Deciphering the role of MYC Gene Variants in Childhood Cerebral Adrenoleukodystrophy (ccALD) using Integrative Transcriptomic and Variant Analysis

Chakresh Kumar Jain*[1], Sarita Maurya[1] , Pankaj Kumar Tripathi[1], Divya Mishra[2]

[1]*Department of Biotechnology, Jaypee Institute of Information Technology, A-10, Sector 62 Noida, Uttar Pradesh, India- 201309*

[2] Department of Biomedical Sciences, Humanitas University, Milan, Italy
[3] Tumor Microenvironment Unit, IRCCS Humanitas Research Hospital, Milan, Italy

[1]**Corresponding author: ckj522@yahoo.com, chakresh.jain@mail.jiit.ac.in*

Abstract

Objective: Childhood cerebral adrenoleukodystrophy (ccALD) is a severe neurodegenerative disorder characterized by rapid demyelination and disruption of the blood–brain barrier (BBB). Although mutations in the *ABCD1* gene are the primary cause of the disease, the downstream molecular mechanisms contributing to disease progression remain poorly understood. This study aimed to identify key regulatory genes and functionally relevant variants associated with ccALD using an integrative transcriptomic and variant analysis approach.

Methods: RNA sequencing (RNA-seq) data derived from brain microvascular endothelial cells (BMECs) of ccALD patients and healthy controls were analyzed to identify differentially expressed genes (DEGs). Gene interaction networks were constructed using Cytoscape to identify hub genes. Variant calling was performed following the GATK Best Practices pipeline, including read preprocessing, alignment, base quality score recalibration, variant calling, filtering, and functional annotation using GATK Funcotator.

Results: Differential expression analysis identified 1,039 upregulated and 744 downregulated genes in ccALD BMECs compared with controls. Network analysis revealed *MYC* as the top hub gene, indicating its central role in disease-associated molecular networks. Variant analysis identified two rare variants within the *MYC* gene: rs2130098148 (C>A), a missense mutation, and rs2130107263 (A>T), a stop-gain mutation. Both variants were associated with pathways related to MAPK signaling, suggesting potential functional consequences for MYC-mediated cellular regulation.

Conclusion: This integrative transcriptomic and variant analysis demonstrates disruption of *MYC* at both the expression and genetic levels in ccALD. These findings suggest that *MYC* may act as a modifier gene contributing to BBB dysfunction and neuroinflammation in ccALD. The study highlights the value of integrative multi-omics approaches in uncovering molecular mechanisms underlying rare neurodegenerative diseases and identifies *MYC* as a potential candidate for further functional investigation.

Keywords: ccALD, MYC, RNA-seq, variant analysis, MAPK, BBB

1. Introduction

The rare X-linked genetic condition known as adrenoleukodystrophy (ALD) mostly affects the adrenal glands and central nervous system (CNS). The peroxisomal ATP-binding cassette (ABC) transporter is encoded by the *ABCD1* gene, and mutations in this gene are the cause.This transporter is essential for the β-oxidation of very long-chain fatty acids (VLCFAs) within peroxisomes**[1]**. In ALD, mutations in the *ABCD1* gene disrupt the normal catabolism of VLCFAs, leading to their accumulation in various tissues, especially the CNS. The buildup of VLCFAs in the white matter of the brain leads to demyelination, neurodegeneration, and progressive cognitive and motor dysfunction**[2]**. The

disorder presents in several forms, with childhood cerebral ALD (ccALD) being the most severe and rapidly progressing subtype. ccALD typically manifests in young children, often between the ages of 4 and 10, with rapid deterioration in cognitive abilities, motor skills, and other neurological functions. This form of the disease is characterized by aggressive demyelination in the CNS, leading to irreversible neurological damage**[3]**. Without early intervention, children with ccALD experience a poor prognosis, with death often occurring within a few years of disease onset. The rapid progression of ccALD makes it one of the most devastating neurodegenerative disorders**[4]**. The *ABCD1* gene mutation is the primary genetic cause of ALD, but the exact mechanisms through which the accumulation of VLCFAs leads to neurodegeneration remain largely unclear. The pathophysiology of ALD involves several interrelated processes, including inflammation, oxidative stress, and mitochondrial dysfunction**[5].** The accumulation of VLCFAs, particularly in the CNS, is toxic to myelin-producing oligodendrocytes, leading to demyelination. In addition, the disruption of lipid metabolism and alterations in immune signaling pathways contribute to the progression of the disease. These complex interactions are not yet fully understood, which underscores the need for comprehensive research to explore the molecular underpinnings of ccALD**[6]**. Recent advances in genomic technologies, particularly RNA sequencing (RNA-seq), have significantly enhanced our ability to explore the molecular mechanisms underlying genetic disorders. RNA-seq offers high-resolution, genome-wide transcriptomic data that can be used to identify differentially expressed genes (DEGs) and provide insights into cellular pathways affected by disease**[7]**. By examining the gene expression profiles in patients with ccALD, researchers can uncover altered pathways involved in disease progression. RNA-seq can also be used to identify specific genetic variants that may play a role in disease pathogenesis. This approach provides an opportunity to investigate the effects of genetic mutations not only on gene expression but also on the downstream cellular processes that drive disease**[8]**.

ALD primarily affects males due to its X-linked inheritance. Female carriers typically remain asymptomatic but may exhibit mild symptoms. The disease manifests in three clinical forms:

ccALD: Onset between ages 4–10, characterized by behavioral changes, learning difficulties, and rapid neurodegeneration due to demyelination. Early diagnosis is critical, as untreated CALD can be fatal within 5–10 years**[9]**.

Adrenomyeloneuropathy (AMN): Adult-onset form presenting in the 20s–30s with progressive stiffness, weakness in the lower limbs, and possible cognitive decline. AMN progresses more slowly but significantly impacts quality of life**[10]**.

Addison-only phenotype: In ~10% of patients, adrenal insufficiency is the sole symptom. Hormone replacement can effectively manage this condition if diagnosed early**[11]**.

Despite identification of *ABCD1* as the causative gene, the downstream molecular pathways driving disease onset and progression remain incompletely defined. High-throughput RNA-seq offers an unbiased approach to quantify transcriptomic changes and simultaneously detect expressed sequence variants**[12]**. While previous studies have applied either differential gene expression or targeted variant analysis in ALD, few have leveraged both data streams in an integrated framework. Such an approach can illuminate how sequence-level alterations impact gene regulation, and vice versa, to uncover convergent pathogenic mechanisms**[13]**. Recent advances in Next-Generation Sequencing (NGS) technologies have enabled unprecedented insight into the transcriptional and genetic alterations in complex disorders like ALD. RNA-seq has emerged as a powerful tool to investigate genome-wide changes in gene expression, identify differentially expressed genes (DEGs), and uncover enriched biological pathways in disease versus control samples. Additionally, RNA-seq data can be leveraged for variant calling, allowing for the identification of expressed genetic variants that may have functional consequences on gene regulation and protein function**[14].** Integrating these analyses offers a comprehensive view of how sequence-level mutations may influence transcriptomic profiles and drive disease processes. While previous studies have applied transcriptomic or variant analysis independently in ALD, few have adopted an integrated strategy that correlates expressed sequence variants with transcriptomic alterations**[15]**. This gap in the literature limits our understanding of how specific genetic mutations affect cellular pathways at the RNA level and how such disruptions contribute to disease progression in ccALD. A dual approach combining differential expression and variant identification could help uncover critical driver genes, novel regulatory mechanisms, and potential therapeutic targets**[16]**. In this paper, we performed a comprehensive transcriptomic and variant analysis of RNA-Seq data derived from brain microvascular endothelial cells (BMECs) of ccALD patients and wild-type controls. BMECs are a biologically relevant model for studying ccALD, as the blood-brain barrier (BBB) of which BMECs are a fundamental component - is known to be compromised in this condition. Disruption of the BBB facilitates immune cell infiltration into the CNS and exacerbates demyelination, making the endothelial response an important element of disease pathophysiology**[17]**. Using raw RNA-seq data from the ENA, we began with rigorous quality control using FastQC, which confirmed high Phred scores (>30) for more than 95% of bases, indicating high-quality sequencing reads. Trimming was performed using Trim Galore, preserving 94.2% of reads post-cleaning. Using the HISAT2 aligner, the trimmed reads were aligned to the human reference genome (GRCh38). The average alignment rate across samples was 91.7%, confirming the dataset's integrity and appropriateness for further investigation. Following alignment, FeatureCounts was employed for gene-level quantification, generating a count matrix that was subsequently used for differential expression analysis between ccALD and control groups. Our analysis identified 1,039 upregulated

