

Exploration of the Mechanism of Xinjiang Rosa laxa Retz.Fruit Extract on IgA Nephropathy by Bioinformatics and Machine Learning Methods

FENG Qianqian

Xinjiang College of Traditional Chinese Medicine

HE Yuan

Xinjiang College of Traditional Chinese Medicine

GUO Ying

Medical College,XinJiang University of Science and Technology,Korla

LI Aoqing

Xinjiang College of Traditional Chinese Medicine

Dilinur Kamil

Xinjiang College of Traditional Chinese Medicine

Li Tian

tianli109@126.com

Xinjiang College of Traditional Chinese Medicine <https://orcid.org/0000-0002-6197-5202>

Research Article

Keywords:

Posted Date: March 20th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-9095922/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at Bioresources and Bioprocessing on June 17th, 2026. See the published version at <https://doi.org/10.1186/s40643-026-01090-9>.

1 **Exploration of the Mechanism of Xinjiang *Rosa laxa***
2 **Retz.Fruit Extract on IgA Nephropathy by Bioinformatics**
3 **and Machine Learning Methods**

4 FENG Qianqian¹, HE Yuan¹, GUO Ying², LI Aoqing¹, Dilinur¹, TIAN Li^{1,3*}
5 (1.College of Traditional Chinese Medicine, Xinjiang Medical University,
6 Urumqi 830017, Xinjiang, China; 2.Medical College, Xinjiang University of Science
7 and Technology, Korla 841000, Xinjiang, China; 3.Institute of Traditional Chinese
8 Medicine, Xinjiang Medical University, Urumqi 830017, China)

9 **Abstract**

10 This study aimed to elucidate the core therapeutic targets and mechanisms of
11 action of rosehip extract (FRLE) in IgA nephropathy (IgAN) by integrating
12 bioinformatics, machine learning, and computational simulation techniques to
13 simultaneously analyze the therapeutic effects of FRLE on IgAN. Through screening
14 differentially expressed genes in IgAN patients and identifying potential target
15 proteins of the core components of FRLE, 118 overlapping targets were ultimately
16 determined as the gene set involved in the synergistic interaction between the two..
17 Through triple machine learning validation, FGR and LDHB were identified as core
18 target proteins, with ellagic acid, rutin, and hyperoside as the core active components.
19 In IgAN patients, LDHB was significantly downregulated, while FGR was
20 significantly upregulated. Both demonstrated excellent diagnostic efficacy and were
21 closely associated with the renal immune microenvironment. Molecular docking and
22 molecular dynamics simulations confirmed that the core components could stably
23 bind to FGR and LDHB. In summary, FRLE targets and modulates FGR and LDHB
24 through its core components, intervening in inflammatory pathways and immune

microenvironment imbalance, thereby providing a theoretical basis for subsequent experiments and clinical translation.

Introduction

Glomerulonephritis(GN) is one of the leading causes of chronic kidney disease and end-stage renal disease worldwide. Current therapeutic approaches are limited by the severe adverse effects associated with long-term use of corticosteroids and immunosuppressants, as well as suboptimal efficacy in some patients, which drives the search for more suitable alternative treatments(Shankar et al. 2025). Xinjiang *Rosa laxa* Retz., a plant rich in natural bioactive compounds with a favorable safety profile, is explicitly documented in the Kazakh Materia Medica for its fruit's therapeutic effect on chronic nephritis. However, the mechanisms by which *Rosa laxa* Retz. fruit ameliorates GN remain to be elucidated. In this study, bioinformatics and machine learning algorithms were employed to predict the potential core targets of *Rosa laxa* fruit extract against GN, namely FGR and LDHB. Molecular docking and molecular dynamics simulations were further conducted to validate these key targets, with a focus on evaluating the binding affinity and stability of its key compounds—ellagic acid, hyperoside, and rutin—with the targets. GO and KEGG enrichment analysis results indicate that the therapeutic effects of Xinjiang rose fruit extract may be associated with regulating pathways related to inflammatory responses, immune cell activation, and glomerular mesangial cell proliferation. Molecular docking results show that the binding energy range of two key targets, FGR and LDHB, with key compounds, ellagic acid, hyperoside, and rutin, is -5 to -8 kcal/mol, indicating

moderate binding affinity between the target proteins and compounds. This suggests certain binding potential between the target proteins and compounds, which can serve as candidate compounds for further optimization in treating IgA nephropathy (IgAN). Stable binding conformations were also observed from molecular dynamics simulations. Therefore, ellagic acid, hyperoside, and rutin in Xinjiang rose fruit extract may improve pathological damage of glomerulonephritis by inhibiting the activity of FGR and LDHB targets, thereby reducing renal inflammatory infiltration and abnormal mesangial cell proliferation.

