Predicting Survival

To test for survival-associated genes based on expression data, unadjusted “Cox proportional hazards” (Cox-PH) regression models were applied. The Cox-PH models were implemented using the R packages “survival” and “survminer”. “Hazard ratios” (HR) were calculated using this, as well as “confidence intervals” (% 95 CI) and “p-values”. HR is a hazard rate ratio that represents the death risk associated with one group compared to another by employing an appropriate gene-expression cutoff. We employed “Kaplan-Meier” (KM) plots and “log-rank tests” to compare survival curves between two risk categories. With HR more than or less than 1 and $p<0.05$, survival related genes were identified. . The Concordance (C) measure was used to evaluate the model's prediction performance. Further the gene voting model was created as mentioned in section 7.2.

8.2.3 Correlation Between Apoptotic and PRR Biomarker genes

To find out the relation between the pathways, we calculated the pearson correlation coefficient (PCC) among PRR and apoptotic biomarker genes using their expression values across different cancers. We chose THCA the top one, MESO the mid one and SKCM the last one. The main objective here to compare the performance of the prognostic models generated using genes from these different pathways across the cancer.

8.2.4 Network Construction Using Apoptotic & PRR Biomarkers

A co-expression network of PRR genes was constructed in order to discover interactions between the genes. To develop the gene co-expression network used here, we used Pearson's correlation coefficient (PCC) for each PRR gene pair and gene expression value to establish statistically significant key genes and therefore design highly accurate risk-prediction models. A correlation matrix was generated for these three cancer THCA, MESO and SKCM with

correlation calculated between all possible pairings of genes based on the expression data. The correlation matrix used to generate network edges using 'Igraph' for strongly correlated pairs of genes (|PCC| > 0.5). To visualize and analyze the gene network, we utilized the 'Cytoscape' software.

8.3 Results

8.3.1 Cancer Specific Prognostic Models:

We also cross checked and match the performance of cancer specific prognostic models. As done in (Arora, Kaur, and Raghava 2021) Table 8.1 depicts the prognostic models across all 33 cancers using 165 apoptotic genes. It is clear from Table 8.3 that THCA has the best performing model using top ten significant apoptotic genes (HR=41.59, P=3.36x10^{-4}).Whereas, SKCM has the least performing model (HR=1.99, p=2.18x10^{-5}). When compared to cancer specific prognostic models created using top ten significant PRR genes as shown in Table 8.4, we obtained that THCA has the best performing model with HR=53.57, p=1.41E-04) whereas, SKCM has got the lowest performing model with HR=1.99, p=2.33E-05. MESO has the best performing model with HR=3.99, p=1.67x10^{-6} using apoptotic genes and HR=4.24, p=9.53E-07 using PRR genes. It is quite clear from both the tables that irrespective of the pathway PRRs or apoptosis the trend is same. To investigate further we have evaluated below mentioned steps.

Table 8.1 The table displays the total number of patient samples ("N"), the number of "BPM" and "GPM" genes, and the top 10 survival linked genes for 33 cancers for "PRR" genes.

Cancer	N	GPM	BPM	Total	Top Genes
ACC	79	27	15	42	*LGALS3BP ,PKD1L2,NLRP4,ACAN,MAPKAPK3,PJA2,TNFAIP3,RNF125,CYLD,NLRP1*
BLCA	404	15	8	23	*IRF3 ,IKBKB,AGER,HHIPL1,CLEC2D,ATRNL1,ALPK1,UBQLN1,CLEC12A,UNC93B1*
BRCA	1091	38	14	52	*NFKBIA ,TNIP1,TNIP2,CLEC2D,NLRP6,REG4,ZCCHC3,XIAP,ANKRD17,USP17L2*
CESC	304	26	12	38	*LOXL2 ,CD6,NR1H4,NLRC5,LILRA4,CD5,NLRC3,HHIPL1,SELL,CLEC10A*
CHOL	36	4	3	7	*C1QBP ,TAB3,CD300LF,SSC4D,DGCR2,OLR1,TMPRSS4,,,*
COAD	297	5	8	13	*MBL1P ,COLEC11,TMPRSS15,DGCR2,BCL10,SCARB2,MAP3K7,TAB1,IRF7,CLEC2A*
DLBC	47	0	0	0	-
ESCA	183	8	4	12	*UNC93B1 ,DHX58,PRSS12,LGR4,FGA,TAB3,ATRN,SFTPA2,HMGB1,CD248*

GBM	160	4	18	22	*CLEC4C ,PIK3C3,SCARA3,RTN4,HSP90B1,CTSB,CTSL,NLRP12,CD248,NOD1*
HNSC	519	32	16	48	*PKD1 ,PRSS12,KLRB1,CD5,AGER,CACTIN,NLRC3,MAPKAPK2,FCRL3,NLRP1*
KICH	65	3	7	10	*TIFA ,PGLYRP2,REG3A,LBP,SFTPD,NR1D1,APPL2,APPL1,IKBKB,SELE*
KIRC	532	52	89	141	*NPLOC4 ,TLR9,TRAF6,APPL1,LGALS3BP,CNPY3,IKBKG,PRKCE,PJA2,PKD1L2*
KIRP	287	8	26	34	*SFTPD ,LBP,RIPK2,CLEC2L,TANK,LYN,UBE2D1,NLRP5,IRF1,STAB2*
LAML	173	18	27	45	*MAP3K1 ,CLEC11A,RFTN1,CD300A,TLR9,ESR1,IRAK1,CLEC18A,CLEC5A,MRC1*
LGG	511	31	112	143	*MYD88 ,DGCR2,CD302,MAP3K1,CLEC18B,CLEC18A,CIITA,IRGM,TAB1,CD69*
LIHC	369	13	15	28	*NLRC3 ,NLRP8,CD5,CLEC3B,TREML4,KLRK1,CLEC2L,FCRL3,NLRP9,ITCH*
LUAD	497	20	18	38	*NLRP10 ,TRIM15,UBE2N,LOXL2,NLRP2,AGER,FADD,NLRC3,CD302,CTSL*
LUSC	488	6	21	27	*CD14 ,RPS27A,FGA,CLEC18C,NLRP12,LOXL2,ESR1,DAB2IP,MAP2K6,S100A1*
MESO	86	18	25	43	*LOXL2 ,LAYN,OTULIN,RTN4,UBE2N,INAVA,CACTIN,KLRB1,PKD1L3,TRIL*
OV	305	22	0	22	*CASP8 ,HMGB1,NLRP4,CD300A,BIRC3,STAB1,HSPD1,LGR4,NLRP12,CACTIN*
PAAD	178	26	27	53	*ITCH ,WDFY1,IRAK4,FADD,PTPRS,SARM1,BIRC2,ERBIN,IRAK2,TMPRSS4*
PCPG	179	0	0	0	-
PRAD	497	3	0	3	*RIOK3 ,PGLYRP1,TREML4*
READ	96	1	2	3	*IRAK1 ,PGBD1,ASGR2*
SARC	257	37	10	47	*RNF125 ,SELE,TICAM1,CLEC10A,SELP,IRF1,HPN,NLRP11,DHX58,LILRA4*
SKCM	449	78	11	89	*TLR4 ,KLRD1,CLEC6A,RSAD2,IFIH1,BIRC3,KLRC1,TLR8,KLRC2,TNIP3*
STAD	413	5	18	23	*NLRP14 ,CAV1,PJA2,NOD2,VCAN,TICAM1,TREML4,S100A9,SELE,REG1A*
TGCT	133	0	0	0	-
THCA	505	10	9	19	*CLEC4C ,PKD1L3,TREML4,KLRB1,LAYN,FADD,CHODL,NOD1,CLEC4D,APPL1*
THYM	119	1	3	4	*PGLYRP1 ,KLRC1,CLEC4C,CD6*
UCEC	541	3	10	13	*CD68 ,NPLOC4,FLOT1,TYRO3,KLRG2,ESR1,KLRC3,CD163L1,HMGB1,PKD1L3*
UCS	57	5	5	10	*NLRP2,KLRF1,TICAM1,MAPKAPK2,NLRP14,LGALS3BP,HSP90B1,TREML4,MAP3K1, SMPDL3B*
UVM	80	0	43	43	*REG3G ,LBP,IRAK1,COLEC11,CD5L,RNF125,SCARB1,CLEC11A,IKBKG,CLEC12A*

Table 8.2 The table displays the total number of patient samples ("N"), the number of "BPM" and "GPM" genes, and the top 10 survival linked genes for 33 cancers for "Apoptotic" genes. Source ~ (Arora, Kaur, and Raghava 2021).

Cancer	N	BPM	GPM	Total	Top Genes
LGG	511	77	17	94	*WEE1,BTG3,BMP2,PLAT,SMAD7,ANXA1,PEA15,CDK2,HSPB1,SOD2*
KIRC	532	50	32	82	*CASP9,F2,TIMP1,IL6,CDC25B,ADD1,CCNA1,BAK1,SLC20A1,TIMP3*
MESO	86	33	15	48	*HMGB2,TOP2A,BRCA1,PLAT,SLC20A1,WEE1,PPP2R5B,MADD,PDCD4,LMNA*
SKCM	449	10	33	43	*TNFSF10,SATB1,DPYD,BIRC3,SOD2,F2R,CYLD,GCH1,CD69,PSEN2*
PAAD	178	34	7	41	*CASP4,TNFSF10,PSEN1,CD44,CASP2,EMP1,TOP2A,DPYD,CCND1,HMGB2*
ACC	79	22	14	36	*TOP2A,PEA15,BRCA1,H1F0,HMGB2,MADD,CDK2,SPTAN1,CYLD,SQSTM1*
BRCA	1091	10	25	35	*PTK2,NEFH,IGF2R,PLAT,DNM1L,XIAP,ETF1,NEDD9,IRF1,RARA*
LAML	173	14	16	30	*PDCD4,ISG20,LMNA,NEDD9,CCND2,PSEN1,HGF,SOD1,ADD1,CD44*
HNSC	519	19	10	29	*CCND1,BMF,CCNA1,BAK1,PSEN1,APP,TIMP1,BCAP31,SLC20A1,TNFRSF12A*
UVM	80	17	12	29	*ERBB3,ISG20,EREG,TIMP3,LEF1,SATB1,TXNIP,PPP2R5B,ERBB2,PTK2*
CESC	304	16	10	26	*EREG,CASP2,MGMT,CD2,IL1B,IGF2R,APP,NEFH,TIMP2,GCH1*
KIRP	287	21	3	24	*BCL2L10,TOP2A,PMAIP1,MCL1,LEF1,PPP2R5B,PEA15,DCN,IRF1,H1F0*
SARC	257	7	16	23	*CTH,RNASEL,GSN,IRF1,SPTAN1,CASP1,BTG2,CFLAR,TNF,CASP2*
BLCA	404	7	15	22	*EMP1,GCH1,HMGB2,GSTM1,CASP7,ANXA1,IFNGR1,ETF1,SLC20A1,AIFM3*
LIHC	369	12	4	16	*MGMT,ETF1,RARA,GPX3,EREG,CD2,DAP3,GPX4,FASLG,CDC25B*
STAD	413	13	3	16	*CAV1,CD44,PDGFRB,DNAJC3,EREG,TGFB2,CTNNB1,DFFA,BCL2L11,CASP6*
LUSC	488	12	3	15	*CD14,BTG3,EREG,CCND2,PTK2,PAK1,ADD1,HSPB1,TIMP3,SMAD7*
LUAD	497	9	5	14	*EREG,VDAC2,BBC3,SLC20A1,BTG2,TOP2A,RELA,CD2,GPX4,ETF1*
ESCA	183	6	7	13	*ENO2,IL18,TOP2A,DAP,BCL2L1,PMAIP1,ISG20,IL1A,TSPO,SATB1*
COAD	297	5	5	10	*BCL10,CASP4,FAS,IL6,GSR,TIMP1,BGN,LUM,ERBB2,BTG2*
OV	305	4	5	9	*DAP,CASP8,EMP1,BIRC3,CASP2,WEE1,PSEN1,NEDD9,SOD1*
THCA	505	4	5	9	*ANXA1,TGFBR3,CLU,PSEN1,TNFRSF12A,GPX4,TIMP3,LEF1,BNIP3L*
KICH	65	6	2	8	*IFNB1,MADD,BIK,GSR,TOP2A,PTK2,DAP3,CLU*
GBM	160	6	1	7	*HSPB1,FDXR,TXNIP,ANKH,EGR3,F2R,IER3*
UCEC	541	5	0	5	*BCL2L1,MCL1,AVPR1A,SLC20A1,ISG20*
UCS	57	2	3	5	*MGMT,HGF,BMF,H1F0,PTK2*
CHOL	36	3	1	4	*PSEN1,BNIP3L,EREG,JUN*
THYM	119	2	2	4	*IER3,SOD2,CD2,LEF1*
PRAD	497	1	1	2	*SATB1,IER3*
READ	96	1	1	2	*BRCA1,DNAJC3*
DLBC	47	0	0	0	-
PCPG	179	0	0	0	-
TGCT	133	0	0	0	-

Table 8.3 The performance of cancer-specific prognostic models using apoptotic genes. Source ~ (Arora, Kaur, and Raghava 2021)

Cancer	HR	p-value	logrank-p	C	%95 CI L	%95 CI U
THCA	41.59	$3.36x10^{-4}$	$3.81x10^{-8}$	0.84	5.42	319.17
UVM	40.50	$5.32x10^{-4}$	$5.12x10^{-7}$	0.85	4.99	328.82
KICH	25.61	$2.27x10^{-3}$	$3.53x10^{-5}$	0.83	3.19	205.6
ACC	22.68	$7.95x10^{-7}$	$1.63x10^{-10}$	0.81	6.57	78.31
THYM	12.53	$2.42x10^{-2}$	$6.98x10^{-3}$	0.79	1.39	112.93
UCEC	10.42	$4.51x10^{-4}$	$1.13x10^{-4}$	0.7	2.81	38.6
CHOL	8.72	$4.75x10^{-4}$	$2.45x10^{-4}$	0.77	2.59	29.4
PRAD	8.42	$4.41x10^{-3}$	$4.20x10^{-3}$	0.65	1.94	36.5
READ	7.45	$6.50x10^{-2}$	$2.56x10^{-2}$	0.72	0.88	62.93
KIRP	5.10	$6.64x10^{-5}$	$1.27x10^{-5}$	0.72	2.29	11.37
LGG	4.99	$2.88x10^{-12}$	$1.54x10^{-13}$	0.72	3.18	7.83
CESC	4.92	$2.14x10^{-8}$	$2.98x10^{-9}$	0.71	2.82	8.6
LIHC	4.58	$7.91x10^{-11}$	$2.24x10^{-11}$	0.7	2.89	7.24
PAAD	4.41	$4.23x10^{-7}$	$1.72x10^{-7}$	0.69	2.48	7.85
COAD	4.08	$5.05x10^{-5}$	$2.42x10^{-5}$	0.67	2.07	8.05
MESO	3.99	$1.67x10^{-6}$	$2.00x10^{-6}$	0.68	2.26	7.03
KIRC	3.96	$5.41x10^{-16}$	$3.03x10^{-17}$	0.68	2.84	5.53
LAML	3.96	$3.92x10^{-12}$	$5.07x10^{-12}$	0.67	2.68	5.84
ESCA	3.80	$2.19x10^{-6}$	$3.32x10^{-6}$	0.65	2.19	6.61
UCS	3.61	$8.77x10^{-4}$	$6.13x10^{-4}$	0.68	1.69	7.67
BRCA	3.45	$2.36x10^{-9}$	$6.76x10^{-10}$	0.67	2.3	5.18
BLCA	3.41	$6.35x10^{-10}$	$3.51x10^{-10}$	0.66	2.31	5.02
STAD	3.35	$2.78x10^{-7}$	$1.39x10^{-7}$	0.64	2.11	5.31
SARC	2.81	$1.32x10^{-5}$	$1.03x10^{-5}$	0.67	1.77	4.48
LUAD	2.76	$6.94x10^{-8}$	$4.82x10^{-8}$	0.63	1.91	3.99
HNSC	2.36	$9.24x10^{-8}$	$5.80x10^{-8}$	0.62	1.72	3.24
LUSC	2.21	$1.26x10^{-6}$	$1.30x10^{-6}$	0.61	1.6	3.04
OV	2.19	$1.38x10^{-6}$	$1.16x10^{-6}$	0.61	1.59	3
GBM	2.07	$3.73x10^{-4}$	$3.22x10^{-4}$	0.61	1.38	3.09
SKCM	1.99	$2.18x10^{-5}$	$2.55x10^{-5}$	0.59	1.45	2.75

Table 8.4 The performance of prognostic models which are cancer specific created using PRR genes.

Cancer	HR	p-value	logrank-p	C	%95 CI lower	%95 CI upper
THCA	53.57	1.41E-04	5.46E-09	0.84	6.89	416.23
KICH	37.79	6.34E-04	3.02E-06	0.86	4.70	303.63
ACC	15.11	1.58E-06	7.39E-09	0.78	4.99	45.80
THYM	12.73	2.27E-02	6.15E-03	0.80	1.43	113.60
CHOL	11.79	1.35E-03	6.08E-05	0.77	2.61	53.31
KIRP	7.56	8.65E-07	6.25E-08	0.75	3.38	16.94
LGG	4.83	1.04E-11	6.28E-13	0.74	3.07	7.61
COAD	4.76	1.81E-06	3.10E-06	0.71	2.51	9.02
UCS	4.39	3.12E-04	2.68E-04	0.65	1.96	9.82
MESO	4.24	9.53E-07	4.90E-07	0.68	2.38	7.54
BRCA	3.97	1.81E-10	1.83E-11	0.64	2.60	6.07
PAAD	3.88	1.82E-06	7.87E-07	0.68	2.22	6.78
KIRC	3.59	3.15E-13	9.69E-15	0.66	2.55	5.06
ESCA	3.55	1.34E-05	2.10E-05	0.65	2.01	6.29
BLCA	3.22	1.34E-08	3.63E-09	0.64	2.15	4.81
LUAD	3.15	1.14E-09	1.39E-09	0.66	2.18	4.55
LIHC	3.07	3.37E-07	8.17E-07	0.66	2.00	4.73
STAD	3.04	1.48E-06	8.76E-07	0.64	1.93	4.79
SARC	2.92	7.16E-06	4.97E-06	0.66	1.83	4.65
LAML	2.89	9.94E-08	6.35E-08	0.61	1.95	4.26
CESC	2.76	1.50E-04	1.12E-04	0.67	1.63	4.68
LUSC	2.49	2.61E-08	2.75E-08	0.61	1.81	3.44
GBM	2.48	8.38E-06	1.42E-05	0.62	1.66	3.70
HNSC	2.48	9.10E-08	2.48E-08	0.61	1.78	3.45

OV	2.13	2.31E-06	2.98E-06	0.60	1.56	2.92
SKCM	1.99	2.33E-05	2.39E-05	0.59	1.45	2.74

8.3.2 Correlation between Apoptotic and PRR genes

The main objective here to compare the performance of the prognostic models generated using genes from these different pathways across the cancer. The study follows by :

(i) THCA: In this cancer we obtained top prognostic model using apoptotic and PRR genes b oth. The model obtained was constructed using ten PRR genes namely *CLEC4C , PKD1L3, T REML4, KLRB1, LAYN, FADD, CHODL, NOD1, CLEC4D, APPL1* in case of cancer specific prognostic model using PRR genes only. While using nine apoptotic genes named *ANXA1, T GFBR3 ,CLU, PSEN1, TNFRSF12A, GPX4, TIMP3, LEF1, BNIP3L* the prognostic model for THCA was developed (Arora, Kaur, and Raghava 2021). We evaluated the correlation betwee n the top ten and top nine PRR and apoptotic genes. We also computed the self-correlation be tween the ten PRR genes and the nine apoptotic genes. Table 8.5 shows the correlation betwe en the biomarker using ten PRR genes and biomarker developed using nine apoptotic genes. F igure 8.1 is the visual representation for same. Figure 8.2 represents the self- correlation plots.

