INFORMATICS JOURNALS

Research Article

Journal of Natural Remedies
DOI: 10.18311/jnr/2026/53272

eISSN: 2320-3358 | Vol 26(7), 2026 | pp. 1765-1779

Deciphering ECM Remodelling in Retinoblastoma through Integrative Transcriptomic and Computational Approaches: Identification of Key Regulators and Natural Inhibitors

G. Ganavi[1], Anuroopa G. Nadh[2], Aswin Mohan[2], Rajesh Raju[2] and J. R. Rani[1*]

[1]Department of Biotechnology, Vignan's Foundation for Science, Technology and Research (Deemed to be University), Guntur - 522213, Andhra Pradesh, India; ranivfstr@gmail.com

[2]Centre for Integrative Omics Data Science, Yenepoya (Deemed to be) University, Mangalore - 575018, Karnataka, India

Abstract:

Background: Retinoblastoma (RB) is the most common intraocular malignancy in children, primarily associated with inactivation of the RB1 gene. Besides genetic alterations, Extracellular Matrix (ECM) remodelling and changes in the tumour microenvironment play critical roles in RB invasion, angiogenesis, and drug resistance. **Aim**: This study aimed to identify key ECM remodelling regulators in RB and evaluate potential natural inhibitors using integrated *in silico* approaches. **Methods:** Transcriptomic profiling from the GEO dataset GSE111168 was analysed to identify differentially expressed ECM-related genes. Machine learning and systems biology approaches were applied to prioritise significant targets. Molecular docking and Molecular Dynamics (MD) simulations were performed to assess the interaction and stability of natural compounds with the top-ranked target, PLOD1. Pharmacokinetic, ADMET, and oral bioavailability analyses were further conducted to evaluate drug-likeness. **Results:** PLOD1 emerged as a major ECM remodelling gene overexpressed in RB. Among the screened phytochemicals, the natural flavonoid luteolin demonstrated the highest docking affinity (score: 119.15) compared to carboplatin (65.96). MD simulations confirmed the stable binding of luteolin in the catalytic pocket of PLOD1, supported by favourable pharmacokinetic and safety profiles. **Conclusion:** The study identifies PLOD1 as a candidate promising therapeutic target in retinoblastoma and luteolin as a potential inhibitor. These findings provide a computational framework for future experimental validation and the development of phytochemical-based therapies in pediatric ocular oncology.

Major Findings: PLOD1 was identified as a key ECM-associated candidate biomarker in retinoblastoma through integrative transcriptomic and systems-level analysis. Among screened phytochemicals, luteolin demonstrated the strongest predicted binding affinity and stable interaction with PLOD1 in docking and molecular dynamics simulations, supporting its potential as a lead compound for further experimental validation.

Keywords: Differentially Expressed Genes, Extracellular Matrix, Luteolin, Molecular Docking, Protein Lysine Oxoglutarate Dioxygenase, Retinoblastoma

(Article Chronicle: Article Received on: 19.11.2025; Revised on: 10.04.2026; Accepted on: 05.06.2026)

1. Introduction

Retinoblastoma (RB) typically affects children under five years of age and arises from biallelic inactivation of the RB1 tumour suppressor gene located on chromosome 13q14. This gene encodes the retinoblastoma protein (pRB), a key regulator of the G1/S cell cycle checkpoint[1]. While survival rates exceed 95% in high-income countries owing to early diagnosis and multidisciplinary management, outcomes remain

*Author for correspondence

poor in middle- and low-income regions due to delayed detection and limited access to specialized care[2].

RB cells employ multiple survival mechanisms, including evasion of apoptosis, disruption of cell-cycle control, and metabolic reprogramming. Hypoxic conditions within the tumour microenvironment activate hypoxia-inducible factors (HIFs), which drive angiogenesis and tumour progression. Autophagy serves as a cytoprotective mechanism, helping tumour cells withstand therapeutic stress and resist chemotherapy-induced damage. Additionally, microRNAs play a critical role in post-transcriptional regulation, influencing tumour proliferation, differentiation, and drug resistance[2].

Current treatment strategies encompass intravenous and intra-arterial chemotherapy, focal therapies such as thermotherapy and cryotherapy, and enucleation in advanced cases. However, these approaches are often associated with severe complications, including vision loss and an increased risk of secondary malignancies. Alternative therapeutic strategies, such as nanotechnology-based drug delivery systems, are emerging as promising approaches to enhance treatment specificity and efficacy while reducing systemic toxicity[3]. *In-silico* drug design, particularly molecular docking approaches, provides a rapid and cost-effective strategy for identifying and optimising potential therapeutics for many disorders, including retinoblastoma, by virtually screening extensive compound libraries and predicting binding affinities before experimental validation[4]. Phytochemicals, the naturally occurring bioactive compounds found in plants, play a pivotal role in modern drug discovery and development[5,6]. These molecules serve as a vast and structurally diverse reservoir of therapeutic candidates, often exhibiting high biological activity and favourable pharmacokinetic profiles. Many conventional drugs, including anticancer, antimicrobial, and cardiovascular agents, have originated directly or indirectly from plant-derived compounds[7,8]. Moreover, their natural origin often correlates with lower toxicity and improved biocompatibility compared to fully synthetic drugs[9,10]. As indicated, advances in computational biology and systems pharmacology have further enhanced the ability to screen, predict, and optimise phytochemical efficacy, accelerating the identification of potent lead compounds[11,12]. Previous studies have successfully applied such methods to target dysregulated molecules in disease conditions[13,14]. In the case of RB, significant knowledge gaps remain, especially in uncovering additional molecular targets and biomarkers involved in disease pathogenesis beyond RB1 mutations, which hinders early detection and targeted intervention[3]. Increasing evidence suggests that Extracellular Matrix (ECM) remodelling plays an important role in retinoblastoma progression by influencing tumour invasion, angiogenesis, and cell–matrix interactions within the retinal microenvironment. Alterations in collagen organisation and basement membrane integrity have been associated with invasive tumour behaviour and metastatic potential in retinoblastoma. Enzymes involved in collagen post-translational modification, including lysyl hydroxylases such as members of the PLOD family, regulate collagen cross-linking and extracellular matrix stiffness, which are critical determinants of tumour cell migration and stromal remodelling. These observations highlight the relevance of investigating ECM-modifying regulators as potential molecular targets for understanding retinoblastoma pathogenesis beyond RB1-driven mechanisms. In this context, the present study employs an in-silico framework to analyse differentially expressed genes in retinoblastoma and to virtually screen natural compounds with documented anticancer properties against prioritised molecular targets.

In this study, a comprehensive integrative approach was employed to identify potential biomarkers and therapeutic targets for retinoblastoma. Gene expression data were retrieved from public repositories and subjected to differential expression analysis, followed by Support Vector Machine (SVM)-based classification and biomarker filtering to prioritise key genes. Protein-protein interaction, pathway, and gene ontology enrichment analyses further highlighted the molecular mechanisms underlying disease progression. Subsequently, molecular docking and molecular dynamics simulations were conducted to assess the binding affinity and stability of selected phytochemicals with critical protein targets. Finally, drug-likeness and ADMET profiling were carried out to evaluate the pharmacokinetic properties and safety of the candidate compounds, thereby integrating computational genomics and structure-based approaches to explore novel therapeutic avenues for RB.