Globally, IgA nephropathy (IgAN) is the most common subtype of primary glomerulonephritis, accounting for approximately 30%-50% of all primary glomerulonephritis patients. Notably, about 20%-30% of IgAN patients progress to end-stage renal disease (ESRD) within 10 to 20 years after diagnosis, making IgAN one of the core glomerulonephritis subtypes that develop into ESRD.(Wheeler et al. 2021).Early diagnosis of IgA nephropathy often involves the frequent oversight of hematuria or proteinuria symptoms. In countries with limited medical resources and inadequate systematic screening mechanisms, the diagnosis of IgA nephropathy is often delayed or misdiagnosed, leading to accelerated disease progression and increased risks of disability and death.(Tonelli et al. 2020).The affected population is predominantly young adults aged 20-40 years, with a higher prevalence in males than females. The male-to-female ratio of cases is approximately 2:1, and male patients exhibit a higher risk of disease progression.(Ghaddar et al. 2024).The pathogenesis of IgA nephropathy has common mechanisms with the majority of glomerulonephritis;

69 both take immune-mediated glomerular inflammation as their foundation, which is
70 caused by immune complex deposition. However, this disease has its pathological
71 characteristic as abnormal immunoglobulin A (IgA) deposition in the glomerular
72 mesangium. Hence, this deposition will later give rise to mesangial cell multiplication,
73 mesangial stromal enlargement and glomerular structural damage, thus finally causing
74 continuous renal function reduction. At present, glucocorticoids combined with
75 immunosuppressants still serve as the main clinical treatment measure for IgA
76 nephropathy. However, long-time application of such medicines can easily arouse bad
77 effects such as osteoporosis, elevated infection risk, and hepatorenal harm. According
78 to statistical data, about 30% of patients show weak reaction to current standard
79 treatment methods, and certain situations can even not effectively stop disease
80 development, thus displaying obvious limitations in treatment effect.(Zhao et al.
81 2023).Furthermore, the morbidity mechanism of IgA nephropathy is complicated,
82 which relates to a pathological network connected with immune regulation disorder
83 and metabolic abnormality. Therefore, relying only on single-target medicines often
84 cannot realize effective therapeutic interference, hence this situation shows a
85 limitation of present treatment strategies. Therefore, there is an urgent need for the
86 development of novel therapeutic agents in clinical practice that can precisely target
87 the intricate pathological mechanisms of IgA nephropathy while ensuring both
88 efficacy and high safety. This is of significant importance for the precise modulation
89 of key pathological processes, slowing disease progression, and improving long-term
90 patient outcomes.

FRLE has been documented in traditional medicine for the direct treatment of nephritis, with a common dosage of 50 – 100 g. Previous research by our group has shown that FRLE is rich in components such as ellagic acid, hyperoside, vitamin C, as well as flavonoids, polyphenols, and triterpenoid acids(Wang et al. 2025).IgAN is characterized primarily by renal immune inflammation triggered by IgA immune complex deposition, abnormal proliferation of mesangial cells, oxidative stress, and renal interstitial fibrosis. Ellagic acid, hyperoside, and vitamin C can synergistically regulate this pathological process: ellagic acid exerts anti-inflammatory effects by inhibiting the STAT3 and NF- κ B signaling pathways and exerts anti-fibrotic effects by downregulating the TGF- β 1 signaling pathway(Huang et al. 2024).hyperoside synergistically contributes to anti-inflammatory and immunomodulatory activities, reducing IgA immune complex deposition; vitamin C provides antioxidant support and assists in immune regulation. Together, these three components form a synergistic "anti-inflammatory – anti-fibrotic – antioxidant" system(Dennis et al. 2017).The preliminary research results of the study group indicate that this drug exhibits no significant hepatotoxicity or nephrotoxicity. Therefore, such drugs are more suitable for long-term maintenance therapy in patients with nephritis requiring prolonged medication. This suggests that Rosalaxa fruit has the potential as a natural therapeutic agent for adjuvant treatment of glomerulonephritis, providing a new research direction for optimizing patient prognosis in clinical practice.