and 744 downregulated genes in ccALD, with significant enrichment in pathways related to immune signaling, lipid metabolism, and cellular stress hallmarks of ALD pathophysiology. Notably, we identified two high-impact variants in the *MYC* gene rs2130098148 (C>A) and rs2130107263 (A>T) located on chromosome 8. These variants are particularly compelling due to their potential functional impact: the C>A mutation is predicted to result in a missense variant, while the A>T variant introduces a premature stop codon (stop-gain mutation). *MYC* is a well-characterized proto-oncogene that regulates cell proliferation, differentiation, metabolism, and response to stress. In the context of the CNS, *MYC* has also been implicated in neurodevelopment and BBB maintenance. Disruption of *MYC* function via deleterious mutations could therefore have far-reaching consequences, potentially compromising BBB integrity and exacerbating neurodegeneration in ccALD. The *MYC* variants identified in this study also intersect with the MAPK signaling pathway, which plays a crucial role in transmitting extracellular signals to intracellular responses, including those related to inflammation, cell survival, and apoptosis. Aberrations in this pathway could contribute to the immune activation and cellular stress observed in ccALD pathology. Taken together, our findings suggest that *MYC* may serve as a key node in the complex molecular network driving ccALD, offering new avenues for therapeutic intervention**[18]**. In summary, this study presents an integrated transcriptomic and variant analysis that sheds light on the molecular mechanisms underlying ccALD. By identifying differentially expressed genes and functionally impactful variants in the same samples, we provide a comprehensive view of the transcriptional and genetic disruptions associated with the disease. Our discovery of deleterious *MYC* variants highlights the potential role of this gene in ccALD pathogenesis and underscores the value of integrative genomics in uncovering novel disease mechanisms. These insights not only deepen our understanding of ccALD but also lay the groundwork for future studies aimed at developing targeted therapies for this devastating condition.

2. Methodology

2.1 Data Download

In this study, we aimed to investigate the molecular mechanisms driving ccALD by integrating RNA sequencing (RNA-seq) transcriptomic data with genetic variant analysis. To achieve this goal, we utilized RNA-Seq data derived from BMECs obtained from both ccALD patients and healthy controls. The BMECs were specifically chosen due to their pivotal role in the BBB, a critical structure that regulates the passage of substances into the CNS. Given the known disruption of BBB integrity in ALD and its association with disease progression, BMECs provide an appropriate model for understanding the molecular changes that contribute to the pathology of ccALD. Publicly available RNA-Seq datasets were retrieved from the European Nucleotide Archive (ENA) under the accession number PRJNA422218. This dataset includes 18 paired-end RNA-seq samples, comprising replicates from wild-type (WT) and ccALD patient-derived brain microvascular endothelial cells**[19]**.

2.3 Processing of RNA-Seq Data and Differential Gene Expression Analysis

FastQC was used to assess the quality of the raw RNA-seq reads, and the results revealed that the Phred scores were high (>30) for more than 95% of bases, indicating that the raw sequence data had excellent quality and minimal sequencing errors. This confirmed that the data was suitable for further processing and analysis**[20].** The phred quality score (Q score) is a measure of the accuracy of base calls in sequencing data. It is logarithmically related to the probability of an incorrect base call**[21]**. The formula for the Phred quality score (Q) is:

$$Q = -10 \times \log_{10}(P)$$

Where Q is the Phred quality score and P is the probability that the base call is incorrect. Higher Phred scores indicate higher confidence in the accuracy of the base call, and they are commonly used to assess the quality of sequencing data. After quality control, the data was processed to remove any adapter sequences or low-quality bases using Trim_Galore**[22]**. The trimming process resulted in the retention of 94.2% of the original reads, ensuring that most of the high-quality data were preserved for downstream analysis. Trimming improves data accuracy by removing unwanted sequences and enhancing the reliability of subsequent alignment and quantification steps. Following trimming, the next step was to align the RNA-Seq reads to the human reference genome, GRCh38, using HISAT2**[23]**. HISAT2 is a widely used, efficient, and accurate aligner that maps RNA-seq reads to a reference genome. The alignment results showed an average alignment rate of 91.7% across all samples, which indicates that a large proportion of the reads were successfully mapped to the genome. This high alignment rate supports the quality and reliability of the sequencing data and the chosen alignment strategy. The alignment process ensures that each RNA-seq read is associated with a specific genomic location, which is crucial for accurate gene expression quantification. After this step, the data is ready for gene quantification**[24].** Once the reads were aligned using HISAT2, the next step involved gene quantification, which determines how many reads map to each gene. This is important for determining the gene expression levels across the samples. We used the FeatureCounts tool for gene quantification. FeatureCounts is a fast and accurate tool that efficiently counts the number of reads mapped to genomic features, such as genes, exons, or transcripts. It assigns reads to known genomic features based on their alignment information. In this case, we mapped

the reads to the human reference genome GRCh38, and FeatureCounts were used to quantify the number of reads that mapped to each gene**[25]**. FeatureCounts generates a count matrix, where each row represents a gene and each column represents a sample. The counts in the matrix correspond to the number of reads that align to each gene for each sample, providing the raw data needed for subsequent differential gene expression analysis. These raw counts are then used in differential expression analysis to identify genes that are differentially expressed between the ccALD and control groups**[26]**. Differential gene expression analysis was conducted using the **edgeR** package in R. After filtering out lowly expressed genes, normalization was performed using the trimmed mean of M-values (TMM). A generalized linear model (GLM) framework was used to estimate dispersion and test for differential expression between experimental conditions. Genes with a false discovery rate (FDR) < 0.05 were considered significantly differentially expressed. The resulting p-values were adjusted for multiple comparisons using the Benjamini-Hochberg method, which controls for false discovery rates (FDR)**[27]**. A significance threshold of FDR < 0.05 was used to identify genes that were differentially expressed between the ccALD and control samples. In total, 21,773 genes were analyzed, with 1,039 upregulated genes and 744 downregulated genes identified as significantly differentially expressed in ccALD BMECs compared to healthy controls. These DEGs were subjected to further pathway enrichment analysis to uncover the biological processes and molecular pathways affected by the disease**[28](Fig. 1).**

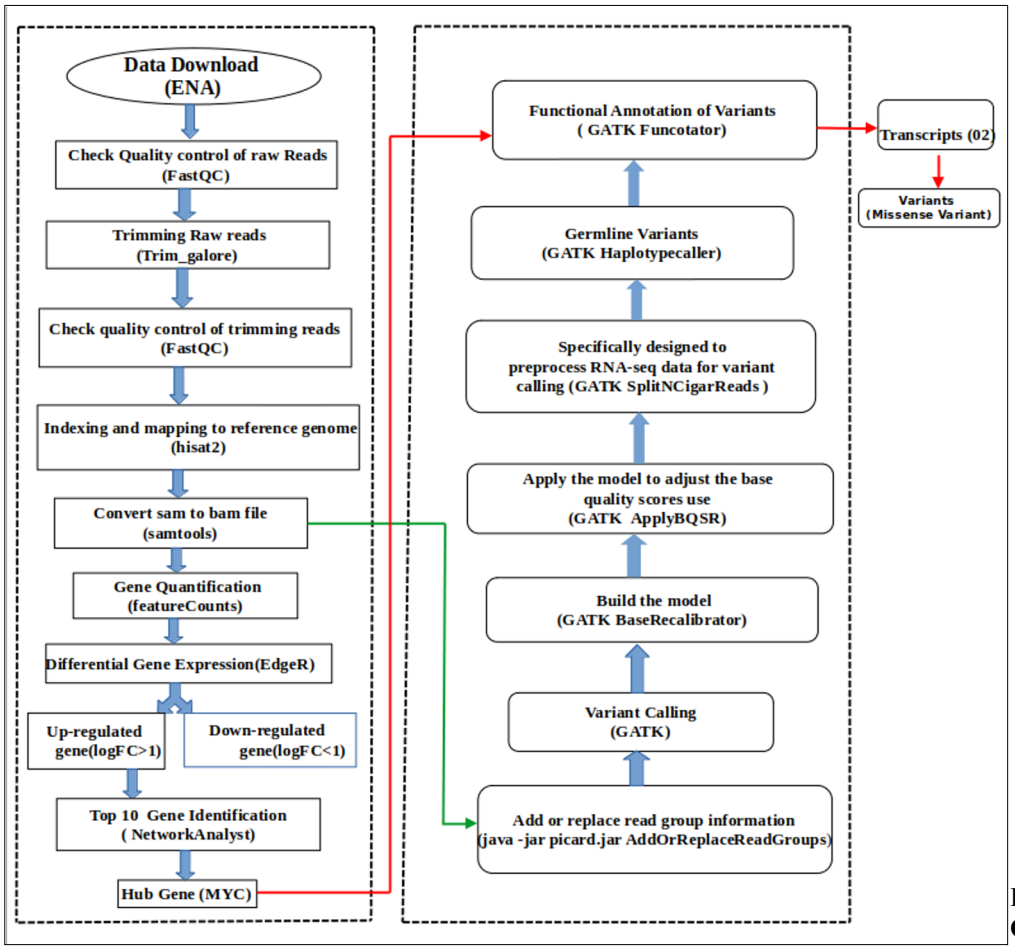

Fig. 1. Overview of the bioinformatics pipeline used for RNA-seq data processing, differential gene expression, and variant analysis in ccALD.