2. Materials and Methods

2.1 Data Retrieval and Pre-processing

Gene expression data related to retinoblastoma were retrieved from the Gene Expression Omnibus (GEO), an open-access repository maintained by the National Centre for Biotechnology Information (NCBI). The dataset analysed in this study, GSE111168, comprises six samples: three from normal retinal tissues and three from retinoblastoma-affected tissues. This microarray-based dataset was chosen to identify Differentially Expressed Genes (DEGs) specific to retinoblastoma.

For differential expression analysis, the GEO2R web application was employed. GEO2R, which utilises the limma package, enables comparison between control and disease groups[15]. Genes with a false discovery rate (FDR) < 0.05 and an absolute log2 fold change > 2 were considered significant. To facilitate interpretation, DEGs were visualised using volcano plots and mean-difference plots.

2.2 SVM-based Classification and Feature Selection for Differentially Expressed Genes

Support Vector Machine (SVM) modelling was employed to classify retinoblastoma gene expression profiles. High-dimensional biological datasets characterised by small sample sizes and complex feature interactions make SVM effective for disease classification and biomarker discovery[16]. The model was trained using the GSE111168 dataset. Preprocessing was performed to identify statistically and biologically significant genes using adjusted *p*-value (padj < 0.05) together with an absolute $\log_2$ fold change threshold ($|\log_2 FC| \geq 2$), consistent with the DEG filtering criteria applied in the differential expression analysis. Each gene was then assigned a binary "Significant" label, which served as the target variable for supervised classification. The dataset was split into training and testing sets at an 8:2 ratio, and features were standardised before model fitting. Due to the limited number of samples available in the GSE111168 dataset (n = 6), the SVM model was applied primarily as an exploratory feature-prioritisation approach rather than as a predictive classification framework. Under such conditions, model performance metrics may be affected by overfitting and should therefore be interpreted cautiously. Consequently, the reported classification accuracy reflects internal dataset behaviour and does not represent externally validated predictive performance. Model performance was evaluated by calculating classification accuracy and visualising the results with a confusion matrix. The model demonstrated high precision and recall, thereby validating its ability to correctly classify statistically significant genes.

2.2.1 Biomarker Filtering and Selection

For validating the performance of the class model, accurate thresholds were applied in selecting the recognised biomarkers. Genes were first filtered based on two essential thresholds. (1) An adjusted p-value of less than 0.05 and (2) an absolute log2 fold change of ≥2. This two-step filtering ensures picking genes that are not only statistically distinct but also exhibit notable biological changes. Based on this optimised dataset, the top 10 of the differentially expressed genes were chosen as candidate biomarkers for Retinoblastoma. This two-stage filtering strategy, combining statistical thresholds (adjusted p-value < 0.05 and $|\log_2 FC| \geq 2$) with SVM-based prioritisation, was implemented to reduce dimensionality before classification and to identify biologically relevant candidate biomarkers. However, due to dataset size constraints, this prioritisation step should be considered exploratory and requires validation using independent transcriptomic cohorts. Differential expression analysis was performed using GEO2R, which utilises the limma package on pre-normalised expression matrices available in the GEO database. Probe-to-gene symbol mapping was conducted using platform annotation files provided with the dataset, and when multiple probes corresponded to the same gene, the probe with the most significant adjusted *p*-value was retained for downstream analysis. As the dataset did not contain multiple experimental batches, batch-effect correction was not required.

2.3 Protein-Protein Interactions

To investigate the molecular events associated with the identified DEGs, Protein–Protein Interaction (PPI) analysis was performed using the STRING database[17]. Only interactions with a high confidence score (>0.7) were considered to ensure reliability. The resulting interaction data were imported into Cytoscape for

network visualisation and analysis[18]. The constructed PPI network highlighted significant interconnections among proteins, indicating coordinated biological activities.

2.4 Pathway Enrichment Analysis

Pathway enrichment analysis was performed to interpret the biological significance of the filtered gene set. The Enrichr platform was used to query multiple repositories, including Reactome 2024, WikiPathways 2024, and KEGG 2021. Statistical methods were applied to identify pathways containing a higher-than-expected number of genes from the selected dataset, thereby highlighting key cellular processes that may be implicated in retinoblastoma. By integrating information across these distinct enrichment databases, the analysis provided insights into how the identified pathways may contribute to biological regulation, gene expression control, and disease mechanisms[19].

2.5 Gene Ontology Analysis

For functional investigation of the identified genes, Gene Ontology (GO) enrichment analysis was performed using the Enrichr database. The three GO categories, Biological Process, Cellular Component, and Molecular Function, were investigated. This was utilised for the classification of genes based on their respective biological roles, cellular localisation, and molecular function, which indicates their functional relevance in disease.

2.6 Molecular Docking

2.6.1 Target Identification

Following the outlined workflow, a set of critical target proteins was identified. The final shortlist of biologically significant targets was refined through a literature-based review, considering their previously established roles in cancer-related pathways, expression profiles in tumour tissues, therapeutic relevance, and supporting evidence from prior studies.

2.6.2 Ligand Identification

Phytochemicals with documented anticancer activity were screened through a systematic literature review. A comprehensive library of natural bioactive compounds was assembled from public chemical databases, including PubChem, ChEMBL, and NPASS (Natural Product Activity and Species Source Database). Only

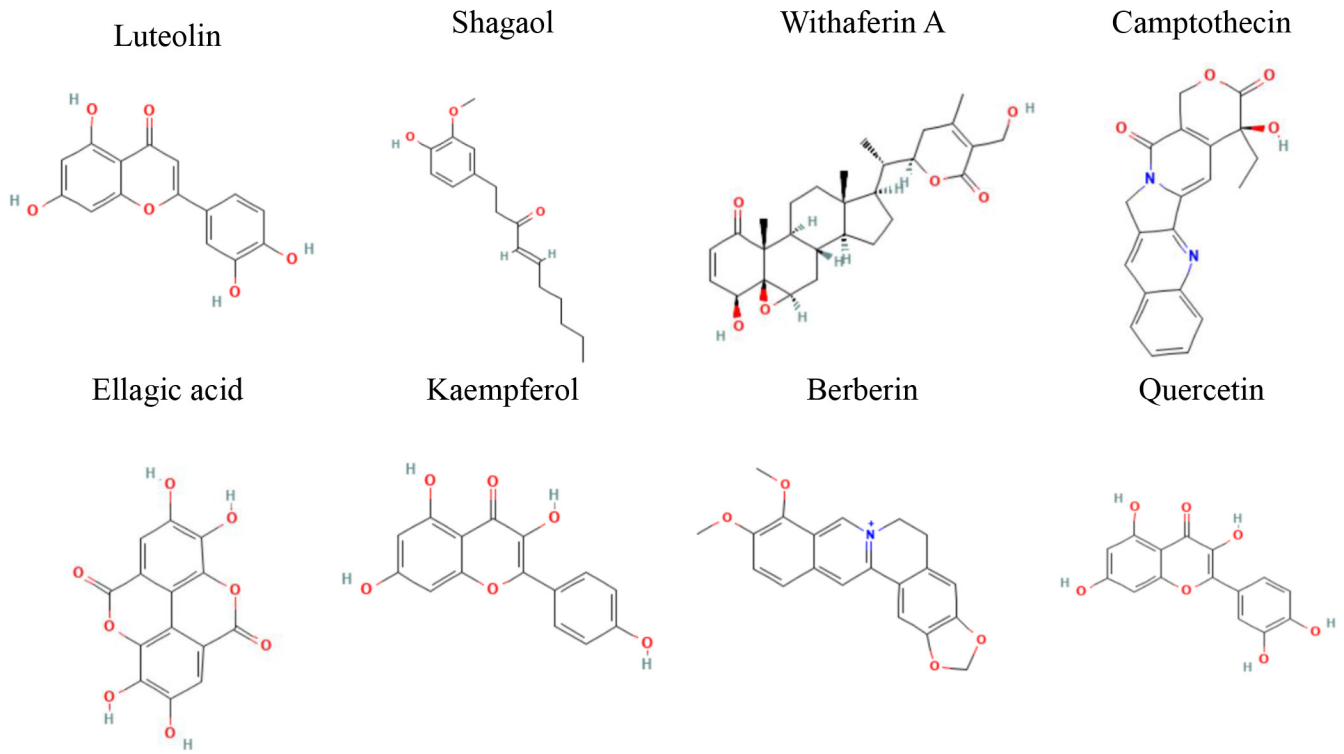


Figure 1. List of phytochemicals selected for virtual screen against selected targets of retinoblastoma.

compounds with experimentally verified chemical structures and known biological activities were considered. Figure 1 shows the 2D structure of selected natural compounds for virtual screening.