Table 8.5 Pearson Correlation between ten PRR biomarker genes and nine apoptotic genes in case of THCA. Genes on x-axis are apoptotic biomarker genes whereas genes on y-axis are PRR genes.

	ANXA1	**TGFBR3**	**CLU**	**PSEN1**	**TNFRSF 12A**	**GPX4**	**TIMP3**	**LEF1**	**BNIP3L**
APPL1	-0.25	0.25	-0.31	0.42	-0.48	-0.59	0.08	-0.02	0.44
TREML4	-0.07	0.01	-0.01	0.01	-0.08	-0.07	0.03	0.22	0.15
LAYN	-0.33	0.20	-0.14	-0.18	-0.26	-0.13	0.20	0.09	0.12
CLEC4C	-0.03	0.12	-0.01	0.02	-0.06	-0.09	-0.05	0.22	0.12
CLEC4D	0.06	0.00	0.02	0.02	0.03	-0.01	-0.18	0.16	0.11
PKD1L3	-0.05	0.16	0.08	-0.07	-0.10	-0.04	0.00	0.17	0.00
KLRB1	0.03	-0.02	0.13	-0.11	-0.01	0.10	-0.17	0.30	-0.01
CHODL	0.11	0.00	0.15	-0.07	0.10	0.03	-0.01	0.11	-0.15
FADD	0.38	-0.33	0.38	-0.12	0.39	0.43	-0.29	-0.07	-0.15
NOD1	0.60	-0.18	0.54	0.11	0.47	0.26	-0.15	0.01	-0.18

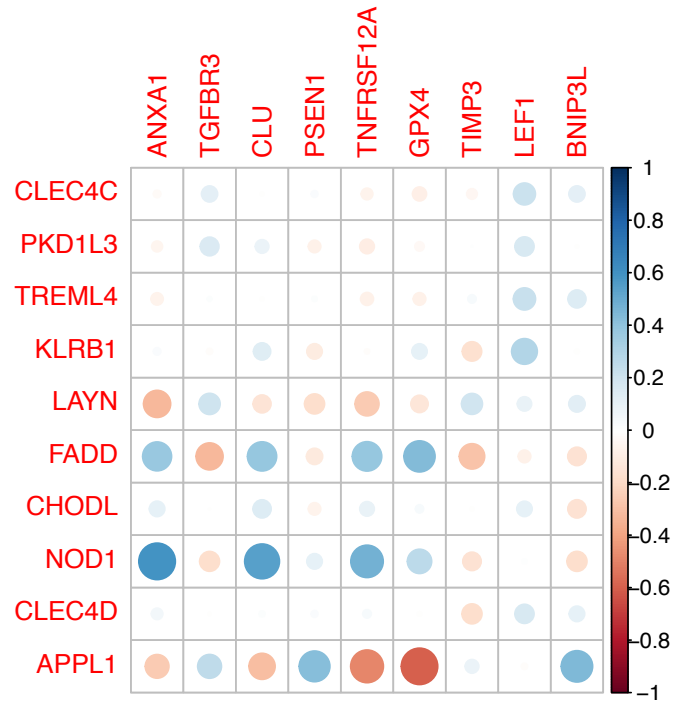
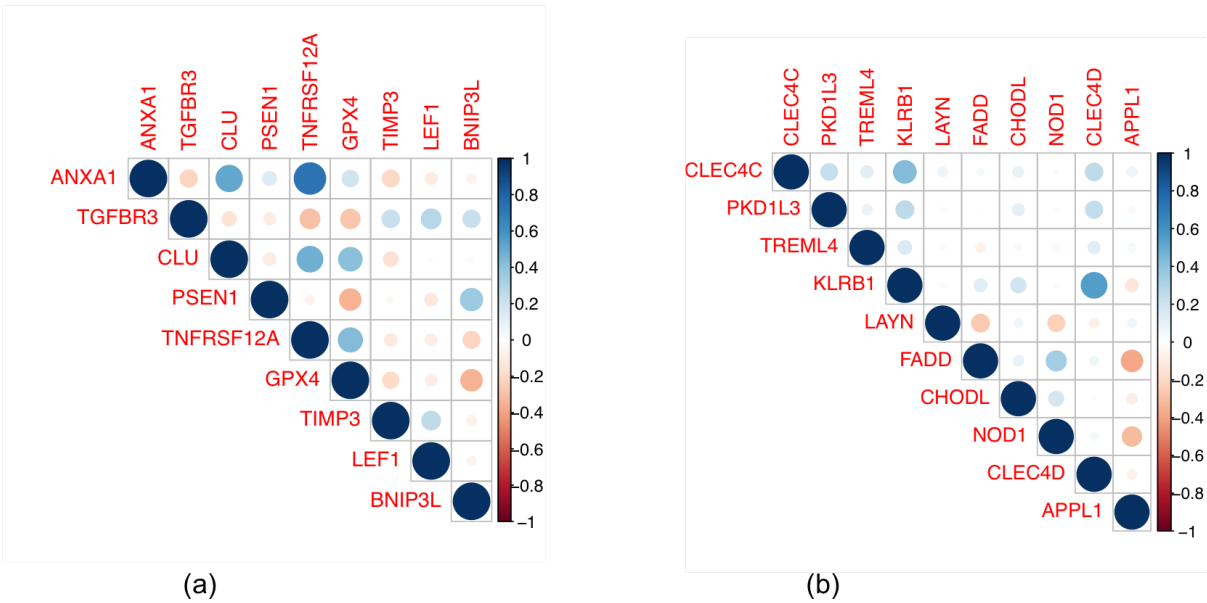


Figure 8.1 Correlation plot between PRR and apoptotic biomarker for THCA



(a) (b)

Figure 8.2 Represents the self-correlation plot of (a) apoptotic biomarker genes (b) PRR biomarker genes for THCA.

(ii) MESO: In this cancer, we found the best prognostic model by employing both apoptotic and PRR biomarker genes. The model obtained was constructed using *LOXL2 ,LAYN, OTULIN, RTN4, UBE2N, INAVA, CACTIN, KLRB1, PKD1L3, TRIL* in case of cancer specific prognostic model using PRR genes only. While using ten apoptotic genes named *HMGB2, TOP2A, BRCA1, PLAT, SLC20A1,WEE1, PPP2R5B, MADD, PDCD4, LMNA* a prognostic model for MESO has been created. We investigated the relationship between the top 10 PRR and apoptotic genes. The self-correlation between the ten PRR genes and the ten apoptotic genes was also calculated. Table 8.6 shows the correlation between the biomarker using ten PRR genes and biomarker developed using ten apoptotic genes. Figure 8.3 is the visual representation for same. Figure 8.4 represents the self- correlation plots.

Table 8.6 Pearson Correlation between ten PRR biomarker genes and nine apoptotic genes in case of MESO. Genes on x-axis are apoptotic biomarker genes whereas genes on y-axis are PRR genes.

	HMGB2	**TOP2A**	**BRCA1**	**PLAT**	**SLC20A1**	**WEE1**	**PPP2R5B**	**MADD**	**PDCD4**	**LMNA**
LOXL2	0.23	0.53	0.48	0.22	0.26	0.22	0.42	-0.34	-0.35	0.09
LAYN	0.18	0.19	0.23	0.14	0.04	0.15	0.52	-0.26	-0.31	-0.04
OTULIN	0.36	0.37	0.40	0.05	0.06	0.20	0.31	-0.06	-0.30	0.07
RTN4	0.15	0.42	0.39	0.47	0.40	0.22	0.36	-0.53	-0.29	0.26
UBE2N	0.28	0.32	0.23	0.30	0.30	0.17	0.27	-0.28	-0.12	0.39
INAVA	0.07	0.17	0.10	-0.01	-0.11	0.01	0.22	-0.15	-0.04	0.00
CACTIN	0.26	0.16	0.26	-0.01	-0.04	0.02	0.22	0.00	-0.14	0.04
KLRB1	0.05	-0.31	-0.24	-0.02	-0.06	-0.04	0.00	-0.16	-0.05	0.06
PKD1L3	-0.23	-0.14	-0.14	-0.13	-0.14	-0.23	-0.26	0.10	0.07	-0.18
TRIL	-0.02	-0.17	-0.23	0.03	-0.09	-0.09	-0.05	0.27	0.09	-0.15

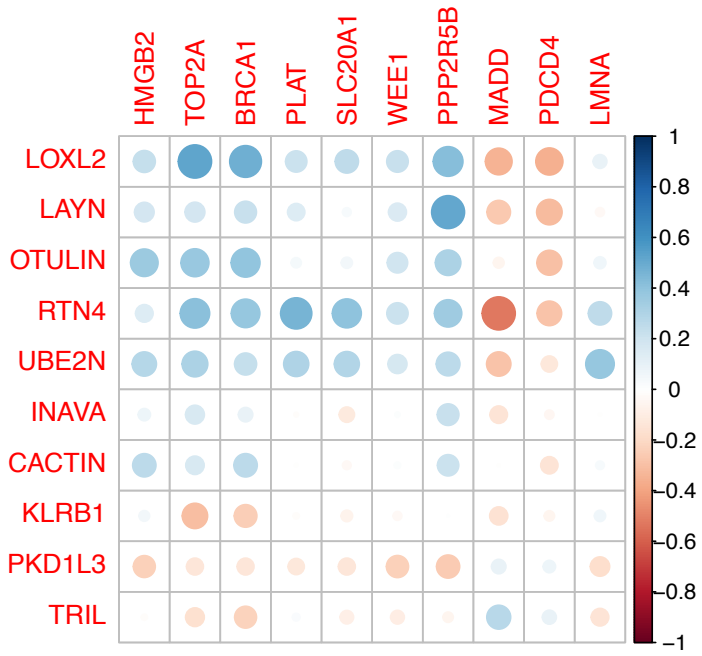
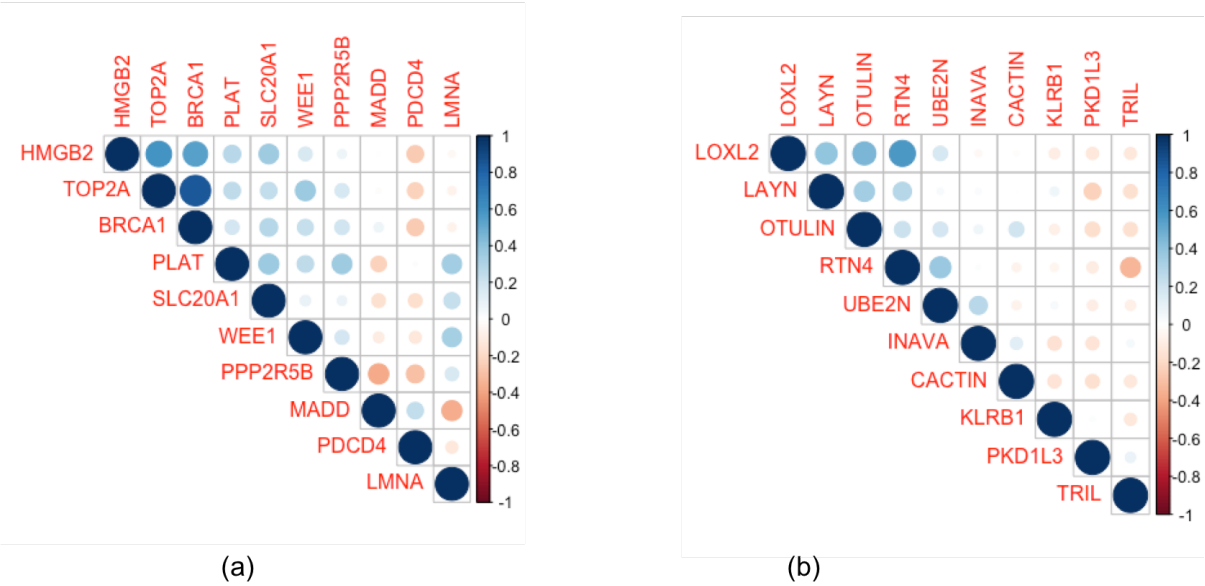


Figure 8.3 Correlation plot between PRR and apoptotic biomarker for MESO



(a) (b)

Figure 8.4 Represents the self-correlation plot of (a) apoptotic biomarker genes (b) PRR biomarker genes for MESO.

(iii) SKCM : In this cancer we obtained top prognostic model using apoptotic and PRR genes both. The model obtained was constructed using ten PRR genes namely *TLR4, KLRD1, CLEC6A, RSAD2, IFIH1, BIRC3, KLRC1, TLR8, KLRC2, TNIP3* in case of cancer specific prognostic model using PRR genes only. While using nine apoptotic genes named *TNFSF10, SATB1, DPYD, BIRC3, SOD2, F2R, CYLD, GCH1, CD69, PSEN2*. The prognostic model for THCA was developed (Arora, Kaur, and Raghava 2021). We evaluated the correlation between the top ten and top ten PRR and apoptotic genes. We also computed the self-correlation between the ten PRR genes and the ten apoptotic genes. Table 8.7 shows the correlation between the biomarker using ten PRR genes and biomarker developed using ten apoptotic genes. Figure 8.5 is the visual representation for same. Figure 8.6 represents the self-correlation plots.

Table 8.7 Pearson Correlation between ten PRR biomarker genes and ten apoptotic genes in case of S KCM. Genes on x-axis are apoptotic biomarker genes whereas genes on y-axis are PRR genes.

	TNFSF10	**SATB1**	**DPYD**	**BIRC3**	**SOD2**	**F2R**	**CYLD**	**GCH1**	**CD69**	**PSEN2**
TLR4	0.52	0.24	0.41	0.33	0.34	0.05	0.53	0.49	0.32	-0.29
KLRD1	0.62	0.04	0.28	0.33	0.49	0.01	0.52	0.82	0.36	-0.12
CLEC6A	0.39	0.07	0.25	0.49	0.38	0.03	0.38	0.51	0.33	-0.18
RSAD2	0.48	0.03	0.24	0.21	0.27	0.01	0.34	0.32	0.17	-0.12
IFIH1	0.53	0.11	0.38	0.35	0.34	0.12	0.49	0.48	0.26	-0.19
BIRC3	0.47	0.18	0.34	1.00	0.37	0.00	0.67	0.43	0.67	-0.20
KLRC1	0.49	0.05	0.22	0.25	0.30	-0.03	0.34	0.54	0.23	-0.11
TLR8	0.71	0.12	0.43	0.37	0.52	0.03	0.59	0.77	0.40	-0.19
KLRC2	0.25	0.24	0.22	0.09	0.13	0.07	0.18	0.28	0.12	-0.12
TNIP3	0.59	0.02	0.32	0.33	0.38	0.02	0.45	0.65	0.33	-0.17

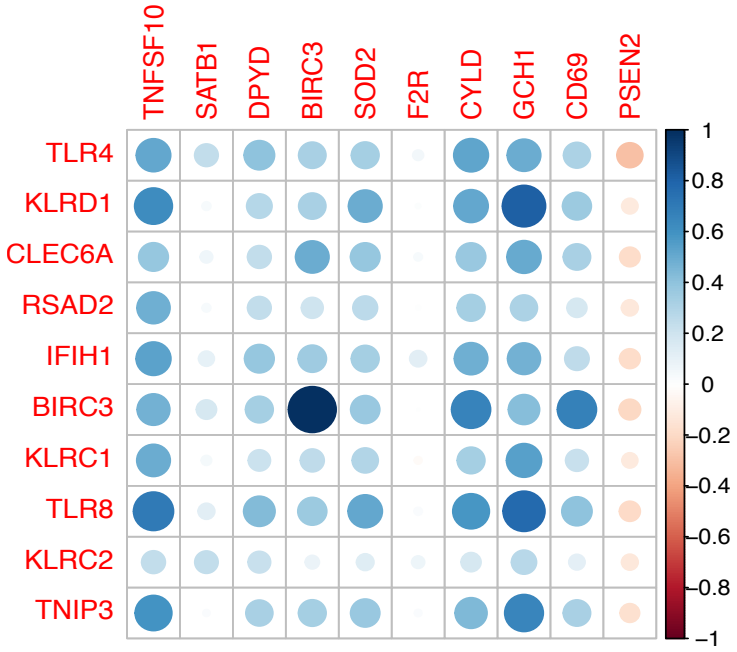
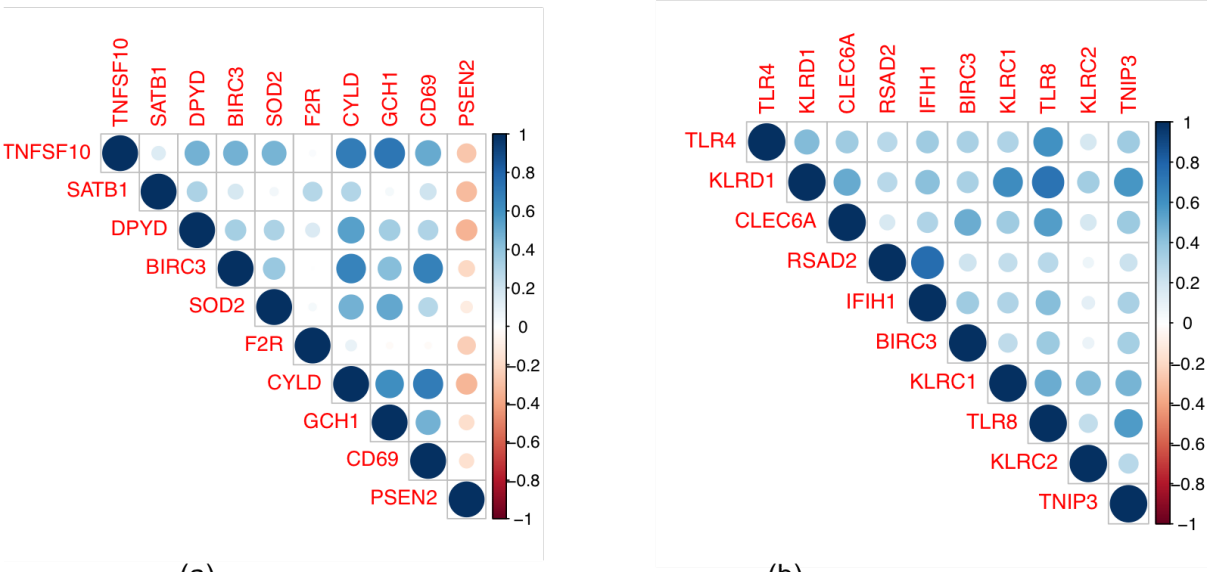


Figure 8.5 Correlation plot between PRR and apoptotic biomarker for SKCM



(a) (b)

Figure 8.6 Represents the self-correlation plot of (a) apoptotic biomarker genes (b) PRR biomarker genes for SKCM

8.3.2.2 Network Analysis Using Apoptotic & PRR Biomarkers

To find relationships between the genes of both the pathway, a co-expression network of PRR and apoptotic biomarker genes was built. We employed Pearson's correlation coefficient (PCC) for each gene pair utilising gene expression value to determine statistically significant important genes and therefore design highly accurate risk-prediction models to create the gene co-expression network used here. We have constructed the networks for the following cancer.