A multidimensional research approach was used to study how FRLE regulates the disease process in IgAN. In IgAN, the core problematic phenomenon is the

113 generation of abnormal IgA1 immune complexes. Therefore, these complexes
114 accumulate within the mesangial region of the kidney, thus triggering inflammatory
115 reactions and organ damage. FGR is a crucial enzyme existing in immune cells; hence,
116 it exerts a function throughout the whole disease progression process, starting from
117 the initiation of immune system abnormalities and ending with the final occurrence of
118 kidney impairment. First, FGR controls signals from Fc receptors and integrins. This
119 control affects how immune cells recognize and take in the immune complexes, and it
120 also makes the inflammatory response bigger. Then, it makes mesangial inflammation
121 worse, causes mesangial cells to grow too much, and damages tubules. It does this by
122 turning on pathways like STAT3 and NOTCH1(Jing et al. 2021).In IgAN patients, the
123 deposition of immune complexes causes a lack of oxygen and inflammation in the
124 kidney. This condition makes LDHB unable to perform its normal function in lactate
125 metabolism. As a result, too much lactate builds up in the kidney. This buildup of
126 lactate then makes the inflammation worse. It also increases damage to the glomerular
127 mesangium. At the same time, it harms the kidney tubules and leads to the growth of
128 fibrous tissue in the kidney. The lactate buildup also affects the metabolism of
129 immune cells. Because of this change, these immune cells start to produce more
130 abnormal IgA1.(Fang et al. 2024).Three core chemical components of FRLE are
131 ellagic acid, hyperoside, and rutin. These compounds all have the ability to target and
132 bind to two key proteins, FGR and LDHB. Molecular docking experiments provided
133 specific binding results. Ellagic acid presented the most powerful binding strength to
134 FGR. Hyperoside exhibited the most stable binding action with LDHB. Rutin showed

fine binding capability toward both two proteins. Further kinetic simulations have confirmed the working mechanism of these components. Therefore, they stabilize the structural form of the FGR and LDHB complex; thus, they also alter the functional activity of these proteins. Hence, this enables the components to produce targeted regulatory effects. The combined function of hyperoside and rutin on both FGR and LDHB is of great importance. This joint targeting has direct connection with the inhibition of pro-inflammatory signaling pathways activation. It also has link with the correction of metabolism and immunity imbalance.

This research makes use of bioinformatics technologies for data acquisition, and these technologies supply high-dimensional data coming from genomics, proteomics, and metabolomics levels. Therefore, researchers apply machine learning algorithms to handle this complicated data; the algorithms include Support Vector Machine Recursive Feature Elimination, Least Absolute Shrinkage and Selection Operator regression, and Random Forest. Hence, these methods can carry out accurate identification of key regulatory targets from the datasets, and finding such targets is of great importance for the comprehension of *Rosa laxa* Retz. fruit extract' s functional mechanisms. One particular algorithm is Support Vector Machine Recursive Feature Elimination: it operates via an iterative process, in which in every step, features with the smallest contribution to disease classification are removed, and this process goes on until it locates the optimal target combination. The Root Mean Square Error is used to measure the prediction accuracy of this model.(Sanches et al. 2024).This research utilizes bioinformatics technologies to carry out data acquisition, and these

technologies provide high-dimensional data that comes from genomics, proteomics, and metabolomics levels. Therefore, researchers employ machine learning algorithms to deal with this complex data; such algorithms include Support Vector Machine Recursive Feature Elimination, Least Absolute Shrinkage and Selection Operator regression, and Random Forest. Hence, these methods can conduct accurate identification of key regulatory targets from the datasets, and the discovery of such targets has great significance for the understanding of *Rosa laxa* Retz. fruit extract's functional mechanisms. Support Vector Machine Recursive Feature Elimination is one specific algorithm: it runs through an iterative process, in which in each step, features that have the smallest contribution to disease classification are removed, and this process continues until it finds the optimal target combination. The prediction accuracy of this model is measured by the Root Mean Square Error.(Zhou et al. 2025).

The fast growth of computational chemistry and bioinformatics has pushed computational simulation technologies into natural product drug discovery. Two important examples are molecular docking and molecular dynamics simulation. Molecular docking technology works by predicting the best binding shape of a ligand inside a receptor's active site. It uses energy calculation parameters to measure how strongly they bind. This provides the main data needed for screening potential therapeutic targets. Molecular dynamics simulation works in a different way. It follows the interaction between receptor and ligand step by step over time. It does this at the atomic level. This method clearly shows how biomolecules change their shape during the process. It also shows how networks of interactions between molecules are

179 formed. And it reveals the main pathways through which they act.(Adediwura et al.
180 2024).Computational simulation techniques have certain advantages over traditional
181 experimental methods. They can perform microscopic analysis at the atomic level.
182 They are also useful in more situations. They are safer to use and cost less money.
183 Traditional experiments have several drawbacks. They need very long time cycles to
184 finish their work. They have extremely high cost levels. Also, they are very difficult to
185 catch dynamic processes at the moment they take place. Therefore, computational
186 techniques can solve these problems in an effective way. Hence, they provide a
187 different method to realize the molecular mechanisms about how drugs do interaction
188 with their target objects.