2.6.3 Molecular Docking

As experimental structural data for the target protein were unavailable, its three-dimensional structure was retrieved from AlphaFold, an AI-driven platform that predicts protein conformations from amino acid sequences[20]. The 3D structures of small molecules and the reference drug carboplatin were obtained from the PubChem database[21]. Carboplatin was included as a reference compound because it is a clinically established chemotherapeutic agent widely used in retinoblastoma treatment. Although its mechanism of action involves DNA cross-linking rather than direct inhibition of PLOD1, its inclusion provides a clinically relevant benchmark to contextualise the relative docking performance of the screened phytochemicals within the present computational framework. Before docking, both protein targets and ligands were prepared using standard protocols in the Discovery Studio (DS) Client software suite[22]. Binding sites were defined based on receptor cavities, and virtual screening was performed using the LibDock module of DS. LibDock is a rigid docking algorithm that maps binding site hot spots by analysing the protein's polar and apolar features and aligns the physicochemical properties of ligands accordingly to predict favourable interactions.

2.7 Molecular Dynamics Simulation

The molecular dynamics simulation was carried out using Gromacs, version 2024.1[23]. For the protein ligand complex, the docked complex was used to generate topologies of protein and ligand. Topologies were generated separately using the pdb2gmx tool and the CGenFF server for protein and ligand, respectively. The system was solvated using the TIP3P water model within a triclinic box, followed by the addition of monovalent counterions (Na^+ and Cl^-) to neutralise the overall charge of the system. Energy minimisation was performed to obtain the most stable conformation of the complex by reducing its potential energy to the lowest possible level using the steepest descent algorithm. Thermostat coupling was done, followed by the two-stage equilibration phase. Equilibration phase 1 - NVT ensemble (Number of particles, Volume and Temperature) and Equilibration phase 2 - NPT ensemble (Number of particles, Pressure and Temperature) was performed for a span of 100 picoseconds. The final production MD was carried out for 100 nanoseconds in 50000000 steps. The trajectory files were analysed for Root Mean Square Deviation (RMSD), Root Mean Square Fluctuation (RMSF), H-Bond, and SASA (Solvent Accessible Surface Area).

2.8 Lipinski Filtering and ADMET

ADMET properties and druglikeness of the promising compounds were screened with the internet-based web server ADMETLab[24]. Lipinski's 'Rule of Five' was employed to calculate the physicochemical properties of the compounds and their potential for oral bioavailability in man[25]. This regulation dictates that to be considered an oral drug candidate, the compound should satisfy the following criteria of Hydrogen bond donors (HBD) ≤ 5, hydrogen bond acceptors (HBA) ≤ 10, molecular weight (MW) < 500 Da, and octanol-water partition coefficient (AlogP) ≤ 5. Furthermore, the ADMET characteristics like solubility, intestinal permeability, blood-brain barrier permeability (BBB), plasma protein binding, inhibition of cytochrome P4502D6, and hepatotoxicity were also predicted by the ADMET descriptors module of ADMETLab. This software provides reliable pharmacological predictions based on the molecular structure of the compounds under investigation.

3. Results and Discussion

3.1 Data Retrieval and Preprocessing

A total of 6,525 differentially expressed genes (DEGs) were identified in the retinoblastoma dataset (GSE111168), comprising 2,250 upregulated and 4,275 downregulated genes when comparing retinoblastoma tissues with normal retinal tissues. To visualise these differences, a volcano plot was generated (Figure 2) to display the distribution of expression changes (Figure 2a), while an MA (mean-difference) plot illustrated overall expression patterns across samples (Figure 2b).

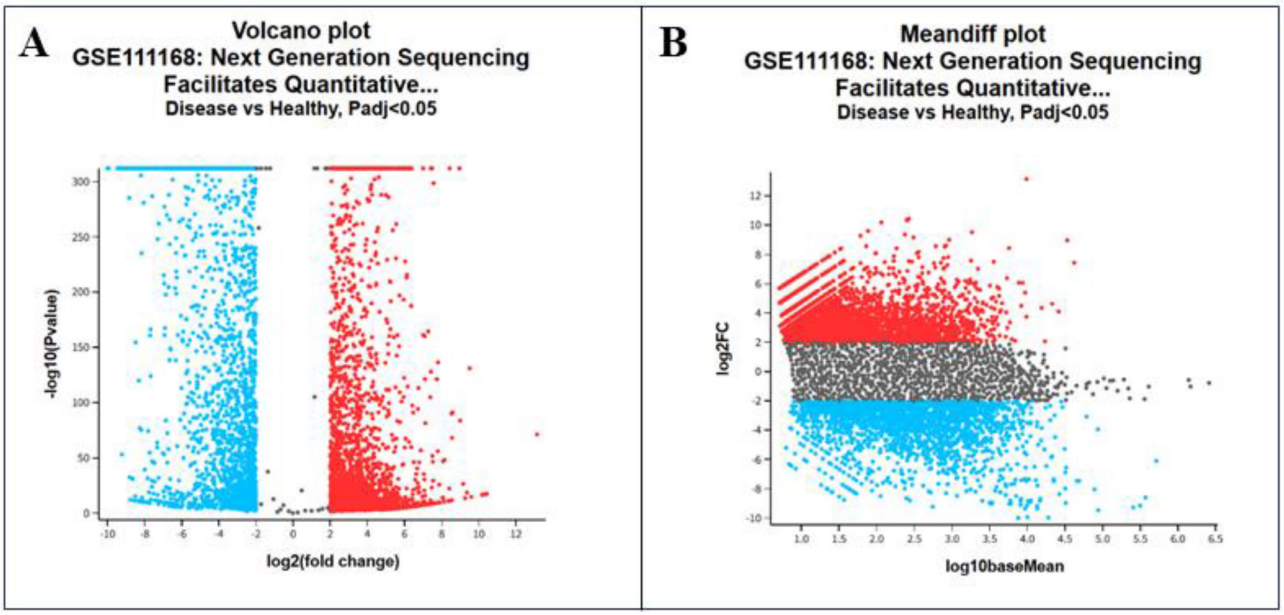


Figure 2. Volcano plot of DEGs between retinoblastoma and normal retina samples. LHS represents the distribution of expression changes. **(Figure 2a).** The x-axis is $\log_2$ fold change, and the y-axis is $-\log_{10}$ p-value. RHS shows the mean difference plot of gene expression variation in retinoblastoma compared to the normal retina. **(Figure 2b).** The x-axis indicates the average expression ($\log_{10}$ base mean), and the y-axis indicates $\log_2$ fold change.