(i) THCA: A correlation matrix was generated from 19X19 PRR and apoptotic genes, with correlation calculated between all possible pairings of genes based on the expression data. The correlation matrix used to generate network edges using 'Igraph' for strongly correlated pairs of genes (|PCC| > 0.5). To visualize and analyze the gene network, we utilized the 'Cytoscape' software. When the effective correlation was set at larger than 0.5, there were 6 nodes and 9 edges only. Figure 8.7 display the co-expression network. *ANXA1* being the hub node has the highest degree and interlinked with *NOD1, CLU, TNFRS12A*. *APPL1* is connected with *GPX4* whereas, *CLEC4D* is connected with *KLRB1.* Among these three genes are PRR genes *(NOD1, CLEC4D and APPL1)*. While rest belongs to apoptosis pathway. It is quite transparent that *ANXA1* being the hub apoptotic gene is highly correlated with *NOD1* and other apoptotic genes.

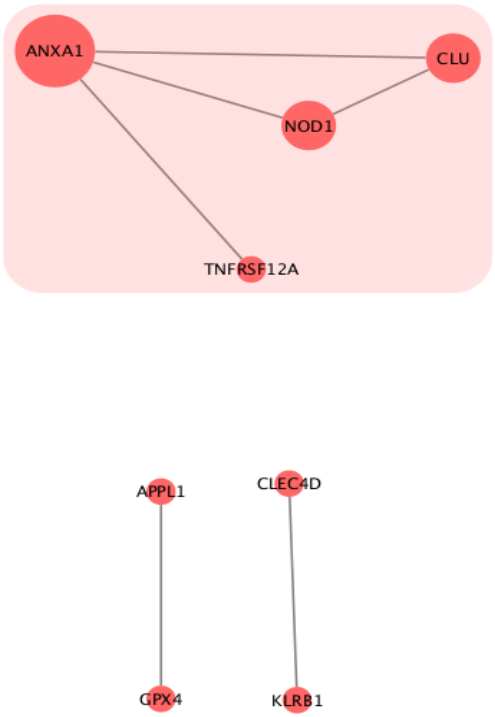


Figure 8.7 Co-expression network between 19 X19 PRR and apoptotic biomarker genes in case of THCA. Bigger node size depicts the higher degree.

(ii) MESO : A correlation matrix of 20X20 PRR and apoptotic genes was created, with correlation computed between all feasible gene pairs based on expression data. When effective correlation cut off |PCC<0.5| was taken we obtained 8 nodes and 7 edges effectively. The co-expression network is shown in Figure 8.8. Among these apoptotic genes like TOP2A is the hub gene connected strongly with other apoptotic genes like *BRCA1, HMGB2, MADD* and PRR genes like *LOXL2, RTN4*. While *LAYN* which is a PRR gene is connected to an apoptotic gene *PPP2R5B*. The network depicts the highly interlinked and correlated apoptotic and PRR genes on the basis of degree.

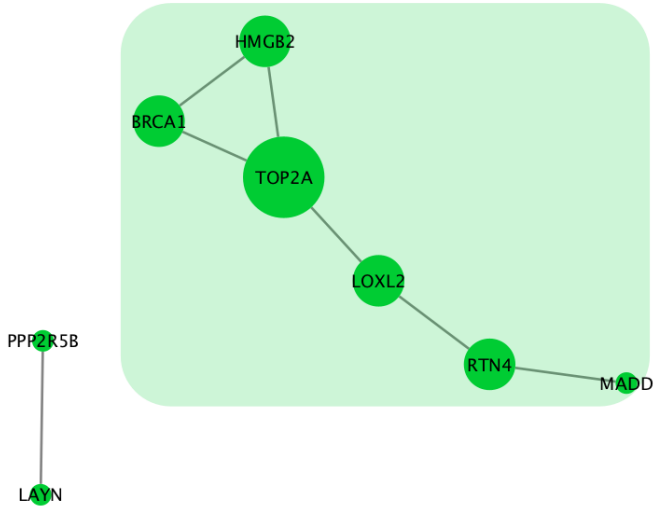


Figure 8.8 Co-expression network between 20 x20 PRR and apoptotic biomarker genes in case of MESO. Bigger node size depicts the higher degree.

(iii) SKCM: A correlation matrix of 20X20 PRR and apoptotic genes was created, with correlation computed between all feasible gene pairs based on expression data. Among both the pathways *BIRC3* gene was common. When effective correlation cut off |PCC<0.5| was taken we obtained 16 nodes and 34 edges effectively. The co-expression network is shown in Figure 8.9. Among these apoptotic genes like CYLD, TNFS10A, GCH1 and PRR gene TLR8 shown higher degree then rest of the genes. The network depicts the highly interlinked and correlated apoptotic and PRR genes on the basis of degree.

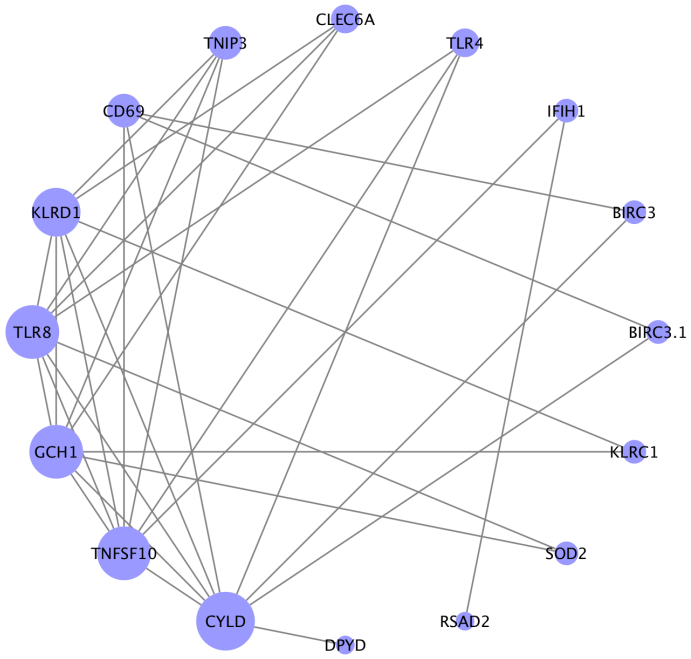


Figure 8.9 Co-expression network between 20x20 PRR and apoptotic biomarker genes in case of SKCM. Bigger node size depicts the higher degree. Edges representing the correlation coefficient.

7.3.4 Hybrid Prognostic Model

After finding out the interconnection between apoptotic and PRR biomarker across cancers, we decided to merge both. We integrated both the biomarker sets for THCA, MESO and SKCM in order to improve the performance of the prognostic models. We ran different combination for each case and used RFE (Recursive Feature Elimination) to select the best set of genes for improved performance. Table 8.8 represents the performance for these mentioned cancer using PRR biomarker, apoptotic biomarker and hybrid (PRR + apoptotic) biomarker genes.

Table 8.8 Performance of prognostic models developed using PRR genes only, apoptotic genes only and hybrid (PRR + Apoptotic) genes in terms of HR for THCA, MESO and SKCM.

Cancer	PRR genes only (HR)	Apoptotic genes only (HR)	Hybrid (PRR+ Apoptotic) (HR)
THCA	53.57	41.59	56.7
MESO	4.24	3.99	4.30

SKCM	1.99	1.99	2.08

8.4 Discussion and Conclusion

Diagnosis at initial stage is very crucial in case of any malignancy especially cancer. A detailed prognosis is frequently necessary for strategic therapeutic intervention planning. The modern oncology research provided lot of omics data that can be helpful for the development of effective prognostic biomarkers and corresponding therapy. Previous research has highlighted critical roles of various genes belong to important pathway like apoptosis and PRR signalling pathway. In the current study, we compared the prognostic biomarkers and their performance for these two pathways in case of THCA, MESO and SKCM. We chose these three on the basis of the result from our prior study. To do so first correlation among the PRR biomarker genes and apoptotic biomarker genes have been taken. Self-correlation was also taken into consideration. Using peasron correlation coefficient with the effective cut-off an edgelist was developed to create network in case of these three cancers. Network analysis and visualization is also done and shown in the study. We developed a hybrid model to improve the prognostic performance of the biomarkers by combining apoptotic and PRR signalling genes and found that the performance increased in the case of THCA, MESO and SKCM. The high correlation among genes from these two different pathways clarifies that these are not independent and have dependency on each other. Overall all this study highlights the high interconnection between cross pathways in case of cancer. It also encourages more study into the possibilities of such a hybrid model combining several pathways for development in cancer therapy and management.

9

Summary

Several initiatives have been undertaken in the last decade to research adaptive immunity. Tremendous exposure like annotation, creating in-silico tool, making usage of biological insight to understand the mechanistic point has been explored well in this arm of immunity. Whereas, adaptive immunity get activated through first line of defense innate immunity and if any malignancy get resolved at first step there would not be any need to go further on another step. But, there is not as much work has been done in innate immunity. Although it has important role in fighting against infection and providing host defense, also it plays a vital role as pro and anti-tumoral molecules. The innate immune molecules requires a proper annotation so thus researcher use them for translational benefits in research and therapies. Several essential regulators have been identified, as well as their involvement in this complex system. In summary, it has been shown that some components and portions of the innate immune system are weakened in cancer cells, causing these injured cells to refuse to die and disseminate the harm to future generations. Because of our current understanding of the pathways, drugs that target these critical components and restore the survival/death balance have been developed. Furthermore, changes in the concentrations or status of innate immune molecule regulators are utilized to predict cancer prognosis and risk. The development of novel prognostic biomarkers/methods for cancer risk assessment, on the other hand, remains a challenge. Likewise, given the importance of numerous clinical aspects in cancer genesis and progression, these prospective techniques should incorporate important elements in order to supplement or replace existing risk prediction systems. The innovative prognostic approaches can be used to provide more precise risk prediction and, as a result, more effective therapy planning.

Chapter2 focuses on the fact that pattern recognition receptors have long been thought to be a minor biological phenomena. It began to change as their role in the infection became obvious, as did the relationship between innate and acquired immunity. Multiple past studies have revealed the detailed mechanism of PRRs that are essential part of innate immune system. Yet specific proper annotation is lacking for the ligands/ agonist corresponding to PRRs. *In-silico* web resources and updated knowledgebase for PRRs for better understanding and designing vaccine adjuvants is required. Also, due to its dual role in cancer PRR can be use as targeted therapy but, utilization of these PRR and their agonist in prognosis of cancers is not explored yet. Furthermore, computational tools and databases that provide updated information and insight of PRR mechanism and use as biomarkers in

cancer are not available.

Chapter 3 mainly describe the update PRRDB 2.0 which is database of pattern recognition receptor. PRRDB 2.0, comprising more than 2700 entries, provides better coverage of all the PRRs and their ligands studied till now. The information obtained from the research articles and the patents is summarized in tabular form under 25 fields in the database. Fields like name, source , origin, role, sequence of receptors and their ligands are provided. TLRs are the most well-studied PRRs, although cytoplasmic PRRs also play an important role in the accumulation of diverse immunological responses, which requires further research. Other PRRs, such as mannose receptors, scavenger receptors, and a few secreted PRRs, also require additional investigation. We have improved the data coverage by incorporating additional fields and also highlighting the role and specificity of PRRs and ligands in eliciting the immune response. The hyper linkage with Swiss-Prot, PDB and PubChem will provide maximum information at a single place. We believe that the updated version will be very helpful to the scientific community. PRRDB 2.0 is freely available at " **https://webs.iiitd.edu.in/raghava/prrdb2/"** as a user-friendly, display compatible interface. The previous version can be accessed at " **http://crdd.osdd.net/raghava/prrdb/**".

Multiple previous studies show the importance of PRRs in various diseases like heart failure, cancer, autoimmune disorders, kidney disease, asthma, atherosclerosis, sepsis, Parkinson's disease, immunodeficiency disorders like chronic granulomatous disease (CGD), and X-linked agammaglobulinemia. Thus, PRRs have seems to have a vital role in the therapeutic research mainly in adjuvant designing. Therefore, it is indispensable to have a profound understanding of biological machinery and functional role of PRRs in our immune system. For that proper annotation of PRRs is much required. Thus in Chapter 4 to aid up to this step we have created a simple web server (**http://webs.iiitd.edu.in/raghava/prrpred/)** that allows users to predict whether or not a particular protein is a pattern recognition receptor. Under prediction, the server's web interface contains two sub-modules: (i) Composition Based and (ii) Evolutionary Information Based. The "Composition Based" module enables a user to find a protein sequence based on amino acid composition. This module also gives the user the choice of using the non-hybrid technique, which is solely AAC-based, or the hybrid method, which

is AAC+BLAST-based. The "Evolutionary Information Based" module assists the user in predicting PRRs based on evolutionary information from a protein sequence. The PSSM-400 composition profile for the input protein sequence is generated and used as a feature vector for prediction in this step. This module, like the composition-based module, supports non-hybrid and hybrid models. The web server was built using a responsive HTML template to adjust to the browsing device. As a result, our web server is compatible with a broad range of devices, such as desktop computers, tablets, and smartphones.

Antibiotic resistance is spreading among microorganisms all across the world, and conventional therapies for drug-resistant pathogens are inadequate. With increased pathogen drug resistance, the concern of a post-antibiotic era needs the development of alternatives to standard antibiotics or small molecule-based medicines. Because of their various therapeutic effects, AMPs represent a class of prospective therapeutics with curative promise. Several organisms' innate immune systems rely significantly on these evolutionarily conserved molecules. Defensins are a subclass of AMPs with several roles and modes of action, making them less likely to be drug resistant. Thus, in Chapter 5 for the annotation of defensins and in order to better serve the scientific community, we created a web server called "DefPred," as well as a standalone version that included our top models. The standalone version is Python-based and provides the user with a plethora of choices. The accompanying server, on the other hand, is user-friendly and compatible with a variety of displays, including laptops, Android mobile phones, iPhones, and iPads. In addition, we have given a stand-alone facility in the form of Docker technology. This standalone programme is incorporated into our "GPSRdocker" package, which may be obtained from the website **https://webs.iiitd.edu.in/gpsrdocker/.** We expect that this work will aid vaccine designers as well as provide a better knowledge of immune defense response.

Immunotherapy is the fourth and most advanced pillar of cancer treatment. Over the last decade, there has been a rise in interest in cancer immunotherapy research because to the significant improvement in patient survival, especially in resistant disease types. The use of PRR genes in cancer prognosis is one of the advanced immune-therapeutics techniques that has been extensively studied in the literature. The purpose of chapter 6 was to look at the changed expression profile of PRR genes in the context of survival prediction in patients with UCEC. The discovery of key biomarker genes having a strong connection with survival

can aid in risk group stratification and prognosis. As a result, the discovered biomarker genes can serve as a strong foundation for the investigation of novel therapeutic techniques in the treatment of UCEC. Our study used a variety of bioinformatics methodologies, including network-based approaches, Cox-proportional hazard (PH) survival studies, and clustering-based approaches, to identify important genes and construct highly accurate risk-prediction models. We also assessed the prognostic significance of different clinicopathological features and investigated the molecular mechanisms linked to the discovered genes in order to uncover relevant therapeutic molecules that could improve the survival of patients with UCEC. Finally a risk classification model based on nine PRR-related genes was devised to evaluate survival outcomes and provide personalised anticancer treatment in patients with UCEC. Our findings also imply that combining the suggested gene signature with clinical staging gives better prognosis than staging information alone. As a result, the findings of our study may serve as a potential change to UCEC's existing risk evaluation method. Overall, this work contributes to a better understanding of the oncogenic function of innate immune receptors in UCEC.

In the past, various attempts have been made, and several cancer-specific prognostic models have been established, which are only relevant to single cancer. This study aimed to find "universal or multi-cancer" prognostic biomarkers and build models for predicting survival risk in different types of cancer patients. To do this, we assessed the prognostic significance of 331 PRR gene expression in 33 cancers in terms of patient overall survival. We discovered prognostic biomarker genes for only 30 different types of cancers initially. HR_{SKCM}=1.99 and HR_{THCA}=53.57 were attained by the cancer-specific prognostic models. Furthermore, a comprehensive study was carried out in order to uncover universal biomarker genes across a wide range of cancers. Our best prognostic model included 12 genes (*UNC93B1, ALPK1, APPL1, CASP8, CD5, CLEC2D, HMGB1, HSP90B1, IKBKG, IRF1, KLRB1, NCAN)* and classified risk groups across 6 cancers (HR_{THYM}=14.03, HR_{UCEC}=0.88). Besides offering a complete assessment of the prognostic potential of PRR signaling genes in diverse cancer types, our research might aid in the development of adaptable risk management and treatment options for cancer patients.

Diagnosis at initial stage is very crucial in case of any malignancy especially cancer. A detailed prognosis is frequently necessary for strategic therapeutic intervention planning.

The modern oncology research provided lot of omics data that can be helpful for the development of effective prognostic biomarkers and corresponding therapy. Previous research has highlighted critical roles of various genes belong to important pathway like apoptosis and PRR signalling pathway. In the current study, we compared the prognostic biomarkers and their performance for these two pathways in case of THCA, MESO and SKCM. We chose these three on the basis of the result from our prior study. To do so first correlation among the PRR biomarker genes and apoptotic biomarker genes have been taken. Self-correlation was also taken into consideration. Using peasron correlation coefficient with the effective cut-off an edgelist was developed to create network in case of these three cancers. Network analysis and visualization is also done and shown in the study. We developed a hybrid model to improve the prognostic performance of the biomarkers by combining apoptotic and PRR signalling genes and found that the performance increased in the case of THCA, MESO and SKCM. The high correlation among genes from these two different pathways clarifies that these are not independent and have dependency on each other. Overall all this study highlights the high interconnection between cross pathways in case of cancer. It also encourages more study into the possibilities of such a hybrid model combining several pathways for development in cancer therapy and management.

Altogether, the research presented in this thesis provides several unique approaches for the proper annotation of innate system molecules. Also, these molecules related signaling genes were utilized to create prognostic biomarker in various cancer. We anticipate that clinicians and researchers will use the findings of our investigations to develop advanced cancer treatment approaches.

Bibliography

Abdolvahab, Mohadeseh Haji, Behrad Darvishi, Mohammad Zarei, Keivan Majidzadeh-A, and Leila Farahmand. 2020. “Interferons: Role in Cancer Therapy.” *Immunotherapy* 12 (11): 833–55. https://doi.org/10.2217/imt-2019-0217.

Aggarwal, Charu C. 2014. *Data Classification: Algorithms and Applications*. 1st ed. Chapman & Hall/CRC.

Agrawal, P., R. Kumar, S.S. Usmani, A. Dhall, S. Patiyal, N. Sharma, H. Kaur, et al. 2019. “GPSRdocker: A Docker-Based Resource for Genomics, Proteomics and Systems Biology.” *BioRxiv*. https://doi.org/10.1101/827766.

Agrawal, Piyush, Sherry Bhalla, Kumardeep Chaudhary, Rajesh Kumar, Meenu Sharma, and Gajendra P S Raghava. 2018. “In Silico Approach for Prediction of Antifungal Peptides.” *Frontiers in Microbiology* 9 (February): 323. https://doi.org/10.3389/fmicb.2018.00323.