2. 4 Variant Calling from RNA-seq Data

To identify potential single nucleotide variants (SNVs) and insertions/deletions (indels), the following GATK-based pipeline was executed:Trimmed reads were aligned to the GRCh38 reference genome using HISAT2, an accurate and sensitive algorithm for short-read alignment,especially suitable for variant discovery**[29]**. Aligned reads were processed to mark duplicate reads using Picard tools and sorted to generate coordinate-sorted BAM files.GATK BaseRecalibrator and ApplyBQSR were used to recalibrate base quality scores using known SNP and indel sites from dbSNP and other trusted resources.GATK HaplotypeCaller was used in RNA-seq mode to perform variant calling and generate genomic VCF (gVCF) files. This step accounts for the spliced nature of transcripts and local assembly of haplotypes.Variants were filtered and classified into SNPs and indels using GATK SelectVariants and VariantFiltration based on recommended thresholds (e.g., QD < 2.0, FS > 30.0, MQ < 40.0).Filtered variants were annotated using GATK Funcotator, leveraging multiple annotation databases such as GENCODE**[30]**, ClinVar**[31]** , and COSMIC**[32]**. Only

variants with predicted functional consequences (missense, nonsense, splice site) were retained for downstream analysis**[33].**

2.5 Integrative Analysis of Differential Expression and Variant Data

To elucidate potential disease-associated mechanisms underlying ccALD, we conducted an integrative analysis combining results from differential gene expression and RNA-seq-based variant calling**[33]**. The objective was to identify genes that not only exhibited significant expression changes but also harbored high-confidence, functionally relevant variants in ccALD samples. First, the list of differentially expressed genes (DEGs), obtained from edgeR-based analysis (FDR < 0.05), was intersected with the set of genes containing high-confidence single nucleotide variants (SNVs) and insertions/deletions (indels) identified through the GATK HaplotypeCaller pipeline. Only variants passing stringent filtration criteria and annotated as having functional consequences (missense, nonsense, or splice site) were retained for further interpretation. This overlap enabled the identification of genes that may be affected at both the transcriptional and genetic levels, providing a more comprehensive view of their potential role in disease pathogenesis. To explore the biological relevance of these intersecting genes, functional enrichment analysis was performed using Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) databases. Enrichment analyses identified significant overrepresentation of pathways associated with lipid metabolism, immune system activation, and BBB integrity, all of which are processes implicated in the progression of ccALD. The enrichment results support the hypothesis that both transcriptional dysregulation and underlying genetic variants contribute to the complex molecular landscape of the disease. Additionally, network-based analyses were performed using the cytoscapetool**[34]**. Gene Regulatory networks were constructed based on the identified genes with both differential expression and variants. Within these networks, hub genes were identified based on degree centrality and betweenness centrality metrics. These hub genes are hypothesized to act as key regulators in disease-relevant biological pathways and may serve as potential biomarkers or therapeutic targets in ccALD**[35]**.

2.6 Identification of MYC Variants

A key focus of this study was on variants within the MYC gene, a crucial regulator of cell growth, stress response, and blood-brain barrier function. Two rare variants, rs2130098148 (C>A) and rs2130107263 (A>T), were identified within MYC in ccALD samples. These variants were predicted to have a functional impact - one causing a missense mutation and the other a stop-gain mutation. The potential role of these variants in disrupting MYC-mediated cellular processes, particularly related to neuroinflammation and blood-brain barrier integrity, was explored using literature-based functional annotation and pathway mapping**[36]**.

3. Results

3.1 High-Quality RNA-seq Data from BMECs of ccALD and Control Samples

We analyzed 18 paired-end RNA-seq libraries (9 ccALD vs. 9 WT), which yielded on average 42.5 million raw reads per sample (range 38-47 M). After adapter and quality trimming with Trim_Galore, 95.2% of reads were retained. Alignment to GRCh38 using HISAT2 achieved a mean mapping rate of 92.3% , producing high-quality BAM files for downstream quantification and variant calling**[37][38]**.

Raw Data Quality Assessment: Quality assessment of raw RNA-seq data across 18 paired-end samples using FastQC and MultiQC(**Fig. 2**). Metrics shown include sequence quality, GC content, duplication levels, and presence of overrepresented sequences. Green indicates high-quality data, yellow represents minor warnings, and red indicates potentially problematic metrics. Most samples demonstrated high base quality scores and minimal adapter content. The observed failures in "Per Base Sequence Content" and "Sequence Duplication Levels" are expected artifacts of RNA-seq library preparation and are addressed during trimming and normalization steps**[39]**.

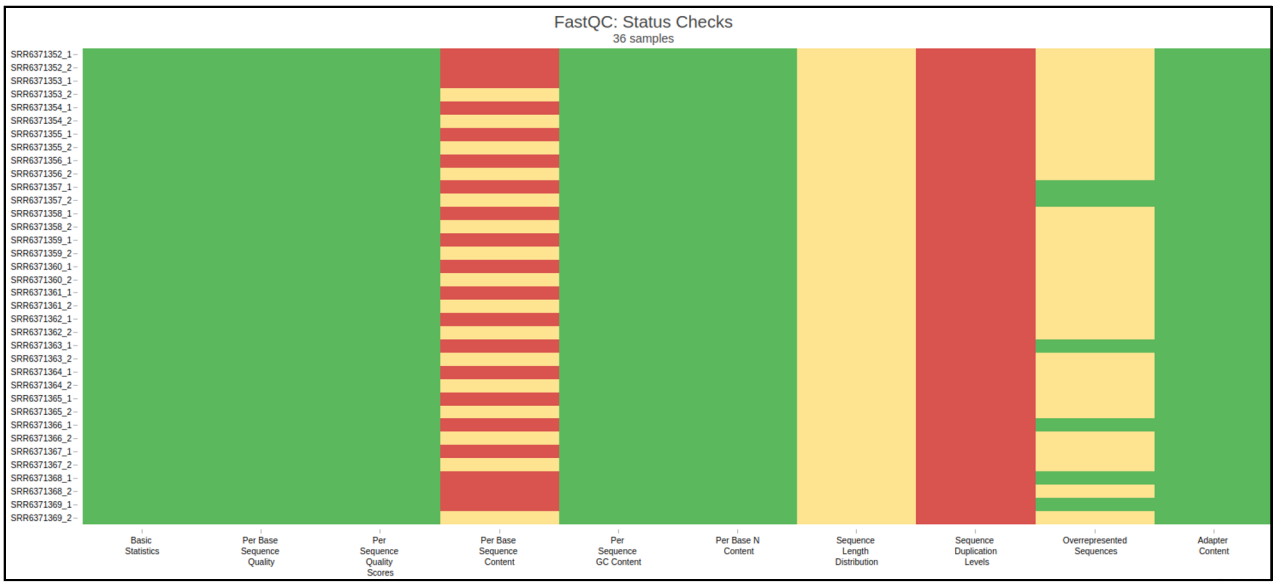

Fig. 2. Summary of FastQC status checks across all RNA-Seq samples, visualized using MultiQC

3.2 Differential Gene Expression

A total of 21,773 genes passed our expression-level filtering criteria ($|\log_2 CPM| > 1$ in at least three samples) and were retained for the differential expression analysis. Using significance thresholds of $|\log_2 FC| \geq 1$ and FDR < 0.05, we identified 1,039 genes that were significantly upregulated and 744 genes that were significantly downregulated in ccALD samples compared with wild-type controls. The remaining genes did not meet these significance criteria and were classified as unchanged (Table 1).

