

**MACHINE LEARNING-DRIVEN INTEGRATIVE
TRANSCRIPTOMIC ANALYSIS TO IDENTIFY
SHARED BIOMARKERS AND MOLECULAR
PATHWAYS BETWEEN HUNTINGTON'S DISEASE
AND COMMON VARIABLE
IMMUNODEFICIENCY**

**Thesis Submitted
in Partial Fulfillment of the Requirements for the
Degree of**

**MASTERS OF TECHNOLOGY
in
BIOINFORMATICS**

**by
PRERNA JOSHI
(Roll No. 24/BIO/07)**

**Under the Supervision of
Dr. ASMITA DAS**



Department of Biotechnology

**DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Shahbad Daulatpur, Main Bawana Road, Delhi-110042, India**

May, 2026

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my supervisor, **Dr. Asmita Das**, for her invaluable guidance, constant encouragement, insightful suggestions, and unwavering support throughout the course of this thesis work. Her supervision, scientific rigor, and motivating presence have played a crucial role in shaping this research and in strengthening my academic growth.

I am sincerely thankful to **Mr. Yuvraj**, **Ms. Simran**, and **Ms. Monika**, Ph.D. scholars in the laboratory, for their generous assistance, helpful discussions, and continuous support during different stages of this work. Their guidance and willingness to help have been truly encouraging.

I would also like to extend my heartfelt thanks to my friends **Yatharth Mishra**, **Arpit Raturi**, **Pallavit Lata**, and **Kanan** for their support, encouragement, and companionship throughout this journey. Their presence made this academic phase more meaningful and manageable.

I owe my deepest gratitude to my parents, **Mr. V.D. Joshi** and **Mrs. Sarita Joshi**, whose unconditional love, sacrifices, and blessings have always been my greatest strength. Their unwavering faith in me and constant encouragement have been the foundation of my achievements. This accomplishment would not have been possible without their endless support, care, and guidance. I remain forever grateful to my parents for their invaluable sacrifices and for always believing in me.

With deep affection, I thank my sister-in-law **Mrs. Shivani Joshi** and my sister **Ms. Akanksha Joshi** for their love, motivation, and constant encouragement. Their support has always been a source of strength for me.

I am grateful to my brothers, **Mr. Shubham Joshi** and **Mr. Tarun Joshi**, for their affection, support, and constant presence in my life. Their motivation and support have always inspired me to move forward with confidence.

Finally, I would like to thank everyone who directly or indirectly contributed to the successful completion of this thesis. I am grateful to all those who supported and encouraged me throughout this academic journey.

PRERNA JOSHI

24/BIO/07



DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)
Shahbad Daulatpur, Main Bawana Road, Delhi-42

CANDIDATE'S DECLARATION

I Prerna Joshi hereby certify that the work which is being presented in the thesis entitled **“Machine Learning-Driven Integrative Transcriptomic Analysis to Identify Shared Biomarkers and Molecular Pathways Between Huntington’s Disease and Common Variable Immunodeficiency”** in partial fulfillment of the requirements for the award of the Degree of Master of Technology, submitted in the Department of Biotechnology, Delhi Technological University is an authentic record of my own work carried out during the period from January 2026 to May 2026 under the supervision of Dr. Asmita Das.

The matter presented in the thesis has not been submitted by me for the award of any other degree of this or any other Institute.

Candidate’s Signature

This is to certify that the student has incorporated all the corrections suggested by the examiner in the thesis and the statement made by the candidate is correct to the best of our knowledge.

Signature of Supervisor



DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)
Shahbad Daulatpur, Main Bawana Road, Delhi-42

CERTIFICATE BY THE SUPERVISOR

Certified that **Prerna Joshi (24/BIO/07)** has carried out their search work presented in this thesis entitled “**Machine Learning-Driven Integrative Transcriptomic Analysis to Identify Shared Biomarkers and Molecular Pathways Between Huntington’s Disease and Common Variable Immunodeficiency**” for the award of **Master of Technology** from Department of Biotechnology, Delhi Technological University, Delhi, under my supervision. The thesis embodies results of original work, and studies are carried out by the student himself and the contents of the thesis do not form the basis for the award of any other degree to the candidate or to anybody else from this or any other University/Institution.

Dr. Asmita Das

Supervisor

Department of Biotechnology
Delhi Technological University

Prof. Yasha Hasija

Head of Department

Department of Biotechnology
Delhi Technological University

Date:

Place:

Machine Learning-Driven Integrative Transcriptomic Analysis to Identify Shared Biomarkers and Molecular Pathways between Huntington's Disease and Common Variable Immunodeficiency

Purna Joshi

ABSTRACT

It is becoming more widely acknowledged that neurodegenerative illnesses are intricate systemic conditions impacted by immunological dysregulation, extracellular matrix remodelling, and modified molecular signalling beyond the central nervous system. This study investigated the relationship between immunological dysfunction and neurodegeneration using a combined Machine Learning (ML) and literature-based paradigm. To find common biomarkers associated with disease, preprocessing, normalisation, supervised machine learning, and Explainable Artificial Intelligence (xAI) were used to examine public transcriptome datasets from Huntington's disease (HD) and Common Variable Immunodeficiency (CVID). Support Vector Machine (SVM) performed better than the other classifiers, with an accuracy of 95.7% on the test set. While functional enrichment analysis identified pathways such as antigen processing and presentation, clathrin-mediated endocytosis, hyaluronan uptake and degradation, and axon guidance, SHAP-based feature ranking identified genes like CSF1, ITGB5, TCP11L2, CDC5L, L3MBTL2, and BTG2 as important contributors to classification, indicating a strong confluence between immune regulation and neuronal biology. Additionally, network analysis revealed that CSF1 and BTG2 might serve as significant regulatory hubs in miRNA–mRNA interactions. Simultaneously, hyaluronan was investigated as a biologically significant extracellular matrix component that affects the stability of the neuronal milieu, immune cell trafficking, and inflammation. According to the reviewed data, hyaluronan metabolism is extremely context-dependent as fragmented hyaluronan can stimulate inflammatory signalling and immunological activation, high-molecular-weight hyaluronan often maintains tissue homeostasis. Due to its impact on glial activation, perineuronal net remodelling, demyelination, and delayed regeneration, dysregulated hyaluronan turnover has been linked to neurodegenerative disease. All of these results point to a single paradigm that explains how immune dysregulation and neurodegeneration interact through shared transcriptional signatures and hyaluronan-associated pathways. Potential biomarkers and molecular pathways that could be helpful for upcoming diagnostic and immunologically guided treatment approaches were found in this investigation.

Keywords- common variable immunodeficiency, huntington's disease, gene expression analysis, machine learning, explainable AI, biomarker identification, enrichment analysis, network analysis, hyaluronan metabolism

TABLE OF CONTENTS

CONTENTS	PAGE NO.
ACKNOWLEDGEMENTS.....	ii
CANDIDATE’S DECLARATION.....	iii
CERTIFICATE	iv
ABSTRACT.....	v
TABLE OF CONTENTS.....	vi-vii
LIST OF TABLES.....	viii
LIST OF FIGURES.....	ix
LIST OF ABBREVIATIONS	x
CHAPTER 1	
INTRODUCTION	1-3
CHAPTER 2	
LITERATURE REVIEW.....	4-9
2.1 Molecular Basis of Huntington’s Disease and the Need for Biomarkers.....	4
2.2 Common Variable Immunodeficiency and Transcriptomic Heterogeneity.....	4-5
2.3 Integrative Transcriptomics and Machine Learning in Biomarker Discovery	5
2.4 Functional Enrichment and Regulatory Network Reconstruction.....	5
2.5 Hyaluronan Metabolism and Biological Functions	6
2.6 Hyaluronan-Mediated Immune Regulation	6-7
2.7 Hyaluronan in the Central Nervous System and Neurodegenerative Disease Pathology	7-8
2.8 Therapeutic Potential of Targeting Hyaluronan Pathways	8
2.9 Emerging Molecular Intersection and Study Rationale	9
CHAPTER 3	
METHODOLOGY	10-14
3.1 Study Design.....	10
3.2 Data Acquisition	10
3.3 Data Preprocessing	11
3.4 Machine Learning Implementation and Model Development	11

3.5 Model Evaluation	11
3.6 Explainable Artificial Intelligence Analysis.....	12
3.7 Functional Enrichment Analysis	12
3.8 miRNA–mRNA Network Analysis.....	13
3.9 Literature-Based Analysis of Hyaluronan Metabolism.....	14
3.10 Reproducibility and Data Interpretation	14
CHAPTER 4	
RESULTS	15-24
4.1 Dataset Composition and Sample Structure	15
4.2 Model Evaluation and Classification Performance.....	16
4.3 Biomarker Identification by SHAP Analysis.....	17
4.3.1 Mean Absolute SHAP Bar Plot with Generic Feature Labels.....	18
4.3.2 Mean Absolute SHAP Bar Plot with Probe/Transcript Labels	19
4.3.3 SHAP Beeswarm Plot.....	20
4.3.4 SHAP Force Plot.....	21
4.3.5 SHAP Waterfall Plot	22
4.4 Functional Enrichment Analysis	23
4.5 miRNA–mRNA Network Analysis.....	24
CHAPTER 5	
DISCUSSION	25-27
CHAPTER 5	
CONCLUSION AND FUTURE SCOPE.....	28-29
REFERENCES	30-32
LIST OF PUBLICATIONS.....	33-35
PLAGIARISM REPORT	36-40

LIST OF TABLES

TABLE NO.	TITLE OF TABLES	PAGE NO.
Table 3.1	The computational workflow and analytical tools utilized for HD-CVID transcriptomic analysis	14
Table 4.1	Composition of GEO Datasets used in the study	15
Table 4.2	Representative SHAP-ranked genes and their mean SHAP values showing the most important classification features in the HD-CVID transcriptomic analysis.	17

LIST OF FIGURES

Figure No.	TITLE OF FIGURES	PAGE No.
Fig. 3.1	GUI of NCBI GEO Database	10
Fig. 3.2	GUI of Enrichr Tool Used for Gene Set Enrichment Analysis	12
Fig. 3.3	GUI of NetworkAnalyst tool	13
Fig. 3.4	NetworkAnalyst interface with gene/protein list input for constructing the network and downstream analysis	13
Fig. 4.1	Optimized Support Vector Machine (SVM) classification performance on the independent test set displayed as confusion matrix in the transcriptomic analysis of the HD–CVID data set	16
Fig. 4.2	SHAP feature-importance bar plot, using generic feature identifiers, of top 20 transcriptomic features ranked by mean absolute SHAP value	18
Fig. 4.3	SHAP feature-importance bar plot showing the top 20 transcript/probe-level features ranked by mean absolute SHAP value	19
Fig. 4.4	SHAP summary beeswarm plot indicating the distribution and direction of contribution of the top transcriptomic features by all samples	20
Fig. 4.5	SHAP force plot displaying individual transcriptomic features contribution to the prediction of a single sample from the trained SVM model	21
Fig. 4.6	SHAP waterfall plot illustrating the contribution of individual transcriptomic features to a single model prediction	22
Fig. 4.7	Enrichment bar graph for Reactome pathways of the genes prioritized by SHAP from the transcriptomic analysis of HD–CVID	23
Fig. 4.8	The miRNA–mRNA interaction network created using the shortlisted biomarkers, based on the SHAP-based feature ranking	24

LIST OF ABBREVIATIONS

AD	Alzheimer's Disease
AI	Artificial Intelligence
BTG2	B-cell Translocation Gene 2
CAG	Cytosine-Adenine-Guanine
CD44	Cluster of Differentiation 44
CDC5L	Cell Division Cycle 5-Like
CEMIP	Cell Migration Inducing Hyaluronidase 1
CNS	Central Nervous System
CSF	Cerebrospinal Fluid Hyaluronan
CSF1	Colony Stimulating Factor 1
CVID	Common Variable Immunodeficiency
ECM	Extracellular Matrix
GAG	Glycosaminoglycan
GEO	Gene Expression Omnibus
HD	Huntington's Disease
HA	Hyaluronan
HMW-HA	Hyaluronan Synthase
HLA	High-Molecular-Weight Hyaluronan
HTT	Huntingtin
IFN	Interferon
IL	Interleukin
ITGB5	Integrin Subunit Beta 5
KIAA1199	Hyaluronan-Binding Protein / CEMIP-Associated Protein
LMW-HA	Low-Molecular-Weight Hyaluronan
LYVE-1	Lymphatic Vessel Endothelial Hyaluronan Receptor 1
ML	Machine Learning
mRNA	Messenger RNA
miRNA	Micro RNA
MS	Multiple Sclerosis
OPC	Oligodendrocyte Progenitor Cell
PI	Primary Immunodeficiency
PNN	Perineuronal Net
RHAMM	Receptor for Hyaluronan-Mediated Motility
RNA	Ribonucleic Acid
SHAP	Shapley Additive Explanations
SVM	Support Vector Machine
TLR	Toll-Like Receptor
TCP11L2	Tubulin Coiled-Coil Protein 11 Like 2
xAI	Explainable Artificial Intelligence

CHAPTER 1

INTRODUCTION

It is generally recognised that neurodegenerative illnesses are complex systemic disorders in which extracellular matrix remodelling, immunological dysregulation, chronic inflammation, and neuronal damage function in tandem rather than as separate pathogenic events. In diseases like Huntington's disease (HD), where blood-derived molecular changes can reflect ongoing disease biology before advanced clinical deterioration becomes evident, this broader perspective has shifted attention from a purely neuron-centric model toward a more integrated neuroimmune framework [1], [3], [4]. Simultaneously, primary immunodeficiency ailments, such as common variable immunodeficiency (CVID) provide a useful model for researching persistent immune imbalance as they are marked by distinct transcriptomic signatures, inflammatory phenotypes, cytokine dysregulation, and impaired antibody production [2], [5], [6]. The idea that immunological dysregulation and neurodegeneration may be connected by common molecular processes is supported by these two disease circumstances taken together.