Agrawal, Piyush, Sumit Kumar, Archana Singh, Gajendra P S Raghava, and Indrakant K Singh. 2019. “NeuroPIpred: A Tool to Predict, Design and Scan Insect Neuropeptides.” *Scientific Reports* 9 (1): 5129. https://doi.org/10.1038/s41598-019-41538-x.

Agrawal, Piyush, and Gajendra P S Raghava. 2018. “Prediction of Antimicrobial Potential of a Chemically Modified Peptide From Its Tertiary Structure.” *Frontiers in Microbiology* 9: 2551. https://doi.org/10.3389/fmicb.2018.02551.

Aithal, Abhijit, Arun Sharma, Shilpy Joshi, Gajendra P S Raghava, and Grish C Varshney. 2012. “PolysacDB: A Database of Microbial Polysaccharide Antigens and Their Antibodies.” *PloS One* 7 (4): e34613. https://doi.org/10.1371/journal.pone.0034613.

Akira, Shizuo. 2011. “Innate Immunity and Adjuvants.” *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 366 (1579): 2748–55. https://doi.org/10.1098/rstb.2011.0106.

Akira, Shizuo, Satoshi Uematsu, and Osamu Takeuchi. 2006. “Pathogen Recognition and Innate Immunity.” *Cell* 124 (4): 783–801. https://doi.org/10.1016/j.cell.2006.02.015.

Allhorn, Svenja, Carsten Boing, Andrea A Koch, Rainer Kimmig, and Isabella Gashaw. 2008. “TLR3 and TLR4 Expression in Healthy and Diseased Human Endometrium.” *Reproductive Biology and Endocrinology : RB&E* 6 (September): 40. https://doi.org/10.1186/1477-7827-6-40.

Alvero, Ayesha B, Rui Chen, Han-Hsuan Fu, Michele Montagna, Peter E Schwartz, Thomas Rutherford, Dan-Arin Silasi, et al. 2009. “Molecular Phenotyping of Human Ovarian

Cancer Stem Cells Unravels the Mechanisms for Repair and Chemoresistance." *Cell Cycle (Georgetown, Tex.)* 8 (1): 158–66. https://doi.org/10.4161/cc.8.1.7533.

Ansari, Hifzur Rahman, Darren R Flower, and G P S Raghava. 2010. "AntigenDB: An Immunoinformatics Database of Pathogen Antigens." *Nucleic Acids Research* 38 (Database issue): D847-53. https://doi.org/10.1093/nar/gkp830.

Arora, Chakit, Dilraj Kaur, and Gajendra P S Raghava. 2021. "Universal and Cross-Cancer Prognostic Biomarkers for Predicting Survival Risk of Cancer Patients from Expression Profile of Apoptotic Pathway Genes." *Proteomics*, October, e2000311. https://doi.org/10.1002/pmic.202000311.

Atanasova, M., A. Patronov, I. Dimitrov, D. R. Flower, and I. Doytchinova. 2013. "EpiDOCK: A Molecular Docking-Based Tool for MHC Class II Binding Prediction." *Protein Engineering Design and Selection* 26 (10): 631–34. https://doi.org/10.1093/protein/gzt018.

Bai, Ling, Wenqian Li, Weijia Zheng, Dongsheng Xu, Naifei Chen, and Jiuwei Cui. 2020. "Promising Targets Based on Pattern Recognition Receptors for Cancer Immunotherapy." *Pharmacological Research* 159 (September): 105017. https://doi.org/10.1016/j.phrs.2020.105017.

Bauer, Johannes H, and Stephen L Helfand. 2006. "New Tricks of an Old Molecule: Lifespan Regulation by P53." *Aging Cell* 5 (5): 437–40. https://doi.org/10.1111/j.1474-9726.2006.00228.x.

Bauer, Johannes H, Stephen L Helfand, Milena Gasco, Shukri Shami, and Tim Crook. 2002. "The P53 Pathway in Breast Cancer." *Breast Cancer Research : BCR* 4 (2): 70–76. https://doi.org/10.1186/bcr426.

Baylor, Norman W, William Egan, and Paul Richman. 2002. "Aluminum Salts in Vaccines--US Perspective." *Vaccine* 20 Suppl 3 (May): S18-23. https://doi.org/10.1016/s0264-410x(02)00166-4.

Bendtsen, Jannick Dyrløv, Lars Juhl Jensen, Nikolaj Blom, Gunnar Von Heijne, and Søren Brunak. 2004. "Feature-Based Prediction of Non-Classical and Leaderless Protein Secretion." *Protein Engineering, Design and Selection*. https://doi.org/10.1093/protein/gzh037.

Bhardwaj, Anshu, Vinod Scaria, Gajendra Pal Singh Raghava, Andrew Michael Lynn, Nagasuma Chandra, Sulagna Banerjee, Muthukurussi V Raghunandanan, et al. 2011.

"Open Source Drug Discovery--a New Paradigm of Collaborative Research in Tuberculosis Drug Development." *Tuberculosis (Edinburgh, Scotland)* 91 (5): 479–86. https://doi.org/10.1016/j.tube.2011.06.004.

Bhasin, Manoj, and G P S Raghava. 2004. "ESLpred: SVM-Based Method for Subcellular Localization of Eukaryotic Proteins Using Dipeptide Composition and PSI-BLAST." *Nucleic Acids Research* 32 (Web Server issue): W414-9. https://doi.org/10.1093/nar/gkh350.

———. 2007. "A Hybrid Approach for Predicting Promiscuous MHC Class I Restricted T Cell Epitopes." *Journal of Biosciences* 32 (1): 31–42. http://www.ncbi.nlm.nih.gov/pubmed/17426378.

Bhasin, Manoj, Harpreet Singh, and G P S Raghava. 2003. "MHCBN: A Comprehensive Database of MHC Binding and Non-Binding Peptides." *Bioinformatics (Oxford, England)* 19 (5): 665–66. http://www.ncbi.nlm.nih.gov/pubmed/12651731.

Bian, Jia, Yuzi Xu, Fei Wu, Qiangwei Pan, and Yunlong Liu. 2020. "Identification of a Five-Gene Signature for Predicting the Progression and Prognosis of Stage I Endometrial Carcinoma." *Oncology Letters* 20 (3): 2396–2410. https://doi.org/10.3892/ol.2020.11798.

Black, Carin, Amy Feng, Sophie Bittinger, Michael Quinn, Deborah Neesham, and Orla McNally. 2016. "Uterine Papillary Serous Carcinoma: A Single-Institution Review of 62 Cases." *International Journal of Gynecological Cancer : Official Journal of the International Gynecological Cancer Society* 26 (1): 133–40. https://doi.org/10.1097/IGC.0000000000000569.

Blasco, Benjamin, Didier Leroy, and David A Fidock. 2017. "Antimalarial Drug Resistance: Linking Plasmodium Falciparum Parasite Biology to the Clinic." *Nature Medicine* 23 (8): 917–28. https://doi.org/10.1038/nm.4381.

Boyanova, Lyudmila, Rumyana Markovska, and Ivan Mitov. 2019. "Multidrug Resistance in Anaerobes." *Future Microbiology* 14 (August): 1055–64. https://doi.org/10.2217/fmb-2019-0132.

Brackett, Craig M, Bojidar Kojouharov, Jean Veith, Kellee F Greene, Lyudmila G Burdelya, Sandra O Gollnick, Scott I Abrams, and Andrei V Gudkov. 2016. "Toll-like Receptor-5 Agonist, Entolimod, Suppresses Metastasis and Induces Immunity by Stimulating an NK-Dendritic-CD8+ T-Cell Axis." *Proceedings of the National Academy of Sciences of*

the United States of America 113 (7): E874-83. https://doi.org/10.1073/pnas.1521359113.

Breuer, Karin, Amir K Foroushani, Matthew R Laird, Carol Chen, Anastasia Sribnaia, Raymond Lo, Geoffrey L Winsor, Robert E W Hancock, Fiona S L Brinkman, and David J Lynn. 2013. "InnateDB: Systems Biology of Innate Immunity and beyond--Recent Updates and Continuing Curation." *Nucleic Acids Research* 41 (Database issue): D1228-33. https://doi.org/10.1093/nar/gks1147.

Bryant, Clare E, and Tom P Monie. 2012. "Mice, Men and the Relatives: Cross-Species Studies Underpin Innate Immunity." *Open Biology* 2 (4): 120015. https://doi.org/10.1098/rsob.120015.

Bubna, Aditya Kumar. 2015. "Imiquimod - Its Role in the Treatment of Cutaneous Malignancies." *Indian Journal of Pharmacology* 47 (4): 354–59. https://doi.org/10.4103/0253-7613.161249.

Buchmann, Kurt. 2014. "Evolution of Innate Immunity: Clues from Invertebrates via Fish to Mammals." *Frontiers in Immunology* 5: 459. https://doi.org/10.3389/fimmu.2014.00459.

Burley, Stephen K, Helen M Berman, Charmi Bhikadiya, Chunxiao Bi, Li Chen, Luigi Di Costanzo, Cole Christie, et al. 2019. "RCSB Protein Data Bank: Biological Macromolecular Structures Enabling Research and Education in Fundamental Biology, Biomedicine, Biotechnology and Energy." *Nucleic Acids Research* 47 (D1): D464–74. https://doi.org/10.1093/nar/gky1004.

Camacho, Christiam, George Coulouris, Vahram Avagyan, Ning Ma, Jason Papadopoulos, Kevin Bealer, and Thomas L. Madden. 2009. "BLAST+: Architecture and Applications." *BMC Bioinformatics*. https://doi.org/10.1186/1471-2105-10-421.

Chan, James K, Johannes Roth, Joost J Oppenheim, Kevin J Tracey, Thomas Vogl, Marc Feldmann, Nicole Horwood, and Jagdeep Nanchahal. 2012. "Alarmins: Awaiting a Clinical Response." *The Journal of Clinical Investigation* 122 (8): 2711–19. https://doi.org/10.1172/JCI62423.

Chang, C-H, C-H Chung, C-C Hsu, H-C Peng, and T-F Huang. 2014. "Inhibitory Effects of Polypeptides Derived from a Snake Venom C-Type Lectin, Aggretin, on Tumor Cell-Induced Platelet Aggregation." *Journal of Thrombosis and Haemostasis : JTH* 12 (4): 540–49. https://doi.org/10.1111/jth.12519.

Chang, Yuchi C, Vandana Madkan, Robert Cook-Norris, Karan Sra, and Stephen Tyring. 2005. “Current and Potential Uses of Imiquimod.” *Southern Medical Journal* 98 (9): 914–20. https://doi.org/10.1097/01.smj.0000176712.01491.98.

Chaplin, David D, Achal Pashine, Nicholas M Valiante, Jeffrey B Ulmer, Gustavo P Amarante-Mendes, Sandy Adjemian, Laura Migliari Branco, Larissa C Zanetti, Ricardo Weinlich, and Karina R Bortoluci. 2018. “Pattern Recognition Receptors and the Host Cell Death Molecular Machinery .” *Frontiers in Immunology* 9 (4 Suppl): 2379. https://doi.org/10.1038/nm1210.

Chaudhary, Kumardeep, Ritesh Kumar, Sandeep Singh, Abhishek Tuknait, Ankur Gautam, Deepika Mathur, Priya Anand, Grish C. Varshney, and Gajendra P.S. Raghava. 2016. “A Web Server and Mobile App for Computing Hemolytic Potency of Peptides.” *Scientific Reports*. https://doi.org/10.1038/srep22843.

Chaudhary, Kumardeep, Gandharva Nagpal, Sandeep Kumar Dhanda, and Gajendra P S Raghava. 2016. “Prediction of Immunomodulatory Potential of an RNA Sequence for Designing Non-Toxic SiRNAs and RNA-Based Vaccine Adjuvants.” *Scientific Reports* 6 (1): 20678. https://doi.org/10.1038/srep20678.

Chauhan, Jagat S, Nitish K Mishra, and Gajendra P S Raghava. 2010. “Prediction of GTP Interacting Residues, Dipeptides and Tripeptides in a Protein from Its Evolutionary Information.” *BMC Bioinformatics* 11 (1): 301. https://doi.org/10.1186/1471-2105-11-301.

Chen, Peigen, Yuebo Yang, Yu Zhang, Senwei Jiang, Xiaomao Li, and Jing Wan. 2020. “Identification of Prognostic Immune-Related Genes in the Tumor Microenvironment of Endometrial Cancer.” *Aging* 12 (4): 3371–87. https://doi.org/10.18632/aging.102817.

Chen, Wei, Guanghua Li, Jianjun Peng, Weigang Dai, Qiao Su, and Yulong He. 2020. “Transcriptomic Analysis Reveals That Heat Shock Protein 90α Is a Potential Diagnostic and Prognostic Biomarker for Cancer.” *European Journal of Cancer Prevention : The Official Journal of the European Cancer Prevention Organisation (ECP)* 29 (4): 357–64. https://doi.org/10.1097/CEJ.0000000000000549.

Chen, Yongqiang, Elizabeth S Henson, Wenyan Xiao, Daniel Huang, Eileen M McMillan-Ward, Sara J Israels, and Spencer B Gibson. 2016. “Tyrosine Kinase Receptor EGFR Regulates the Switch in Cancer Cells between Cell Survival and Cell Death Induced by Autophagy in Hypoxia.” *Autophagy* 12 (6): 1029–46.

https://doi.org/10.1080/15548627.2016.1164357.

Coffman, Robert L, Alan Sher, and Robert A Seder. 2010. “Vaccine Adjuvants: Putting Innate Immunity to Work.” *Immunity* 33 (4): 492–503. https://doi.org/10.1016/j.immuni.2010.10.002.

Coll-de la Rubia, Eva, Elena Martinez-Garcia, Gunnar Dittmar, Antonio Gil-Moreno, Silvia Cabrera, and Eva Colas. 2020. “Prognostic Biomarkers in Endometrial Cancer: A Systematic Review and Meta-Analysis.” *Journal of Clinical Medicine* 9 (6). https://doi.org/10.3390/jcm9061900.

Cordoba, Santiago Rodriguez de, Agustin Tortajada, Claire L Harris, and B Paul Morgan. 2012. “Complement Dysregulation and Disease: From Genes and Proteins to Diagnostics and Drugs.” *Immunobiology* 217 (11): 1034–46. https://doi.org/10.1016/j.imbio.2012.07.021.

Cronin, Kathleen A, Andrew J Lake, Susan Scott, Recinda L Sherman, Anne-Michelle Noone, Nadia Howlader, S Jane Henley, et al. 2018. “Annual Report to the Nation on the Status of Cancer, Part I: National Cancer Statistics.” *Cancer* 124 (13): 2785–2800. https://doi.org/10.1002/cncr.31551.

Cui, Minghua, Quanzhu Chen, Chaojun He, Nan Wang, Yong Yu, Ziyang Sun, Zhenhua Lin, et al. 2019. “A Single Nucleotide Polymorphism CTSB Rs12898 Is Associated with Primary Hepatic Cancer in a Chinese Population.” *International Journal of Clinical and Experimental Pathology* 12 (8): 3063–69.

D’Agostini, Cartesio, Francesca Pica, Giuseppe Febbraro, Sandro Grelli, Carlo Chiavaroli, and Enrico Garaci. 2005. “Antitumour Effect of OM-174 and Cyclophosphamide on Murine B16 Melanoma in Different Experimental Conditions.” *International Immunopharmacology* 5 (7–8): 1205–12. https://doi.org/10.1016/j.intimp.2005.02.013.

D’Aloisio, Vera, Paolo Dognini, Gillian A Hutcheon, and Christopher R Coxon. 2021. “PepTherDia: Database and Structural Composition Analysis of Approved Peptide Therapeutics and Diagnostics.” *Drug Discovery Today*, February. https://doi.org/10.1016/j.drudis.2021.02.019.

D’Souza, R S, K G Bhat, D Sailaja, D V Babji, T K Bandiwadekar, and R M Katgalkar. 2013. “Analysis of Expression and Localization of TLR-2 by Immunofluorescent Technique in Healthy and Inflammed Oral Tissues.” *J Clin Diagn Res* 7 (12): 2683–2780. https://doi.org/10.7860/JCDR/2013/6745.3745.

Damasdi, Miklos, Krisztina Kovacs, Nelli Farkas, Ferenc Jakab, and Gyula Kovacs. 2017. "Down-Regulation of Toll-like Receptor TLR4 Is Associated with HPV DNA Integration in Penile Carcinoma." *Anticancer Research* 37 (10): 5515–19. https://doi.org/10.21873/anticanres.11982.

Debatin, K M. 2004. "Apoptosis Pathways in Cancer and Cancer Therapy." *Cancer Immunol Immunother* 53 (3): 153–59. https://doi.org/10.1007/s00262-003-0474-8.

Degn, S E, and S Thiel. 2013. "Humoral Pattern Recognition and the Complement System." *Scandinavian Journal of Immunology* 78 (2): 181–93. https://doi.org/10.1111/sji.12070.

Demaria, Olivier, Stephanie Cornen, Marc Daeron, Yannis Morel, Ruslan Medzhitov, and Eric Vivier. 2019. "Harnessing Innate Immunity in Cancer Therapy." *Nature* 574 (7776): 45–56. https://doi.org/10.1038/s41586-019-1593-5.

Dhall, Anjali, Sumeet Patiyal, Neelam Sharma, Salman Sadullah Usmani, and Gajendra P S Raghava. 2021. "Computer-Aided Prediction and Design of IL-6 Inducing Peptides: IL-6 Plays a Crucial Role in COVID-19." *Briefings in Bioinformatics* 22 (2): 936–45. https://doi.org/10.1093/bib/bbaa259.

Dhanda, Sandeep Kumar, Sudheer Gupta, Pooja Vir, and G P S Raghava. 2013. "Prediction of IL4 Inducing Peptides." *Clinical & Developmental Immunology* 2013 (December): 263952. https://doi.org/10.1155/2013/263952.

Dhanda, Sandeep Kumar, Swapnil Mahajan, Sinu Paul, Zhen Yan, Haeuk Kim, Martin Closter Jespersen, Vanessa Jurtz, et al. 2019. "IEDB-AR: Immune Epitope Database-Analysis Resource in 2019." *Nucleic Acids Research* 47 (W1): W502–6. https://doi.org/10.1093/nar/gkz452.

Dhanda, Sandeep Kumar, Salman Sadullah Usmani, Piyush Agrawal, Gandharva Nagpal, Ankur Gautam, and Gajendra P S Raghava. 2017. "Novel in Silico Tools for Designing Peptide-Based Subunit Vaccines and Immunotherapeutics." *Briefings in Bioinformatics* 18 (3): 467–78. https://doi.org/10.1093/bib/bbw025.

Dhanda, Sandeep Kumar, Pooja Vir, and Gajendra P S Raghava. 2013. "Designing of Interferon-Gamma Inducing MHC Class-II Binders." *Biology Direct* 8 (December): 30. https://doi.org/10.1186/1745-6150-8-30.

Dostert, Melanie, Corrie R Belanger, and Robert E W Hancock. 2019. "Design and Assessment of Anti-Biofilm Peptides: Steps Toward Clinical Application." *Journal of Innate Immunity* 11 (3): 193–204. https://doi.org/10.1159/000491497.