189 This research adopts a number of methods to investigate how FRLE could
190 possibly conduct treatment on glomerulonephritis and its working mechanisms; the
191 methods cover bioinformatics, machine learning, and computational simulation. First,
192 bioinformatics analysis searches out potential targets for FRLE, thus helping to
193 confirm which biological targets and signaling pathways may have involvement in the
194 process. Next, through screening huge volumes of biological data, machine learning
195 algorithms select out those core regulatory targets that hold the closest linkage with
196 the disease' s development, and this step is of key importance for the comprehension
197 of FRLE' s working ways. After that, molecular docking technology carries out
198 inspection on the binding degree between FRLE' s main chemical components and
199 these core targets, so as to assess their binding affinity. Finally, molecular dynamics
200 simulations perform verification on the stability of these binding conformations; they

construct molecular models and observe the interaction processes between compounds and targets across time, and therefore confirm if the binding possesses stability and specificity.

Methods

Data Processing and Differential Expression Gene Analysis

The original raw gene count data coming from GSE127136 dataset were input into R software to conduct analysis; hereby, 0 was used to replace missing values of genes inside the dataset. Therefore, after we utilized the DESeq2 software package to filter genes with low expression, analysis on differential gene expression was then carried out. Genes with significant inter-group expression differences were identified based on corrected p-values (padj) < 0.05 and $|\log_2(\text{Fold Change})| \geq 1$. Results were validated using visualization tools to generate volcano plots of differentially expressed genes. A standardized expression matrix was constructed through variance stabilization transformation (VST). To ensure the reliability of the results, differential expression analysis was repeated using different methods, with the limma software package employed for linear models and empirical Bayesian testing. Gene expression heatmaps were created using the pheatmap software package, visualizing the top five genes with increased expression and the top five genes with decreased expression.

Potential targets of FRLE against nephritis

Using the PubChem database, we searched for the 12 major compounds identified in FRLE and obtained their simplified SMILES representations. Subsequently, the SMILES representations of these 12 major compounds were

uploaded to the SwissTargetPrediction platform (<http://swisstargetprediction.ch/>). By analyzing the chemical structures of the compounds and integrating existing target information, we efficiently identified potential targets. Through the construction of a Venn diagram, we visualized the overlapping genes between FRLE-related targets and differentially expressed genes (DEGs) in IgAN. This approach revealed targets that are simultaneously involved in the pathological process of IgAN and the pharmacological effects of FRLE. These overlapping targets may represent potential therapeutic targets for FRLE's anti-nephritis effects.

GO and KEGG Analysis

GO and KEGG analyses of the intersecting genes were performed using the R software package.(Yu et al. 2012). This tool is designed for functional enrichment analysis and gene set interpretation, capable of performing functional feature analysis on encoded and non-encoded genomic data from thousands of species, with annotation information sourced from the GO Knowledge Base.(Ashburner et al. 2000) and the KEGG pathway database(Kanehisa et al. 2000). Enrichment significance was assessed via a hypergeometric test (using the Benjamini-Hochberg correction method, adjusted p-value < 0.05). The hypergeometric test was employed to estimate the probability of observing genes annotated with a specific GO term within the test set. The Benjamini-Hochberg method was applied to control the false discovery rate and correct for multiple testing. The top 10 enriched cellular components, molecular functions, biological processes, and KEGG pathways were visualized using SRplot(Tang et al. 2023).

GSVA Analyse

This study first analyzed the gene expression profiles of the GSE127136 dataset, categorizing the samples into an IgAN group and a NC group to identify DEGs between the two groups. Subsequently, the potential target genes of RLFE compounds were obtained. These targets were intersected with the aforementioned DEGs to derive a core gene set for investigation. Based on this core gene set, GSVA was employed in conjunction with the Hallmark gene sets from the MSigDB database(Liberzon et al. 2015) to calculate pathway activity scores for each sample(Hänzelmann et al. 2013).The differential pathways were subsequently analyzed using the limma package. Pathway changes were assessed by t-values (cut-off value=1.96, corresponding to $P < 0.05$). Finally, pathway activity heatmaps and differential pathway bar plots were visualized using the pheatmap and ggplot2 packages, respectively.