Huntington's disease is an autosomal dominant, progressive neurological disorder characterized by an increase in the number of CAG repeats in the HTT gene. Although the disease is typically characterised by cognitive decline, mental symptoms, and motor dysfunction, its molecular pathology goes much beyond mutant huntingtin accumulation [3], [4]. Transcriptional dysregulation, mitochondrial malfunction, inflammatory activation, and synaptic damage are linked to HD, emphasising that the disease progresses via several biological layers [1], [3], [4]. The search for reliable blood-based biomarkers continues to be a top priority for early diagnosis, prognosis, and therapeutic monitoring because there is currently no proven disease-modifying therapy [1], [4], [10].

Recurrent infections, reduced immunoglobulin production, and widespread immunological imbalance are the hallmarks of CVID, the most prevalent symptomatic primary immunodeficiency [5], [6]. Patients often have granulomatous inflammation, autoimmune illness, and aberrant cytokine signalling in addition to humoral deficiencies, indicating that CVID represents systemic immunological failure rather than a single isolated problem [5], [6]. Blood expression profiling is useful in identifying biologically significant disease subgroups since transcriptomic investigations have shown that CVID has unique inflammatory patterns, including interferon-associated signatures [2]. CVID offers a pertinent framework for understanding immune-linked molecular mechanisms that are shared with HD as immunological disorders are becoming more widely acknowledged as contributors to neurodegenerative pathology [2], [5], [6].

The expanding accessibility of public transcriptome resources has made it feasible to use computational methods to study disease biology at the genome-wide level. Gene expression datasets allow for comparison across seemingly disparate illnesses and can

identify disease-associated pathways that are not evident from clinical observation alone [1], [2], [10]. Because HD and CVID are reflected in peripheral blood samples, which can capture systemic molecular changes that may be related to interference of immune dysregulation and neurodegeneration, transcriptome-wide analysis is especially well suited for studying these disorders [1], [2]. In addition, these datasets offer a useful basis for route mapping, regulatory network reconstruction, and biomarker development.

With gene expression data, due to its high dimensionality, noise and nonlinearity, machine learning has proven a valuable tool for the quest of transcriptomic biomarkers [7]. Logistic Regression, Support Vector Machine (SVM), Random Forest and other supervised classifiers are more effective than many traditional methods at finding discriminative patterns in such datasets [7], [12]. One of the important aspects of SVM is its effectiveness in regions with high feature density, where the feature space has a much larger number of features than samples, such as in omics applications [7], [12].

However, biological interpretation requires more than just strong predictive performance. In order to connect computational results to disease causes, genes that have been selected in biomedical research must also be explicable [8], [9]. Hence, Explainable Artificial Intelligence (xAI) has become more and more useful in omics studies to convert predictions into interpretable feature rankings [8]. In particular, Shapley Additive Explanations (SHAP) can pinpoint the genes that have the most impact on categorisation and offer a rational framework for allocating feature-level contributions to model predictions [9]. This supports mechanistic interpretation rather than merely statistical classification in transcriptome analysis by enabling biomarker prioritisation while maintaining biological transparency [8], [9].

Once candidate genes have been pinpointed, functional enrichment and network analysis are needed to place the candidates in the context of relevant biological pathways. Enrichment tools such as Enrichr enable statistically rigorous interpretation of gene sets in terms of pathway [10]. A systems-level understanding on gene regulation and interaction is offered by network-based platforms like NetworkAnalyst, which make it possible to identify hub genes, miRNA–mRNA linkages, and more comprehensive regulatory modules [11], [12]. Since the same genes may be involved in cytokine signalling, intracellular trafficking, and neurodevelopmental processes, these methods are especially helpful for studying illnesses involving both immune and neural pathways [10]–[12].

Hyaluronan metabolism provides another crucial mechanistic insight on the relationship between immune dysregulation and neurodegeneration concurrently with the finding of transcriptome biomarkers. One of the main glycosaminoglycans of the extracellular matrix, hyaluronan is now understood to be a dynamic signalling molecule that regulates inflammation, tissue repair, immune cell migration, and cellular communication in addition to being a structural polymer [13]–[15]. Its biological effects are highly dependent on the size and context of the polymer: fragmented low-molecular-weight hyaluronan can act as a pro-inflammatory signal, while high-molecular-weight hyaluronan is typically linked to homeostasis and anti-inflammatory activity [13], [14].

Therefore, the equilibrium between hyaluronan synthesis and degradation is important from a biological perspective. Hyaluronidases and oxidative processes break down

hyaluronan, which is produced at the plasma membrane by hyaluronan synthases [13]. Fragments produced during degradation can trigger the release of cytokines, the recruitment of immune cells, and the activation of inflammatory genes by activating receptors such as CD44, RHAMM, and Toll-like receptors [14], [15]. Due to these characteristics, hyaluronan is a potential cause of chronic inflammatory conditions and a significant facilitator of tissue injury responses. Dysregulated hyaluronan turnover may link peripheral immunological problems to central nervous system dysfunction because immune activation and matrix remodelling are intimately related [14], [15].

In the CNS, hyaluronan is an important component of the specialised extracellular matrix that surrounds neurones and glial cells [17], [19]. It aids in the development and upkeep of perineuronal nets, which control extracellular diffusion, synaptic stability, and neuronal plasticity [17], [18], [19]. Several neurodegenerative diseases, such as multiple sclerosis and Alzheimer's disease, have been linked to altered hyaluronan metabolism, which may hinder remyelination, encourage glial activation, and interfere with tissue regeneration [16], [18], [21]. As hyaluronan has a well-established neurological significance, its dysregulation is increasingly seen as an active part of disease pathogenesis rather than a passive structural consequence [16]-[21].

This connection is further reinforced by hyaluronan's inflammatory relevance. Hyaluronan fragments have the ability to function as endogenous danger-associated molecular patterns, which can alter the local tissue environment and trigger innate immune responses [14], [15]. Astrocytes and microglia in the brain respond to hyaluronan-mediated cues that may be involved in persistent neuroinflammation and extracellular matrix remodelling [15], [16]. These alterations provide a biologically plausible way by which peripheral immune disruption can influence central pathology, making them especially intriguing in disorders where immunological failure coexists with brain susceptibility [14], [16], [21]. Hyaluronan metabolism is therefore an important potential pathway in the broader investigation of immune-associated neurodegeneration.

This study employed both transcriptomic and literature-based methods to investigate the molecular connection between immune dysregulation and neurodegeneration using an integrated framework. utilising machine learning, explainable AI, enrichment analysis, and network analysis, shared biomarker signatures, regulatory interactions, and enriched biological pathways were found utilising publicly accessible HD and CVID expression datasets [1], [2], [7]-[12]. Simultaneously, hyaluronan metabolism was assessed as a key extracellular matrix mechanism that could influence neuronal susceptibility and inflammatory signalling [13]-[21]. By integrating these perspectives, the research hopes to shed light on how immune disorders, altered gene expression, and extracellular matrix remodelling interact to affect the biology of neurodegenerative diseases and to aid in the creation of new diagnostic and immunologically guided treatment approaches.

CHAPTER 2

LITERATURE REVIEW

2.1 Molecular Basis of Huntington's Disease and the Need for Biomarkers

The genetic cause of Huntington's disease (HD) as a hereditary neurodegenerative disorder has been known for many years, and the presence of a CAG repeat expansion in the HTT gene is the root cause of the disorder, but the pathological underpinning of HD is now clearly demonstrated in the literature to be much more complex than just toxicity of the HTT mutant protein huntingtin. A large number of studies have shown that in HD, transcriptional dysregulation, inflammatory activation, mitochondrial impairment, synaptic changes, and changes in cellular stress responses occur [1], [3], [4]. These molecular abnormalities can be also observed in peripheral blood, thus blood-based transcriptomic profiling is a useful strategy for biomarker discovery [1], [4], [5].

Strong support for the peripheral biomarkers in HD has been provided by studies demonstrating that blood expression profiles are able to differentiate disease from control and that they can reflect disease-associated biology prior to the onset of severe clinical deterioration [1]. These are particularly important biomarkers in the context of a progressive disorder where early diagnosis, prognosis and monitoring of therapy are still important clinical challenges [3], [4]. As a result, the literature has shifted away from the quest for single-gene markers and towards a more comprehensive transcriptomic approach that can provide insight into the complexity of HD biology [4], [5]. Such a change justifies transcriptome-based comparative studies in the context of HD and justifies the integration of a computational analysis in the present work.

2.2 Common Variable Immunodeficiency and Transcriptomic Heterogeneity

The most prevalent symptomatic primary immunodeficiency (PI) is common variable immunodeficiency (CVID), which manifests as broad immunological imbalance, recurrent infections, and a lack or failure of antibody generation [2], [5], [6]. However, the literature is growing to highlight that in CVID the failure of humoral immunity is not the only feature. Common features of patients are auto-immune reactions, granulomatous disease, chronic inflammation and inappropriately regulated cytokine signaling, suggesting a systemic disease of immune regulation, rather than a single defect [5], [6]. This clinical diversity makes CVID an important disease model for the study of inflammatory and immune mediated transcriptomic signatures [2], [5], [6].

Established transcriptomic studies have demonstrated that the expression pattern in peripheral blood in CVID can stratify immunological subgroups and identify inflammatory endotypes [2]. Transcriptome profiling can reflect the biologically significant immunological diversity among patient groups, such as the interferon signature in inflammatory CVID [2]. These results are significant for neurodegenerative research since immunological dysfunction has now been

demonstrated to contribute to the progression of a number of illnesses that were previously believed to be neuron-centric [1], [2], [5].

2.3 Integrative Transcriptomics and Machine Learning in Biomarker Discovery

The interpretation of high-dimensional transcriptomic data is challenging with traditional statistical approaches, especially when disease-related patterns are nonlinear or distributed across many weak contributors [7]. In this context, machine learning has emerged as an ever more valuable asset in biomarker discovery for its ability to classify samples, deal with large feature spaces, and extract discriminative molecular patterns without oversimplifying the biological phenomenon [7], [8]. The algorithm we use, Support Vector Machine (SVM), is well established in the context of genomic and transcriptomic studies, and has been shown to be effective in cases with a high number of variables compared to the number of samples, a typical situation in expression data [7], [12].

However, in biomedical research, predicting accuracy isn't enough. The features selected for a model also have to be explainable in order to be linked to the mechanism instead of being considered a black-box predictor [8], [9]. Explainable Artificial Intelligence (xAI), especially SHAP-based analysis, offers a methodical approach to measure how each variable influences model choices [9]. SHAP is useful for transcriptomic research as it is able to generate a ranked list of candidate biomarkers while maintaining interpretability while maintaining a classifier [8], [9]. This is particularly relevant in disease contexts like HD or CVID where the biological signal can exist in a small subset of genes within a much larger and noisy expression background.

2.4 Functional Enrichment and Regulatory Network Reconstruction

After shortlist of potential candidate genes have been identified, pathway enrichment analysis is critical to see whether the genes share coherent biological pathways. Many tools are available because they can perform the statistically robust interpretation of gene lists using curated pathway and functional databases like Enrichr [10]. Enrichment analysis is particularly valuable in the field of neuro-immune studies as apparently small biomarker sets are typically associated with larger, underlying pathways that involve immune signaling, vesicle trafficking, cytoskeletal regulation, and extracellular matrix remodeling [10].