3.2 SVM-based Classification and Feature Selection for Differentially Expressed Genes

The SVM model proved highly effective in classifying significant genes associated with retinoblastoma. The model achieved an accuracy of 99.77%, underscoring its reliability. As shown in the confusion matrix (Figure 3), the true positive rate was high, with 4,331 genes correctly classified as significant and 1,222 genes correctly identified as non-significant. Only 13 false positives were detected, and notably, no false negatives were reported, highlighting the robustness of the model in terms of both precision and sensitivity. These results demonstrate the ability of SVM to handle high-dimensional biological data efficiently and to reliably separate positive from negative gene classes. The strong performance of the model suggests its potential utility as a robust classifier for biological datasets, providing valuable insights into molecular processes underlying retinoblastoma and supporting future biomarker discovery efforts.

3.2.1 Biomarker Filtering and Selection

A two-step filtering strategy was applied to identify reliably Differentially Expressed Genes (DEGs) in retinoblastoma. Genes meeting both statistical and biological significance thresholds, adjusted p-value < 0.05, and absolute log2 fold change ≥ 2, were retained. From this filtered dataset, the top 10 DEGs showing the most pronounced expression differences were shortlisted as potential biomarkers. All selected genes exhibited highly significant adjusted p-values and were strongly upregulated, with log2 fold change values ranging from 2.23 to 5.50. These candidate biomarkers are summarised in Table 1.

All the above-listed genes showed highly significant differential expression, indicating robust statistical confidence in their altered expression patterns between retinoblastoma and normal retinal tissues. Several of these genes are functionally associated with extracellular matrix (ECM) remodelling and cell adhesion, which are critical processes in tumour progression. For instance, COL16A1 (Collagen Type XVI Alpha 1) and PLOD1 (Procollagen-Lysine,2-Oxoglutarate 5-Dioxygenase 1)

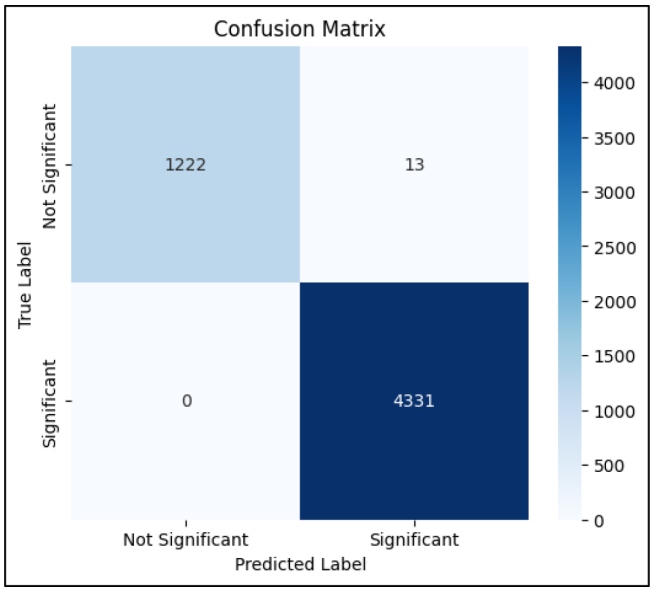


Figure 3. Support Vector Machine (SVM) model classification performance confusion matrix for identifying statistically significant genes in retinoblastoma.

are directly involved in collagen biosynthesis and ECM organisation, supporting their potential role in modifying the tumour microenvironment. The obtained results also correlate with these, as both exhibited particularly high log2 fold changes (4.4 and 4.9, respectively), highlighting their marked upregulation in RB tissues. MFAP2 (Microfibril Associated Protein 2) and HSPG2 (Heparan Sulfate Proteoglycan 2) also contribute to ECM structure and signalling. Genes such as PADI2 (Peptidyl Arginine Deiminase 2) and ID3 (Inhibitor of DNA Binding 3) are implicated in epigenetic regulation, cell differentiation, and proliferation, processes that are often dysregulated in cancer. MYOM3 (Myomesin 3) also showed substantial alteration in its expression, although its precise role in RB requires further functional characterisation. ASAP3 (ArfGAP with SH3 Domain, Ankyrin Repeat and PH Domain 3) and TCEA3 (Transcription Elongation Factor A3) may also influence signalling and transcriptional regulation in tumour cells.

Collectively, these genes represent a functionally diverse set of potential biomarkers, reflecting alterations in ECM remodelling, transcriptional regulation, and cellular signalling pathways. Their differential expression highlights possible molecular mechanisms driving RB pathogenesis and provides a foundation for future studies on diagnostic and therapeutic targeting.

Table 1. Top 10 differentially expressed genes (DEGs) identified as possible biomarkers of retinoblastoma

S.NO	Gene ID	Symbol	Adjusted P value<0.05	log2fold change ≥2
1	1953	MEGF2	0.0	2.894993
2	5351	COL16A1	0.0	4.441444
3	4237	MFAP2	0.0	3.898023
4	11240	PADI2	0.0	3.540274
5	3339	HSPG2	0.0	3.428351
6	6920	TCEA3	0.0	3.386397
7	55616	ASAP3	0.0	2.237982
8	3399	ID3	0.0	3.697884
9	127294	MYOM3	0.0	3.497084
10	1307	PLOD1	0.0	4.967358

*DEGs were filtered using adjusted p-value < 0.05 and $\log_2$ fold change ≥ 2.

3.3 Protein-Protein Interactions

To further elucidate the molecular mechanisms underlying the Differentially Expressed Genes (DEGs), a protein–protein interaction (PPI) network was generated using the STRING database with a high-confidence interaction threshold (>0.7). The interaction data were subsequently imported into Cytoscape for visualisation and in-depth network analysis. Notably, PLOD1 and COL16A1 exhibited a direct interaction, indicating a potential functional association between these genes (Figure 4). Other genes, such as PADI2, ID3, MYOM3, MFAP2, HSPG2, ASAP3, and CELSR3, did not demonstrate direct high-confidence interactions within this network subset; they remain biologically significant. Together, these observations highlight possible co-regulatory mechanisms involving extracellular matrix remodelling and cell adhesion processes.

3.4 Pathway Enrichment Analysis

Table 2 summarises the pathway enrichment analysis of Differentially Expressed Genes (DEGs) in retinoblastoma, highlighting the key molecular pathways in which these genes are involved, along with their biological significance. The analysis integrates results from three major databases: Reactome 2024, WikiPathways 2024, and KEGG 2021, providing a

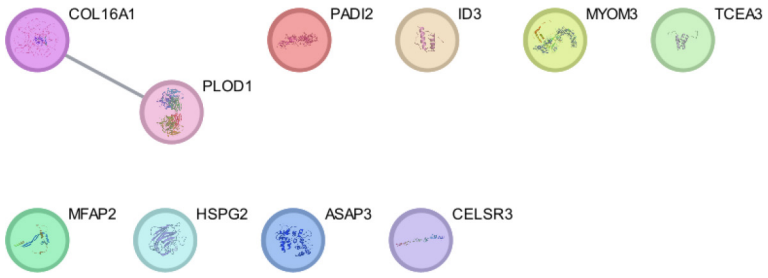


Figure 4. Protein-protein interaction (PPI) network of selected DEGs of retinoblastoma. PPI was predicted by STRING and visualised in Cytoscape. Only high-confidence interactions (score > 0.7) are displayed

comprehensive overview of the functional contexts of the DEGs. Reactome 2024 pathways primarily Emphasise Extracellular Matrix (ECM) organisation and collagen-related processes. For example, PLOD1 and COL16A1 are involved in Extracellular Matrix Organisation ($p = 0.000010$) and Collagen Biosynthesis ($p = 0.000489$), reflecting their critical role in maintaining ECM structure and connective tissue integrity. Other pathways, including Integrin-Cell Surface Interactions and ECM Degradation, link PLOD1 and HSPG2 to cell–ECM adhesion, signalling, and dynamic remodelling, which are key events in tumour invasion and progression.