Machine Learning Analysis

Based on the intersection genes, core regulatory genes were screened using three machine learning algorithms, and a lasso regression model was constructed with the glmnet software package.(Friedman et al. 2010), The optimal regularization parameter (λ) was selected through 10-fold cross-validation to identify key genes with non-zero coefficients. A random forest classification model was constructed using the randomForest package(Qing et al. 2025) with gene importance assessed through the mean Gini coefficient (MeanDecreaseGini). Fifteen core genes were screened and visualized as bubble plots. The SVM-RFE function was employed to analyze model

accuracy and error rates across different feature quantities via 10-fold cross-validation, thereby determining the optimal feature set. Feature ranking bar charts and performance trend plots were generated. Finally, the screening results from the three methods were integrated using a Venn diagram and consistency analysis to validate the reliability of core biomarkers. custom SVM-RFE function was employed, combined with 10-fold cross-validation to analyze model accuracy and error rates across different feature quantities, thereby determining the optimal feature set. Feature ranking bar charts and performance trend plots were also generated. Finally, the screening results from the three methods were integrated using a Venn diagram and consistency analysis to validate the reliability of core biomarkers.

Differential analysis of core genes, immune infiltration

Based on the identification of two key genes, FGR and LDHB, through three machine learning methods, their expression levels were extracted from the two sample groups. We utilized box plots and violin plots to visualize the expression levels and distribution characteristics of differential genes. The Wilcoxon rank sum test was employed to analyze whether the intergroup differences were statistically significant. Additionally, the receiver operating characteristic (ROC) curve was plotted, and the area under the curve (AUC) along with its 95% confidence interval were calculated to evaluate diagnostic efficacy. Single-sample gene set enrichment analysis (ssGSEA) was employed to compute immune cell infiltration scores. Spearman correlation analysis was used to investigate the association between FGR and LDHB expression and immune cell infiltration, with results visualized via heatmaps and scatter

plots(Ren et al. 2025). Finally, immunohistochemical images from the Human Protein Atlas (HPA) database were utilized to validate differences in protein expression of the two genes between normal and diseased renal tissues.

Molecular docking

The three main 3D structures of tannic acid, luteolin, and rutin in FRLE were retrieved and downloaded using the PubChem database. The ligand-protein files were processed with AutoDockTools 1.5.7 software, including ligand hydrogenation and docking type selection, and the final ligand files were exported in PDBQT format.(Che et al. 2024). Utilize D3Pocket (<https://www.d3pharma.com>) to predict and identify potential binding pockets of receptor proteins, score these pockets, and select regions with higher pocket scores as targets for subsequent molecular docking. Perform molecular docking using Autodock Vina, adopting the conformation with the lowest binding energy from each receptor-ligand docking result as the stable conformation to comprehensively evaluate ligand conformation and its interaction with the protein binding pocket. Finally, visualize the docking results using PyMOL software to intuitively demonstrate the interaction patterns between the ligand and the receptor protein.

Molecular Dynamics Simulation

Based on the optimal binding conformations of tannic acid, tannins, and rutin with the target proteins FGR and LDHB obtained through molecular docking, we performed molecular dynamics simulations of the optimal conformations using the Gromacs software. The specific steps included ligand/protein preprocessing, complex

construction, system solvation, ion neutralization, energy minimization, and NVT/NPT equilibrium, followed by a 20-nanosecond dynamics simulation. Multidimensional trajectory analyses, including RMSD, RMSF, and FES, were employed to validate the dynamic binding stability and interaction patterns of the complexes(Yuan et al. 2025).

Result

Prediction of biological targets of FRLE in the treatment of IgAN

The volcano plot visually presents the magnitude of gene expression changes and their statistical significance, providing a systematic summary of the distribution characteristics of DEGs. A total of 1350 DEGs were identified, comprising 721 upregulated genes and 629 downregulated genes (Fig. 1A) . A heatmap was used to display the top 15 upregulated and downregulated genes (Fig. 1B). The SwissTargetPrediction platform was employed to predict the target proteins of the core chemical components in FRLE, including tilianin, hyperoside, quinic acid, rutin, oleanolic acid, ellagic acid, kaempferol, kaempferol-3-O-rutinoside, euscaphic acid, isoquercitrin, pomolic acid, and astragalin. This analysis identified 817 potential target proteins. Cross-referencing these target proteins with the DEGs associated with IgA nephropathy revealed 118 overlapping targets (Fig. 1C).