The network-based analysis offers a further level of interpretation to demonstrate the interaction between genes, miRNAs and other regulatory molecules. In order to identify hub genes and regulatory modules not evident from enrichment results alone, platforms like NetworkAnalyst integrates expression information with interaction data [11], [12]. The latest release of NetworkAnalyst 3.0 added further such functions, offering a visual analytics platform for meta-analysis and gene expression profiling [28]. In this context, such methods are useful to determine if the shortlisted genes are under common regulatory programs that could be contributing to their ability to be strong discriminators in the machine learning model.

2.5 Hyaluronan Metabolism and Biological Functions

Hyaluronan (HA) is a large glycosaminoglycan of the extracellular matrix that is structurally unique because it is a linear, non-sulfated polysaccharide with repeating units of β -1,3-N-acetyl-D-glucosamine and β -1,4-D-glucuronic acid. HA is synthesized at the plasma membrane, unlike most other components of the extracellular matrix, by hyaluronan synthase enzymes, and extruded into the extracellular space during HA polymerization [13], [15]. In this mode of synthesis, HA can be deposited quickly into tissues, providing contribution to matrix hydration, biomechanical integrity and cellular organization.

In addition to its structural function, hyaluronan is now known to be a biologically active signaling molecule. It binds to cell surface receptors like CD44, LYVE-1, RHAMM, which affect cell proliferation, migration, differentiation and inflammatory activity [14], [15]. HA is highly dependent on its function with regard to size of the molecule. Higher molecular weight HA is typically linked to tissue homeostasis and anti-inflammatory effect while the degradation by enzymatic activity or oxidative stress leads to formation of lower molecular weight HA fragments which are known to have pro-inflammatory and immunostimulatory effects [13], [14]. This duality, which depends on size, is crucial to the understanding of the role of HA in physiological maintenance and in the development of disease.

HA is involved in many processes in the nervous system, such as tissue patterning, axonal guidance, neurite extension and synaptic stabilization. The HA during development acts as a permissive extracellular matrix to allow for neuronal migration and network organization, and in the adult nervous system it plays a role in stabilizing structure and controlling plasticity [17], [19]. Thus, imbalance in HA metabolism may impact the neural microenvironment and impact processes vital to normal brain function.

2.6 Hyaluronan-Mediated Immune Regulation

Increasingly, hyaluronan is not simply considered as a matrix molecule but an important regulator of immune signaling. Tissue damage, long-term inflammation, and oxidative stress are responsible for the degradation of HA and the production of endogenous danger-associated molecular patterns (DAMPs). These fragments can interact with receptors such as Toll-like receptors, CD44, and RHAMM which activate downstream signaling pathways that control cytokine production, leukocyte recruitment, and inflammatory gene expression [14], [15]. In this manner, HA establishes extracellular matrix remodeling and immune activation.

Immunological activity of HA is very context dependent too. It has been noted that intact high-molecular-weight HA tends to inhibit the activation of leukocytes and to facilitate resolution of inflammation, while fragmented HA has the opposite effect and can trigger innate immune system responses and extend pro-inflammatory signals [14], [15]. This equilibrium is especially crucial in diseased conditions that result from chronic tissue stress and lead to chronic matrix turnover. In these environments, an unbalanced HA metabolism can contribute to a local immune environment that is more conducive to inflammation than healing.

The role of HA in the CNS is particularly relevant as matrix-derived signals actively regulate the activity of CNS glial cells. Microglia and astrocytes express HA receptors and are capable of both producing and degrading HA [17]. Chronic inflammation leads to matrix remodeling by astrocytes around the perivascular and perineuronal space and high levels of HA deposition can add to the immunologically altered microenvironment [17], [24]. For instance, in multiple sclerosis lesions, CD44 positive astrocytes and HA degrading enzymes can initiate a vicious cycle of abnormal matrix turnover and may hinder immune resolution [16], [24], [25]. Excess HA may also prevent the proliferation of regulatory T cells via CD44-mediated signalling, suggesting that abnormal HA can maintain autoimmunity and alter the balance of the adaptive immune system [14], [25].

In summary, all of these observations suggest that hyaluronan plays a crucial role in innate and adaptive immune responses. Its metabolism can connect with immunological processes in a dynamic way, which can lead to either a protective or a pathogenic immunological reaction, depending on the molecular environment [14], [20].

2.7 Hyaluronan in the Central Nervous System and Neurodegenerative Disease Pathology

HA is a very abundant component of the extracellular matrix in the central nervous system, which is highly specialized [17], [19]. The brain extracellular matrix contains glycosaminoglycans and proteoglycans instead of the large amounts of fibrillar collagen found in most other peripheral tissues and this composition is important for molecular diffusion, synaptic communication and cellular stability [17], [19]. Another special architecture in which HA plays a role is the perineuronal net, a condensed matrix that regulates synaptic plasticity and stabilizes neuronal circuits [17], [18], [19].

Many changes in HA metabolism have been associated with neurodegenerative pathology. In MS, the dysregulated HA turnover is linked to white-matter lesions, impaired remyelination and oligodendrocyte progenitor maturation inhibition [16]. High molecular weight HA is seen in abundance in chronic lesions, while more fragmented HA is seen in more active inflammatory areas and may maintain glial activation and prevent repair [16], [8]. Increased CSF HA concentration also has been associated with disease activity, which suggests that CSF HA may be a marker of disease burden and disease inflammatory state [12]. These results highlight the important role that abnormal HA metabolism plays in lesion persistence and failure to repair, even though it is not simply a consequence of tissue damage.

HA deposition has been seen in the extracellular matrix of the hippocampus in Alzheimer's disease and around the borders of amyloid plaques [4], [18]. Surrounding plaques, the reactive astrocytes can up-regulate CD44, which in turn leads to an increase in the local production of HA and a positive feedback loop of matrix accumulation [4], [13]. At the same time, fragments of HA and its degradation products can stimulate the activation of microglia and astrocytes, via Toll-like receptor pathways, amplifying neuroinflammatory responses [14], [26]. In contrast, in intact perineuronal nets, neurons are more resistant to the toxicity caused by amyloids and tau, suggesting that HA-rich matrix structures can afford some degree of

neuroprotection [18]. This equilibrium of HA accumulation, fragmentation and matrix integrity seems to play a role in vulnerability in neurodegenerative diseases.

Extracellular matrix remodeling is now known to be a common trait of neurodegeneration, in general [21]. Excessive HA fragmentation over prolonged periods may worsen inflammatory signalling and neuronal stress, while an imbalance in HA metabolism may result in glial scarring, impaired synaptic function and inadequate neuronal repair [15], [16], [18], [24], [25]. Hence, HA is not only a passive, structural component of disease microenvironment, but also a biologically active one.

2.8 Therapeutic Potential of Targeting Hyaluronan Pathways

Various therapeutic approaches have been proposed in the literature that are based on the hyaluronan axis. A strategy is to control the activity of HAS to prevent excessive HA production in damaged or inflamed neural tissue. Decreasing HA synthesis or modifying HA molecular size distribution might be beneficial to restore a more favorable repair environment, as abnormal HA deposition can lead to changes in the structure of the ECM and affect neuronal regeneration [13], [22].

A second pathway of therapy is to block HA degradation. Hyaluronidase-mediated cleavage creates low molecular weight fragments which act as DAMPs and stimulate inflammatory receptors, including TLR2 and TLR4 [13], [14]. Thus, preventing excessive fragmentation could result in the diminished neuroinflammatory signaling and the activation of astrocytes and microglia [3], [8]. For disorders in which the presence of fragmented HA is a cause of lesion persistence, controlling the degradation of HA may be as critical as controlling the synthesis of HA.

A third approach is to aim at the HA receptors. CD44 is involved in cell adhesion, leukocyte migration and activation of the immune system within the extracellular matrix [3], [14]. It is likely that modulation of this pathway can affect recruitment of inflammatory cells and immune responses in the CNS. Experimental research has demonstrated that disrupting CD44 signalling has the potential to slow the inflammatory activity of the cells in a number of disease contexts making it a therapeutic target [3], [26]. The possibility of selectively targeting receptor-level communication of HA will be explored, given that CD44 is also required for communications between glial cells and immune cells.

Besides pharmacological strategies, HA-based biomaterials have been a focus of interest in the field of neural tissue engineering and regenerative medicine. In addition to promoting neuronal growth, HA hydrogels and scaffolds can create a permissive microenvironment for repair and regulate inflammatory responses in a tissue compatible manner [15], [16]. The extracellular matrix of the brain is naturally rich in HA making it an intriguing candidate for use in central nervous system repair strategies [22], [23]. These studies demonstrate the importance of HA not just as a biomarker associated molecule but also as a platform for therapeutic design.

2.9 Emerging Molecular Intersection and Study Rationale

Evidence from the literature suggests that, together, HA plays a role at the crossroads of extracellular matrix biology, immune regulation, and neurodegeneration. It can be synthesized, degraded, and interacts with receptors to regulate signaling pathways in health and disease, and its size determines its biological effect [13]-[20]. Dysregulated HA metabolism has been linked to chronic inflammation, impaired repair, and altered neuronal resilience in neurodegenerative disorders, and in immune-mediated disorders, modulation of the behavior of leukocytes and cytokine signaling [14], [16], [18], [24], [25], [27].

Although this is substantial mechanistic evidence, there are few studies that have combined HA biology with the discovery of transcriptomic biomarkers in a comparative context of both neurodegenerative and immunodeficient disorders. A significant opportunity is therefore present for systems level analysis. To test this hypothesis, the present study employs transcriptomic and machine learning techniques to further investigate if there is a shared molecular signature between HA and CVID, and if so, whether this signature can help to elucidate the interface between immune dysregulation and neurodegeneration.

CHAPTER 3

METHODOLOGY

3.1 Study Design

In order to explore common molecular signatures between Huntington's disease (HD) and common variable immunodeficiency (CVID) and to place these signatures within the biology of immune dysregulation, neurodegeneration, and hyaluronan metabolism, this study was created as an integrative computational framework. The workflow combined explainable artificial intelligence, pathway enrichment, supervised machine learning, public transcriptome data analysis, and regulatory network reconstruction. Simultaneously, the function of hyaluronan as an extracellular matrix mediator of inflammatory and neurobiological processes was interpreted using a systematic literature-based synthesis. In order to transform statistical classification output into mechanistic understanding, the pipeline was designed to progress from raw expression data to interpretable biomarkers and then to biological context at the route and network levels [7]-[12], [28].

3.2 Data Acquisition

The retrieval of the publicly available human peripheral blood transcriptomic datasets was done from the NCBI Gene Expression Omnibus (GEO). GSE1751 and GSE51405 were chosen because they are relevant to HD and CVID, respectively [30], [31]. The GSE1751 dataset consisted of three classes-control, presymptomatic HD, and symptomatic HD – while GSE51405 consisted of control-HC, CVID-N, and CVID-Y classes. Systemic disease-associated molecular patterns were investigated, which were appropriate for both datasets, since they were both from peripheral blood, thus reflecting immune-neurodegenerative crosstalk. This choice of these datasets allowed for a cross-condition analysis of the transcriptomic signatures for two clinically distinct but biologically similar diseases.

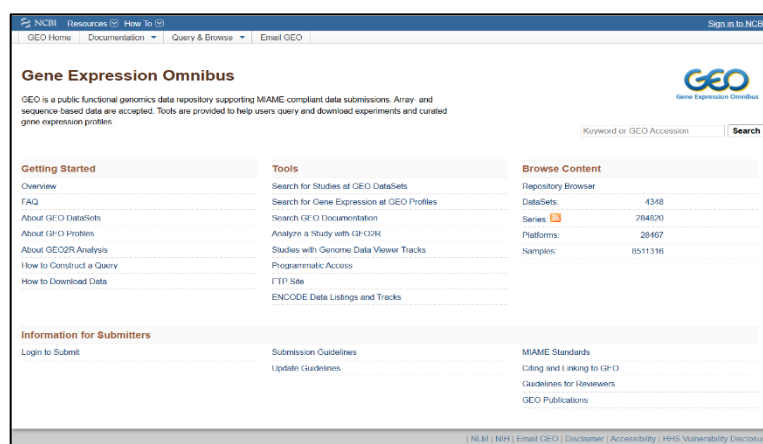


Fig. 3.1 GUI of NCBI GEO Database

3.3 Data Preprocessing

To make the raw gene expression matrices consistent, comparable and suitable for downstream machine learning analysis, they have been pre-processed. The matrices were transposed to make each row a sample and each column a feature defined by genes in the row. The data were then reviewed for missing, inconsistent or incomplete data and any attributes with a high percentage of missing data were deleted to ensure data integrity. Labels for class were assigned by combining the metadata of the related data set. The labels were then encoded numerically and converted to these categorical labels to be used by supervised learning algorithms. This was done to reduce scale-related bias of features and then z-score standardization was performed to make all genes contribute equally to the training of the model. This pre-processing pipeline was required for the technical variation not to overwhelm a biological variation in gene expression [7].