Dowling, Jennifer K, and Ashley Mansell. 2016. “Toll-like Receptors: The Swiss Army Knife of Immunity and Vaccine Development.” *Clinical & Translational Immunology* 5 (5): e85. https://doi.org/10.1038/cti.2016.22.

Fakhri, Sajad, Fatemeh Abbaszadeh, Masoumeh Jorjani, and Mohammad Hossein Pourgholami. 2021. “The Effects of Anticancer Medicinal Herbs on Vascular Endothelial Growth Factor Based on Pharmacological Aspects: A Review Study.” *Nutrition and Cancer* 73 (1): 1–15. https://doi.org/10.1080/01635581.2019.1673451.

Farrugia, Mark, and Byron Baron. 2017. “The Role of Toll-Like Receptors in Autoimmune Diseases through Failure of the Self-Recognition Mechanism.” *International Journal of Inflammation* 2017: 8391230. https://doi.org/10.1155/2017/8391230.

Fearon, D T, and R M Locksley. 1996. “The Instructive Role of Innate Immunity in the Acquired Immune Response.” *Science (New York, N.Y.)* 272 (5258): 50–53. http://www.ncbi.nlm.nih.gov/pubmed/8600536.

Franchi, Luigi, Neil Warner, Kyle Viani, and Gabriel Nuñez. 2009. “Function of Nod-like Receptors in Microbial Recognition and Host Defense.” *Immunological Reviews* 227 (1): 106–28. https://doi.org/10.1111/j.1600-065X.2008.00734.x.

Fu, Limin, Beifang Niu, Zhengwei Zhu, Sitao Wu, and Weizhong Li. 2012. “CD-HIT: Accelerated for Clustering the next-Generation Sequencing Data.” *Bioinformatics (Oxford, England)* 28 (23): 3150–52. https://doi.org/10.1093/bioinformatics/bts565.

Fukata, Masayuki, Anli Chen, Arunan S Vamadevan, Jason Cohen, Keith Breglio, Suneeta Krishnareddy, David Hsu, et al. 2007. “Toll-like Receptor-4 Promotes the Development of Colitis-Associated Colorectal Tumors.” *Gastroenterology* 133 (6): 1869–81. https://doi.org/10.1053/j.gastro.2007.09.008.

Galli, Stephen J, Niels Borregaard, and Thomas A Wynn. 2011. “Phenotypic and Functional Plasticity of Cells of Innate Immunity: Macrophages, Mast Cells and Neutrophils.” *Nature Immunology* 12 (11): 1035–44. https://doi.org/10.1038/ni.2109.

Gao, Qing-Bin, Hongyu Zhao, Xiaofei Ye, and Jia He. 2012a. “Prediction of Pattern Recognition Receptor Family Using Pseudo-Amino Acid Composition.” *Biochemical and Biophysical Research Communications* 417 (1): 73–77. https://doi.org/10.1016/j.bbrc.2011.11.057.

———. 2012b. “Prediction of Pattern Recognition Receptor Family Using Pseudo-Amino Acid Composition.” *Biochemical and Biophysical Research Communications* 417 (1):

73–77. https://doi.org/10.1016/j.bbrc.2011.11.057.

———. 2012c. “Prediction of Pattern Recognition Receptor Family Using Pseudo-Amino Acid Composition.” *Biochemical and Biophysical Research Communications* 417 (1): 73–77. https://doi.org/10.1016/j.bbrc.2011.11.057.

Garg, Aarti, and Gajendra P.S. Raghava. 2008a. “A Machine Learning Based Method for the Prediction of Secretory Proteins Using Amino Acid Composition, Their Order and Similarity-Search.” *In Silico Biology*.

Garg, Aarti, and Gajendra P S Raghava. 2008b. “ESLpred2: Improved Method for Predicting Subcellular Localization of Eukaryotic Proteins.” *BMC Bioinformatics* 9 (1): 503. https://doi.org/10.1186/1471-2105-9-503.

Garlapati, Srinivas, Marina Facci, Monika Polewicz, Stacy Strom, Lorne A Babiuk, George Mutwiri, Robert E W Hancock, Melissa R Elliott, and Volker Gerdts. 2009. “Strategies to Link Innate and Adaptive Immunity When Designing Vaccine Adjuvants.” *Veterinary Immunology and Immunopathology* 128 (1–3): 184–91. https://doi.org/10.1016/j.vetimm.2008.10.298.

Gogas, Helen, John Ioannovich, Urania Dafni, Catherine Stavropoulou-Giokas, Konstantina Frangia, Dimosthenis Tsoutsos, Petros Panagiotou, et al. 2006. “Prognostic Significance of Autoimmunity during Treatment of Melanoma with Interferon.” *The New England Journal of Medicine* 354 (7): 709–18. https://doi.org/10.1056/NEJMoa053007.

Golkar, Zhabiz, Omar Bagasra, and Donald Gene Pace. 2014. “Bacteriophage Therapy: A Potential Solution for the Antibiotic Resistance Crisis.” *Journal of Infection in Developing Countries* 8 (2): 129–36. https://doi.org/10.3855/jidc.3573.

Gottwald, Leszek, Piotr Pluta, Janusz Piekarski, Michal Spych, Katarzyna Hendzel, Katarzyna Topczewska-Tylinska, Dariusz Nejc, Robert Bibik, Jerzy Korczynski, and Aleksandra Cialkowska-Rysz. 2010. “Long-Term Survival of Endometrioid Endometrial Cancer Patients.” *Archives of Medical Science : AMS* 6 (6): 937–44. https://doi.org/10.5114/aoms.2010.19305.

Goutagny, Nadege, Yann Estornes, Uzma Hasan, Serge Lebecque, and Christophe Caux. 2012. “Targeting Pattern Recognition Receptors in Cancer Immunotherapy.” *Targeted Oncology* 7 (1): 29–54. https://doi.org/10.1007/s11523-012-0213-1.

Gowing, Stephen D, Simon C Chow, Jonathan J Cools-Lartigue, Crystal B Chen, Sara Najmeh, Henry Y Jiang, France Bourdeau, et al. 2017. “Gram-Positive Pneumonia

Augments Non-Small Cell Lung Cancer Metastasis via Host Toll-like Receptor 2 Activation." *International Journal of Cancer* 141 (3): 561–71. https://doi.org/10.1002/ijc.30734.

Gregg, Kelsey A, Erin Harberts, Francesca M Gardner, Mark R Pelletier, Corinne Cayatte, Li Yu, Michael P McCarthy, Jason D Marshall, and Robert K Ernst. 2017. "Rationally Designed TLR4 Ligands for Vaccine Adjuvant Discovery." *MBio* 8 (3). https://doi.org/10.1128/mBio.00492-17.

Grimmig, Tanja, Niels Matthes, Katharina Hoeland, Sudipta Tripathi, Anil Chandraker, Martin Grimm, Romana Moench, et al. 2015. "TLR7 and TLR8 Expression Increases Tumor Cell Proliferation and Promotes Chemoresistance in Human Pancreatic Cancer." *International Journal of Oncology* 47 (3): 857–66. https://doi.org/10.3892/ijo.2015.3069.

Grimmig, Tanja, Romana Moench, Jennifer Kreckel, Stephanie Haack, Felix Rueckert, Roberta Rehder, Sudipta Tripathi, et al. 2016. "Toll Like Receptor 2, 4, and 9 Signaling Promotes Autoregulative Tumor Cell Growth and VEGF/PDGF Expression in Human Pancreatic Cancer." *International Journal of Molecular Sciences* 17 (12). https://doi.org/10.3390/ijms17122060.

Gupta, Sudheer, Hifzur Rahman Ansari, Ankur Gautam, Gajendra PS Raghava, and Gajendra P S Raghava. 2013. "Identification of B-Cell Epitopes in an Antigen for Inducing Specific Class of Antibodies." *Biology Direct* 8 (1): 27. https://doi.org/10.1186/1745-6150-8-27.

Gupta, Sudheer, Pallavi Kapoor, Kumardeep Chaudhary, Ankur Gautam, Rahul Kumar, and Gajendra P S Raghava. 2015. "Peptide Toxicity Prediction." *Methods in Molecular Biology (Clifton, N.J.)* 1268: 143–57. https://doi.org/10.1007/978-1-4939-2285-7_7.

Haghparast, Alireza, Amin Zakeri, and Mahboubeh Ebrahimian and Mohammad Ramezani. 2016. "Targeting Pattern Recognition Receptors (PRRs) in Nano- Adjuvants: Current Perspectives." *Current Bionanotechnology (Discontinued)*. https://doi.org/http://dx.doi.org/10.2174/2213529402666160601125159.

Haider, Tanweer, Rahul Tiwari, Suresh Prasad Vyas, and Vandana Soni. 2019. "Molecular Determinants as Therapeutic Targets in Cancer Chemotherapy: An Update." *Pharmacology & Therapeutics* 200 (August): 85–109. https://doi.org/10.1016/j.pharmthera.2019.04.011.

Han, Mi, Hua Jing Yang, and Qin Lin. 2019. “KLHL14, an Ovarian and Endometrial-Specific Gene, Is over-Expressed in Ovarian and Endometrial Cancer.” *Mathematical Biosciences and Engineering : MBE* 17 (2): 1702–17. https://doi.org/10.3934/mbe.2020089.

Hasuo, H, and T Akasu. 2001. “Activation of Inhibitory Pathways Suppresses the Induction of Long-Term Potentiation in Neurons of the Rat Lateral Septal Nucleus.” *Neuroscience* 105 (2): 343–52. https://doi.org/10.1016/s0306-4522(01)00195-6.

Hilchie, Ashley L, Kelli Wuerth, and Robert E W Hancock. 2013. “Immune Modulation by Multifaceted Cationic Host Defense (Antimicrobial) Peptides.” *Nature Chemical Biology* 9 (12): 761–68. https://doi.org/10.1038/nchembio.1393.

Hoving, J. Claire, Gillian J. Wilson, and Gordon D. Brown. 2014a. “Signalling C-Type Lectin Receptors, Microbial Recognition and Immunity.” *Cellular Microbiology* 16 (2): 185–94. https://doi.org/10.1111/cmi.12249.

Hoving, J Claire, Gillian J Wilson, and Gordon D Brown. 2014b. “Signalling C-Type Lectin Receptors, Microbial Recognition and Immunity.” *Cellular Microbiology* 16 (2): 185–94. https://doi.org/10.1111/cmi.12249.

Ikebe, Mio, Yoshiki Kitaura, Masafumi Nakamura, Haruo Tanaka, Akio Yamasaki, Shuntaro Nagai, Junji Wada, et al. 2009. “Lipopolysaccharide (LPS) Increases the Invasive Ability of Pancreatic Cancer Cells through the TLR4/MyD88 Signaling Pathway.” *Journal of Surgical Oncology* 100 (8): 725–31. https://doi.org/10.1002/jso.21392.

Ikehata, Naoki, Masakatsu Takanashi, Takafumi Satomi, Masato Watanabe, On Hasegawa, Michihide Kono, Ai Enomoto, Daichi Chikazu, and Masahiko Kuroda. 2018. “Toll-like Receptor 2 Activation Implicated in Oral Squamous Cell Carcinoma Development.” *Biochemical and Biophysical Research Communications* 495 (3): 2227–34. https://doi.org/10.1016/j.bbrc.2017.12.098.

Inglefield, Jon R, Calin Dan Dumitru, Sefik S Alkan, Sheila J Gibson, Kenneth E Lipson, Mark A Tomai, Chris J Larson, and John P Vasilakos. 2008. “TLR7 Agonist 852A Inhibition of Tumor Cell Proliferation Is Dependent on Plasmacytoid Dendritic Cells and Type I IFN.” *Journal of Interferon & Cytokine Research : The Official Journal of the International Society for Interferon and Cytokine Research* 28 (4): 253–63. https://doi.org/10.1089/jir.2007.0097.

Iwasaki, Akiko, and Ruslan Medzhitov. 2015. “Control of Adaptive Immunity by the Innate

Immune System." *Nature Immunology* 16 (4): 343–53. https://doi.org/10.1038/ni.3123.

Jain, Aakanksha, and Chandrashekhar Pasare. 2017. "Innate Control of Adaptive Immunity: Beyond the Three-Signal Paradigm." *Journal of Immunology (Baltimore, Md. : 1950)* 198 (10): 3791–3800. https://doi.org/10.4049/jimmunol.1602000.

Jespersen, Martin Closter, Bjoern Peters, Morten Nielsen, and Paolo Marcatili. 2017. "BepiPred-2.0: Improving Sequence-Based B-Cell Epitope Prediction Using Conformational Epitopes." *Nucleic Acids Research* 45 (W1): W24–29. https://doi.org/10.1093/nar/gkx346.

Jiang, Ninghong, Feng Xie, Qisang Guo, Ming-Qing Li, Jingjing Xiao, and Long Sui. 2017. "Toll-like Receptor 4 Promotes Proliferation and Apoptosis Resistance in Human Papillomavirus-Related Cervical Cancer Cells through the Toll-like Receptor 4/Nuclear Factor-KappaB Pathway." *Tumour Biology : The Journal of the International Society for Oncodevelopmental Biology and Medicine* 39 (6): 1010428317710586. https://doi.org/10.1177/1010428317710586.

Jiang, S, L Wang, M Huang, Z Jia, T Weinert, E Warkentin, C Liu, et al. 2017. "DM9 Domain Containing Protein Functions As a Pattern Recognition Receptor with Broad Microbial Recognition Spectrum." *Front Immunol* 8: 1607. https://doi.org/10.3389/fimmu.2017.01607.

Jiang, Yu, Mengxin Geng, and Liping Bai. 2020. "Targeting Biofilms Therapy: Current Research Strategies and Development Hurdles." *Microorganisms* 8 (8). https://doi.org/10.3390/microorganisms8081222.

Jouhi, Lauri, Hesham Mohamed, Antti Makitie, Satu Maria Remes, Caj Haglund, Timo Atula, and Jaana Hagstrom. 2017. "Toll-like Receptor 5 and 7 Expression May Impact Prognosis of HPV- Positive Oropharyngeal Squamous Cell Carcinoma Patients." *Cancer Immunology, Immunotherapy : CII* 66 (12): 1619–29. https://doi.org/10.1007/s00262-017-2054-3.

Kaiser, W J, H Sridharan, C Huang, P Mandal, J W Upton, P J Gough, C A Sehon, R W Marquis, J Bertin, and E S Mocarski. 2013. "Toll-like Receptor 3-Mediated Necrosis via TRIF, RIP3, and MLKL." *J Biol Chem* 288 (43): 31268–79. https://doi.org/10.1074/jbc.M113.462341.

Kang, Xinyue, Fanyi Dong, Cheng Shi, Shicai Liu, Jian Sun, Jiaxin Chen, Haiqi Li, Hanmei Xu, Xingzhen Lao, and Heng Zheng. 2019. "DRAMP 2.0, an Updated Data Repository

of Antimicrobial Peptides." *Scientific Data* 6 (1): 148. https://doi.org/10.1038/s41597-019-0154-y.

Kaundal, Rakesh, and Gajendra P.S. Raghava. 2009. "RSLpred: An Integrative System for Predicting Subcellular Localization of Rice Proteins Combining Compositional and Evolutionary Information." *Proteomics*. https://doi.org/10.1002/pmic.200700597.

Kaur, Bani Preet, and Elizabeth Secord. 2019. "Innate Immunity." *Pediatric Clinics of North America* 66 (5): 905–11. https://doi.org/10.1016/j.pcl.2019.06.011.

Kaur, D., C. Arora, and G.P.S. Raghava. 2019. "A Hybrid Model for Predicting Pattern Recognition Receptors Using Evolutionary Information." *BioRxiv*. https://doi.org/10.1101/846469.

Kaur, D., S. Patiyal, N. Sharma, S.S. Usmani, and G.P.S. Raghava. 2019. "PRRDB 2.0: A Comprehensive Database of Pattern-Recognition Receptors and Their Ligands." *Database : The Journal of Biological Databases and Curation* 2019. https://doi.org/10.1093/database/baz076.

Kaur, Dilraj, Chakit Arora, and Gajendra P S Raghava. 2020. "A Hybrid Model for Predicting Pattern Recognition Receptors Using Evolutionary Information." *Frontiers in Immunology* 11: 71. https://doi.org/10.3389/fimmu.2020.00071.

Kaur, Dilraj, Sumeet Patiyal, Neelam Sharma, Salman Sadullah Usmani, and Gajendra P S Raghava. 2019a. "PRRDB 2.0: A Comprehensive Database of Pattern-Recognition Receptors and Their Ligands." *Database : The Journal of Biological Databases and Curation* 2019 (January). https://doi.org/10.1093/database/baz076.

———. 2019b. "PRRDB 2.0: A Comprehensive Database of Pattern-Recognition Receptors and Their Ligands." *Database : The Journal of Biological Databases and Curation* 2019 (January). https://doi.org/10.1093/database/baz076.

Kaur, Harpreet, and G P S Raghava. 2004. "Prediction of Alpha-Turns in Proteins Using PSI-BLAST Profiles and Secondary Structure Information." *Proteins* 55 (1): 83–90. https://doi.org/10.1002/prot.10569.

Kawai, Taro, and Shizuo Akira. 2009. "The Roles of TLRs, RLRs and NLRs in Pathogen Recognition." *International Immunology* 21 (4): 317–37. https://doi.org/10.1093/intimm/dxp017.

———. 2010. "The Role of Pattern-Recognition Receptors in Innate Immunity: Update on Toll-like Receptors." *Nature Immunology* 11 (5): 373–84.

https://doi.org/10.1038/ni.1863.

Kelley, Lawrence A, Stefans Mezulis, Christopher M Yates, Mark N Wass, and Michael J E Sternberg. 2015. “The Phyre2 Web Portal for Protein Modeling, Prediction and Analysis.” *Nature Protocols* 10 (6): 845–58. https://doi.org/10.1038/nprot.2015.053.

Kelly, Michael G, Ayesha B Alvero, Rui Chen, Dan-Arin Silasi, Vikki M Abrahams, Serena Chan, Irene Visintin, Thomas Rutherford, and Gil Mor. 2006. “TLR-4 Signaling Promotes Tumor Growth and Paclitaxel Chemoresistance in Ovarian Cancer.” *Cancer Research* 66 (7): 3859–68. https://doi.org/10.1158/0008-5472.CAN-05-3948.

Kennedy, M N, G E Mullen, C A Leifer, C Lee, A Mazzoni, K N Dileepan, and D M Segal. 2004. “A Complex of Soluble MD-2 and Lipopolysaccharide Serves as an Activating Ligand for Toll-like Receptor 4.” *J Biol Chem* 279 (33): 34698–704. https://doi.org/10.1074/jbc.M405444200.

Kim, Sunghwan, Jie Chen, Tiejun Cheng, Asta Gindulyte, Jia He, Siqian He, Qingliang Li, et al. 2019. “PubChem 2019 Update: Improved Access to Chemical Data.” *Nucleic Acids Research* 47 (D1): D1102–9. https://doi.org/10.1093/nar/gky1033.

Kim, Youn H, Michael Girardi, Madeleine Duvic, Timothy Kuzel, Brian K Link, Lauren Pinter-Brown, and Alain H Rook. 2010. “Phase I Trial of a Toll-like Receptor 9 Agonist, PF-3512676 (CPG 7909), in Patients with Treatment-Refractory, Cutaneous T-Cell Lymphoma.” *Journal of the American Academy of Dermatology* 63 (6): 975–83. https://doi.org/10.1016/j.jaad.2009.12.052.