Functional enrichment analysis of the 118 overlapping target proteins indicated their primary enrichment in inflammation- and immunity-related pathways, such as the integrin signaling pathway and the NF- κ B signaling pathway (Fig. 1D). The integrin signaling pathway contained the highest number of enriched target proteins, establishing it as one of the core pathways through which FRLE intervenes in IgA

334 nephropathy. Through analysis of the GO result(Fig. 1E), it was demonstrated that on
335 the biological process level, these overlapping genes participated in
336 immune-inflammatory processes, for instance leukocyte migration and inflammatory
337 response regulation. At the level of cellular component, these overlapping genes were
338 mainly positioned at adhesion-related constructions such as integrin complexes.
339 Therefore, at the molecular function level, these overlapping genes mostly carried out
340 regulation of signal transduction via integrin binding and cytokine receptor binding.
341 (Fig. 1F).

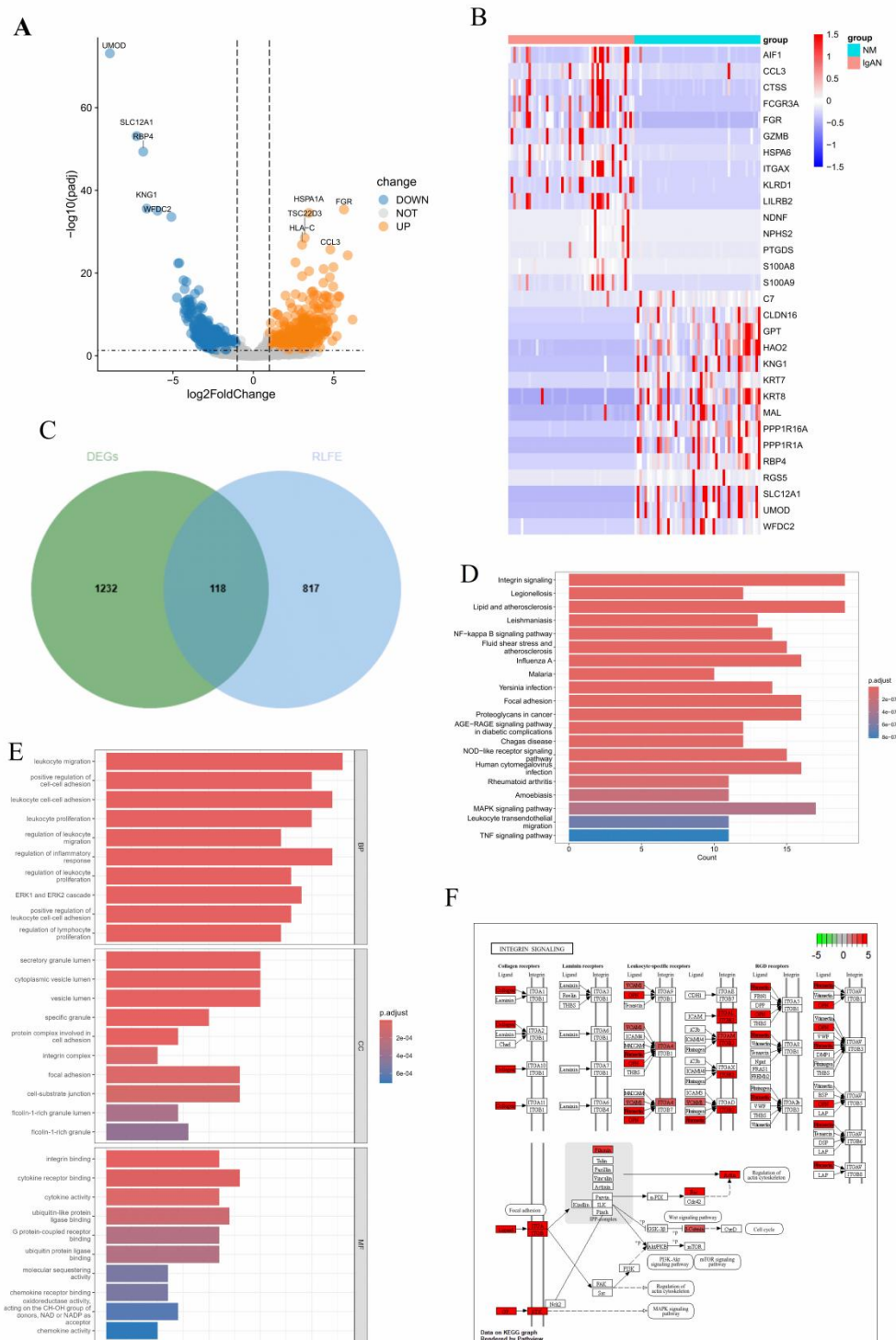


Fig. 1 Analysis of differential genes, target protein intersections, and functional enrichment following FRLE intervention in IgA nephropathy. (A: Volcano plot of differentially expressed genes (DEGs) in IgA nephropathy; B: Heatmap showing the expression of core differentially expressed genes in the normal group (NM) versus the IgAN group; C: Venn diagram illustrating the intersection between IgA nephropathy DEGs and potential target proteins of RLFE; D: Bar chart of KEGG pathway