3.4 Machine Learning Implementation and Model Development

The datasets were split in an 80:20 ratio between training and testing sets after preprocessing. The results were obtained by using a fixed random state while partitioning to ensure the reproducibility. Several supervised machine learning algorithms were tested such as Logistic Regression, Decision Tree, Random Forest, K-Nearest Neighbours (KNN), Support Vector Machine (SVM), Gaussian Naïve Bayes and XGBoost. To add a new benchmark for comparison between tree-based and margin-based classifiers, XGBoost was added to the model ensemble [32]. SVM was chosen for further investigation due to its impressive ability to effectively process high-dimensional transcriptomic data and its high generalization power in the separation of classes that are biologically distinct. The final trained SVM model was then used for interpretability analysis and biomarker discovery. This method is similar to previous studies demonstrating the suitability of SVM classification for large feature spaces and a low sample size, such as an omics dataset [7], [12].

3.5 Model Evaluation

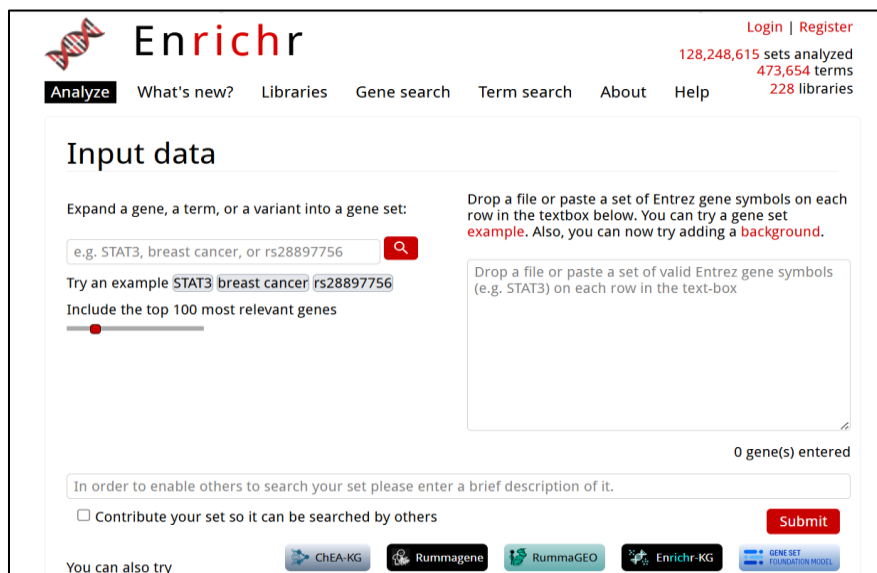
Standard classification measures such as accuracy, precision, recall and F1 score were used to evaluate model performance. These measures gave a multi-dimensional perspective of the performance of the classifiers, including both correct prediction and a balance between sensitivity and specificity. Further, a confusion matrix was created to represent the true positives, true negative, false positive, and false negative. This framework enabled the model to be assessed not only with respect to its prediction power, but also its power to provide robust separation of disease-associated transcriptomic signatures. An independent test set was used to reduce the possibility of overfitting and to provide a clearer indication of the generalization capabilities [19].

3.6 Explainable Artificial Intelligence Analysis

The Shapley Additive Explanations (SHAP) was used as an explainable artificial intelligence framework to enhance the interpretability of the trained SVM model [9]. SHAP gives a contribution value to each feature, which helps to explain the effects of the feature on the model prediction, giving both global and local interpretability. The overall importance of the genes in the classification was ranked using their mean absolute SHAP values. The contributions of individual genes were then summarized as summary plots and force plotted as force plots across the data set and for individual predictions. It was crucial that this explainability step is included because, in transcriptomic analysis, interpretation of the biological aspect is needed along with prediction performance. SHAP allowed for the transformation of a successful classifier into a biomarker ranking system that is biologically informative [8], [9].

3.7 Functional Enrichment Analysis

Functional enrichment analysis was performed on the genes ranked by SHAP with the help of Enrichr [10]. Overrepresented biological processes and signaling pathways were identified by analyzing a curated pathway database, Reactome and related databases with the selected gene-set. The step was required to see if the biomarkers were coalescing into immune, transport or neurobiological pathways or if they were just random statistical features. The first layer of biological interpretation was obtained by enrichment analysis, which resulted in the correlation of the shortlisted genes with the mechanisms: antigen processing and presentation, clathrin mediated endocytosis, hyaluronan related processes, axon guidance. Therefore, the gene list was converted to a pathway hypothesis of immune-neurodegenerative crosstalk [10].



The screenshot shows the Enrichr web application interface. At the top, there is a navigation bar with the Enrichr logo, a DNA helix icon, and links for 'Login | Register'. Below this, a statistics bar indicates '128,248,615 sets analyzed', '473,654 terms', and '228 libraries'. The main section is titled 'Input data' and contains two input areas. The left area is for expanding a gene, term, or variant into a gene set, with a search bar containing 'e.g. STAT3, breast cancer, or rs28897756' and a search button. Below this, there is a section for 'Try an example' with 'STAT3', 'breast cancer', and 'rs28897756' entered, and a slider for 'Include the top 100 most relevant genes'. The right area is for dropping a file or pasting a set of Entrez gene symbols, with a text box and a 'Drop a file or paste a set of valid Entrez gene symbols (e.g. STAT3) on each row in the text-box' instruction. Below the input areas, there is a checkbox for 'Contribute your set so it can be searched by others' and a 'Submit' button. At the bottom, there is a section 'You can also try' with links to ChEA-KG, Rummagene, RummageO, Enrichr-KG, and GENE SET FOUNDATION MODEL.

Fig. 3.2 GUI of Enrichr Tool Used for Gene Set Enrichment Analysis

3.8 miRNA–mRNA Network Analysis

The shortlisted genes were then analysed using NetworkAnalyst to gain insight into the post-transcriptional regulation of the genes [11], [12]. This platform allowed the miRNA–mRNA interaction networks to be reconstructed and the identification of key hubs that could control the behavior of biomarkers. NetworkAnalyst 3.0 was especially helpful for incorporating expression profiling with visual network exploration and meta-analysis [28]. The structure of the network was analyzed, and hub genes and connected regulatory nodes were identified that might affect the disease-relevant pathways.

This allowed for a system-level perspective in the transcriptomic analysis, revealing how transcriptional biomarkers can be influenced by miRNA-mediated regulation [11], [12], [28].

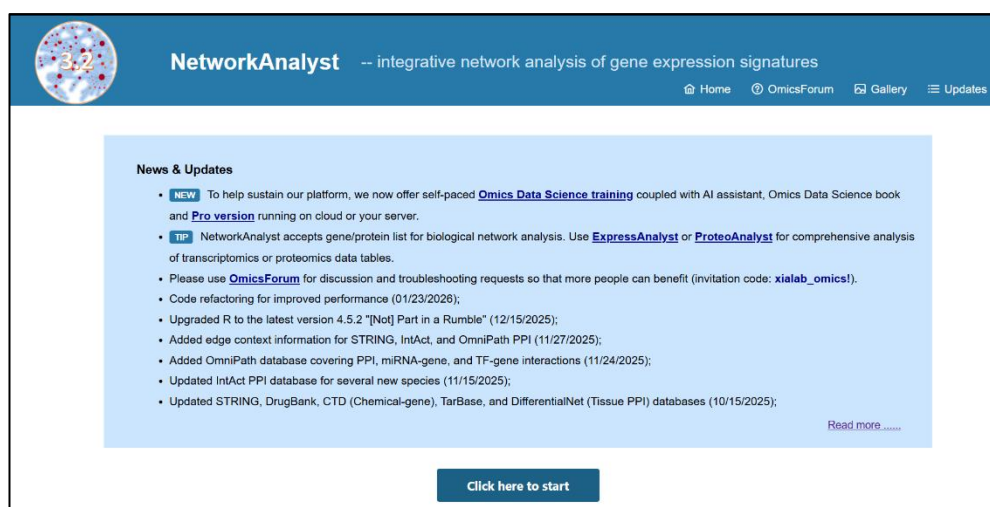


Fig. 3.3 GUI of NetworkAnalyst tool.

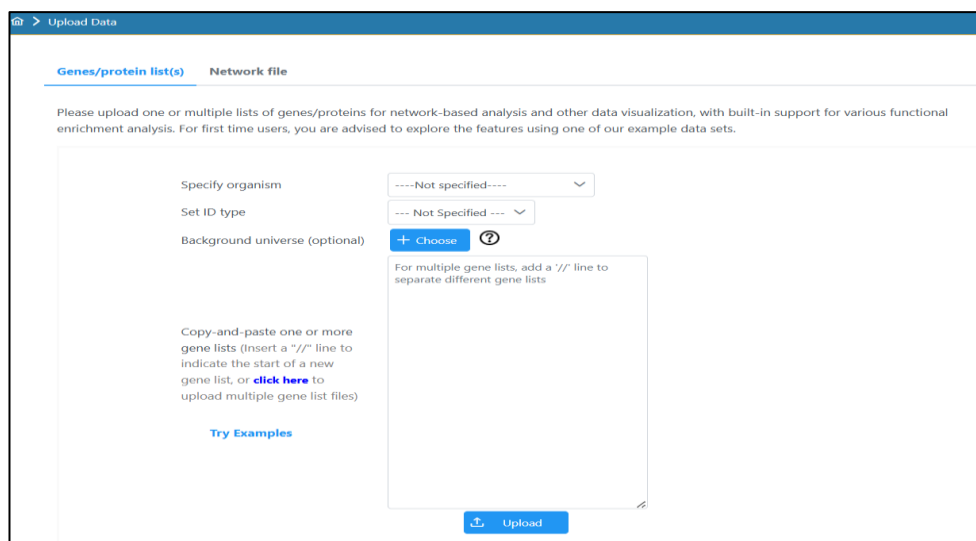


Fig. 3.4 NetworkAnalyst interface with gene/protein list input for constructing the network and downstream analysis

3.9 Literature-Based Analysis of Hyaluronan Metabolism

A literature-based approach was employed to distill evidence for the hyaluronan-based aspect, which focused on the metabolism of hyaluronan, extracellular matrix remodeling, immune signaling, and neurodegenerative disease pathology. Hyaluronan biosynthesis and degradation, receptor-mediated signaling, biology of the perineuronal net, neuroinflammatory signaling and disease-associated changes in extracellular matrix were investigated in the context of peer-reviewed studies. The review was theme-based, so that the literature collected could be put together in a limited number of steps and become part of a unified mechanistic framework that connects immune dysregulation with neuronal vulnerability. To supplement the transcriptomic analysis and to aid interpretation of hyaluronan related pathways identified in the biomarker study, this literature-based synthesis was used.

3.10 Reproducibility and Data Interpretation

Where appropriate, all analyses have been conducted in an easily replicable computational environment with fixed randomization parameters [33]. The integrated workflow was designed to maintain transparency throughout the pipeline, starting from raw data preprocessing up to the ranking of biomarkers and interpretation of the involved pathways. The biological conclusions were drawn by combining machine learning output, SHAP explanation, enrichment result, network topology and literature evidence, instead of using any single analytical method for drawing conclusions [7]-[16].

Table 3.1 The computational workflow and analytical tools utilized for HD-CVID transcriptomic analysis

STEP	TOOL/METHOD	PURPOSE
1	Preprocessing, normalization and z-score scaling	Cleaning of the data, elimination of scale bias, and standardization of gene expression features
2	Supervised machine learning classifiers	Comparative categorisation of control and disease samples
3	Support Vector Machine (SVM)	The final prediction model chosen for the search for biomarkers
4	SHAP (Shapley Additive Explanations)	Interpretability of the trained model both locally and globally
5	Enrichr	Functional enrichment analysis of prioritized genes
6	NetworkAnalyst	Reconstruction of the miRNA-mRNA network identification of regulatory hubs

CHAPTER 4

RESULTS

4.1 Dataset Composition and Sample Structure

The analysis was performed using two publicly available GEO datasets. There were 14 control, 5 presymptomatic and 12 symptomatic HD samples in the HD dataset (GSE1751). There were 29 CVID-Y, 30 CVID-N and 24 control-HC samples within the CVID dataset, GSE51405. The datasets were suitable for cross-condition analysis of systemic transcriptomic changes observed in immune dysregulation and neurodegeneration since both were peripheral blood. Both datasets were also biological classes with sufficient variation to identify biomarkers, and also provided an appropriate framework for machine learning classification in this structure.