WikiPathways 2024 highlights additional processes such as ID Signalling (ID3; $p = 0.007973$), which regulates cell growth and differentiation, and Fibroblast Metabolism and Type I Collagen Synthesis (PLOD1; $p < 0.016$), indicating roles in ECM production, repair, and tissue structural support.

KEGG 2021 pathways further underscore PLOD1's involvement in Lysine Degradation ($p = 0.031064$), which is directly related to collagen cross-linking and ECM quality. HSPG2 participates in ECM-Receptor Interaction ($p = 0.043148$), facilitating structural support and signalling, whereas ID3 contributes to TGF-Beta Signalling ($p = 0.046028$), relevant to fibrosis, immune modulation, and oncogenic processes.

Overall, this table illustrates that several of the top DEGs, particularly PLOD1, COL16A1, HSPG2, and ID3, converge on pathways involved in ECM remodelling, collagen synthesis, cell adhesion, and signalling. These processes are crucial for retinoblastoma progression, suggesting that these genes may serve as both functional biomarkers and potential therapeutic targets.

3.5 Gene Ontology Analysis

Gene Ontology enrichment analysis of the differentially expressed genes revealed a strong involvement of Extracellular Matrix (ECM) remodelling, collagen modification, and cell adhesion processes. PLOD1 emerged as a central gene driving biological processes such as extracellular matrix organisation, peptidyl-lysine hydroxylation, protein hydroxylation, and regulation of focal adhesion and cell-matrix adhesion. These findings suggest that alterations in collagen maturation and ECM assembly may contribute to changes in tissue architecture and tumour microenvironment dynamics. Supporting evidence from previous studies indicates that PLOD1 is frequently upregulated in cancers characterised by extracellular matrix remodelling and enhanced collagen cross-linking, where it contributes to tumour invasion and metastasis. Similarly, ECM-associated mechanisms have been implicated in retinoblastoma progression, suggesting that the observed upregulation of PLOD1 in the present dataset is biologically plausible and warrants further validation using larger transcriptomic cohorts. HSPG2 was implicated in the regulation of amyloid fibril formation and basement membrane organisation, indicating potential roles in proteoglycan-mediated ECM stabilisation and amyloid-related pathways. Cellular component analysis highlighted enrichment in collagen-containing ECM, basement membrane, and various organelle lumens, consistent with active processing, trafficking, and secretion of ECM proteins. Molecular function enrichment further revealed enzymatic activities, including hydrolase and hydroxylase activity (PADI2, PLOD1) and receptor binding functions (HSPG2), as well as protein homodimerization, suggesting that these genes not only contribute structurally but may also modulate

Table 2. Pathway enrichment analysis parameters of DEGs of retinoblastoma

Pathway	Pathway Involved	P-Value	Genes Involved	Biological Insight
Reactome 2024	Extracellular Matrix Organization	0.000010	PLOD1, COL16A1, HSPG2	Involved in ECM structure and tissue stability
	Collagen Biosynthesis	0.000489	COL16A1, PLOD1	Key process in connective tissue formation
	Integrin-6(A)RMSDCell Surface Interactions	0.000786	PLOD1, HSPG2	Mediates cell-ECM adhesion and signalling
	Collagen Formation	0.000880	PLOD1	Essential for ECM integrity and repair
	ECM Degradation	0.002110	PLOD1, HSPG2	Linked to ECM remodelling and disease progression
WikiPathways 2024	ID Signaling	0.007973	ID3	Regulates cell growth and differentiation
	Fibroblast Metabolism	0.015889	PLOD1	Supports ECM production and repair
	Type I Collagen Synthesis	0.016382	PLOD1	Impacts bone strength and collagen stability
KEGG 2021	Lysine Degradation	0.031064	PLOD1	Related to collagen cross-linking and ECM quality
	ECM-Receptor Interaction	0.043148	HSPG2	Facilitates structural support and signalling
	TGF-Beta Signaling	0.046028	ID3	Involved in fibrosis, immune response, and cancer

*Pathways, p-values, involved genes, and biological insights are provided.

Table 3. Enrichment for Gene Ontology (GO) terms of degrees categorised into Biological Process (BP), Cellular Component (CC), and Molecular Function (MF)

GO Functions	Term	GO ID	p-value	q-value	Overlap Genes
Biological Function	Extracellular Matrix Organization	GO:0030198	0.000488	0.020506	PLOD1
	Peptidyl-Lysine Hydroxylation	GO:0017185	0.001399	0.025168	PLOD1
	Negative Regulation of Amyloid Fibril Formation	GO:1905907	0.002398	0.025168	HSPG2
	Protein Hydroxylation	GO:0018126	0.002398	0.025168	PLOD1
	Regulation of Amyloid Fibril Formation	GO:1905906	0.003196	0.025168	HSPG2
	Positive Regulation of Focal Adhesion Assembly	GO:0051894	0.003595	0.025168	PLOD1
	Positive Regulation of Cell-Substrate Junction Organisation	GO:0150117	0.004592	0.027554	PLOD1
	Collagen Fibril Organization	GO:0030199	0.008970	0.040833	PLOD1
	Positive Regulation of Cell-Matrix Adhesion	GO:0001954	0.008970	0.040833	PLOD1
	Regulation of Focal Adhesion Assembly	GO:0051893	0.010559	0.040833	PLOD1
Cellular Function	Collagen-Containing Extracellular Matrix	GO:0062023	0.000027	0.000256	PLOD1, HSPG2
	Basement Membrane	GO:0005604	0.000037	0.000256	COL16A1, HSPG2
	Intracellular Organelle Lumen	GO:0070013	0.011720	0.054692	PLOD1, HSPG2
	Lysosomal Lumen	GO:0043202	0.017091	0.055033	HSPG2
	Golgi Lumen	GO:0005796	0.019655	0.055033	HSPG2
	Vacuolar Lumen	GO:0005775	0.032206	0.075147	HSPG2
Molecular Function	Hydrolase Activity, Acting on Carbon-Nitrogen	GO:0016813	0.004492	0.034733	PADI2
	Low-Density Lipoprotein Particle Receptor Binding	GO:0050750	0.011938	0.034733	HSPG2
	Lipoprotein Particle Receptor Binding	GO:0070325	0.014409	0.034733	HSPG2
	Nuclear Estrogen Receptor Binding	GO:0030331	0.017367	0.034733	PADI2
	Protein Homodimerization Activity	GO:0042803	0.043344	0.069351	PADI2, MYOM3

signalling within the tumour microenvironment. Collectively, these results underscore the pivotal role of ECM remodelling and post-translational modifications in shaping the structural and functional landscape of the tumour, potentially facilitating invasion and progression (Table 3).