enrichment analysis for the overlapping target proteins; E: Bar chart of GO functional enrichment analysis for the overlapping target proteins; F: Visualization of the integrin signaling pathway.)

Validation results from Gene Set Variation Analysis (GSVA) of patient transcriptomic data showed significant upregulation of inflammatory pathways, including the NF- κ B pathway, in IgA nephropathy samples, which strongly corroborated the target protein enrichment findings. The final results indicate that the active components of FRLE can exert a potential therapeutic effect on IgA nephropathy by targeting and modulating the integrin signaling pathway and its downstream inflammatory pathways, such as NF- κ B and MAPK, thereby correcting the pathway imbalances in patients (Fig. 2A, B).

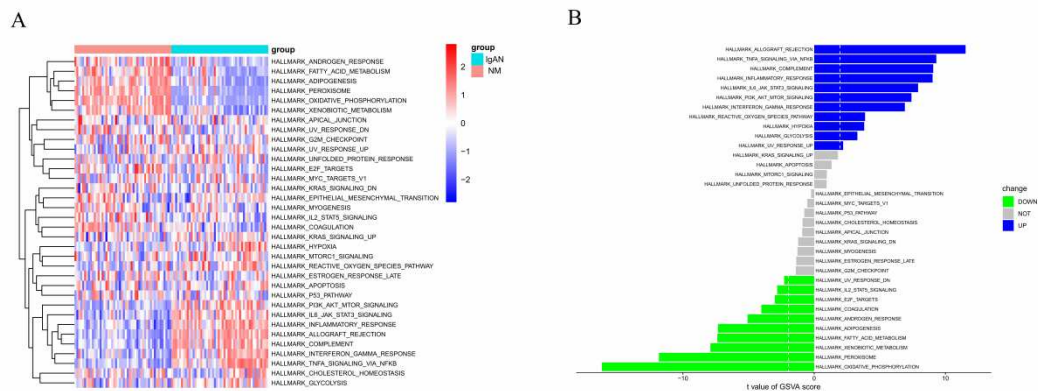


Fig. 2 Differences in GSVA Pathway Activity Between the IgAN Group and the Normal Group (NM)(A: Heatmap of pathway activity in the IgAN group versus the normal group (NM); B: Plot of GSVA pathway enrichment results for IgAN samples.)

FGR and LDHB are core target proteins

To precisely identify the core active components and key targets of FRLE in intervening in IgAN, three machine learning algorithms—LASSO regression, RF, and SVM-RFE—were employed for screening and validation. LASSO regression determined the optimal parameters via coefficient trajectory plots and deviance curves,

369 screening out 25 core features (Fig. 3A, B, C). SVM-RFE and Random Forest were
370 further applied for validation: the former screened 20 optimal features (Fig. 3D, E),
371 while the latter selected 23 optimal feature targets (Fig. 3F, G, H), all of which
372 pointed to inflammation- and adhesion-related proteins. A Venn diagram of the
373 multi-algorithm cross-validation revealed that two core targets—FGR and LDHB—
374 were identified as the common intersection among the three algorithms (Fig. 3K). By
375 intersecting the active components of FRLE with the key genes obtained from the
376 three algorithms, the core components in FRLE for treating IgAN were analyzed. The
377 results indicated that ellagic acid, rutin, and hyperoside exhibited the closest
378 associations with the core targets screened by each algorithm (Fig. 3I, J). This
379 approach precisely pinpointed the core active components and key targets of FRLE in
380 intervening in IgAN, providing a clear material basis and target direction for
381 subsequent in vitro and in vivo experimental validation.

the SVM-RFE feature elimination process; G: SVM-RFE cross-validation performance curve; H: Top feature ranking plot selected by SVM-RFE; I: Chord diagram illustrating the association between core active components and the screened targets; J: Heatmap showing the associations between targets screened by different algorithms and active components; K: Venn diagram of targets screened by the three algorithms.)