Kim, Young Keun, Jeon Soo Shin, and Moon H Nahm. 2016. “NOD-Like Receptors in Infection, Immunity, and Diseases.” *Yonsei Medical Journal* 57 (1): 5–14. https://doi.org/10.3349/ymj.2016.57.1.5.

Komada, Takanori, and Daniel A Muruve. 2019. “The Role of Inflammasomes in Kidney Disease.” *Nature Reviews. Nephrology* 15 (8): 501–20. https://doi.org/10.1038/s41581-019-0158-z.

Koo, Hyun, Raymond N Allan, Robert P Howlin, Paul Stoodley, and Luanne Hall-Stoodley. 2017. “Targeting Microbial Biofilms: Current and Prospective Therapeutic Strategies.” *Nature Reviews. Microbiology* 15 (12): 740–55. https://doi.org/10.1038/nrmicro.2017.99.

Korb, Martin, Aistair G Rust, Vesteinn Thorsson, Christophe Battail, Bin Li, Daehee Hwang, Kathleen A Kennedy, et al. 2008. “The Innate Immune Database (IIDB).” *BMC*

Immunology 9 (1): 7. https://doi.org/10.1186/1471-2172-9-7.

Krieg, Arthur M. 2007. "Development of TLR9 Agonists for Cancer Therapy." *The Journal of Clinical Investigation* 117 (5): 1184–94. https://doi.org/10.1172/JCI31414.

Krol, E, T Mentzel, D Chinchilla, T Boller, G Felix, B Kemmerling, S Postel, et al. 2010. "Perception of the Arabidopsis Danger Signal Peptide 1 Involves the Pattern Recognition Receptor AtPEPR1 and Its Close Homologue AtPEPR2." *J Biol Chem* 285 (18): 13471–79. https://doi.org/10.1074/jbc.M109.097394.

Kumar, Manish, M. Michael Gromiha, and Gajendra P.S. Raghava. 2011. "SVM Based Prediction of RNA-Binding Proteins Using Binding Residues and Evolutionary Information." *Journal of Molecular Recognition*. https://doi.org/10.1002/jmr.1061.

Kumar, Manish, Michael M Gromiha, and Gajendra P S Raghava. 2007. "Identification of DNA-Binding Proteins Using Support Vector Machines and Evolutionary Profiles." *BMC Bioinformatics* 8 (1): 463. https://doi.org/10.1186/1471-2105-8-463.

Kumar, V. 2019. "Toll-like Receptors in the Pathogenesis of Neuroinflammation." *Journal of Neuroimmunology* 332 (July): 16–30. https://doi.org/10.1016/j.jneuroim.2019.03.012.

Kumar, Vinod, Piyush Agrawal, Rajesh Kumar, Sherry Bhalla, Salman Sadullah Usmani, Grish C. Varshney, and Gajendra P.S. Raghava. 2018. "Prediction of Cell-Penetrating Potential of Modified Peptides Containing Natural and Chemically Modified Residues." *Frontiers in Microbiology*. https://doi.org/10.3389/fmicb.2018.00725.

Kumari, S Ramya, Ritesh Badwaik, Vijayaraghavan Sundararajan, and V K Jayaraman. 2012. "Defensinpred: Defensin and Defensin Types Prediction Server." *Protein and Peptide Letters* 19 (12): 1318–23. https://doi.org/10.2174/092986612803521594.

Lanki, Mira, Hanna Seppanen, Harri Mustonen, Jaana Hagstrom, and Caj Haglund. 2019. "Toll-like Receptor 1 Predicts Favorable Prognosis in Pancreatic Cancer." *PloS One* 14 (7): e0219245. https://doi.org/10.1371/journal.pone.0219245.

Lata, Sneh, Nitish K Mishra, and Gajendra P S Raghava. 2010. "AntiBP2: Improved Version of Antibacterial Peptide Prediction." *BMC Bioinformatics* 11 Suppl 1 (January): S19. https://doi.org/10.1186/1471-2105-11-S1-S19.

Lata, Sneh, and Gajendra P S Raghava. 2008a. "PRRDB: A Comprehensive Database of Pattern-Recognition Receptors and Their Ligands." *BMC Genomics* 9 (1): 180. https://doi.org/10.1186/1471-2164-9-180.

———. 2008b. "PRRDB: A Comprehensive Database of Pattern-Recognition Receptors and

Their Ligands." *BMC Genomics* 9 (1): 180. https://doi.org/10.1186/1471-2164-9-180.

Lata, Sneh, B K Sharma, and G P S Raghava. 2007. "Analysis and Prediction of Antibacterial Peptides." *BMC Bioinformatics* 8 (July): 263. https://doi.org/10.1186/1471-2105-8-263.

Lathwal, A, C Arora, and G P S Raghava. 2019. "Prediction of Risk Scores for Colorectal Cancer Patients from the Concentration of Proteins Involved in Mitochondrial Apoptotic Pathway." *PLoS One* 14 (9): e0217527. https://doi.org/10.1371/journal.pone.0217527.

Lathwal, Anjali, Rajesh Kumar, Dilraj kaur, and Gajendra P S Raghava. 2021. "In Silico Model for Predicting IL-2 Inducing Peptides in Human." *BioRxiv*, January, 2021.06.20.449146. https://doi.org/10.1101/2021.06.20.449146.

Laurie, Scott A, and Glenwood D Goss. 2013. "Role of Epidermal Growth Factor Receptor Inhibitors in Epidermal Growth Factor Receptor Wild-Type Non-Small-Cell Lung Cancer." *Journal of Clinical Oncology : Official Journal of the American Society of Clinical Oncology* 31 (8): 1061–69. https://doi.org/10.1200/JCO.2012.43.4522.

Lee, Chun Geun, Carla A Da Silva, Jae-Young Lee, Dominik Hartl, and Jack A Elias. 2008. "Chitin Regulation of Immune Responses: An Old Molecule with New Roles." *Current Opinion in Immunology* 20 (6): 684–89. https://doi.org/10.1016/j.coi.2008.10.002.

Lees, Emily A, Jessica L Forbester, Sally Forrest, Leanne Kane, David Goulding, and Gordon Dougan. 2019. "Using Human Induced Pluripotent Stem Cell-Derived Intestinal Organoids to Study and Modify Epithelial Cell Protection Against Salmonella and Other Pathogens." *Journal of Visualized Experiments : JoVE*, no. 147 (May). https://doi.org/10.3791/59478.

Lehrer, R I, A K Lichtenstein, and T Ganz. 1993. "Defensins: Antimicrobial and Cytotoxic Peptides of Mammalian Cells." *Annual Review of Immunology* 11: 105–28. https://doi.org/10.1146/annurev.iy.11.040193.000541.

Lehrer, Robert I, Charles L Bevins, and Tomas Ganz. 2005. "Defensins and Other Antimicrobial Peptides and Proteins." Edited by Jiri Mestecky, Michael E Lamm, Jerry R McGhee, John Bienenstock, Lloyd Mayer, and Warren Strober. *Mucosal Immunology*, 95–110. https://doi.org/10.1016/B978-012491543-5/50010-3.

Leppanen, Joni, Olli Helminen, Heikki Huhta, Joonas H Kauppila, Joel Isohookana, Kirsi-Maria Haapasaari, Petri Lehenkari, Juha Saarnio, and Tuomo J Karttunen. 2017. "High Toll-like Receptor (TLR) 9 Expression Is Associated with Better Prognosis in Surgically Treated Pancreatic Cancer Patients." *Virchows Archiv : An International Journal of*

Pathology 470 (4): 401–10. https://doi.org/10.1007/s00428-017-2087-1.

Li, Baihui, Bailu Zhang, Xuezhou Wang, Ziqing Zeng, Ziqi Huang, Lin Zhang, Feng Wei, Xiubao Ren, and Lili Yang. 2020. "Expression Signature, Prognosis Value, and Immune Characteristics of Siglec-15 Identified by Pan-Cancer Analysis." *Oncoimmunology* 9 (1): 1807291. https://doi.org/10.1080/2162402X.2020.1807291.

Li, P, H Ren, Y Zhang, and Z Zhou. 2018. "Fifteen-Gene Expression Based Model Predicts the Survival of Clear Cell Renal Cell Carcinoma." *Medicine (Baltimore)* 97 (33): e11839. https://doi.org/10.1097/MD.0000000000011839.

Lian, Yao, Ze Chi Huang, Meng Ge, and Xian Ming Pan. 2015. "An Improved Method for Predicting Linear B-Cell Epitope Using Deep Maxout Networks." *Biomedical and Environmental Sciences : BES* 28 (6): 460–63. https://doi.org/10.3967/bes2015.065.

Liao, Xia, Yang Bu, Zihan Xu, Fengan Jia, Fan Chang, Junrong Liang, Qingan Jia, and Yi Lv. 2020. "WISP1 Predicts Clinical Prognosis and Is Associated With Tumor Purity, Immunocyte Infiltration, and Macrophage M2 Polarization in Pan-Cancer." *Frontiers in Genetics* 11: 502. https://doi.org/10.3389/fgene.2020.00502.

Lin, Yi-Tzu, Amanda Verma, and Conrad P Hodgkinson. 2012. "Toll-like Receptors and Human Disease: Lessons from Single Nucleotide Polymorphisms." *Current Genomics* 13 (8): 633–45. https://doi.org/10.2174/138920212803759712.

Locatelli, M C, G Facendola, G Pizzocaro, L Piva, C Pegoraro, E B Pallavicini, A Signaroldi, et al. 1999. "Subcutaneous Administration of Interleukin-2 and Interferon-Alpha 2b in Advanced Renal Cell Carcinoma: Long-Term Results." *Cancer Detection and Prevention* 23 (2): 172–76. https://doi.org/10.1046/j.1525-1500.1999.09915.x.

Loo, Yueh-Ming, and Michael Gale. 2011. "Immune Signaling by RIG-I-like Receptors." *Immunity* 34 (5): 680–92. https://doi.org/10.1016/j.immuni.2011.05.003.

Manjang, Kalifa, Shailesh Tripathi, Olli Yli-Harja, Matthias Dehmer, Galina Glazko, and Frank Emmert-Streib. 2021. "Prognostic Gene Expression Signatures of Breast Cancer Are Lacking a Sensible Biological Meaning." *Scientific Reports* 11 (1): 156. https://doi.org/10.1038/s41598-020-79375-y.

Marelli, Giulia, Antonio Sica, Luca Vannucci, and Paola Allavena. 2017. "Inflammation as Target in Cancer Therapy." *Current Opinion in Pharmacology* 35 (August): 57–65. https://doi.org/10.1016/j.coph.2017.05.007.

Mariathasan, Sanjeev, David S Weiss, Kim Newton, Jacqueline McBride, Karen O'Rourke,

Meron Roose-Girma, Wyne P Lee, Yvette Weinrauch, Denise M Monack, and Vishva M Dixit. 2006. “Cryopyrin Activates the Inflammasome in Response to Toxins and ATP.” *Nature* 440 (7081): 228–32. https://doi.org/10.1038/nature04515.

Martinon, Fabio, Kimberly Burns, and Jürg Tschopp. 2002. “The Inflammasome: A Molecular Platform Triggering Activation of Inflammatory Caspases and Processing of ProIL-Beta.” *Molecular Cell* 10 (2): 417–26. https://doi.org/10.1016/s1097-2765(02)00599-3.

Maryam, Lubna, Salman Sadullah Usmani, and Gajendra P S Raghava. 2021. “Computational Resources in the Management of Antibiotic Resistance: Speeding up Drug Discovery.” *Drug Discovery Today*, April. https://doi.org/10.1016/j.drudis.2021.04.016.

Mathur, Deepika, Sandeep Singh, Ayesha Mehta, Piyush Agrawal, and Gajendra P S Raghava. 2018. “In Silico Approaches for Predicting the Half-Life of Natural and Modified Peptides in Blood.” *PloS One* 13 (6): e0196829. https://doi.org/10.1371/journal.pone.0196829.

Matsumoto, Misako, Yohei Takeda, Megumi Tatematsu, and Tsukasa Seya. 2017. “Toll-Like Receptor 3 Signal in Dendritic Cells Benefits Cancer Immunotherapy.” *Frontiers in Immunology* 8: 1897. https://doi.org/10.3389/fimmu.2017.01897.

Mauldin, Ileana S, Nolan A Wages, Anne M Stowman, Ena Wang, Walter C Olson, Donna H Deacon, Kelly T Smith, et al. 2016. “Topical Treatment of Melanoma Metastases with Imiquimod, plus Administration of a Cancer Vaccine, Promotes Immune Signatures in the Metastases.” *Cancer Immunology, Immunotherapy : CII* 65 (10): 1201–12. https://doi.org/10.1007/s00262-016-1880-z.

Mearini, Ettore, Giulia Poli, Giovanni Cochetti, Andrea Boni, Maria Giulia Egidi, and Stefano Brancorsini. 2017. “Expression of Urinary MiRNAs Targeting NLRs Inflammasomes in Bladder Cancer.” *OncoTargets and Therapy* 10: 2665–73. https://doi.org/10.2147/OTT.S132680.

Meher, Prabina Kumar, Tanmaya Kumar Sahu, Varsha Saini, and Atmakuri Ramakrishna Rao. 2017. “Predicting Antimicrobial Peptides with Improved Accuracy by Incorporating the Compositional, Physico-Chemical and Structural Features into Chou’s General PseAAC.” *Scientific Reports* 7 (February): 42362. https://doi.org/10.1038/srep42362.

Mehta, Divya, Priya Anand, Vineet Kumar, Anshika Joshi, Deepika Mathur, Sandeep Singh, Abhishek Tuknait, et al. 2014. “ParaPep: A Web Resource for Experimentally Validated Antiparasitic Peptide Sequences and Their Structures.” *Database : The Journal of Biological Databases and Curation* 2014. https://doi.org/10.1093/database/bau051.

Miao, Edward A, Dat P Mao, Natalya Yudkovsky, Richard Bonneau, Cynthia G Lorang, Sarah E Warren, Irina A Leaf, and Alan Aderem. 2010. “Innate Immune Detection of the Type III Secretion Apparatus through the NLRC4 Inflammasome.” *Proceedings of the National Academy of Sciences of the United States of America* 107 (7): 3076–80. https://doi.org/10.1073/pnas.0913087107.

Milani, A, D Sangiolo, F Montemurro, M Aglietta, and G Valabrega. 2013. “Active Immunotherapy in HER2 Overexpressing Breast Cancer: Current Status and Future Perspectives.” *Annals of Oncology : Official Journal of the European Society for Medical Oncology* 24 (7): 1740–48. https://doi.org/10.1093/annonc/mdt133.

Mileo, Anna Maria, and Stefania Miccadei. 2016. “Polyphenols as Modulator of Oxidative Stress in Cancer Disease: New Therapeutic Strategies.” *Oxidative Medicine and Cellular Longevity* 2016: 6475624. https://doi.org/10.1155/2016/6475624.

Mogensen, Trine H. 2009. “Pathogen Recognition and Inflammatory Signaling in Innate Immune Defenses.” *Clinical Microbiology Reviews* 22 (2): 240–73, Table of Contents. https://doi.org/10.1128/CMR.00046-08.

Monkkonen, Teresa, and Jayanta Debnath. 2018. “Inflammatory Signaling Cascades and Autophagy in Cancer.” *Autophagy* 14 (2): 190–98. https://doi.org/10.1080/15548627.2017.1345412.

Mookherjee, Neeloffer, Marilyn A Anderson, Henk P Haagsman, and Donald J Davidson. 2020. “Antimicrobial Host Defence Peptides: Functions and Clinical Potential.” *Nature Reviews. Drug Discovery* 19 (5): 311–32. https://doi.org/10.1038/s41573-019-0058-8.

Morice, Philippe, Alexandra Leary, Carien Creutzberg, Nadeem Abu-Rustum, and Emile Darai. 2016. “Endometrial Cancer.” *Lancet (London, England)* 387 (10023): 1094–1108. https://doi.org/10.1016/S0140-6736(15)00130-0.

Morimoto, Natsuki, Tomoya Kono, Masahiro Sakai, and Jun-Ichi Hikima. 2021. “Inflammasomes in Teleosts: Structures and Mechanisms That Induce Pyroptosis during Bacterial Infection.” *International Journal of Molecular Sciences* 22 (9). https://doi.org/10.3390/ijms22094389.

Mortaz, Esameil, Ian M Adcock, Payam Tabarsi, Ilad Alavi Darazam, Masoud Movassaghi, Johan Garssen, Hamidreza Jamaati, and Aliakbar Velayati. 2017. “Pattern Recognitions Receptors in Immunodeficiency Disorders.” *European Journal of Pharmacology* 808 (August): 49–56. https://doi.org/10.1016/j.ejphar.2017.01.014.

Morton, Jennifer P, Paul Timpson, Saadia A Karim, Rachel A Ridgway, Dimitris Athineos, Brendan Doyle, Nigel B Jamieson, et al. 2010. “Mutant P53 Drives Metastasis and Overcomes Growth Arrest/Senescence in Pancreatic Cancer.” *Proceedings of the National Academy of Sciences of the United States of America* 107 (1): 246–51. https://doi.org/10.1073/pnas.0908428107.

Mosley, Jonathan D, Mark D Benson, J Gustav Smith, Olle Melander, Debby Ngo, Christian M Shaffer, Jane F Ferguson, et al. 2018. “Probing the Virtual Proteome to Identify Novel Disease Biomarkers.” *Circulation* 138 (22): 2469–81. https://doi.org/10.1161/CIRCULATIONAHA.118.036063.

Mullen, Lisa M, Giselle Chamberlain, and Sandra Sacre. 2015. “Pattern Recognition Receptors as Potential Therapeutic Targets in Inflammatory Rheumatic Disease.” *Arthritis Research & Therapy* 17 (1): 122. https://doi.org/10.1186/s13075-015-0645-y.

Musa, Aliyu, Laleh Soltan Ghoraie, Shu-Dong Zhang, Galina Glazko, Olli Yli-Harja, Matthias Dehmer, Benjamin Haibe-Kains, and Frank Emmert-Streib. 2018. “A Review of Connectivity Map and Computational Approaches in Pharmacogenomics.” *Briefings in Bioinformatics* 19 (3): 506–23. https://doi.org/10.1093/bib/bbw112.

Nagpal, Gandharva, Kumardeep Chaudhary, Piyush Agrawal, and Gajendra P S Raghava. 2018. “Computer-Aided Prediction of Antigen Presenting Cell Modulators for Designing Peptide-Based Vaccine Adjuvants.” *Journal of Translational Medicine* 16 (1): 181. https://doi.org/10.1186/s12967-018-1560-1.

Nagpal, Gandharva, Sudheer Gupta, Kumardeep Chaudhary, Sandeep Kumar Dhanda, Satya Prakash, and Gajendra P S Raghava. 2015. “VaccineDA: Prediction, Design and Genome-Wide Screening of Oligodeoxynucleotide-Based Vaccine Adjuvants.” *Scientific Reports* 5 (1): 12478. https://doi.org/10.1038/srep12478.

Nagpal, Gandharva, Salman Sadullah Usmani, Sandeep Kumar Dhanda, Harpreet Kaur, Sandeep Singh, Meenu Sharma, and Gajendra P S Raghava. 2017. “Computer-Aided Designing of Immunosuppressive Peptides Based on IL-10 Inducing Potential.” *Scientific Reports* 7 (1): 42851. https://doi.org/10.1038/srep42851.