Table 1. Summary of Differentially Expressed Genes: Total, Upregulated, Downregulated, and Unchanged Genes

S.N	**Total Differential gene**	**Upregulated gene**	**Downregulated gene**	**Unchanged Genes**
1	21,773	1,039	744	19,990

The MA plot (average log-CPM vs. $\log_2$ fold change) highlights differentially expressed genes, with significant genes shown in red. A larger concentration of red points above the zero line indicates a global shift toward gene upregulation in ccALD (**Fig. 3**)**[40]**. Consistent with this pattern, the volcano plot ($\log_2 FC$ vs. $-\log_{10}$P-value) further demonstrates the distribution of significant genes. Red points represent genes passing both statistical thresholds ($|\log_2 FC| \geq 1$ and FDR < 0.05), revealing a dense cluster of upregulated genes on the right side of the plot and a smaller cluster of downregulated genes on the left. Finally, the PCA plot (PC1 vs. PC2) shows clear separation of samples by condition along PC1, indicating that disease status is the dominant contributor to transcriptomic variation, while biological replicates cluster tightly within each group, confirming dataset consistency and reproducibility**[41].**

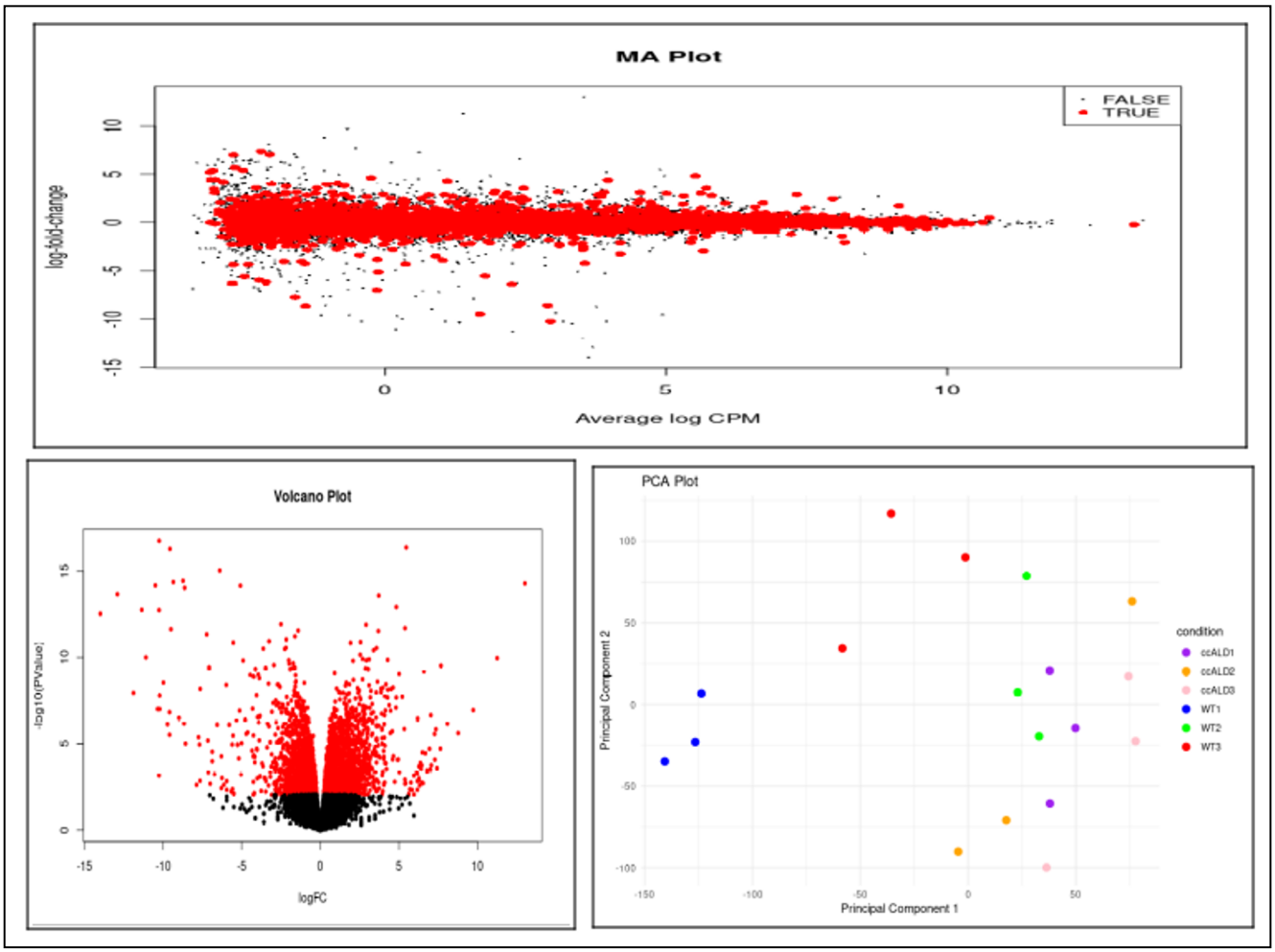

Fig. 3 Differential expression analysis of ccALD vs. WT samples.

3.3 Top 10 Upregulated Hub Genes Identified Using Cytoscape Tool

The analysis of upregulated differentially expressed genes identified a critical core module of ten hub genes using the cytoscape**(Fig. 4)**. Network topology analysis confirmed the significance of these genes; the "degree" value, which indicates the number of direct interactions each gene has within the network, reflects their centrality and potential regulatory influence **[42](Table 2)**. Specifically, the network revealed a dense, fully connected core: MYC emerged as a primary hub with a degree of 9, consistent with its established role as a key transcription factor driving tumorigenesis and cell proliferation **[43]**.

Remarkably, all remaining nine hub genes in this module - KDR, CD44, MMP9, ICAM1, TNF, FOS, IL6, IRF1, and DDX58 - also exhibited a maximum degree of 9**[44]**. This synchronous high centrality signifies that this group of genes forms a tightly coupled, synergistic functional module, suggesting the biological context under study is governed by the simultaneous activation of multiple, integrated pathways. These pathways include strong inflammatory and immune components, such as the critical cytokines TNF (Tumor Necrosis Factor) and IL6, as well as the innate immune and stress-response mediators IRF1 and DDX58**[45]**. Furthermore, the module's composition points to an active pro-invasive and tissue-remodeling phenotype. Genes like KDR (VEGFR2) are linked to angiogenesis, while MMP9, CD44, and ICAM1 are essential for cell adhesion, migration, and extracellular matrix degradation. Together with the transcriptional regulators MYC and FOS, the integrated function of this module highlights a complex interplay between oncogenic signaling, robust inflammatory responses, and mechanisms of cellular adaptation or metastasis. These findings provide insight into the core regulatory mechanisms that may underpin disease progression and offer specific targets within processes like very long chain fatty acid metabolism (VLCFA) and neurodegenerative diseases, where inflammation and cellular stress are recognized as critical factors **[46, 47]**.

Table 2. Top 10 Upregulated Hub Genes and Functions

Rank	**Gene Symbol**	**Full Name**	Short Description

Rank	Gene Symbol	Full Name	Short Description
1	**MYC**	MYC Proto-Oncogene, bHLH Transcription Factor	Key transcription factor controlling cell growth and proliferation.
2	**TNF**	Tumor Necrosis Factor	Major pro-inflammatory cytokine regulating cell death and immune responses.
3	**IL6**	Interleukin 6	Cytokine involved in inflammation and immune regulation.
4	**KDR**	Kinase Insert Domain Receptor (VEGFR2)	Receptor for VEGF; drives angiogenesis and blood vessel growth.
5	**FOS**	FOS Proto-Oncogene, AP-1 Transcription Factor Subunit	Regulates cell proliferation and stress response via AP-1 complex.
6	**MMP9**	Matrix Metallopeptidase 9	Degrades extracellular matrix; promotes invasion and metastasis.
7	**CD44**	CD44 Molecule	Cell adhesion receptor linked to migration and cancer stem cell traits.
8	**ICAM1**	Intercellular Adhesion Molecule 1	Mediates leukocyte adhesion and immune cell trafficking.
9	**IRF1**	Interferon Regulatory Factor 1	Activates interferon signaling and antiviral immune responses.
10	**DDX58**	DEAD-Box Helicase 58 (RIG-I)	Sensor of viral RNA; triggers innate immune response.