Table 4.1 Composition of GEO Datasets used in the study

DATASET	CLASS LABEL	CLASS DESCRIPTION	SAMPLE COUNT
GSE1751	CONTROL	Healthy people without Huntington's disease	14
	PRESYMPTOMIC HD	People carrying the HD-associated mutation but who do not yet have overt clinical motor symptoms	05
	SYMPTOMATIC HD	People with clinically manifested Huntington's disease having the characteristic neurological and motor symptoms associated with progression of the disease.	12
GSE51405	CONTROL-HC	Healthy controls without common variable immunodeficiency	24
	CVID-N	CVID patients without inflammatory or autoimmune complications, representing the non-inflammatory subgroup	30
	CVID-Y	CVID patients with inflammatory manifestations and immune dysregulation, representing the inflammatory subgroup	29

4.2 Model Evaluation and Classification Performance

The Support Vector Machine (SVM) model was optimized and evaluated to achieve the most important classification outcome among the evaluated machine learning models. The model performed well with an overall accuracy of 95.7% on independent test set, showing good discrimination between diseased and control samples. The confusion matrix revealed these results: 6 True Negatives, 16 True Positives, 1 False Positive, 0 False Negative, with high sensitivity being obtained for diseased samples. The class-wise evaluation, on the other hand, revealed a precision of 0.94 and a recall of 1.00, indicating that the classifier maintained its high precision and recall of the disease-associated transcriptomic patterns even though it was working towards minimising missed positive cases. These results show that the model was able to distinguish HD, CVID, and control profiles based on features derived from the transcriptome.

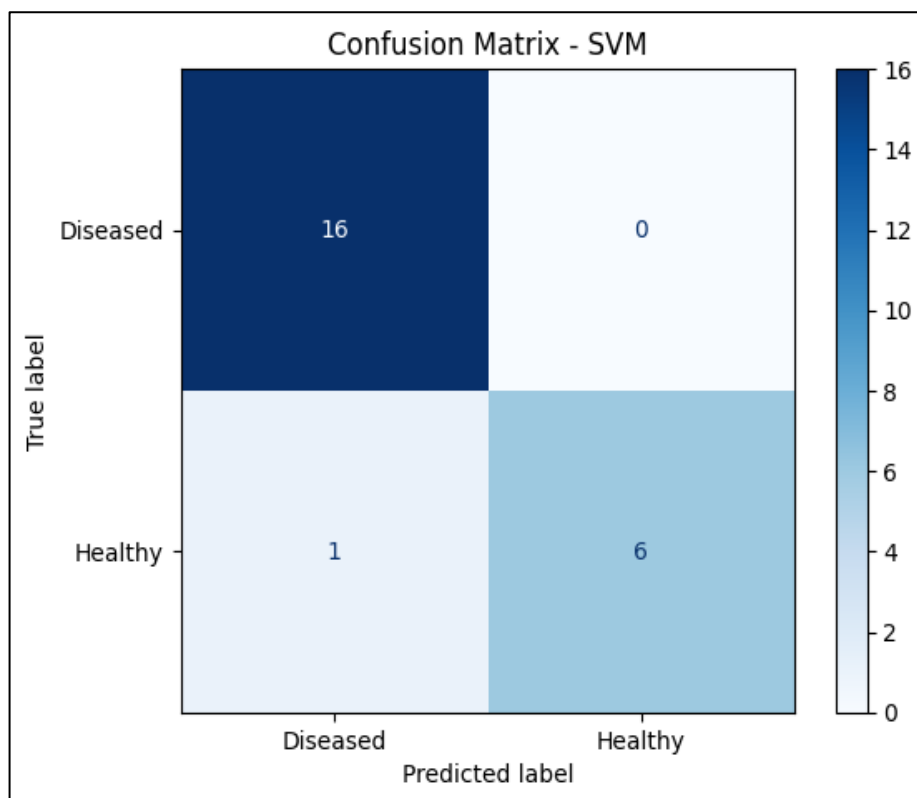


Fig. 4.1 Optimized Support Vector Machine (SVM) classification performance on the independent test set displayed as confusion matrix in the transcriptomic analysis of the HD–CVID data set

4.3 Biomarker Identification by SHAP Analysis

The trained SVM model was used to rank the features by using SHAP analysis. The top reported biomarkers include CSF1, ITGB5, TCP11L2, CDC5L, L3MBL2, and BTG2. Of these, CSF1 and BTG2 were found in the top contributors ranking. We generated model interpretations—top-20 feature list, summary plots and waterfall plots—using the SHAP analysis.

Table 4.2 Representative SHAP-ranked genes and their mean SHAP values showing the most important classification features in the HD-CVID transcriptomic analysis

Gene	Gene Symbol	Mean_SHAP
ILMN_1659852	CSF1	0.92158204
ILMN_1653646	FLJ32310	0.9203993
201124_at	ITGB5	0.873532
ILMN_1654677	TCP11L2	0.5408923
ILMN_1696867	C3orf48	0.21954623
ILMN_1686662	L3MBTL2	0.20887491
ILMN_1678705	PURB	0.19182794
ILMN_1656546	GPR114	0.14863548
ILMN_1670565	ZNF143	0.12651321
ILMN_1658427	CDC5L	0.12383916
200026_at	RPL34	0.115184225
ILMN_1652735	INPP1	0.114695
ILMN_1681591	ARPC4	0.10867162
ILMN_1673322	LYVE1	0.102337964
ILMN_1652549	BTG2	0.08733735
200821_at	LAMP2	0.08316912
200600_at	MSN	0.0822303
200929_at	TMED10	0.07674451
ILMN_1674984	AMPH	0.07059094
ILMN_1685112	NECAB3	0.06937082

4.3.1 Mean Absolute SHAP Bar Plot with Generic Feature Labels

The SHAP bar plot lists the top 20 features based on their mean absolute SHAP value, using generic labels. The top four are Feature 12686, Feature 12051, Feature 247 and Feature 12144. They are markedly longer than the rest of the features in their bars. The bar lengths decrease rapidly after the first four features and the last features make up a lower, continuous tail. The plot shows a ranking of the important variables according to the average model impact magnitude.

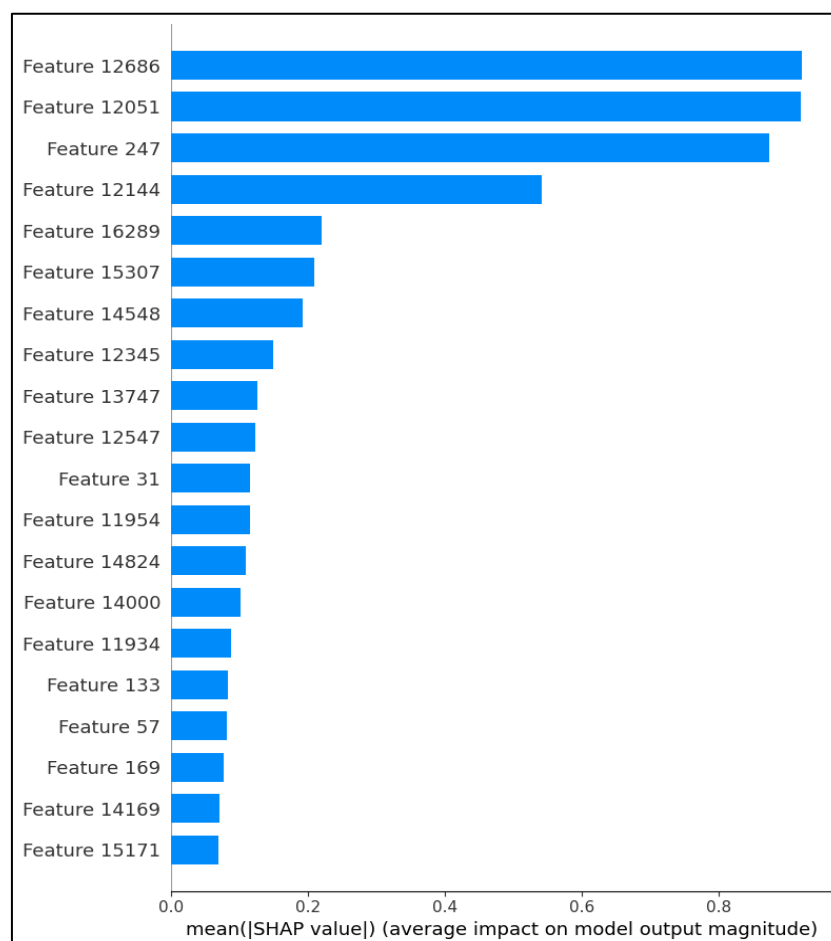


Fig. 4.2 SHAP feature-importance bar plot, using generic feature identifiers, of top 20 transcriptomic features ranked by mean absolute SHAP value

4.3.2 Mean Absolute SHAP Bar Plot with Probe/Transcript Labels

The same ranking format using probe or transcript identifiers is shown by the SHAP bar plot. The top features are ILMN_1659852, ILMN_1653646, 201124_at, and ILMN_1654677. The mean absolute SHAP values of the first two features are very high, with 201124_at coming close behind. The remaining probes' feature importance values decrease in order after these probes. The features are shown from most important to least important at the plot-level.

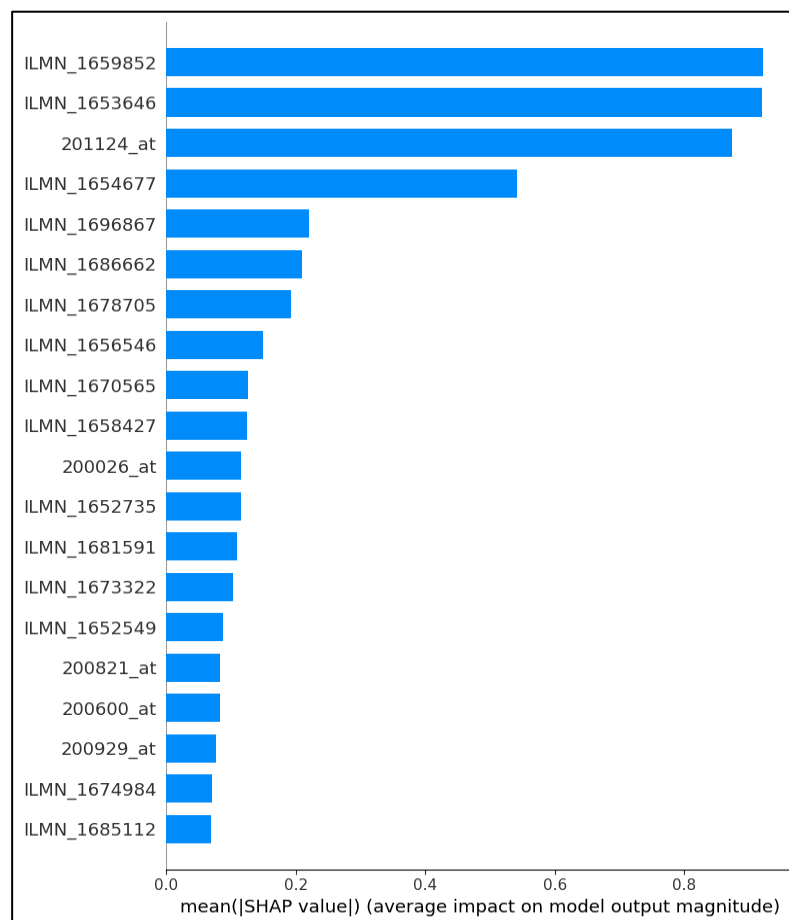


Fig. 4.3 SHAP feature-importance bar plot showing the top 20 transcript/probe-level features ranked by mean absolute SHAP value

4.3.3 SHAP Beeswarm Plot

The distribution of feature effects across all samples is shown by the SHAP beeswarm plot. Every point is a sample; the x value is the SHAP value; the color scale indicates the intensity of the value of the feature. The top features in the plot are ILMN_1659852, ILMN_1653646, 201124_at, and ILMN_1654677. While the less ranked features are more clustered around zero, the wider ranged features have the largest SHAP values in the sample set. The plot also displays positive and negative SHAP values for various features, with the points being colored on either side of the zero line.