Net, J B van der, A C Janssens, J C Defesche, J J Kastelein, E J Sijbrands, and E W Steyerberg. 2009. “Usefulness of Genetic Polymorphisms and Conventional Risk Factors to Predict Coronary Heart Disease in Patients with Familial Hypercholesterolemia.” *Am J Cardiol* 103 (3): 375–80. https://doi.org/10.1016/j.amjcard.2008.09.093.

Ni, Jun, Yun Peng, Fu-Lan Yang, Xun Xi, Xing-Wei Huang, and Chun He. 2018. “Overexpression of CLEC3A Promotes Tumor Progression and Poor Prognosis in Breast Invasive Ductal Cancer.” *OncoTargets and Therapy* 11: 3303–12. https://doi.org/10.2147/OTT.S161311.

Novak, Natalija, Chun-Feng Yu, Thomas Bieber, and Jean-Pierre Allam. 2008. “Toll-like Receptor 7 Agonists and Skin.” *Drug News & Perspectives* 21 (3): 158–65.

Nuti, Ramya, Nerella S Goud, A Prasanth Saraswati, Ravi Alvala, and Mallika Alvala. 2017. “Antimicrobial Peptides: A Promising Therapeutic Strategy in Tackling Antimicrobial Resistance.” *Current Medicinal Chemistry* 24 (38): 4303–14. https://doi.org/10.2174/0929867324666170815102441.

O’ Donovan, Daniel H, Yumeng Mao, and Deanna A Mele. 2019. “The Next Generation of Pattern Recognition Receptor Agonists: Improving Response Rates in Cancer Immunotherapy.” *Current Medicinal Chemistry* 26 (June). https://doi.org/10.2174/0929867326666190620103105.

Ochwang’i, Dominic O, Charles N Kimwele, Jemimah A Oduma, Peter K Gathumbi, James M Mbaria, and Stephen G Kiama. 2014. “Medicinal Plants Used in Treatment and Management of Cancer in Kakamega County, Kenya.” *Journal of Ethnopharmacology* 151 (3): 1040–55. https://doi.org/10.1016/j.jep.2013.11.051.

Olive, Colleen. 2012. “Pattern Recognition Receptors: Sentinels in Innate Immunity and Targets of New Vaccine Adjuvants.” *Expert Review of Vaccines* 11 (2): 237–56. https://doi.org/10.1586/erv.11.189.

Othieno, Joseph O, Obadiah Njagi, and Allan Azegele. 2020. “Opportunities and Challenges in Antimicrobial Resistance Behavior Change Communication.” *One Health (Amsterdam, Netherlands)* 11 (December): 100171. https://doi.org/10.1016/j.onehlt.2020.100171.

Pahari, Susanta, Gurpreet Kaur, Mohammad Aqdas, Shikha Negi, Deepyan Chatterjee, Hilal Bashir, Sanpreet Singh, and Javed N Agrewala. 2017. “Bolstering Immunity through

Pattern Recognition Receptors: A Unique Approach to Control Tuberculosis." *Frontiers in Immunology* 8 (August): 906. https://doi.org/10.3389/fimmu.2017.00906.

Palm, Noah W, and Ruslan Medzhitov. 2009. "Pattern Recognition Receptors and Control of Adaptive Immunity." *Immunological Reviews* 227 (1): 221–33. https://doi.org/10.1111/j.1600-065X.2008.00731.x.

"Pan-Cancer Analysis of Whole Genomes." 2020. *Nature* 578 (7793): 82–93. https://doi.org/10.1038/s41586-020-1969-6.

Pande, Akshara, Sumeet Patiyal, Anjali Lathwal, Chakit Arora, Dilraj Kaur, Anjali Dhall, Gaurav Mishra, et al. 2019. "Computing Wide Range of Protein/Peptide Features from Their Sequence and Structure." *BioRxiv*. https://doi.org/10.1101/599126.

Panneerselvam, P, L P Singh, V Selvarajan, W J Chng, S B Ng, N S Tan, B Ho, J Chen, and J L Ding. 2013. "T-Cell Death Following Immune Activation Is Mediated by Mitochondria- Localized SARM." *Cell Death and Differentiation* 20 (3): 478–89. https://doi.org/10.1038/cdd.2012.144.

Park, Ga Bin, Yoon Hee Chung, and Daejin Kim. 2017. "Induction of Galectin-1 by TLR-Dependent PI3K Activation Enhances Epithelial-Mesenchymal Transition of Metastatic Ovarian Cancer Cells." *Oncology Reports* 37 (5): 3137–45. https://doi.org/10.3892/or.2017.5533.

Parkin, J, and B Cohen. 2001. "An Overview of the Immune System." *Lancet (London, England)* 357 (9270): 1777–89. https://doi.org/10.1016/S0140-6736(00)04904-7.

Paterson, Suzanna, Satwik Kar, Seng Kuong Ung, Zoe Gardener, Emma Bergstrom, Stephanie Ascough, Mohini Kalyan, et al. 2021. "Innate-like Gene Expression of Lung-Resident Memory CD8+ T-Cells During Experimental Human Influenza: A Clinical Study." *American Journal of Respiratory and Critical Care Medicine*, July. https://doi.org/10.1164/rccm.202103-0620OC.

Paveley, Ross A, Sarah A Aynsley, Joseph D Turner, Claire D Bourke, Stephen J Jenkins, Peter C Cook, Luisa Martinez-Pomares, and Adrian P Mountford. 2011. "The Mannose Receptor (CD206) Is an Important Pattern Recognition Receptor (PRR) in the Detection of the Infective Stage of the Helminth Schistosoma Mansoni and Modulates IFNγ Production." *International Journal for Parasitology* 41 (13–14): 1335–45. https://doi.org/10.1016/j.ijpara.2011.08.005.

Pohlmann, S, J Zhang, F Baribaud, Z Chen, G J Leslie, G Lin, A Granelli-Piperno, R W

Doms, C M Rice, and J A McKeating. 2003. “Hepatitis C Virus Glycoproteins Interact with DC-SIGN and DC-SIGNR.” *J Virol* 77 (7): 4070–80. https://doi.org/10.1128/jvi.77.7.4070-4080.2003.

Poli, Giulia, Giovanni Cochetti, Andrea Boni, Maria Giulia Egidi, Stefano Brancorsini, and Ettore Mearini. 2017. “Characterization of Inflammasome-Related Genes in Urine Sediments of Patients Receiving Intravesical BCG Therapy.” *Urologic Oncology* 35 (12): 674.e19-674.e24. https://doi.org/10.1016/j.urolonc.2017.08.004.

Postovit, Lynne, Christian Widmann, Peng Huang, and Spencer B Gibson. 2018. “Harnessing Oxidative Stress as an Innovative Target for Cancer Therapy.” *Oxidative Medicine and Cellular Longevity*. United States. https://doi.org/10.1155/2018/6135739.

Pouwels, S D, I H Heijink, N H T ten Hacken, P Vandenabeele, D V Krysko, M C Nawijn, and A J M van Oosterhout. 2014. “DAMPs Activating Innate and Adaptive Immune Responses in COPD.” *Mucosal Immunology* 7 (2): 215–26. https://doi.org/10.1038/mi.2013.77.

Prado, Samira Bernardino Ramos do, Victor Costa Castro-Alves, Gabrielle Fernandez Ferreira, and João Paulo Fabi. 2019. “Ingestion of Non-Digestible Carbohydrates From Plant-Source Foods and Decreased Risk of Colorectal Cancer: A Review on the Biological Effects and the Mechanisms of Action.” *Frontiers in Nutrition* 6 (May): 72. https://doi.org/10.3389/fnut.2019.00072.

Prasad, Suhanya V, Krzysztof Fiedoruk, Tamara Daniluk, Ewelina Piktel, and Robert Bucki. 2019. “Expression and Function of Host Defense Peptides at Inflammation Sites.” *International Journal of Molecular Sciences* 21 (1). https://doi.org/10.3390/ijms21010104.

Prasad, Suhanya V, Krzysztof Fiedoruk, Tamara Daniluk, Ewelina Piktel, Robert Bucki, Hyun Koo, Raymond N Allan, et al. 2020. “UniProt: The Universal Protein Knowledgebase.” Edited by Qing-Xiang Amy Sang. *Nucleic Acids Research* 11 (1): 148. https://doi.org/10.3389/fmicb.2017.02409.

Qin, S, L P Dong, B Bai, and H C Xue. 2019. “[Influence of Toll-like Receptor 7 on CD8(+) T Lymphocytes in Patients with Breast Cancer].” *Zhonghua Yi Xue Za Zhi* 99 (20): 1562–66. https://doi.org/10.3760/cma.j.issn.0376-2491.2019.20.009.

Qin, X Q, N Tao, A Dergay, P Moy, S Fawell, A Davis, J M Wilson, and J Barsoum. 1998. “Interferon-Beta Gene Therapy Inhibits Tumor Formation and Causes Regression of

Established Tumors in Immune-Deficient Mice." *Proceedings of the National Academy of Sciences of the United States of America* 95 (24): 14411–16. https://doi.org/10.1073/pnas.95.24.14411.

Qureshi, Abid, Himani Tandon, and Manoj Kumar. 2015. "AVP-IC50 Pred: Multiple Machine Learning Techniques-Based Prediction of Peptide Antiviral Activity in Terms of Half Maximal Inhibitory Concentration (IC50)." *Biopolymers* 104 (6): 753–63. https://doi.org/10.1002/bip.22703.

Quyun, Chen, Zhiyun Ye, Sheng-Cai Lin, and Biaoyang Lin. 2010. "Recent Patents and Advances in Genomic Biomarker Discovery for Colorectal Cancers." *Recent Patents on DNA & Gene Sequences* 4 (2): 86–93. https://doi.org/10.2174/187221510793205764.

Raj, Periathamby Antony, and Andrew R Dentino. 2002. "Current Status of Defensins and Their Role in Innate and Adaptive Immunity." *FEMS Microbiology Letters* 206 (1): 9–18. https://doi.org/10.1111/j.1574-6968.2002.tb10979.x.

Rasmussen, Michael, Emilio Fenoy, Mikkel Harndahl, Anne Bregnballe Kristensen, Ida Kallehauge Nielsen, Morten Nielsen, and Søren Buus. 2016. "Pan-Specific Prediction of Peptide–MHC Class I Complex Stability, a Correlate of T Cell Immunogenicity." *The Journal of Immunology* 197 (4): 1517–24. https://doi.org/10.4049/jimmunol.1600582.

Ren, Chong, Runsang Pan, Lisong Hou, Huaping Wu, Junkang Sun, Wenguang Zhang, Xiaobin Tian, and Houping Chen. 2020. "Suppression of CLEC3A Inhibits Osteosarcoma Cell Proliferation and Promotes Their Chemosensitivity through the AKT1/MTOR/HIF1alpha Signaling Pathway." *Molecular Medicine Reports* 21 (4): 1739–48. https://doi.org/10.3892/mmr.2020.10986.

Rohatiner, A Z S, W M Gregory, B Peterson, E Borden, P Solal-Celigny, A Hagenbeek, R I Fisher, et al. 2005. "Meta-Analysis to Evaluate the Role of Interferon in Follicular Lymphoma." *Journal of Clinical Oncology : Official Journal of the American Society of Clinical Oncology* 23 (10): 2215–23. https://doi.org/10.1200/JCO.2005.06.146.

Ronkainen, Hanna, Pasi Hirvikoski, Saila Kauppila, Katri S Vuopala, Timo K Paavonen, Katri S Selander, and Markku H Vaarala. 2011. "Absent Toll-like Receptor-9 Expression Predicts Poor Prognosis in Renal Cell Carcinoma." *Journal of Experimental & Clinical Cancer Research : CR* 30 (September): 84. https://doi.org/10.1186/1756-9966-30-84.

Royse, Kathryn E, Liang Chen, David H Berger, Michael M Ittmann, Hashem B El-Serag,

Courtney J Balentine, David Y Graham, et al. 2017. “Expression of Pattern Recognition Receptor Genes and Mortality in Patients with Colorectal Adenocarcinoma.” *International Journal of Molecular Epidemiology and Genetics* 8 (2): 8–18.

Saha, Sudipto, Manoj Bhasin, and Gajendra PS Raghava. 2005. “Bcipep: A Database of B-Cell Epitopes.” *BMC Genomics* 6 (1): 79. https://doi.org/10.1186/1471-2164-6-79.

Saha, Sudipto, and G. P. S. Raghava. 2004. “BcePred: Prediction of Continuous B-Cell Epitopes in Antigenic Sequences Using Physico-Chemical Properties.” In , 197–204. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-30220-9_16.

Saha, Sudipto, and G. P.S. Raghava. 2006. “AlgPred: Prediction of Allergenic Proteins and Mapping of IgE Epitopes.” *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkl343.

Sayers, Samantha, Guerlain Ulysse, Zuoshuang Xiang, and Yongqun He. 2012. “Vaxjo: A Web-Based Vaccine Adjuvant Database and Its Application for Analysis of Vaccine Adjuvants and Their Uses in Vaccine Development.” *Journal of Biomedicine & Biotechnology* 2012: 831486. https://doi.org/10.1155/2012/831486.

“Scikit.” n.d.

Seebah, Shalin, Anita Suresh, Shaowei Zhuo, Yong How Choong, Hazel Chua, Danny Chuon, Roger Beuerman, and Chandra Verma. 2007. “Defensins Knowledgebase: A Manually Curated Database and Information Source Focused on the Defensins Family of Antimicrobial Peptides.” *Nucleic Acids Research* 35 (Database issue): D265-8. https://doi.org/10.1093/nar/gkl866.

Sela-Culang, Inbal, Yanay Ofran, and Bjoern Peters. 2015. “Antibody Specific Epitope Prediction-Emergence of a New Paradigm.” *Current Opinion in Virology* 11 (April): 98–102. https://doi.org/10.1016/j.coviro.2015.03.012.

Semple, Fiona, and Julia R Dorin. 2012. “Beta-Defensins: Multifunctional Modulators of Infection, Inflammation and More?” *Journal of Innate Immunity* 4 (4): 337–48. https://doi.org/10.1159/000336619.

Sepehri, Zahra, Zohre Kiani, Farhad Kohan, Seyed Moayed Alavian, and Saeid Ghavami. 2017. “Toll like Receptor 4 and Hepatocellular Carcinoma; A Systematic Review.” *Life Sciences* 179 (June): 80–87. https://doi.org/10.1016/j.lfs.2017.04.025.

Setiawan, Veronica Wendy, Hannah P Yang, Malcolm C Pike, Susan E McCann, Herbert Yu, Yong-Bing Xiang, Alicja Wolk, et al. 2013. “Type I and II Endometrial Cancers: Have They Different Risk Factors?” *Journal of Clinical Oncology : Official Journal of the*

American Society of Clinical Oncology 31 (20): 2607–18. https://doi.org/10.1200/JCO.2012.48.2596.

Seung, Kwonjune J, Salmaan Keshavjee, and Michael L Rich. 2015. "Multidrug-Resistant Tuberculosis and Extensively Drug-Resistant Tuberculosis." *Cold Spring Harbor Perspectives in Medicine* 5 (9): a017863. https://doi.org/10.1101/cshperspect.a017863.

Sever, R, and J S Brugge. 2015. "Signal Transduction in Cancer." *Cold Spring Harb Perspect Med* 5 (4). https://doi.org/10.1101/cshperspect.a006098.

Sharma, Neelam, Sumeet Patiyal, Anjali Dhall, Akshara Pande, Chakit Arora, and Gajendra P S Raghava. 2020. "AlgPred 2.0: An Improved Method for Predicting Allergenic Proteins and Mapping of IgE Epitopes." *Briefings in Bioinformatics*, November. https://doi.org/10.1093/bib/bbaa294.

Shekarian, T, S Valsesia-Wittmann, J Brody, M C Michallet, S Depil, C Caux, and A Marabelle. 2017. "Pattern Recognition Receptors: Immune Targets to Enhance Cancer Immunotherapy." *Annals of Oncology : Official Journal of the European Society for Medical Oncology* 28 (8): 1756–66. https://doi.org/10.1093/annonc/mdx179.

Shen, Yujie, Shikun Dong, Jinhui Liu, Liqing Zhang, Jiacheng Zhang, Han Zhou, and Weida Dong. 2020. "Identification of Potential Biomarkers for Thyroid Cancer Using Bioinformatics Strategy: A Study Based on GEO Datasets." *BioMed Research International* 2020: 9710421. https://doi.org/10.1155/2020/9710421.

Sheyhidin, Ilyar, Gulnaz Nabi, Ayshamgul Hasim, Rui-Ping Zhang, Julaiti Ainiwaer, Hong Ma, and Hua Wang. 2011. "Overexpression of TLR3, TLR4, TLR7 and TLR9 in Esophageal Squamous Cell Carcinoma." *World Journal of Gastroenterology* 17 (32): 3745–51. https://doi.org/10.3748/wjg.v17.i32.3745.

Shirota, Hidekazu, Debra Tross, and Dennis M Klinman. 2015. "CpG Oligonucleotides as Cancer Vaccine Adjuvants." *Vaccines* 3 (2): 390–407. https://doi.org/10.3390/vaccines3020390.

Siegel, Rebecca L, Kimberly D Miller, and Ahmedin Jemal. 2020. "Cancer Statistics, 2020." *CA: A Cancer Journal for Clinicians* 70 (1): 7–30. https://doi.org/10.3322/caac.21590.

Silva, Joas L da, Shashank Gupta, Kenneth N Olivier, and Adrian M Zelazny. 2020. "Antimicrobial Peptides against Drug Resistant Mycobacterium Abscessus." *Research in Microbiology* 171 (5–6): 211–14. https://doi.org/10.1016/j.resmic.2020.03.001.

Singh, H, and G P Raghava. 2001. "ProPred: Prediction of HLA-DR Binding Sites."

Bioinformatics (Oxford, England) 17 (12): 1236–37. http://www.ncbi.nlm.nih.gov/pubmed/11751237.

Singh, Harinder, Rahul Kumar, Sandeep Singh, Kumardeep Chaudhary, Ankur Gautam, and Gajendra P.S. Raghava. 2016. "Prediction of Anticancer Molecules Using Hybrid Model Developed on Molecules Screened against NCI-60 Cancer Cell Lines." *BMC Cancer*. https://doi.org/10.1186/s12885-016-2082-y.

Singh, Harinder, Sandeep Singh, Deepak Singla, Subhash M. Agarwal, and Gajendra P.S. Raghava. 2015. "QSAR Based Model for Discriminating EGFR Inhibitors and Non-Inhibitors Using Random Forest." *Biology Direct*. https://doi.org/10.1186/s13062-015-0046-9.

Singh, Harpreet, and G P S Raghava. 2003. "ProPred1: Prediction of Promiscuous MHC Class-I Binding Sites." *Bioinformatics (Oxford, England)* 19 (8): 1009–14. http://www.ncbi.nlm.nih.gov/pubmed/12761064.

Singh, P K, H P Jia, K Wiles, J Hesselberth, L Liu, B A Conway, E P Greenberg, et al. 1998. "Production of Beta-Defensins by Human Airway Epithelia." *Proceedings of the National Academy of Sciences of the United States of America* 95 (25): 14961–66. https://doi.org/10.1073/pnas.95.25.14961.