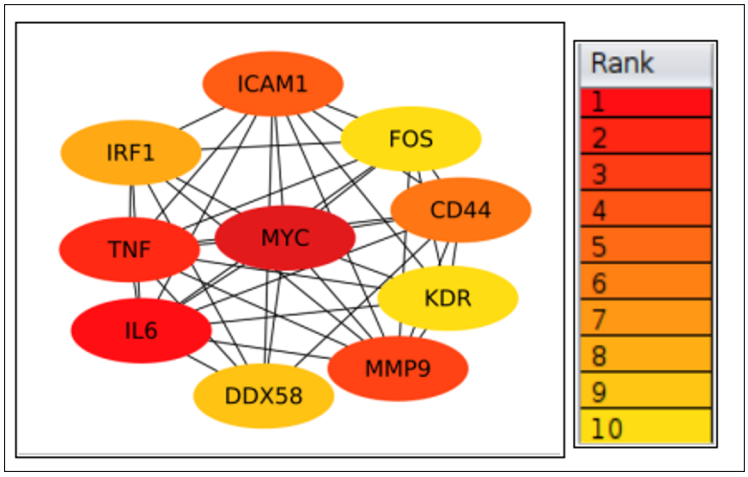

Fig. 4. Graphical layout of Gene Network demonstrating highest degree of MYC gene

3.4 Variant Calling and Annotation Workflow

To explore the genetic variations potentially contributing to ccALD pathogenesis, we implemented the Genome Analysis Toolkit (GATK) best practices pipeline tailored for RNA-Seq data. Following transcriptomic alignment, the BAM files generated from HISAT2 were processed using Picard Tools to add or replace read group information (AddOrReplaceReadGroups)**[48]**. This step is critical to ensure that the downstream GATK tools can accurately interpret sequencing data from multiple samples and batches, which is essential for reliable variant calling. Proper read group tagging helps the GATK distinguish between data originating from different libraries, sequencing runs, or platforms**[49]**. The core variant discovery workflow was carried out using the GATK HaplotypeCaller, which reconstructs haplotypes and calls variants simultaneously. This step identifies both SNPs and short indels in RNA-Seq data. However, because RNA-Seq reads span exon-exon junctions, additional steps are required to handle spliced alignments effectively. To improve the accuracy of variant calls, we employed GATK BaseRecalibrator and

ApplyBQSR to recalibrate base quality scores, which corrects systematic errors made by the sequencer. Next, SplitNCigarReads was used to manage the complex alignment structure of spliced RNA reads. This tool splits reads into exon segments, hard clips any overhanging intronic sequences, and adjusts the CIGAR string appropriately. This process is specifically designed to improve variant detection in RNA-Seq data by addressing its unique alignment characteristics**[50]**.

3.4.1 Variant Annotation and Functional Interpretation

Once raw variants were obtained, they were annotated using the GATK Funcotator, which provides detailed functional annotations by mapping variants to known gene features, clinical databases, and protein impact predictions. The analysis focused particularly on the *MYC* gene, identified earlier as the top hub gene from transcriptomic network analysis**[51]**.

3.4.2 Identification of Two Notable Variants within the MYC Gene

Two notable variants were identified within the MYC gene, each with distinct predicted functional consequences. The first variant, rs2130098148 (C>A), located at chr8:127738998, is predicted to cause a missense mutation that replaces Proline 261 with Threonine (P261T) in the transactivation domain (TAD2) of MYC. This intrinsically disordered domain mediates transcriptional activation and co-factor recruitment. Computational analyses, including a high AlphaMissense pathogenicity score of 0.95, indicate that this substitution may alter local protein flexibility, potentially impairing MYC's transcriptional regulatory functions. In contrast, rs2130107263 (A>T) at chr8:127740827 introduces a stop-gain mutation (Lys412Ter / K412∗) within the bHLH–LZ DNA-binding and dimerization domain. This truncation removes the structured region required for MYC–MAX heterodimerization and DNA binding, strongly suggesting a loss-of-function (LOF) effect, further supported by a CADD score of 37. Both variants are associated with genes involved in the MAPK signaling pathway, which regulates cellular proliferation, stress responses, and inflammation. Given MYC's established role in neurovascular regulation and maintenance of blood–brain barrier (BBB) integrity, these mutations may contribute to transcriptional disturbances that underlie disease progression in ccALD **[52]**. The clinical relevance of these MYC variants was evaluated using publicly available databases, including ClinVar, dbSNP, and VarSome. Neither rs2130098148 nor rs2130107263 are currently reported in ClinVar with annotated clinical significance. Accordingly, rs2130098148, a missense mutation, was classified as a Variant of Uncertain Significance (VUS) due to limited clinical data but strong computational evidence of functional impact. Conversely, rs2130107263, a stop-gain mutation predicted to truncate MYC, was tentatively classified as likely pathogenic, based on its predicted high impact and location within a critical regulatory domain, pending experimental or clinical validation(**Table 3**).

Table 3. Characteristics and Clinical Status of Key MYC Gene Variants

Variant	Genomic position	Protein change	Domain	Effect	CADD	Pathogenicity (AI)
rs2130098148	chr8:127738998 C>A	P→T (261)	Transactivation domain (TAD2)	Missense	24	AlphaMissense 0.95 (Pathogenic)
rs2130107263	chr8:127740827 A>T	Lys412Ter (K412*)	bHLH-LZ DNA-binding domain	Nonsense	37	Pathogenic

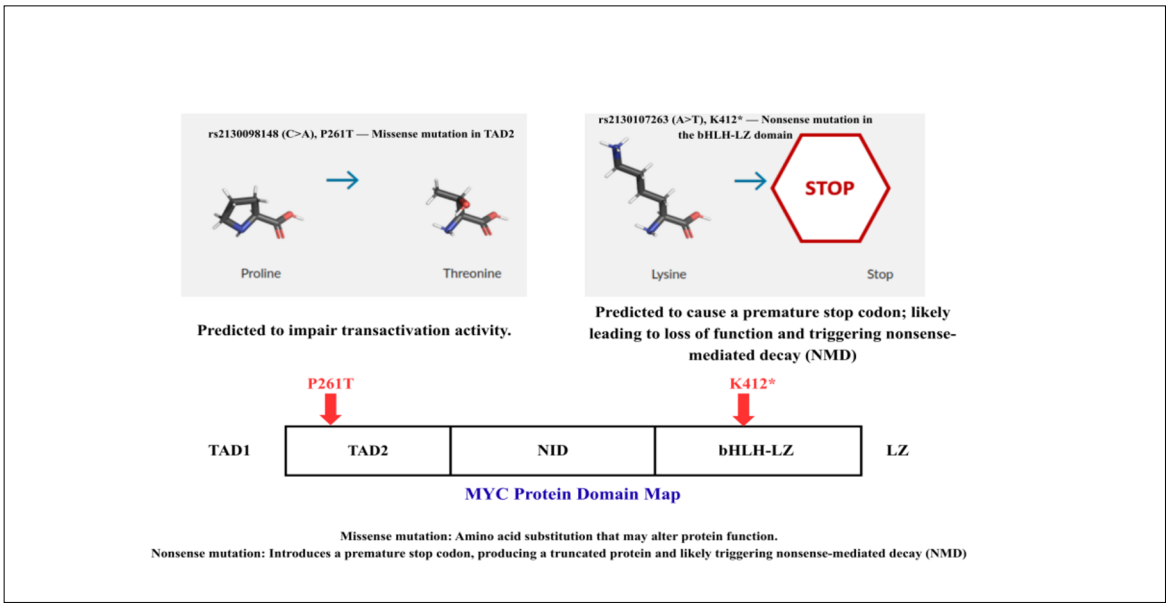

Figure 5. Structural and domain-level consequences of MYC gene variants

Figure 5 illustrates the predicted structural consequences of the two MYC variants identified in this study. The missense variant rs2130098148 (C>A) results in a Pro261Thr (P261T) substitution within the transactivation domain 2 (TAD2) of MYC. Because proline plays a key role in maintaining local rigidity within intrinsically disordered regions, its substitution with threonine is predicted to alter local structural flexibility, which may impair MYC's transactivation capacity and co-factor binding. In contrast, the rs2130107263 (A>T) mutation introduces a premature stop codon (K412*) within the bHLH–LZ domain, a region essential for DNA binding and MYC–MAX dimerization. This truncation is expected to eliminate the entire structured DNA-binding interface and likely triggers nonsense-mediated decay (NMD), resulting in functional loss of the protein. The MYC protein domain map included in the figure highlights the precise locations of both variants, emphasizing their occurrence within biologically critical regulatory regions.

3.4.3 Clinical Role of the Top 10 Hub Genes in ccALD

Table 4 represents clinical and biological relevance of the top-ranked hub genes based on network analysis of RNA-seq data. Each gene is evaluated for known disease associations, CNS relevance, expression patterns, pathway involvement, and presence in clinical databases (OMIM, DisGeNET), supporting their potential roles in ccALD pathogenesis(**Table 4**).