3.6 Molecular Docking

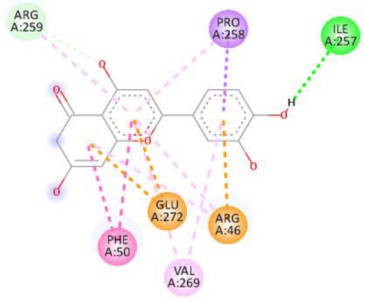
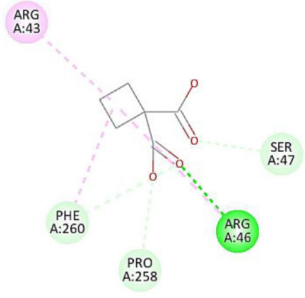
Among the shortlisted differentially expressed genes, PLOD1 was selected as the primary docking target based on multiple prioritisation criteria. Specifically, PLOD1 showed one of the highest expression changes (log_2FC = 4.96) and was significantly enriched in extracellular matrix organisation and collagen biosynthesis pathways identified through Reactome, KEGG, and WikiPathways analyses. Gene Ontology results further indicated its involvement in collagen fibril organisation, protein hydroxylation, and focal adhesion regulation, highlighting its functional importance in tumour microenvironment remodelling. In addition, protein–protein interaction analysis demonstrated its association with other ECM-related regulators such as COL16A1, supporting its coordinated role in extracellular matrix restructuring. Based on its biological relevance and central role in ECM remodelling processes implicated in tumour progression, PLOD1 was selected for subsequent molecular docking and simulation studies.

3.6.1 Target Selection and Preparation

Since an experimentally resolved crystal structure of PLOD1 is not currently available, the three-dimensional structure was obtained from the AlphaFold database (ID: AF-Q02809-F1-v4). The model exhibited a high per-residue confidence score (pLDDT = 92.79%), supporting its suitability for structure-based computational analysis, although predicted structures may still contain local uncertainties in binding-site geometry. The structure exhibited a per-residue model confidence score of 92.79 %, as determined by the predicted local distance difference test (pLDDT), indicating a high likelihood of accurate structural prediction.

From the molecular docking analysis, it was observed that, among the eight screened compounds, Luteolin demonstrated a higher docking score than the reference compound carboplatin. This comparison provides a relative computational benchmark rather than implying a shared mechanism of inhibition, as carboplatin primarily acts through DNA cross-linking rather than direct

Table 4. Docking score and interacting residues of PLOD1 with the selected ligands

Complex	Docking score	2D interaction	Interacting residues
PLOD1-Luteolin	119.15		Arg46 Phe50 Ile257 Pro258 Arg259 Val267 Glu272
PLOD1-Carboplatin (control)	65.96		Arg46 Ser47 Pro258 Phe260

The nature of the interactions is illustrated below. The residues involved in conventional hydrogen bonds are highlighted in Table 4.

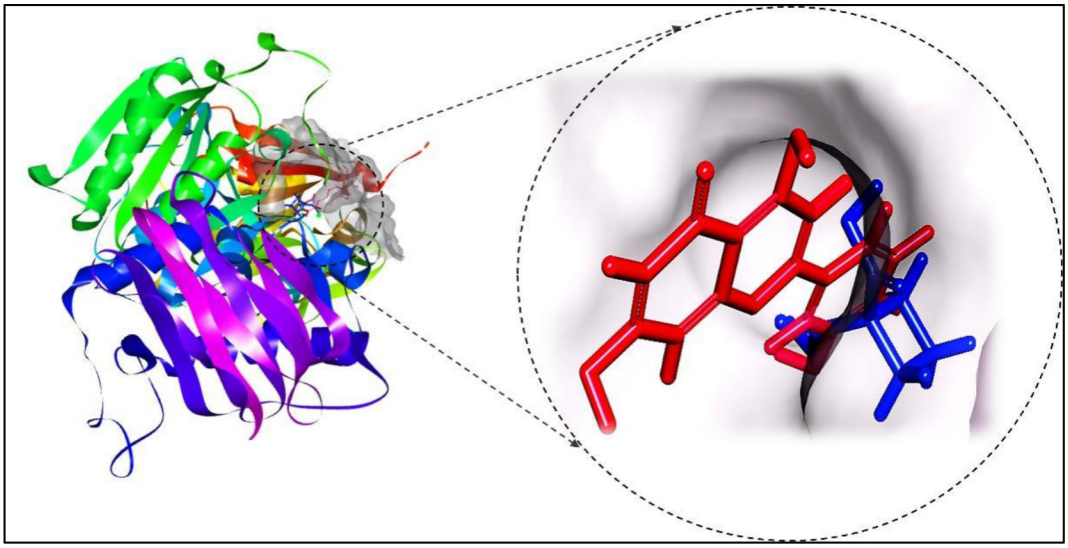


Figure 5. Binding mode of luteolin and carboplatin (control) in the binding cleft of PLOD1. LHS represents the 3D image of the target-ligand bound complex. The protein secondary structure is represented in ribbon and ligands in stick models. RHS represents a zoomed visualisation of the same. Luteolin is shown in red and carboplatin is shown in blue colours.

enzymatic targeting of PLOD1. This comparison with carboplatin is intended to provide a clinical reference context rather than imply a shared mechanism of inhibition with PLOD1, since carboplatin primarily acts through DNA cross-linking in tumour cells. The docking parameters of the target protein with luteolin and the reference drug, carboplatin, are shown in Table 4, and their binding mode is depicted in Figure 5.

3.7 Molecular Dynamics Simulations

Molecular dynamics simulations were performed to evaluate the conformational stability and interaction dynamics of PLOD1 in its free and luteolin-bound states (Figure 6). The RMSD trajectories (Figure 6(A)) revealed that the protein-ligand complex stabilised early and maintained equilibrium throughout the 100 ns simulation, with deviations consistently lower than those of the free protein. This indicates that luteolin binding confers overall structural stability. Residue-wise RMSF analysis (Figure 6(B)) showed that both systems exhibited similar fluctuation patterns, with the highest flexibility localised to the N-terminal region; however, the luteolin-bound complex demonstrated reduced fluctuations across most residues, particularly in structured regions, suggesting local stabilisation. Hydrogen bond analysis (Figure 6(C)) demonstrated that luteolin maintained 1-3 stable hydrogen bonds throughout the trajectory, with intermittent peaks up to 4-5, reflecting persistent and specific ligand–protein interactions. The Solvent-Accessible Surface Area (SASA) profiles (Figure 6(D)) further confirmed that the protein–ligand complex exhibited slightly reduced SASA compared to the free protein, implying tighter packing and decreased solvent exposure upon ligand association. Collectively, these results highlight that luteolin binding enhances the structural stability of PLOD1 while maintaining a compact conformation, supporting its potential role as a stable inhibitor.

3.8 Lipinski Filtering and ADMET

Physicochemical and ADMET profiles of Luteolin and the control drug were evaluated to quantify drug-likeness and safety attributes. Both compounds exhibited safety measures based on the toxicity prediction parameters. No hERG channel blocking potential, human hepatotoxicity, Drug-Induced Liver Injury (DILI), AMES mutagenicity, skin sensitisation, carcinogenicity, eye corrosion, eye irritation, or respiratory toxicity was predicted, suggesting a potentially favourable predicted toxicity profile for both compounds based on computational estimation (Table 5).