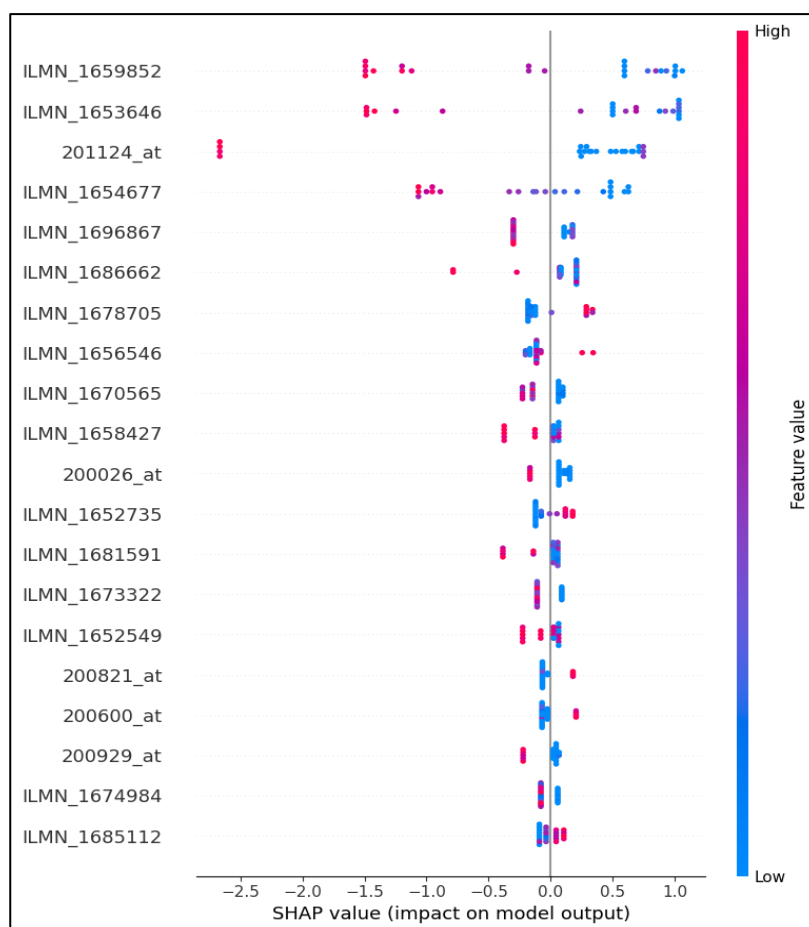


Fig. 4.4 SHAP summary beeswarm plot indicating the distribution and direction of contribution of the top transcriptomic features by all samples

4.3.4 SHAP Force Plot

The SHAP local explanation plot illustrates the contribution of individual transcriptomic features to the prediction of a single sample. The model output is moving towards the positive class, starting from about -0.85 and ending at 3.73. The red section of the plot indicates features that boost the prediction score, while the blue section indicates features that lower the prediction score. The prediction is mainly influenced by a small group of transcriptomic features, including ILMN_1678705, ILMN_1654677, 201124_at, ILMN_1659852, ILMN_1653646, and ILMN_1652574. The result here illustrates how SHAP works on the sample level and how the trained model is able to make an individual prediction.

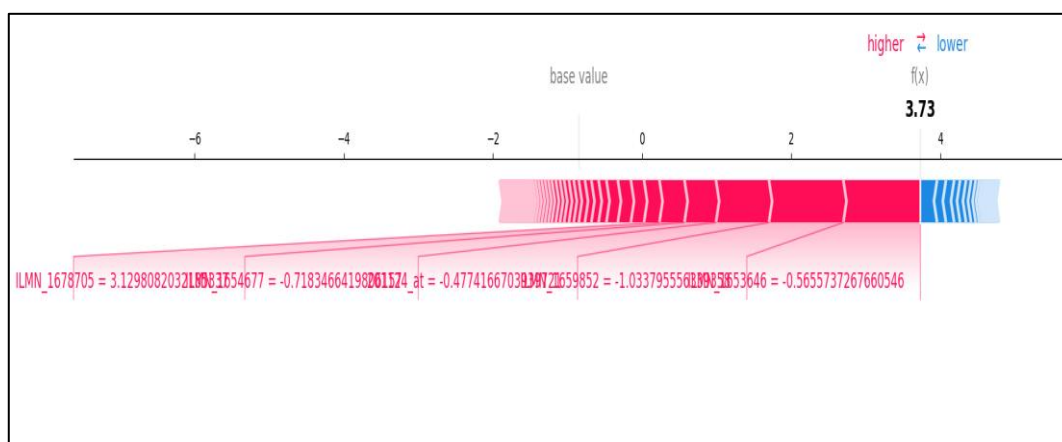


Fig. 4.5 SHAP force plot displaying individual transcriptomic features contribution to the prediction of a single sample from the trained SVM model

4.3.5 SHAP Waterfall Plot

Each feature's contribution to a single model's prediction is shown in the SHAP waterfall plot. The initial value is close to -0.85 and the final value is close to 3.73. The largest positive contributions are from ILMN_1653652 (+1.03), ILMN_1659854 (+1.00), 201129_at (+0.71), ILMN_1654685 (+0.42), and ILMN_1678707 (+0.34). One of the features (-0.23) is ILMN_1652574. This plot also highlights the contribution of 16,373 other features that have a combined effect of +0.75.

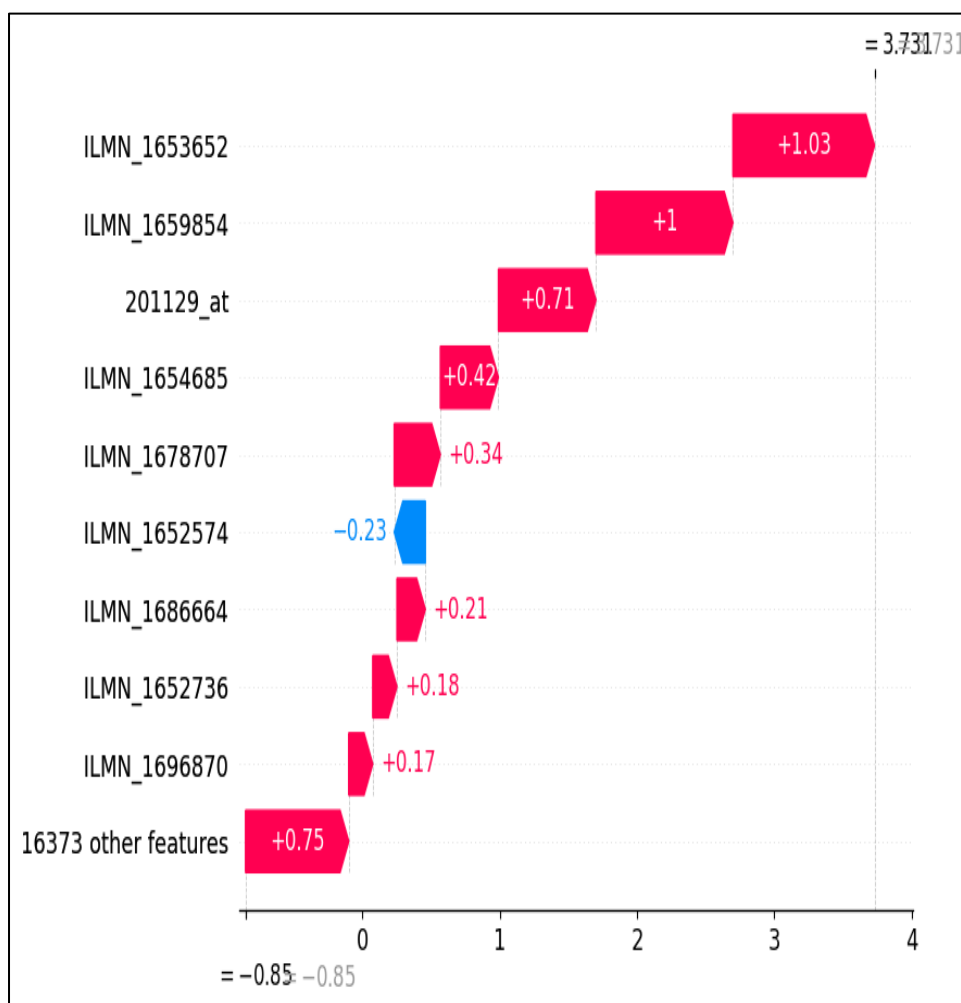


Fig. 4.6 SHAP waterfall plot illustrating the contribution of individual transcriptomic features to a single model prediction

4.4 Functional Enrichment Analysis

Pathway enrichment of SHAP-ranked genes was performed using Enrichr. The most prominent pathways found to be enriched were antigen processing and presentation, clathrin-mediated endocytosis, hyaluronan uptake and degradation, and axon guidance. The results of the enrichment were shown in a ranked bar plot of biological pathways, in which the above pathways were among the highest-ranking results.

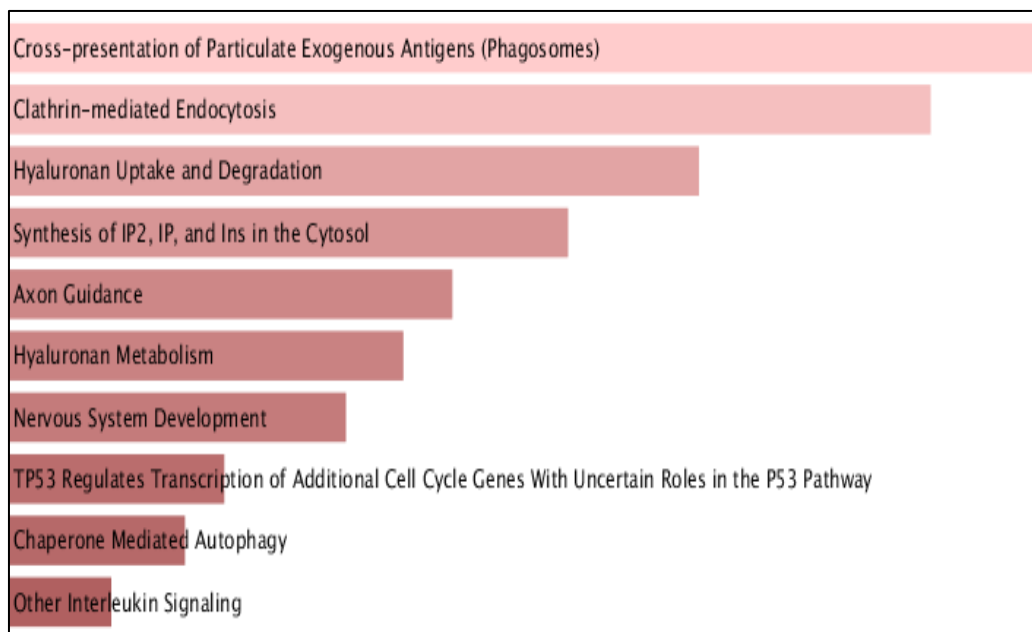


Fig. 4.7 Enrichment bar graph for Reactome pathways of the genes prioritized by SHAP from the transcriptomic analysis of HD-CVID

Network analysis was conducted with the aid of NetworkAnalyst. BTG2 and CSF1 were found to be prominent hubs in the miRNA–mRNA interactions network. The network contained several transcription factors and miRNAs associated with the shortlisted genes and miR-22 and miR-155 were highlighted as inflammatory miRNAs associated with CSF1. This interaction network was represented as a hub-centered network, in which BTG2 and CSF1 were located in a central position.

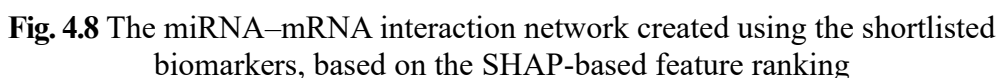


Fig. 4.8 The miRNA–mRNA interaction network created using the shortlisted biomarkers, based on the SHAP-based feature ranking

CHAPTER 5

DISCUSSION

The present findings suggest that there is more than just a clinical theme shared between Huntington's disease and common variable immunodeficiency, as there is a measurable peripheral molecular signature shared between the two. The pipeline for analysis included preprocessing, supervised classification, SHAP-based feature ranking, pathway enrichment, and miRNA–mRNA network reconstruction to narrow down the high-dimensional expression data into a small set of interpretable biomarkers. The framework was designed specifically to highlight common biomarkers and key pathways between the two disorders, and the findings support this goal by revealing similar key genes and pathways emerging across several layers of analysis.

The Support Vector Machine proved to be particularly noteworthy. 95.7% of accuracy was achieved by the model on the independent test set, with 16 true positives, 6 true negatives, 1 false positive, and no false negatives. It is significant in a biomarker oriented biomedical context since the absence of false negative implies that the classifier maintained a good sensitivity on disease samples. The high precision and perfect recall, along with a good F1 score, demonstrates the strength of the transcriptomic signal, which was not weak or diffuse, but structured enough to be captured by a margin-based classifier. This also motivates the choice of SVM as a better model than the other ones tested for downstream interpretation.

The machine learning output was translated into a biologically meaningful ranking of the candidate biomarkers, using the SHAP analysis. The top genes that were consistently highlighted were CSF1, ITGB5, TCP11L2, CDC5L, L3MBL2 and BTG2, with CSF1 and BTG2 being the most significant in the classifier. This is biologically plausible because CSF1 is a gene that plays a role in immune regulation and macrophage related signalling and BTG2 is a gene that plays a role in stress response and antiproliferative transcriptional control. They are consistently the top ranked features in the explainability list, which implies that they are not random statistical features but rather likely important nodes in the disease related transcriptomic program.