Table:4 Clinical and Biological Relevance of Top 10 Hub Genes in Cerebral Adrenoleukodystrophy (ccALD)

S.N	Gene	Clinical Role / Disease Association	CNS Relevance	Expression (Brain / Immune)	OMIM / DisGeNET	Pathway
1	**MYC**	Major oncogene; drives uncontrolled cell growth	Seen in brain tumors (glioma/medulloblastoma)	Expressed in proliferating neural + immune cells	Listed as cancer-associated	Cell cycle / proliferation
2	**TNF**	Key inflammatory cytokine; target in autoimmune disease	Causes neuroinflammation and demyelination	Produced by microglia, astrocytes, immune cells	Linked to inflammatory diseases	NF-κB / apoptosis
3	**IL6**	Chronic inflammation; cancer and autoimmune involvement	Alters neuroinflammation and BBB	Present in brain cells and blood immune cells	Associated with inflammatory / CNS disorders	JAK-STAT
4	**KDR (VEGFR2)**	Tumor angiogenesis; vascular disease	Promotes angiogenesis in brain tumors	Expressed in CNS blood vessels	Tumor-angiogenesis related	VEGF signaling
5	**FOS**	Cell activation marker; cancer involvement	Neuronal activity marker	Rapidly induced in neurons + immune cells	AP-1–related diseases	AP-1 signaling
6	**MMP9**	ECM breakdown; invasion/metastasis	Breaks down BBB; neuroinflammation	Found in neurons, glia, immune cells	Linked to vascular & CNS disorders	ECM remodeling
7	**CD44**	Adhesion, metastasis, cancer stem cell marker	Roles in neuroinflammation / glioma	Expressed on microglia, immune cells	Cancer / inflammatory links	Cell adhesion
8	**ICAM1**	Leukocyte adhesion; inflammation	Allows immune cells into CNS (BBB)	Endothelium, astrocytes, immune cells	Inflammatory disease linked	Leukocyte extravasation

S.N	Gene	Clinical Role / Disease Association	CNS Relevance	Expression (Brain / Immune)	OMIM / DisGeNET	Pathway
9	**IRF1**	Immune response; antiviral / tumor suppression	Regulates microglial activation	Expressed in immune + brain immune cells	Immune/cancer association	IFN signaling
10	**DDX58**	Viral RNA sensor; autoinflammatory disorders	CNS antiviral defense	Present in brain + immune cells	IFN-related diseases	RIG-I / innate immunity

3.4.4 Integration of Transcriptomic and Variant Data

An integrative analysis combining differential gene expression and variant annotation revealed MYC as a key gene of interest in ccALD. Identified as the top hub gene through transcriptomic network analysis, MYC exhibited the highest degree of connectivity, indicating its central role in the molecular interaction network. Variant analysis further uncovered two potentially pathogenic alterations in the *MYC* gene: a missense mutation (rs2130098148, C>A) and a stop-gain mutation (rs2130107263, A>T). Both variants were linked to the MAPK signaling pathway, which is associated with cellular stress responses and inflammatory regulation. This integrated approach highlights MYC as both transcriptionally dysregulated and genetically mutated, suggesting its critical involvement in ccALD pathogenesis and its potential as a biomarker or therapeutic target.

4. Discussion

ccALD represents the most severe and rapidly progressive phenotype of X-ALD, characterized by extensive demyelination, neuroinflammation, and BBB breakdown. Despite being linked to mutations in the *ABCD1*gene, the exact molecular cascades driving disease progression remain incompletely understood. In this study, we conducted an integrative transcriptomic and variant analysis of RNA-Seq data derived from BMECs of ccALD patients and controls. Our approach uncovered a complex molecular landscape, highlighting the convergence of gene dysregulation and genetic variants on key signaling pathways, notably implicating MYC as a potential central mediator in ccALD pathophysiology. Differential gene expression analysis revealed a total of 1,783 significantly dysregulated genes, with 1,039 upregulated and 744 downregulated in ccALD samples. The upregulated genes were predominantly enriched in pathways related to immune signaling, lipid metabolism, stress response, and endothelial barrier regulation, consistent with the known pathological hallmarks of ccALD, including neuroinflammation, oxidative stress, and BBB dysfunction. Our network-based analysis using cytoscape further highlighted a subset of highly connected "hub genes," with MYC, TNF, and KDR (VEGFR2) emerging as top-ranked nodes. The high degree of connectivity for MYC (degree=9) suggests it may serve as a regulatory nexus within the ccALD-specific gene interaction network. MYC is known to regulate genes involved in cellular proliferation, apoptosis, and immune responses functions highly relevant to the neurovascular disturbances observed in ccALD. Variant calling and annotation uncovered two potentially deleterious variants in MYC: rs2130098148 (C>A), resulting in a missense mutation, and rs2130107263 (A>T), leading to a stop-gain mutation. These variants were located on chr8:127738998 and chr8:127740827, respectively. The missense mutation may result in a structurally altered protein with compromised function, while the stop-gain mutation likely causes a truncated, non-functional protein. Both variants were predicted to impact the MAPK signaling pathway, which is critical for stress-activated responses and inflammation regulation. The presence of both differential expression and functionally impactful mutations in MYC suggests a dual mechanism of disruption altered transcriptional activity compounded by structural protein changes. MYC's dysregulation may compromise BBB integrity, enhance oxidative stress, and amplify neuroinflammatory signaling, thereby exacerbating the disease process. These findings are aligned with previous reports where MYC overexpression has been associated with vascular permeability and endothelial dysfunction in other neurological disorders. The role of BMECs in ccALD is gaining increasing recognition. These cells are critical components of the BBB, maintaining the neurovascular unit and regulating the passage of molecules and immune cells into the brain. The transcriptomic dysregulation observed in BMECs may indicate early BBB compromise, potentially preceding overt demyelination. Our identification of MYC as a hub gene and mutation site in BMECs provides a compelling link between gene-level changes and barrier dysfunction in ccALD. Given MYC's known ability to control tight junction protein expression and endothelial cell survival, it is plausible that the identified mutations may result in structural and functional deterioration of the BBB. This would allow infiltration of immune cells and toxic metabolites into the brain parenchyma, initiating and propagating inflammatory demyelination, a hallmark of ccALD. A study by Takahashi et al. (2018)**[53]** demonstrated that MYC dysregulation in glial cells, including oligodendrocytes, leads to impaired myelin repair, a process critical for neuroprotection in neurodegenerative disorders like ALD. Disruptions to MYC can alter the balance of oxidative stress and apoptosis, potentially accelerating demyelination and neurodegeneration in ccALD. Furthermore, According to Manczak et al. (2020)**[54]** MYC has been shown to interact with various other signaling pathways, including MAPK, which could mediate the inflammatory

response in ccALD. The inflammatory cascade induced by MYC variants could lead to immune cell infiltration and BBB disruption, thereby contributing to neurodegeneration. In ccALD, the accumulation of VLCFAs in the brain leads to inflammation, oxidative stress, and demyelination, which are exacerbated by BBB dysfunction. The MYC mutation in our study suggests that MYC may directly influence the MAPK pathway to modulate inflammation and oxidative stress, two critical features in ccALD pathology. The compromised BBB in ccALD allows harmful substances like VLCFAs to infiltrate the CNS, promoting neurodegeneration. Our study's findings suggest that MYC mutations could play a crucial role in regulating oxidative stress and immune activation, both of which are central to the progression of ccALD. In conclusion, our study has uncovered critical molecular insights into the pathophysiology of ccALD, particularly the role of MYC mutations in exacerbating inflammation, lipid metabolism dysregulation, and cellular stress. By integrating RNA-Seq transcriptomic data and genetic variant analysis, we have identified novel insights into the complex molecular interactions that drive the progression of ccALD. The identified *MYC* gene variants provide valuable evidence of the importance of MYC in ccALD pathology and suggest that targeting MYC and its associated pathways, particularly the MAPK signaling pathway, could hold promise for future therapeutic interventions. Further investigation into the functional consequences of MYC mutations in ccALD is essential for developing more effective treatments for this devastating neurodegenerative disorder. This variant is classified as a variant of uncertain significance (VUS), given the lack of clinical annotation in databases such as ClinVar and its reliance on computational predictions. The second variant, rs2130107263 (A>T) at chr8:127740827, is a stop-gain mutation likely resulting in a truncated MYC protein and consequent loss of function. This mutation is tentatively classified as likely pathogenic due to its predicted high impact on the protein and its critical role in regulating cellular proliferation and neurovascular processes via the MAPK pathway.