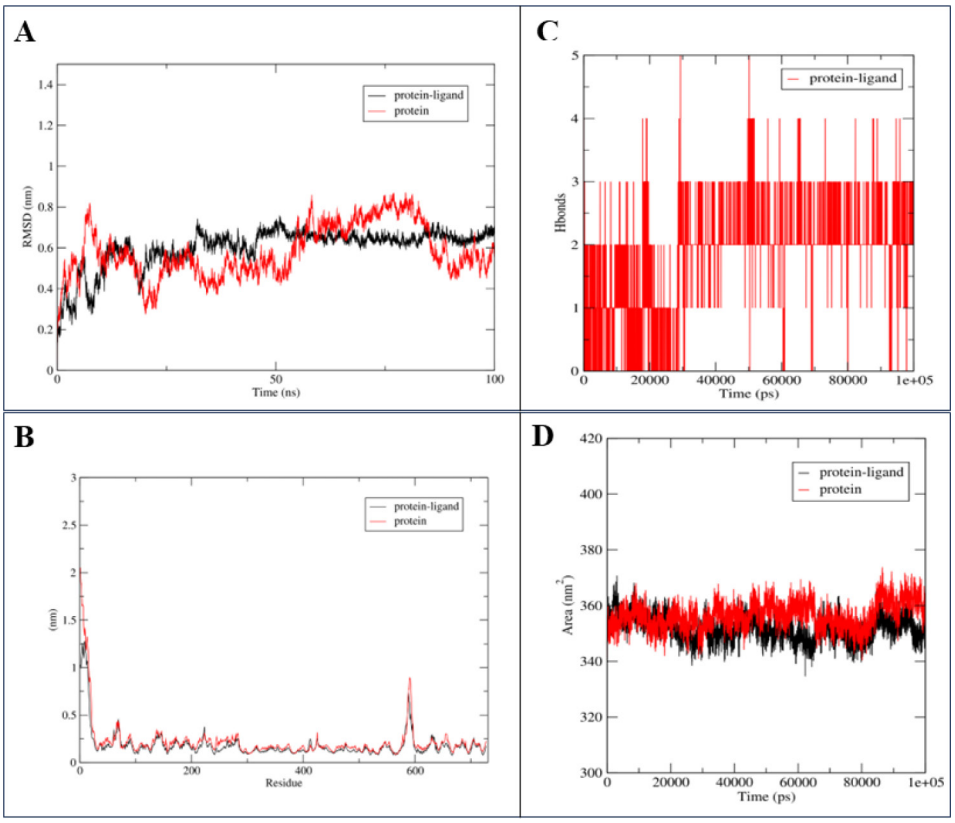


Figure 6. Molecular simulation of PLOD1 in its free and luteolin-bound states. **(A).** RMSD analysis of the PLOD1(red line) and the PLOD1-luteolin complex (black line) over a 100-nanosecond molecular dynamics simulation. **(B).** RMSF analysis of the PLOD1(red line) and the PLOD1-luteolin complex (black line) over a 100-nanosecond molecular dynamics simulation. **(C).** Hydrogen bonds formed between the protein and the ligand over a 100 ns dynamic simulation. **(D).** Solvent-accessible surface area (SASA) of PLOD1(red line) and the PLOD1-luteolin complex (black line) over a 100-nanosecond molecular dynamics simulation.

Table 5. ADMETLab-based toxicity and safety profile of candidate ligands

Ligand	Carboplatin (Standard)	Luteolin
hERG (channel blocking potential)	---	---
H-H (Human Hepatotoxicity)	---	---
DILI (Drug-Induced Liver Injury)	---	---
AMES Toxicity	---	---
FDA MDD	---	---
Skin Sensitization	---	---
Carcinogenicity	---	---
Eye Corrosion	---	---
Eye Irritation	---	---
Respiratory Toxicity	---	---

---: negative; +++: positive

Drug-likeness was assessed using Lipinski's Rule of Five. Both carboplatin and Luteolin demonstrated compliance, with molecular masses of 194.15 Da and 286.24 Da, LogP values of -1.11 and 1.73, H-bond donors (HBD) of 1 and 4, respectively, and H-bond acceptors (HBA) of 6 in both compounds, with no Lipinski violations (Table 6).

These findings indicate potential oral bioavailability for the test compounds. Although Carboplatin data were incomplete in this set, it is a well-established clinically approved drug and presumed to meet drug-likeness criteria. Overall, luteolin demonstrated predicted drug-like properties with no major computational toxicity alerts; however, these results represent preliminary in-silico estimates and require validation through experimental pharmacological studies.

Table 6. Lipinski's rule of five parameters for luteolin and carboplatin

Property	Luteolin	Carboplatin (Standard)	Follows Rule
Molecular Weight (Da)	286.24	194.15	Yes
LogP (Octanol-water)	1.73	-1.11	Yes
Hydrogen Bond Donors (HBD)	4	1	Yes
Hydrogen Bond Acceptors (HBA)	6	6	Yes
Lipinski Violations	0	0	Yes
Overall, Rule Compliance	Yes	Yes	Yes

3.9 Study Limitations

Despite the integrative computational framework employed in this study, several limitations should be considered while interpreting the findings. First, the transcriptomic analysis was performed using a relatively small dataset (GSE111168; n = 6), which may limit statistical robustness and contribute to optimistic performance estimates in the SVM classification. Accordingly, the reported classification accuracy should be interpreted primarily as a feature-prioritisation outcome rather than a definitive predictive model. Second, the absence of independent external validation datasets restricts confirmation of the reproducibility of the identified biomarkers. Third, the three-dimensional structure of PLOD1 used for docking analysis was obtained from AlphaFold predictions rather than experimentally resolved crystallographic data, which may influence structural precision at the binding interface. In addition, molecular docking, molecular dynamics simulations, and ADMET predictions represent computational estimations and require experimental validation through biochemical and cellular assays. Furthermore, quantitative binding free-energy estimation methods such as MM-PBSA or MM-GBSA were not performed in the present study and may provide additional insight into ligand–protein interaction stability in future investigations. Finally, although PLOD1 emerged as a key ECM-associated regulator in this study, retinoblastoma progression involves complex multi-gene regulatory networks, and additional ECM-related targets should be explored in future investigations. Collectively, further validation using larger transcriptomic cohorts and experimental studies will be necessary to confirm the therapeutic relevance of luteolin-PLOD1 interactions in retinoblastoma. Future studies should validate these findings experimentally using retinoblastoma cell-line models to assess the biological effects of luteolin on tumour proliferation and extracellular matrix regulation. In addition, biochemical assays such as collagen hydroxylation or enzymatic activity measurements targeting PLOD1 will be important to confirm the predicted inhibitory interactions observed in the present computational analysis.

4. Conclusion

This study identified differentially expressed genes in retinoblastoma, with a subset of highly upregulated genes (PLOD1, COL16A1, HSPG2, ID3) driving extracellular matrix remodelling, collagen biosynthesis, and cell adhesion. SVM classification achieved 99.77% accuracy, while PPI and pathway analyses highlighted coordinated roles of PLOD1 in tumour progression. Molecular docking and dynamics simulations demonstrated that luteolin binds stably to PLOD1, enhancing structural stability with favourable drug-likeness and minimal predicted toxicity. Based on *in silico* assessments, PLOD1 represents a candidate diagnostic biomarker, and luteolin emerges as a promising therapeutic candidate, requiring further experimental validation. Future studies should validate these targets experimentally and assess luteolin's efficacy in preclinical RB models to inform targeted therapies and improve clinical outcomes.

5. Acknowledgement

The authors gratefully acknowledge the resources and academic support provided by Vignan's Foundation for Science, Technology Research/ Department of Biotechnology.