This is reinforced by the visualizations of the SHAP. Both the bar plots reveal a very focused importance profile, with a clear drop following the top variables, suggesting that the model relies on a short and concentrated predictive signature and less on a large yet poorly differentiated gene set. The beeswarm plot depicts that the direction of the expression change is also important along with the magnitude of the expressions, as the strongest features have both positive and negative effects depending on the feature value. The waterfall plot then shows how a single sample is pushed away from the baseline to the final prediction by a small number of strong features, and the rest of the features make a smaller contribution overall. Combined, this indicates the utility of SHAP in this context: not just for identifying important genes, but also how these genes are influencing classification at the population and single-sample levels.

The most important mechanistic context to the study is added by the pathway enrichment results. The antigen processing and presentation is also relevant to the immune dysregulation aspect of CVID and aligns with the general concept that neurodegenerative conditions are associated with peripheral and central immune activation. Clathrin-mediated endocytosis has an interesting connection to receptor trafficking, vesicle recycling and signalling in both immune cells and neurons. Axon guidance introduces a direct neuronal aspect and indicates that the common signature could be involved in pathways that relate to connectivity, migration and neuroimmune signaling. Both of these pathways form a framework in which immune and neuronal biology converge and are not mutually exclusive.

The enrichment of hyaluronan uptake and degradation is the most distinctive finding. Hyaluronan metabolism plays a key role in extracellular matrix remodeling, tissue homeostasis and inflammatory signaling. It is particularly relevant, because hyaluronan is known to have a context-dependent nature: High-molecular-weight (HMW) hyaluronan is associated with homeostasis and anti-inflammatory properties, while degraded or fragmented HA with immune activation and stress signaling. The presence of hyaluronan uptake and degradation pathways, in the SHAP prioritized gene set, indicates that the classifier is not only reflecting immune related transcriptional changes but may also be reflecting extracellular matrix biology that may affect tissue microenvironment and neuronal vulnerability.

The review focuses on the role of dysregulation of hyaluronan metabolism in affecting glial cell activation, stability of the perineuronal nets, cytokine signaling and CNS repair. It also draws attention to the fact that fragments of hyaluronan are inflammatory mediators that bind to receptors like CD44, RHAMM and Toll-like receptors. The convergence point between the literature-based and data-driven components of the thesis is the fact that uptake and degradation of hyaluronan has emerged from the transcriptomic enrichment analysis. So, the hyaluronan axis is validated independently by both the literature synthesis and the machine learning results and it is one of the most interesting and biologically significant of the entire work.

An additional regulatory layer is added to this research by the miRNA–mRNA network analysis. BTG2 and CSF1 were identified as prominent hubs in the interaction network, BTG2 is associated with stress-related miRNAs while CSF1 is associated with inflammatory miRNAs such as miR-22 and miR-155. This is significant because it implies that the biomarker signal is both transcriptionally and post-transcriptionally controlled, with miRNAs possibly influencing the critical genes' activity and abundance. From the biological perspective, the network appears to have a multi-layered regulation of immune and stress-response pathways, a type of design that would be expected in a disease interface comprising neurodegeneration and chronic inflammation.

Together, these findings support the notion that the HD–CVID overlap observed in blood is not due to random noise at the level of individual genes, but is rather driven by immune linked transcriptional modules. Given the overlap of CSF1 and BTG2 in the SHAP ranking, enrichment context, and network hubs, these genes could potentially be recognized as convergent points where immune signaling, stress response and transcriptional control all converge. This is also supported by the hyaluronan related enrichment, which provides a clear connection to the biology of neuroinflammation and remodelling of the extracellular matrix. These data-

classifications are not only more than a classification; they are a mechanistic hypothesis of immune-neurodegenerative crosstalk that can be tested in future experimental work.

The study also demonstrates the usefulness of interpretable machine learning for discovery of transcriptomic biomarkers. A high accuracy classifier was useful but the real power of the present analysis was the ability to translate the model into biologically interpretable features using SHAP and then contextualize the model within pathways and networks. This combination connects prediction and mechanism and the results more pertinent for the field of translational research. Concurrently, the results are hypothesis-generating, and should be confirmed in independent cohorts and orthogonal experimental systems prior to any claim of a biomarker or pathway being definitive.

Overall, the discussion is in line with a coherent immune-neurodegenerative framework in which, blood-based transcriptomic signatures, explainable AI, enrichment analysis, and network reconstruction all converge to a few common biologically meaningful biomarkers and pathways. The major result is not only the discovery of CSF1 and BTG2, but the proof of principle that two clinically different disorders can be analyzed with a similar molecular architecture by a single integrated analytical strategy. By relating the transcriptome results to extracellular matrix remodelling and inflammatory signalling, the hyaluronan pathway adds depth to this framework and makes a physiologically significant connection between the two sections of the study.

CHAPTER 5

CONCLUSION AND FUTURE SCOPE

Although increasing evidence indicated that both Huntington's disease (HD) and common variable immunodeficiency (CVID) represented immune dysfunction and peripheral blood gene expression alterations that could impact neurodegeneration, the key problem of this study was a lack of an integrated, interpretable transcriptomic framework linking these two disorders through common molecular signatures. The project was designed to fill this gap, with the aim of merging machine learning-based transcriptomic classification with explainable artificial intelligence (XAI), pathway enrichment and miRNA–mRNA network analysis as well as embedding them within the larger context of hyaluronan metabolism and the immune–neurodegenerative crosstalk.

The study showed the potential for high accuracy classification of disease and control states using transcriptomic data in these blood samples, and the ability to identify a small number of biologically relevant biomarkers. Based on the classifiers evaluated, Support Vector Machine gave the best results with high sensitivity and 95.7% accuracy on the independent test set. Conversely, SHAP-based explainability narrowed the high-dimensional expression space to a smaller set of biomarkers, including CSF1, ITGB5, TCP11L2, CDC5L, L3MBTL2 and BTG2, which played a key role in the classification. In particular, CSF1 and BTG2 had the highest model influence, suggesting that the observed transcriptome difference between HD, CVID and control samples is mainly driven by immune-regulatory and stress-response programs.

An important result of the study was that the biomarkers were not found to be isolated statistical traits, but to be part of coherent biological pathways. Functional enrichment suggested shared molecular pathways involved in antigen processing and presentation, clathrin-mediated endocytosis, hyaluronan uptake and degradation, and axon guidance, which suggested the involvement of immune activation, intracellular trafficking, extracellular matrix remodeling, and neural signaling. This is especially important as the increased uptake and degradation of hyaluronan provides a direct mechanistic link between the transcriptomic observations and the existing literature on hyaluronan as a regulator of extracellular matrix integrity, inflammation and stability of the neural microenvironment. This was further strengthened using the miRNA–mRNA network analysis, which identified BTG2 and CSF1 as regulatory hubs that were associated with disease-related stress and inflammatory miRNAs, suggesting a multi-layered model of post-transcriptional regulation in disease-associated gene regulation.

Overall, the results demonstrate that there is a measurable peripheral transcriptomic signature for HD and CVID that is mediated by immune-linked molecular remodeling, and not random gene-level noise. Findings from four different methods highlighted by their convergence: high classifier performance, biomarker prioritisation using SHAP values, pathway enrichment, and network analysis, showed that interpretable machine learning can reveal common disease biology, which is statistically significant and

mechanistically informative. Immunological pathway enrichment (CSF1 and BTG2) implies a potential interplay between immune signaling, transcriptional stress responses and extracellular matrix remodeling that can impact neurodegenerative vulnerability. The results provide a solid backbone for developing hypotheses for future biomarker development and immunologically informed therapeutic exploration.

Future studies should validate the biomarkers identified in independent and larger cohorts, validate the transcript level results with orthogonal experimental approaches such as qPCR and/or RNA-seq, and investigate if the same molecular signature is preserved at the protein and epigenetic level. Further studies on function are also required to elucidate the role of CSF1, BTG2 and hyaluronan associated pathways in immune-neurodegenerative crosstalk within disease-relevant systems. Such efforts will help ascertain if the biomarkers found here can be converted into a strong diagnostic marker, therapeutic targets, or prognostic indicator for a neurodegenerative disorder.

REFERENCES

- [1] F. Borovecki *et al.*, “Genome-wide expression profiling of human blood reveals biomarkers for Huntington’s disease,” *Proceedings of the National Academy of Sciences*, vol. 102, no. 31, pp. 11023–11028, Jul. 2005, doi: 10.1073/pnas.0504921102.
- [2] J. Park *et al.*, “Interferon signature in the blood in inflammatory common variable immune deficiency,” *PLoS ONE*, vol. 8, no. 9, p. e74893, Sep. 2013, doi: 10.1371/journal.pone.0074893.
- [3] C. A. Ross and S. J. Tabrizi, “Huntington’s disease: from molecular pathogenesis to clinical treatment,” *The Lancet Neurology*, vol. 10, no. 1, pp. 83–98, Dec. 2010, doi: 10.1016/s1474-4422(10)70245-3.
- [4] D. W. Weir, A. Sturrock, and B. R. Leavitt, “Development of biomarkers for Huntington’s disease,” *The Lancet Neurology*, vol. 10, no. 6, pp. 573–590, May 2011, doi: 10.1016/s1474-4422(11)70070-9.
- [5] F. A. Bonilla *et al.*, “International Consensus Document (ICON): Common Variable Immunodeficiency Disorders,” *The Journal of Allergy and Clinical Immunology in Practice*, vol. 4, no. 1, pp. 38–59, Nov. 2015, doi: 10.1016/j.jaip.2015.07.025.
- [6] H. Chapel, “Common variable immunodeficiency disorders (CVID) — diagnoses of exclusion, especially combined immune defects,” *The Journal of Allergy and Clinical Immunology in Practice*, vol. 4, no. 6, pp. 1158–1159, Nov. 2016, doi: 10.1016/j.jaip.2016.09.006.
- [7] S. Ng, S. Masarone, D. Watson, and M. R. Barnes, “The benefits and pitfalls of machine learning for biomarker discovery,” *Cell and Tissue Research*, vol. 394, no. 1, pp. 17–31, Jul. 2023, doi: 10.1007/s00441-023-03816-z.
- [8] M. R. Karim *et al.*, “Explainable AI for bioinformatics: methods, tools and applications,” *Briefings in Bioinformatics*, vol. 24, no. 5, Jul. 2023, doi: 10.1093/bib/bbad236.
- [9] S. Lundberg and S.-I. Lee, “A unified approach to interpreting model predictions | Proceedings of the 31st International Conference on Neural Information Processing Systems,” *Guide Proceedings*, doi: 10.5555/3295222.3295230.
- [10] M. V. Kuleshov *et al.*, “Enrichr: a comprehensive gene set enrichment analysis web server 2016 update,” *Nucleic Acids Research*, vol. 44, no. W1, pp. W90–W97, May 2016, doi: 10.1093/nar/gkw377.
- [11] J. Xia, M. J. Benner, and R. E. W. Hancock, “NetworkAnalyst - integrative approaches for protein–protein interaction network analysis and visual exploration,” *Nucleic Acids Research*, vol. 42, no. W1, pp. W167–W174, May 2014, doi: 10.1093/nar/gku443.
- [12] J. Xia, E. E. Gill, and R. E. W. Hancock, “NetworkAnalyst for statistical, visual and network-based meta-analysis of gene expression data,” *Nature Protocols*, vol. 10, no. 6, pp. 823–844, May 2015, doi: 10.1038/nprot.2015.052.