Slattery, Martha L, Jennifer S Herrick, Lila E Mullany, Wade S Samowitz, John R Sevens, Lori Sakoda, and Roger K Wolff. 2017. "The Co-Regulatory Networks of Tumor Suppressor Genes, Oncogenes, and MiRNAs in Colorectal Cancer." *Genes, Chromosomes & Cancer* 56 (11): 769–87. https://doi.org/10.1002/gcc.22481.

Sochocka, Marta, and Zofia Blach-Olszewska. 2005. "[Mechanisms of innate immunity]." *Postepy higieny i medycyny doswiadczalnej (Online)* 59: 250–58.

Solanki, Subhash Singh, Parul Singh, Poonam Kashyap, Manish Singh Sansi, and Syed Azmal Ali. 2021. "Promising Role of Defensins Peptides as Therapeutics to Combat against Viral Infection." *Microbial Pathogenesis* 155 (June): 104930. https://doi.org/10.1016/j.micpath.2021.104930.

Solinhac, B. 1985. "[Augmentation materials in surgery. Can ridge height be increased in edentulous patients?]." *Revue d'odonto-stomatologie* 14 (4): 247–50.

Song, Isabelle Jingyi, Yoon Mee Yang, Sayaka Inokuchi-Shimizu, Yoon Seok Roh, Ling Yang, and Ekihiro Seki. 2018. "The Contribution of Toll-like Receptor Signaling to the Development of Liver Fibrosis and Cancer in Hepatocyte-Specific TAK1-Deleted

Mice.” *International Journal of Cancer* 142 (1): 81–91. https://doi.org/10.1002/ijc.31029.

Song, Mengqiu, Ann M Bode, Zigang Dong, and Mee-Hyun Lee. 2019. “AKT as a Therapeutic Target for Cancer.” *Cancer Research* 79 (6): 1019–31. https://doi.org/10.1158/0008-5472.CAN-18-2738.

Sudheendra, U S, Vishnu Dhople, Aritreyee Datta, Rajiv K Kar, Charles E Shelburne, Anirban Bhunia, and Ayyalusamy Ramamoorthy. 2015. “Membrane Disruptive Antimicrobial Activities of Human Beta-Defensin-3 Analogs.” *European Journal of Medicinal Chemistry* 91 (February): 91–99. https://doi.org/10.1016/j.ejmech.2014.08.021.

Sun, Yunliang, Congshan Wu, Jianxia Ma, Yu Yang, Xiaohua Man, Hongyu Wu, and Shude Li. 2016. “Toll-like Receptor 4 Promotes Angiogenesis in Pancreatic Cancer via PI3K/AKT Signaling.” *Experimental Cell Research* 347 (2): 274–82. https://doi.org/10.1016/j.yexcr.2016.07.009.

Taghavi, Mehdi, Alireza Khosravi, Esmaeil Mortaz, Donya Nikaein, and Seyyed Shamsadin Athari. 2017. “Role of Pathogen-Associated Molecular Patterns (PAMPS) in Immune Responses to Fungal Infections.” *European Journal of Pharmacology* 808 (August): 8–13. https://doi.org/10.1016/j.ejphar.2016.11.013.

Tai, Kenneth P, Karishma Kamdar, Jason Yamaki, Valerie V Le, Dat Tran, Patti Tran, Michael E Selsted, Andre J Ouellette, and Annie Wong-Beringer. 2015. “Microbicidal Effects of Alpha- and Theta-Defensins against Antibiotic- Resistant Staphylococcus Aureus and Pseudomonas Aeruginosa.” *Innate Immunity* 21 (1): 17–29. https://doi.org/10.1177/1753425913514784.

Takeuchi, Osamu, and Shizuo Akira. 2010. “Pattern Recognition Receptors and Inflammation.” *Cell* 140 (6): 805–20. https://doi.org/10.1016/j.cell.2010.01.022.

Tandelilin, Eduardus. 2010. “Scholar (6).” *Portofolio Dan Investasi: Teori Dan Aplikasi, Kanisius*.

Tang, D, R Kang, C B Coyne, H J Zeh, and M T Lotze. 2012. “PAMPs and DAMPs: Signal 0s That Spur Autophagy and Immunity.” *Immunol Rev* 249 (1): 158–75. https://doi.org/10.1111/j.1600-065X.2012.01146.x.

Tang, Juan, Guoxin Lin, Wallace Y Langdon, Lijian Tao, and Jian Zhang. 2018. “Regulation of C-Type Lectin Receptor-Mediated Antifungal Immunity.” *Frontiers in Immunology* 9

(February): 123. https://doi.org/10.3389/fimmu.2018.00123.

Tartey, Sarang, and Osamu Takeuchi. 2017. “Pathogen Recognition and Toll-like Receptor Targeted Therapeutics in Innate Immune Cells.” *International Reviews of Immunology* 36 (2): 57–73. https://doi.org/10.1080/08830185.2016.1261318.

Thakur, Nishant, Abid Qureshi, and Manoj Kumar. 2012. “AVPpred: Collection and Prediction of Highly Effective Antiviral Peptides.” *Nucleic Acids Research* 40 (Web Server issue): W199-204. https://doi.org/10.1093/nar/gks450.

The UniProt Consortium. 2017. “UniProt: The Universal Protein Knowledgebase.” *Nucleic Acids Research* 45 (D1): D158–69. https://doi.org/10.1093/nar/gkw1099.

Ting, Darren Shu Jeng, Roger W Beuerman, Harminder S Dua, Rajamani Lakshminarayanan, and Imran Mohammed. 2020. “Strategies in Translating the Therapeutic Potentials of Host Defense Peptides.” *Frontiers in Immunology* 11: 983. https://doi.org/10.3389/fimmu.2020.00983.

Tuomela, Johanna, Jouko Sandholm, Peeter Karihtala, Joanna Ilvesaro, Katri S Vuopala, Joonas H Kauppila, Saila Kauppila, et al. 2012. “Low TLR9 Expression Defines an Aggressive Subtype of Triple-Negative Breast Cancer.” *Breast Cancer Research and Treatment* 135 (2): 481–93. https://doi.org/10.1007/s10549-012-2181-7.

Turvey, Stuart E, and David H Broide. 2010. “Innate Immunity.” *The Journal of Allergy and Clinical Immunology* 125 (2 Suppl 2): S24-32. https://doi.org/10.1016/j.jaci.2009.07.016.

Usmani, Salman Sadullah, Gursimran Bedi, Jesse S Samuel, Sandeep Singh, Sourav Kalra, Pawan Kumar, Anjuman Arora Ahuja, Meenu Sharma, Ankur Gautam, and Gajendra P S Raghava. 2017. “THPdb: Database of FDA-Approved Peptide and Protein Therapeutics.” Edited by Qing-Xiang Amy Sang. *PloS One* 12 (7): e0181748. https://doi.org/10.1371/journal.pone.0181748.

Usmani, Salman Sadullah, Rajesh Kumar, Sherry Bhalla, Vinod Kumar, and Gajendra P S Raghava. 2018. “In Silico Tools and Databases for Designing Peptide-Based Vaccine and Drugs.” *Advances in Protein Chemistry and Structural Biology* 112: 221–63. https://doi.org/10.1016/bs.apcsb.2018.01.006.

Usmani, Salman Sadullah, Rajesh Kumar, Vinod Kumar, Sandeep Singh, and Gajendra P S Raghava. 2018. “AntiTbPdb: A Knowledgebase of Anti-Tubercular Peptides.” *Database : The Journal of Biological Databases and Curation* 2018 (January).

https://doi.org/10.1093/database/bay025.

Vasou, Andri, Nazife Sultanoglu, Stephen Goodbourn, Richard E Randall, and Leondios G Kostrikis. 2017. “Targeting Pattern Recognition Receptors (PRR) for Vaccine Adjuvantation: From Synthetic PRR Agonists to the Potential of Defective Interfering Particles of Viruses.” *Viruses* 9 (7): 186. https://doi.org/10.3390/v9070186.

Verma, Ruchi, Grish C. Varshney, and G. P S Raghava. 2010. “Prediction of Mitochondrial Proteins of Malaria Parasite Using Split Amino Acid Composition and PSSM Profile.” *Amino Acids*. https://doi.org/10.1007/s00726-009-0381-1.

Vita, Randi, Swapnil Mahajan, James A Overton, Sandeep Kumar Dhanda, Sheridan Martini, Jason R Cantrell, Daniel K Wheeler, Alessandro Sette, and Bjoern Peters. 2019. “The Immune Epitope Database (IEDB): 2018 Update.” *Nucleic Acids Research* 47 (D1): D339–43. https://doi.org/10.1093/nar/gky1006.

Waghu, Faiza Hanif, Ram Shankar Barai, Pratima Gurung, and Susan Idicula-Thomas. 2016. “CAMPR3: A Database on Sequences, Structures and Signatures of Antimicrobial Peptides.” *Nucleic Acids Research* 44 (D1): D1094-7. https://doi.org/10.1093/nar/gkv1051.

Wang, Chao, Shouqiang Cao, Ying Yan, Qiao Ying, Tao Jiang, Ke Xu, and Anhua Wu. 2010. “TLR9 Expression in Glioma Tissues Correlated to Glioma Progression and the Prognosis of GBM Patients.” *BMC Cancer* 10 (August): 415. https://doi.org/10.1186/1471-2407-10-415.

Wang, Dingding, Hong Zhang, Xiaolian Fang, Dingfang Cao, and Honggang Liu. 2020. “Pan-Cancer Analysis Reveals the Role of Long Non-Coding RNA LINC01614 as a Highly Cancer-Dependent Oncogene and Biomarker.” *Oncology Letters* 20 (2): 1383–99. https://doi.org/10.3892/ol.2020.11648.

Wang, Lan, Jie Yin, Xuefei Wang, Miaomiao Shao, Fangfang Duan, Weicheng Wu, Peike Peng, et al. 2016. “C-Type Lectin-Like Receptor 2 Suppresses AKT Signaling and Invasive Activities of Gastric Cancer Cells by Blocking Expression of Phosphoinositide 3-Kinase Subunits.” *Gastroenterology* 150 (5): 1183-1195.e16. https://doi.org/10.1053/j.gastro.2016.01.034.

Wang, Y, F Ren, P Chen, S Liu, Z Song, and X Ma. 2018. “Identification of a Six-Gene Signature with Prognostic Value for Patients with Endometrial Carcinoma.” *Cancer Med* 7 (11): 5632–42. https://doi.org/10.1002/cam4.1806.

Wang, Zi-Hao, Yun-Zheng Zhang, Yu-Shan Wang, and Xiao-Xin Ma. 2019. "Identification of Novel Cell Glycolysis Related Gene Signature Predicting Survival in Patients with Endometrial Cancer." *Cancer Cell International* 19: 296. https://doi.org/10.1186/s12935-019-1001-0.

Wei, L, Z Jin, S Yang, Y Xu, Y Zhu, and Y Ji. 2018. "TCGA-Assembler 2: Software Pipeline for Retrieval and Processing of TCGA/CPTAC Data." *Bioinformatics* 34 (9): 1615–17. https://doi.org/10.1093/bioinformatics/btx812.

Weigel, Brenda J, Sarah Cooley, Todd DeFor, Daniel J Weisdorf, Angela Panoskaltsis-Mortari, Wei Chen, Bruce R Blazar, and Jeffrey S Miller. 2012. "Prolonged Subcutaneous Administration of 852A, a Novel Systemic Toll-like Receptor 7 Agonist, to Activate Innate Immune Responses in Patients with Advanced Hematologic Malignancies." *American Journal of Hematology* 87 (10): 953–56. https://doi.org/10.1002/ajh.23280.

Won, Haejung, Dayson Moreira, Chan Gao, Priyanka Duttagupta, Xingli Zhao, Edwin Manuel, Don Diamond, et al. 2017. "TLR9 Expression and Secretion of LIF by Prostate Cancer Cells Stimulates Accumulation and Activity of Polymorphonuclear MDSCs." *Journal of Leukocyte Biology* 102 (2): 423–36. https://doi.org/10.1189/jlb.3MA1016-451RR.

Wu, Hao-Xiang, Zi-Xian Wang, Qi Zhao, Dong-Liang Chen, Ming-Ming He, Lu-Ping Yang, Ying-Nan Wang, et al. 2019. "Tumor Mutational and Indel Burden: A Systematic Pan-Cancer Evaluation as Prognostic Biomarkers." *Annals of Translational Medicine* 7 (22): 640. https://doi.org/10.21037/atm.2019.10.116.

Xiang, Zuoshuang, Thomas Todd, Kim P Ku, Bethany L Kovacic, Charles B Larson, Fang Chen, Andrew P Hodges, et al. 2008. "VIOLIN: Vaccine Investigation and Online Information Network." *Nucleic Acids Research* 36 (Database issue): D923-8. https://doi.org/10.1093/nar/gkm1039.

Yang, C, L Wang, Z Jia, Q Yi, Q Xu, W Wang, C Gong, C Liu, and L Song. 2017. "Two Short Peptidoglycan Recognition Proteins from Crassostrea Gigas with Similar Structure Exhibited Different PAMP Binding Activity." *Dev Comp Immunol* 70: 9–18. https://doi.org/10.1016/j.dci.2016.12.009.

Yang, De, Gonzalo de la Rosa, Poonam Tewary, and Joost J Oppenheim. 2009. "Alarmins Link Neutrophils and Dendritic Cells." *Trends in Immunology* 30 (11): 531–37.

https://doi.org/10.1016/j.it.2009.07.004.

Yang, P, H An, X Liu, M Wen, Y Zheng, Y Rui, and X Cao. 2010. “The Cytosolic Nucleic Acid Sensor LRRFIP1 Mediates the Production of Type I Interferon via a Beta-Catenin-Dependent Pathway.” *Nat Immunol* 11 (6): 487–94. https://doi.org/10.1038/ni.1876.

Yang, Q, X Li, H Chen, Y Cao, Q Xiao, Y He, J Wei, and J Zhou. 2017. “IRF7 Regulates the Development of Granulocytic Myeloid-Derived Suppressor Cells through S100A9 Transrepression in Cancer.” *Oncogene* 36 (21): 2969–80. https://doi.org/10.1038/onc.2016.448.

Yin, Yuanqin, Fei Li, Songlin Li, Jingjing Cai, Jing Shi, Youhong Jiang, L Sun, J J Dai, W F Hu, and J Wang. 2016. “Expression of Toll-like Receptors in Hepatic Cirrhosis and Hepatocellular Carcinoma.” *Genetics and Molecular Research : GMR* 15 (2): 469–79. https://doi.org/10.4238/gmr.15027419.

Yoshida, Sumito, Hiroaki Shime, Yohei Takeda, Jin-Min Nam, Ken Takashima, Misako Matsumoto, Hiroki Shirato, Masanori Kasahara, and Tsukasa Seya. 2018. “Toll-like Receptor 3 Signal Augments Radiation-Induced Tumor Growth Retardation in a Murine Model.” *Cancer Science* 109 (4): 956–65. https://doi.org/10.1111/cas.13543.

Yuan, Qingchen, Na Sun, Jiayu Zheng, Yingxuan Wang, Xiaole Yan, Wuqian Mai, Yuhua Liao, and Xiao Chen. 2019. “Prognostic and Immunological Role of FUN14 Domain Containing 1 in Pan-Cancer: Friend or Foe?” *Frontiers in Oncology* 9: 1502. https://doi.org/10.3389/fonc.2019.01502.

Yue, Yanqiu, Tao Zhou, Yanjing Gao, Zongli Zhang, Li Li, Lin Liu, Wenna Shi, Lihui Su, and Baoquan Cheng. 2017. “High Mobility Group Box 1/Toll-like Receptor 4/Myeloid Differentiation Factor 88 Signaling Promotes Progression of Gastric Cancer.” *Tumour Biology : The Journal of the International Society for Oncodevelopmental Biology and Medicine* 39 (3): 1010428317694312. https://doi.org/10.1177/1010428317694312.

Zambetti, Lia Paola, Federica Laudisi, Ginevra Licandro, Paola Ricciardi-Castagnoli, and Alessandra Mortellaro. 2012. “The Rhapsody of NLRPs: Master Players of Inflammation...and a Lot More.” *Immunologic Research* 53 (1–3): 78–90. https://doi.org/10.1007/s12026-012-8272-z.

Zhang, Xiaoli, Shiyong Liu, and Anna Tramontano. 2017. “RBPPred: Predicting RNA-Binding Proteins from Sequence Using SVM.” *Bioinformatics*. https://doi.org/10.1093/bioinformatics/btw730.

Zhao, Ning, Maozu Guo, Kuanquan Wang, Chunlong Zhang, and Xiaoyan Liu. 2020. "Identification of Pan-Cancer Prognostic Biomarkers Through Integration of Multi-Omics Data." *Frontiers in Bioengineering and Biotechnology* 8: 268. https://doi.org/10.3389/fbioe.2020.00268.

Zhou, Hongyu, Chufan Zhang, Haoran Li, Lihua Chen, and Xi Cheng. 2020. "A Novel Risk Score System of Immune Genes Associated with Prognosis in Endometrial Cancer." *Cancer Cell International* 20: 240. https://doi.org/10.1186/s12935-020-01317-5.

Zhou, Min, Weinan Jiang, Jiayang Xie, Wenjing Zhang, Zhemin Ji, Jingcheng Zou, Zihao Cong, et al. 2020. "Development and Challenges of Antimicrobial Peptides for Therapeutic Applications." *ChemMedChem* 16 (1): 309–15. https://doi.org/10.3390/antibiotics9010024.

Zhou, Min, Weinan Jiang, Jiayang Xie, Wenjing Zhang, Zhemin Ji, Jingcheng Zou, Zihao Cong, Ximian Xiao, Jiawei Gu, and Runhui Liu. 2021. "Peptide-Mimicking Poly(2-Oxazoline)s Displaying Potent Antimicrobial Properties." *ChemMedChem* 16 (2): 309–15. https://doi.org/10.1002/cmdc.202000530.

Zhu, Gengzheng, Yao Xu, Xiaohong Cen, Kutty Selva Nandakumar, Shuwen Liu, and Kui Cheng. 2018. "Targeting Pattern-Recognition Receptors to Discover New Small Molecule Immune Modulators." *European Journal of Medicinal Chemistry* 144 (January): 82–92. https://doi.org/10.1016/j.ejmech.2017.12.026.

Zuo, Yong-Chun, and Qian-Zhong Li. 2009. "Using Reduced Amino Acid Composition to Predict Defensin Family and Subfamily: Integrating Similarity Measure and Structural Alphabet." *Peptides* 30 (10): 1788–93. https://doi.org/10.1016/j.peptides.2009.06.032.

Zuo, Yongchun, Yu Chang, Shenghui Huang, Lei Zheng, Lei Yang, and Guifang Cao. 2019. "IDEF-PseRAAC: Identifying the Defensin Peptide by Using Reduced Amino Acid Composition Descriptor." *Evolutionary Bioinformatics Online* 15: 1176934319867088. https://doi.org/10.1177/1176934319867088.

Zuo, Yongchun, Yang Lv, Zhuying Wei, Lei Yang, Guangpeng Li, and Guoliang Fan. 2015. "IDPF-PseRAAAC: A Web-Server for Identifying the Defensin Peptide Family and Subfamily Using Pseudo Reduced Amino Acid Alphabet Composition." *PloS One* 10 (12): e0145541. https://doi.org/10.1371/journal.pone.0145541.