6. References

1. Hatem S, Mohammed D, Ezzat N. Nanotechnology-based strategies overcoming the challenges of retinoblastoma: a comprehensive overview and future perspectives. Futur J Pharm Sci. 2024; 10(1):14. https://fjps.springeropen.com/articles/10.1186/s43094-024-00587-4
2. Pareek A, Kumar D, Pareek A, Gupta MM, Jeandet P, Ratan Y, *et al*. Retinoblastoma: An update on genetic origin, classification, conventional to next-generation treatment strategies. Heliyon. 2024; 10(12):32844. https://doi.org/10.1016/j.heliyon.2024.e32844
3. Soebagjo HD, Fatmariyanti S, Sugianto P, Purwanto B, Bintoro UY, Kusumowidagdo ER. Detection of the calcium and ATP role in apoptosis of retinoblastoma culture cells through caspase-3 expression. Res J Pharm Technol. 2019; 12(3):1307-1314. https://doi.org/10.5958/0974-360X.2019.00219.1
4. Dubey S, Bhardwaj S, Parbhakaran P, Singh E. *In silico* prediction of pyrazoline derivatives as antimalarial agents. Asian Journal of Pharmaceutical Research. 2022; 119-124. https://doi.org/10.52711/2231-5691.2022.00018
5. Manju M, Aziz A, Rehman S. Qualitative and quantitative analysis of phytochemicals in some medicinal plants of western Himalayas. Res J Pharm Technol. 2022; 1419-1423. https://doi.org/10.52711/0974-360X.2022.00235
6. Thakur M, Sharma K, Mehta S, Rai S, Sharma I, Tripathi A. Phytochemicals, antimicrobial and antioxidant potential of methanolic extract of *Berberis aristata* roots. Res J Pharm Technol. 2020; 13(12):5763-5767. https://doi.org/10.5958/0974-360X.2020.01004.5
7. Qureshi MS, Reddy AV, Kumar GS, Nousheen L. Ethnobotanical study on medicinal plants used by traditional health practitioners and villagers of Garhphulghar Gram, Chhattisgarh, India. Asian Journal of Pharmaceutical Research. 2017; 7(2):98-105. https://doi.org/10.5958/2231-5691.2017.00017.X
8. Kokate AA, Kumavat RS, Vakhariya RR, Mohite SK. Review of multi-dimensional therapeutic strategies and potential anti- inflammatory medicinal plants. Asian Journal of Research in Pharmaceutical Sciences. 2023; 119-124. https://doi.org/10.52711/2231-5659.2023.00022
9. Debnath S, Chakravorty R, Devi D. A review on the role of medicinal plants in the immune system. Asian Journal of Pharmacy and Technology. 2020; 10(4):273-277. https://doi.org/10.5958/2231-5713.2020.00045.8
10. Kumar M, Sharma P, Sharma D, Kumar Verma K. A review article on pharmacological evaluation of ashwagandha highlighting its healthcare claims, safety, and toxicity aspects. Asian Journal of Research in Pharmaceutical Sciences. 2024; 159-166. https://doi.org/10.52711/2231-5659.2024.00025
11. Otuokere IE, Amaku FJ, Alisa CO. *In Silico* geometry optimisation, excited-state properties of (2 e)- n -hydroxy-3-[3-(phenylsulfamoyl) phenyl] prop-2-enamide (belinostat) and its molecular docking studies with Ebola virus glycoprotein. Asian Journal of Pharmaceutical Research. 2015; 5(3):131-137. https://doi.org/10.5958/2231-5691.2015.00020.9
12. Kurian T. Computational molecular docking analysis of six nitrogen-containing heterocyclic compounds from a chemical database: exploring their potential as anti-Alzheimer's agents and validation through redocking studies. Asian Journal of Pharmaceutical Research. 2024; 121-124. https://doi.org/10.52711/2231-5691.2024.00020
13. Ganesh L, B D, Manikkoth S, G Nadh A. Anxiolytic activity of coconut water by GABA modulation as evidenced by pharmacological experiments and molecular docking studies. Res J Pharm Technol. 2023: pp. 3544-3548. https://doi.org/10.52711/0974-360X.2023.00585
14. Soumya SJ, Arya KR, Abhinand CS, Nadh AG, Rani JR, Oommen OV, *et al.* Emerging paradigms in natural products-based drug discovery in: biodiversity and business. cham: Springer Nature Switzerland. 2024; 231-246. https://doi.org/10.1007/978-3-031-71674-4_14
15. Kalyan GU, Pranathi Reddy D, Chandrakanth G, Pooja B, Anitha V, Vivek D. Gene association disease prediction by the geo2r tool. in: 2023 international conference on evolutionary algorithms and soft computing techniques (EASCT). IEEE. 2023; 1-6. https://doi.org/10.1109/EASCT59475.2023.10392802
16. Haifeng Wang, Dejin Hu. Comparison of SVM and LS-SVM for regression. in: 2005 international conference on neural networks and brain. IEEE. 2005; 279-283. https://doi.org/10.1109/ICNNB.2005.1614615
17. Szklarczyk D, Gable AL, Nastou KC, Lyon D, Kirsch R, Pyysalo S, *et al.* The STRING database in 2021: customizable protein-protein networks, and functional characterisation of user-uploaded gene/measurement sets. Nucleic Acids Res. 2021; 49(D1):605-612. https://doi.org/10.1093/nar/gkaa1074
18. Saito R, Smoot ME, Ono K, Ruscheinski J, Wang PL, Lotia S, *et al.* A travel guide to Cytoscape plugins. Nat Methods. 2012; 9(11):1069-1076. https://doi.org/10.1038/nmeth.2212
19. Tipney H, Hunter L. An introduction to effective use of enrichment analysis software. Hum Genomics. 2010; 4(3):202. https://doi.org/10.1186/1479-7364-4-3-202
20. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, *et al.* Highly accurate protein structure prediction with AlphaFold. Nature. 2021; 596(7873):583-589. https://doi.org/10.1038/s41586-021-03819-2
21. Kim S, Chen J, Cheng T, Gindulyte A, He J, He S, *et al.* PubChem 2023 update. Nucleic Acids Res. 2023; 51(D1): D1373-1380. https://doi.org/10.1093/nar/gkac956

22. Nadh AG, Revikumar A, Sudhakaran PR, Nair AS. Identification of potential lead compounds against BACE1 through in-silico screening of phytochemicals of *Medhya rasayana* plants for Alzheimer's disease management. Comput Biol Med. 2022; 145:105422. https://doi.org/10.1016/j.compbiomed.2022.105422
23. Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJC. GROMACS: Fast, flexible, and free. J Comput Chem. 2005; 26(16):1701-1718. https://doi.org/10.1002/jcc.20291
24. Xiong G, Wu Z, Yi J, Fu L, Yang Z, Hsieh C, *et al*. ADMETlab 2.0: an integrated online platform for accurate and comprehensive predictions of ADMET properties. Nucleic Acids Res. 2021; 49(W1):W5-14. https://doi.org/10.1093/nar/gkab255
25. Lipinski CA. Lead- and drug-like compounds: the rule-of-five revolution. Drug Discov Today Technol. 2004; 1(4):337-341. https://doi.org/10.1016/j.ddtec.2004.11.007

Author Contributions

Conceptualisation R.J.R.; methodology, R.J.R. and G.G.; software R.R. and A.M.; validation, R.J.R. and A.G.N.; formal analysis G.G. and R.J.R.; investigation, R.J.R. and G.G.; resources, R.R. and A.M.; data curation, A.M.; writing - original draft preparation, G.G.; writing - review and editing, R.J.R. and A.G.N.; visualisation, G.G. and A.M.; supervision, R.J.R. and A.G.N. All authors have read and agreed to the published version of the manuscript