- [13] R. Stern, A. A. Asari, and K. N. Sugahara, "Hyaluronan fragments: An information-rich system," *European Journal of Cell Biology*, vol. 85, no. 8, pp. 699–715, Jul. 2006, doi: 10.1016/j.ejcb.2006.05.009.
- [14] A. C. Petrey and C. A. De La Motte, "Hyaluronan, a crucial regulator of inflammation," *Frontiers in Immunology*, vol. 5, p. 101, Mar. 2014, doi: 10.3389/fimmu.2014.00101.
- [15] B. P. Toole, "Hyaluronan: from extracellular glue to pericellular cue," *Nature Reviews. Cancer*, vol. 4, no. 7, pp. 528–539, Jul. 2004, doi: 10.1038/nrc1391.
- [16] S. A. Back *et al.*, "Hyaluronan accumulates in demyelinated lesions and inhibits oligodendrocyte progenitor maturation," *Nature Medicine*, vol. 11, no. 9, pp. 966–972, Aug. 2005, doi: 10.1038/nm1279.
- [17] A. Peters and L. S. Sherman, "Diverse roles for hyaluronan and hyaluronan receptors in the developing and adult nervous system," *International Journal of Molecular Sciences*, vol. 21, no. 17, p. 5988, Aug. 2020, doi: 10.3390/ijms21175988.
- [18] M. Morawski, G. Brückner, C. Jäger, G. Seeger, and T. Arendt, "Neurons associated with aggrecan-based perineuronal nets are protected against tau pathology in subcortical regions in Alzheimer's disease," *Neuroscience*, vol. 169, no. 3, pp. 1347–1363, May 2010, doi: 10.1016/j.neuroscience.2010.05.022.
- [19] E. Syková and C. Nicholson, "Diffusion in brain extracellular space," *Physiological Reviews*, vol. 88, no. 4, pp. 1277–1340, Oct. 2008, doi: 10.1152/physrev.00027.2007.
- [20] S. Amor, F. Puentes, D. Baker, and P. Van Der Valk, "Inflammation in neurodegenerative diseases," *Immunology*, vol. 129, no. 2, pp. 154–169, Jan. 2010, doi: 10.1111/j.1365-2567.2009.03225.x.
- [21] P. Pintér and A. Alpár, "The role of extracellular matrix in human neurodegenerative diseases," *International Journal of Molecular Sciences*, vol. 23, no. 19, p. 11085, Sep. 2022, doi: 10.3390/ijms231911085.
- [22] G. Jensen, J. L. Holloway, and S. E. Stabenfeldt, "Hyaluronic acid biomaterials for central nervous system regenerative medicine," *Cells*, vol. 9, no. 9, p. 2113, Sep. 2020, doi: 10.3390/cells9092113.
- [23] X. Wang, J. He, Y. Wang, and F.-Z. Cui, "Hyaluronic acid-based scaffold for central neural tissue engineering," *Interface Focus*, vol. 2, no. 3, pp. 278–291, Mar. 2012, doi: 10.1098/rsfs.2012.0016.
- [24] M. Marella, L. Jadin, G. A. Keller, B. J. Sugarman, G. I. Frost, and H. M. Shepard, "KIAA1199 expression and hyaluronan degradation colocalize in multiple sclerosis lesions," *Glycobiology*, vol. 28, no. 12, pp. 958–967, Jul. 2018, doi: 10.1093/glycob/cwy064.
- [25] A. Peters *et al.*, "Distinct chemical structures inhibit the CEMIP hyaluronidase and promote oligodendrocyte progenitor cell maturation," *Journal of Biological Chemistry*, vol. 300, no. 12, p. 107916, Oct. 2024, doi: 10.1016/j.jbc.2024.107916.
- [26] A. Konopka *et al.*, "Cleavage of hyaluronan and CD44 adhesion molecule regulate astrocyte morphology via RAC1 signalling," *PLoS ONE*, vol. 11, no. 5, p. e0155053, May 2016, doi: 10.1371/journal.pone.0155053.
- [27] Q. Zhang *et al.*, "High level of serum and cerebrospinal fluid of heparan sulfate and hyaluronic acid might be a biomarker of severity of neuromyelitis optica,"

- Frontiers in Immunology*, vol. 12, p. 705536, Jul. 2021, doi: 10.3389/fimmu.2021.705536.
- [28] G. Zhou, O. Soufan, J. Ewald, R. E. W. Hancock, N. Basu, and J. Xia, "NetworkAnalyst 3.0: a visual analytics platform for comprehensive gene expression profiling and meta-analysis," *Nucleic Acids Research*, vol. 47, no. W1, pp. W234–W241, Mar. 2019, doi: 10.1093/nar/gkz240.
 - [29] L. S. Sherman and S. A. Back, "A 'GAG' reflex prevents repair of the damaged CNS," *Trends in Neurosciences*, vol. 31, no. 1, pp. 44–52, Dec. 2007, doi: 10.1016/j.tins.2007.11.001.
 - [30] "GEO Accession viewer."
<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE1751>
 - [31] "GEO Accession viewer."
<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE51405>
 - [32] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016, pp. 785–794, doi: 10.1145/2939672.2939785
 - [33] PedregosaFabian *et al.*, "SciKit-Learn: Machine Learning in Python," *Journal of Machine Learning Research*, Nov. 2011, doi: 10.5555/1953048.2078195.

LIST OF PUBLICATIONS

1. “**International Conference on Biomedical Innovations in Bioinformatics and Biotechnology (ICBIBB-26)**” organized by National Institute of Engineering and Research (NIER), held on 15th March 2026 at Ahmedabad, India.



National Institute for
**Engineering
Research**

ICBIBB-26

**International Conference on Biomedical Innovations in
Bioinformatics and Biotechnology (ICBIBB-26)**

Date : 02-03-2026

Acceptance Letter

Dear Author(s)

We are pleased to inform you that your paper has been accepted by the review committee for Oral Presentation at the **International Conference on Biomedical Innovations in Bioinformatics and Biotechnology (ICBIBB-26)**

Paper ID: NIER_09351

Article Title: "Hyaluronan Metabolism at the Interface of Immune Dysregulation and Neurodegeneration"

Author's Name: Prerna Joshi, Asmita Das

This conference will be held in **15th March 2026 on Ahmedabad - India**

Please register as soon as possible in order to secure your participation:
<https://nier.in/conf/fee-details.php?id=101014508>

Kindly make the payment and email us a screenshot of the successful payment confirmation, including your name and the title of your paper, to complete your registration.

Sincerely,



Dr. Tara Srivastava
Director
National Institute For Engineering And Research (NIER)



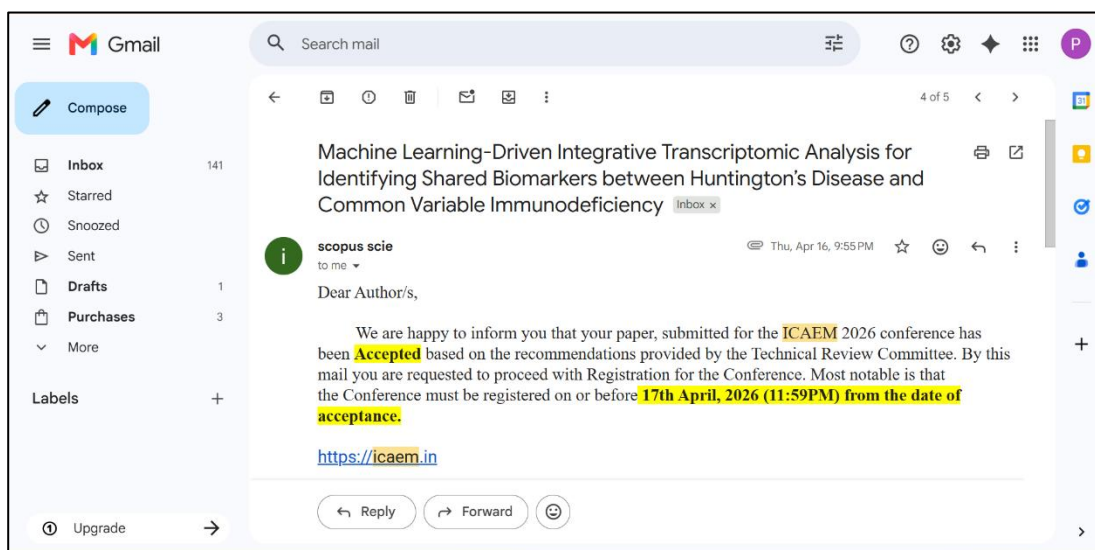
 +91-9677007228

 info@nier.in

 www.nier.in



2. “3rd International Conference on Advances in Engineering and Medical Sciences 2026 (ICAEM 2026)” organized by International School of Technology and Sciences for Women, held on 17th & 18th April 2026 at Andhra Pradesh, India.



PLAGIARISM REPORT

Prerna MTech Thesis

ORIGINALITY REPORT

3 %	2 %	2 %	1 %
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	www.mdpi.com Internet Source	1 %
2	"Adolescent Brain Cognitive Development Neurocognitive Prediction", Springer Science and Business Media LLC, 2019 Publication	<1 %
3	www.science.gov Internet Source	<1 %
4	Submitted to University of Witwatersrand Student Paper	<1 %
5	uef.sav.sk Internet Source	<1 %
6	www.irjet.net Internet Source	<1 %
7	www.ijltemas.in Internet Source	<1 %
8	Rachael I Scahill, Paul Zeun, Katherine Osborne-Crowley, Eileanoir B Johnson et al. "Biological and clinical characteristics of gene carriers far from predicted onset in the Huntington's disease Young Adult Study (HD-YAS): a cross-sectional analysis", The Lancet Neurology, 2020 Publication	<1 %
9	A. Punitha, S. Syedakbar, S. Jeyasudha. "Intelligent and Sustainable Systems: AI,	<1 %

Green IoT, and Adaptive Automation in
Electrical and Communication Technologies -
Proceedings of International Conference in
Pathway of Electrical, Automation,
Communication, and Electronics (PEACE-
2025)", CRC Press, 2026

Publication

- 10 Wengeng Cao, Zhuo Zhang, Yu Fu, Lihua Zhao, Yu Ren, Tian Nan, Huaming Guo. "Prediction of arsenic and fluoride in groundwater of the North China Plain using enhanced stacking ensemble learning", Water Research, 2024

Publication

- 11 ai.jmir.org <1 %

Internet Source

- 12 Harold Ellis. "French's Index of Differential Diagnosis an A-Z", CRC Press, 2016 <1 %

Publication

- 13 repositories.lib.utexas.edu <1 %

Internet Source

- 14 islandscholar.ca <1 %

Internet Source

- 15 Thierry Vincent, Nadir Mehti. "Extracellular matrix in bone marrow can mediate drug resistance in myeloma", Leukemia & Lymphoma, 2009 <1 %

Publication

- 16 Yong-Hong Liao, Stuart A. Jones, Ben Forbes, Gary P. Martin, Marc B. Brown. "Hyaluronan: Pharmaceutical Characterization and Drug Delivery", Drug Delivery, 2008 <1 %

Publication

17 www.ijprems.com <1 %
Internet Source

18 www.ncbi.nlm.nih.gov <1 %
Internet Source

19 Nicholas T. Seyfried. "Fourier transform mass spectrometry to monitor hyaluronan-protein interactions: use of hydrogen/deuterium amide exchange", Rapid Communications in Mass Spectrometry, 01/30/2007 <1 %
Publication

20 Srinivas Sethi, Bibhudatta Sahoo, Deepak Tosh, Suwendra Kumar Jayasingh, Sourav Kumar Bhoi. "Computing, Communication and Intelligence - Proceedings of 2 International Conference on Cutting Edge Technology in Computing, Communication and Intelligence (ICCTCCI) – 2025", CRC Press, 2026 <1 %
Publication

Exclude quotes Off

Exclude matches Off

Exclude bibliography Off



Page 1 of 32 - Cover Page

Submission ID: trn:oid::1:3578852425

Prerna Mtech

Prerna MTech Thesis



Quick Submit



Quick Submit



Delhi Technological University

Document Details

Submission ID

trn:oid::1:3578852425

30 Pages

Submission Date

May 25, 2026, 10:30 PM GMT+5:30

8,457 Words

Download Date

May 26, 2026, 12:51 PM GMT+5:30

52,554 Characters

File Name

Mtech_Thesis_Final.docx

File Size

1.0 MB



Page 1 of 32 - Cover Page

Submission ID: trn:oid::1:3578852425



*% detected as AI

AI detection includes the possibility of false positives. Although some text in this submission is likely AI generated, scores below the 20% threshold are not surfaced because they have a higher likelihood of false positives.

Caution: Review required.

It is essential to understand the limitations of AI detection before making decisions about a student's work. We encourage you to learn more about Turnitin's AI detection capabilities before using the tool.

Disclaimer

Our AI writing assessment is designed to help educators identify text that might be prepared by a generative AI tool. Our AI writing assessment may not always be accurate (i.e., our AI models may produce either false positive results or false negative results), so it should not be used as the sole basis for adverse actions against a student. It takes further scrutiny and human judgment in conjunction with an organization's application of its specific academic policies to determine whether any academic misconduct has occurred.





DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)
Shahbad Daulatpur, Main Bawana Road, Delhi-42

PLAGIARISM VERIFICATION

Title of the Thesis- Machine Learning-Driven Integrative Transcriptomic Analysis to Identify Shared Biomarkers and Molecular Pathways between Huntington's Disease and Common Variable Immunodeficiency

Total Pages 50 Name of the Student Prerna Joshi

Supervisor

Dr. Asmita Das

Department- Department of Biotechnology

This is to report that the above thesis was scanned for similarity detection. Process and outcome is given below:

Software used: Turnitin Similarity Index: 3%, Total Word Count: 8457

Date: _____

Candidate's Signature

Signature of Supervisor