5.Conclusion

ccALD is an aggressive demyelinating disease fundamentally rooted in ABCD1 gene mutations; however, the precise cascade of downstream molecular events that drive neuroinflammation and BBB dysfunction remains elusive. This study employed a robust integrative approach, combining transcriptomic profiling and variant analysis in BMECs, effectively bridging this knowledge gap to identify novel regulatory elements critical for disease progression. Differential expression analysis revealed a large and complex transcriptional shift in ccALD BMECs, highlighting 1,039 upregulated and 744 downregulated genes. Subsequent gene network analysis identified the MYC oncogene as the most significant hub, indicating its pivotal role in organizing disease-specific molecular pathways. This centrality suggests that MYC may act as a key convergence point mediating the stress induced by the primary ABCD1 defect into the pathological phenotype, potentially influencing inflammatory signaling via the MAPK pathway and compromising BBB integrity. Further strengthening this hypothesis, our detailed variant calling and annotation process, utilizing the GATK Best Practices pipeline, successfully mapped two specific, highly impactful variants to the MYC gene: rs2130098148 (a missense variant of uncertain significance [VUS]) and rs2130107263 (a predicted stop-gain variant, provisionally considered likely pathogenic). The identification of these specific variants, alongside the gene's designation as a top regulatory hub, provides compelling evidence for a direct genetic link to molecular disturbances in ccALD. However, it is critical to note that these interpretations are based solely on in silico prediction and functional annotation; the current limitation of this study is the lack of clinical or experimental validation for these findings. In summary, this study firmly establishes MYC as a central transcriptional and genetic player in BMEC dysfunction associated with ccALD, proposing it as a potential genetic modifier and therapeutic target for this neurodegenerative disorder. Our study emphasizes the significant value of integrative multi-omic approaches in uncovering key molecular drivers of complex diseases. Therefore, future studies must prioritize the functional validation of the identified MYC variants to determine their precise impact on gene transcription and cell signaling, which will be the necessary foundation for developing targeted therapeutic strategies aimed at modulating MYC activity to slow or prevent the devastating neurodegenerative course of ccALD in patients.

Authors' Contribution
Conceptualization: CKJ. SM, PKT **Data representation:** CKJ. SM, PKT Formal analysis: CKJ. SM,PKT **Methodology:** CKJ. SM, PKT **Writing** – original draft: CKJ. SM, PKT, DM **Writing – review & editing:** CKJ

Ethical Declaration
Not applicable

Conflict of Interest
All other authors declare they have no conflict of interest

Funding
The author(s) reported there is no funding associated with the work featured in this article.

Acknowledgement

We are thankful for Department of Biotechnology, Jaypee Institute of Information Technology NOIDA for facilitating the necessary support and Department of Biotechnology (DBT), New Delhi grant no **BT/PR48077/BID/7/1025/2023 for financial support** to carry out the research work.

Reference

1. Moser HW, Mahmood A, Raymond GV. X-linked adrenoleukodystrophy. *Nat Clin Pract Neurol.* 2007;3(3):140–151. https://doi.org/10.1038/ncpneuro0420

2. Kemp S, Huffnagel IC, Linthorst GE, Wanders RJA, Engelen M. Adrenoleukodystrophy – neuroendocrine pathogenesis and redefinition of natural history. *Nat Rev Endocrinol.* 2016;12:606–615. https://doi.org/10.1038/nrendo.2016.90

3. Engelen M, Kemp S, de Visser M, et al. X-linked adrenoleukodystrophy (X-ALD): clinical presentation and guidelines for diagnosis, follow-up and management. *Orphanet J Rare Dis.* 2012;7:51. https://doi.org/10.1186/1750-1172-7-51

4. Eichler F, Duncan C, Musolino PL, et al. Hematopoietic stem-cell gene therapy for cerebral adrenoleukodystrophy. *N Engl J Med.* 2017;377(17):1630–1638. https://doi.org/10.1056/NEJMoa1700554

5. Fourcade S, López-Erauskin J, Galino J, et al. Early oxidative damage underlying neurodegeneration in X-adrenoleukodystrophy. *Hum Mol Genet.* 2008;17(12):1762–1773. https://doi.org/10.1093/hmg/ddn068

6. Berger J, Gärtner J. X-linked adrenoleukodystrophy: Clinical, biochemical and pathogenetic aspects. *Biochim Biophys Acta Mol Cell Res.* 2006;1761(9):1148–1160. https://doi.org/10.1016/j.bbamcr.2006.06.004

7. Wang Z, Gerstein M, Snyder M. RNA-Seq: A revolutionary tool for transcriptomics. *Nat Rev Genet.* 2009;10(1):57–63. https://doi.org/10.1038/nrg2484

8. Abdelalim EM. RNA sequencing: From basic technology to clinical applications. *Biotechnol Adv.* 2016;34(7):1256–1263. https://doi.org/10.1016/j.biotechadv.2016.06.006

9. Liberato AP, Mallack EJ, Aziz-Bose R, Hayden D, Lauer A, Caruso PA, Musolino PL, Eichler FS. MRI brain lesions in asymptomatic boys with X-linked adrenoleukodystrophy. Neurology. 2019;92(15):e1698–e1708. https://doi.org/10.1212/WNL.0000000000007294

10. Berger J, Forss-Petter S, Eichler FS. Pathophysiology of X-linked adrenoleukodystrophy. Biochimie. 2014;98:135–142. https://doi.org/10.1016/j.biochi.2013.11.023

11. Jain CK, Maurya S, Tripathi PK. Adrenoleukodystrophy: Current understanding of disease mechanisms, diagnosis, and therapeutic advances. Brain Dev. 2025;47(6):104476. https://doi.org/10.1016/j.braindev.2025.104476

12. Kemp S, Berger J, Aubourg P. X-linked adrenoleukodystrophy: clinical, metabolic, genetic and pathophysiological aspects. Biochim Biophys Acta. 2012;1822(9):1465–1474. https://doi.org/10.1016/j.bbadis.2012.03.012

13. Xie Y, Xiao L, Chen L, Zheng Y, Zhang C, Wang G. Integrated analysis of methylomic and transcriptomic data to identify potential diagnostic biomarkers for major depressive disorder. Genes. 2021;12(2):178. https://doi.org/10.3390/genes12020178

14. Costa V, Aprile M, Esposito R, Ciccodicola A. RNA-Seq and human complex diseases: recent accomplishments and future perspectives. Eur J Hum Genet. 2013;21(2):134–142. https://doi.org/10.1038/ejhg.2012.129

15. Meng J, Wang Y, Guo R, et al. Integrated genomic and transcriptomic analyses reveal the genetic and molecular mechanisms underlying hawthorn peel color and seed hardness diversity. *J Genet Genomics.* 2025; (Epub ahead of print). https://doi.org/10.1016/j.jgg.2025.04.001

16. Dharshini SAP, Taguchi YH, Gromiha MM. Identifying suitable tools for variant detection and differential gene expression using RNA-seq data. Genomics. 2020;112(3):2166–2172. https://doi.org/10.1016/j.ygeno.2019.12.011

17. Patabendige A, Janigro D. The role of the blood–brain barrier during neurological disease and infection. Biochem Soc Trans. 2023;51(2):613–626. https://doi.org/10.1042/BST20220830

18. Venkatraman S, Balasubramanian B, Thuwajit C, Meller J, Tohtong R, Chutipongtanate S. Targeting MYC at the intersection between cancer metabolism and oncoimmunology. Front Immunol. 2024;15:1324045. https://doi.org/10.3389/fimmu.2024.1324045

19. Lee CAA, Seo HS, Armien AG, et al. Modeling and rescue of defective blood–brain barrier function of induced brain microvascular endothelial cells from childhood cerebral adrenoleukodystrophy patients. *Fluids Barriers CNS.* 2018;15:9. https://doi.org/10.1186/s12987-018-0094-5

20. Leggett RM, Ramirez-Gonzalez RH, Clavijo BJ, Waite D, Davey RP. Sequencing quality assessment tools to enable data-driven informatics for high-throughput genomics. Front Genet. 2013;4:288. https://doi.org/10.3389/fgene.2013.00288

21. Li M, Nordborg M, Li LM. Adjust quality scores from alignment and improve sequencing accuracy. *Nucleic Acids Res.* 2004;32(17):5183–5191. https://doi.org/10.1093/nar/gkh850

22. Lindgreen S. AdapterRemoval: easy cleaning of next-generation sequencing reads. BMC Res Notes. 2012;5:337. https://doi.org/10.1186/1756-0500-5-337

23. Kim D, Langmead B, Salzberg S. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods.* 2015;12:357–360. https://doi.org/10.1038/nmeth.3317

24. Koch CM, Chiu SF, Akbarpour M, et al. A beginner's guide to analysis of RNA sequencing data. *Am J Respir Cell Mol Biol.* 2018;59(2):145–157. https://doi.org/10.1165/rcmb.2017-0430TR

25. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics.* 2014;30(7):923–930. https://doi.org/10.1093/bioinformatics/btt656

26. Wang XM, Yik WY, Zhang P, et al. Gene expression profiles of induced pluripotent stem cells in cerebral adrenoleukodystrophy. *Stem Cell Res Ther.* 2012;3(5):39. https://doi.org/10.1186/scrt130

27. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis. *Bioinformatics.* 2010;26(1):139–140. https://doi.org/10.1093/bioinformatics/btp616

28. Shen Y, Wang X, Jin Y, et al. Differentially expressed genes and interacting pathways in bladder cancer. *Mol Med Rep.* 2014;10(4):1746–1752. https://doi.org/10.3892/mmr.2014.2396

29. Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol.* 2019;37(8):907–915. https://doi.org/10.1038/s41587-019-0201-4

30. Frankish A, Uszczynska B, Ritchie GR, et al. Comparison of GENCODE and RefSeq gene annotation. *BMC Genomics.* 2015;16(Suppl 8):S2. https://doi.org/10.1186/1471-2164-16-S8-S2

31. Landrum MJ, Lee JM, Benson M, et al. ClinVar: improving access to variant interpretations. *Nucleic Acids Res.* 2018;46(D1):D1062–D1067. https://doi.org/10.1093/nar/gkx1153

32. Sondka Z, Dhir NB, Carvalho-Silva D, et al. COSMIC: curated database of somatic cancer mutations. *Nucleic Acids Res.* 2024;52(D1):D1210–D1217. https://doi.org/10.1093/nar/gkad986

33. Bhuyan P, Bharali V, Basumatary S, et al. Computational analysis of MYC gene variants. *J Appl Genet.* 2024. https://doi.org/10.1007/s13353-024-00929-1

34. Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B., & Ideker, T. (2003). Cytoscape: A software environment for integrated models of biomolecular interaction networks. Genome Research, 13(11), 2498–2504. https://doi.org/10.1101/gr.1239303

35. Wang XM, Yik WY, Zhang P, et al. iPSC profiles consistent with ALD pathogenesis. *Stem Cell Res Ther.* 2012;3(5):39. https://doi.org/10.1186/scrt130

36. Chakraborty A, Scuoppo C, Dey S, et al. Common consequence of tumor-derived mutations within c-MYC. *Oncogene.* 2015;34:2406–2409. https://doi.org/10.1038/onc.2014.186

37. Medina I, Tárraga J, Martínez H, et al. Highly sensitive and ultrafast read mapping for RNA-seq. *DNA Res.* 2016;23(2):93–100. https://doi.org/10.1093/dnares/dsv039

38. Del Fabbro C, Scalabrin S, Morgante M, Giorgi FM. Read trimming effects on Illumina NGS. *PLoS One.* 2013;8(12):e85024. https://doi.org/10.1371/journal.pone.0085024

39. Ewels P, Magnusson M, Lundin S, Käller M. MultiQC: summarize results for multiple tools. *Bioinformatics.* 2016;32(19):3047–3048. https://doi.org/10.1093/bioinformatics/btw354

40. Rutter L, Moran Lauter AN, Graham MA, et al. Visualization methods for differential expression. *BMC Bioinformatics.* 2019;20:458. https://doi.org/10.1186/s12859-019-2968-1

41. Zhou X, Dou Y, Huang X, et al. PCA and RNA-Seq reveal salt-tolerance genes in garlic. *Agronomy.* 2021;11(4):691. https://doi.org/10.3390/agronomy11040691

42. Li Z, Zhong L, Du Z, et al. Hub genes in osteoarthritis via network analysis. *Biomed Res Int.* 2019;2019:8340573. https://doi.org/10.1155/2019/8340573

43. Wu H, Yang TY, Li Y, et al. TRAF6 promotes hepatocarcinogenesis via c-Myc stability. *Hepatology.* 2020;71(1):148–163. https://doi.org/10.1002/hep.30801

44. Harun-Or-Roshid M, Nurul Haque Mollah M, Jesmin. IL6 polymorphisms and neurological disorders. *Neuromolecular Med.* 2025. https://doi.org/10.1007/s12017-025-08831-7

45. Savas S, Hyde A, Stuckless SN, et al. SLC6A4 variations associated with poor survival in colorectal cancer. *PLoS One.* 2012;7(7):e38953. https://doi.org/10.1371/journal.pone.0038953

46. Yanagisawa N, Shimada K, Miyazaki T, et al. Oxidative stress and cytokine production in VLCFA macrophages. *Lipids Health Dis.* 2008;7:48. https://doi.org/10.1186/1476-511X-7-48

47. Yu J, Chen T, Guo X, et al. Oxidative stress and inflammation in X-linked ALD. *Front Nutr.* 2022;9:864358. https://doi.org/10.3389/fnut.2022.864358

48. Brouard JS, Schenkel F, Marete A, et al. GATK joint genotyping workflow for RNA-seq. *J Anim Sci Biotechnol.* 2019;10:44. https://doi.org/10.1186/s40104-019-0359-0

49. Koboldt DC. Best practices for variant calling in clinical sequencing. *Genome Med.* 2020;12:91. https://doi.org/10.1186/s13073-020-00791-w

50. De Souza VBC, Jordan BT, Tseng E, et al. Transforming alignment files improves performance in long-read RNA variant callers. *Genome Biol.* 2023;24:91. https://doi.org/10.1186/s13059-023-02923-y

51. Lee HG, Casadesus G, Nunomura A, et al. Neuronal MYC expression causes neurodegeneration. *Am J Pathol.* 2009;174(3):891–897. https://doi.org/10.2353/ajpath.2009.080583

52. Zaytseva O, Kim N, Quinn LM. MYC in brain development and cancer. *Int J Mol Sci.* 2020;21(20):7742. https://doi.org/10.3390/ijms21207742

53. Lee HG, Casadesus G, Nunomura A, Zhu X, Castellani RJ, Richardson SL, Perry G, Felsher DW, Petersen RB, Smith MA. The neuronal expression of MYC causes a neurodegenerative phenotype in a novel transgenic mouse. Am J Pathol. 2009;174(3):891–897. https://doi.org/10.2353/ajpath.2009.080583

54. Bachstetter AD, Xing B, de Almeida L, et al. Microglial p38α MAPK is a key regulator of proinflammatory cytokine upregulation induced by toll-like receptor ligands or β-amyloid. J Neuroinflammation. 2011;8:79. https://doi.org/10.1186/1742-2094-8-79

Data Availability

RNA-Seq datasets used in this study are publicly available from the European Nucleotide Archive (ENA),PRJNA422218(***https://www.ebi.ac.uk/ena/browser/view/PRJNA422218***) and the associated metadata from the **Gene Expression Omnibus (GEO), GSE108012**. The data were originally published in the study **"Transcriptomic analysis of cerebral endothelial cells from patients with adrenoleukodystrophy"**in **BMC Neuroscience**.

Abbreviation

ABCD1 – ATP-Binding Cassette Subfamily D Member 1

ALD – Adrenoleukodystrophy

AMN – Adrenomyeloneuropathy

BAM – Binary Alignment Map

BBB – Blood–Brain Barrier

BMEC – Brain Microvascular Endothelial Cells

ccALD – Childhood Cerebral Adrenoleukodystrophy

ClinVar – Clinical Variant Database

COSMIC – Catalogue of Somatic Mutations in Cancer

CPM – Counts Per Million

Cytoscape – Biological Network Visualization Software

DEG – Differentially Expressed Gene

DEGs – Differentially Expressed Genes

edgeR – Empirical Analysis of Digital Gene Expression Data in R

FASTQ – Fast Quality Sequence Format

FastQC – Fast Quality Control (Sequencing Quality Assessment Tool)

featureCounts – Read Assignment to Genomic Features

FPKM – Fragments Per Kilobase per Million Mapped Reads

GATK – Genome Analysis Toolkit

GENCODE – Genome Annotation Consortium

GEO – Gene Expression Omnibus

HISAT2 – Hierarchical Indexing for Spliced Alignment of Transcripts 2

iPSC – Induced Pluripotent Stem Cells

iPSC-BMECs – iPSC-Derived Brain Microvascular Endothelial Cells

MAPK – Mitogen-Activated Protein Kinase

MultiQC – Multi-tool QC Result Summarizer

MYC – Myelocytomatosis Oncogene (Transcription Factor)

NGS – Next-Generation Sequencing

PCA – Principal Component Analysis

RefSeq – Reference Sequence Database

RNA-Seq – RNA Sequencing

ROS – Reactive Oxygen Species

SNP – Single Nucleotide Polymorphism

TPM – Transcripts Per Million

VLCFA – Very-Long-Chain Fatty Acids