Differential Analysis of Core Target Proteins

Through differential expression analysis between patients with IgA nephropathy and normal controls, it was found that the core gene LDHB was significantly downregulated in samples from IgA nephropathy patients, while FGR was significantly upregulated (Fig. 4A, B). Further diagnostic efficacy evaluation revealed that both LDHB (AUC=0.938) and FGR (AUC=0.907) exhibited excellent disease-discriminating ability (Fig. 4C, D). These findings not only confirm the relevance of these core targets in IgA nephropathy but also suggest their potential as diagnostic biomarkers. Furthermore, they provide supporting evidence for the mechanism by which FRLE exerts its therapeutic effects, namely by targeting and modulating these core inflammatory targets to correct pathway imbalances.

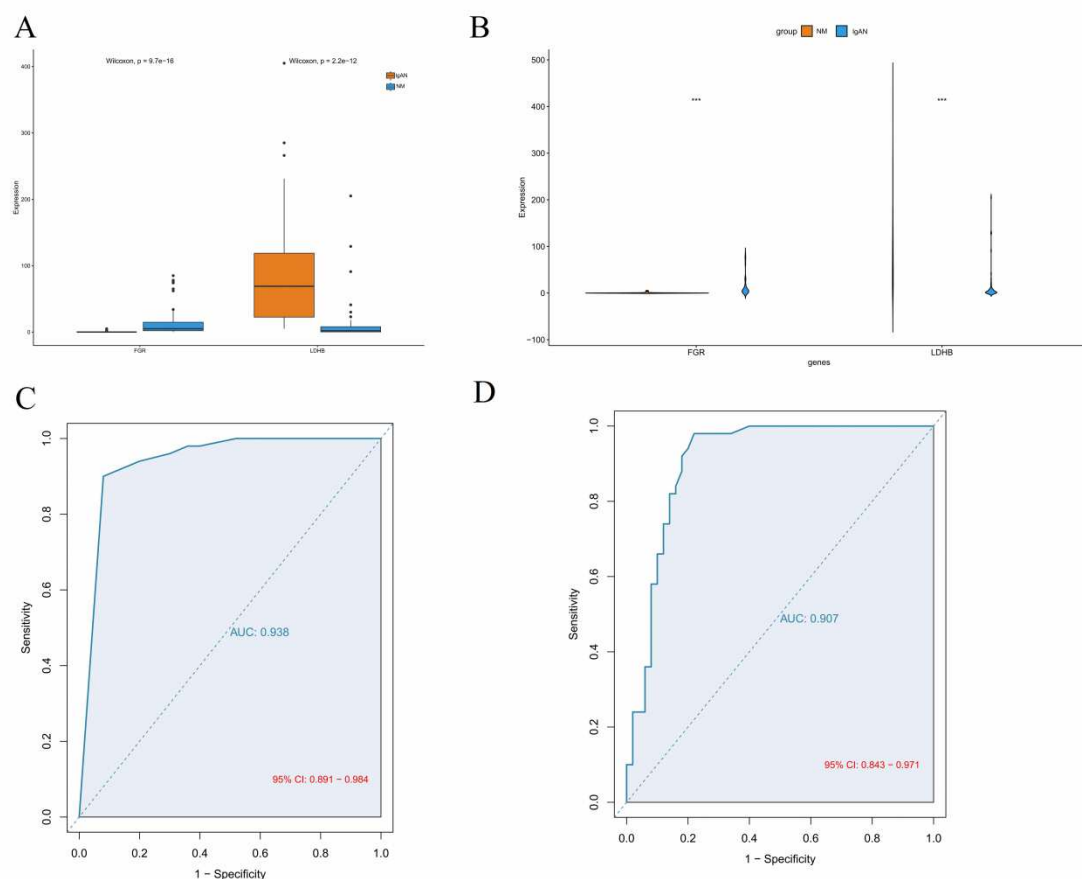


Fig. 4 Expression Differences and Diagnostic Performance Evaluation of the Core Target Proteins FGR and LDHB in IgA Nephropathy (A: Box plot of FGR and LDHB expression in the IgAN group and the NM group; B: Violin plot of FGR and LDHB expression in the IgAN group and the NM group; C: Receiver operating characteristic (ROC) curve for LDHB; D: Receiver operating characteristic (ROC) curve for FGR)

Regulatory Role of Immunomicroenvironment and Protein Level Verification of Core Target FGR and LDHB

Based on the differential expression characteristics of the aforementioned core targets FGR upregulated in IgA nephropathy and LDHB downregulated in IgA nephropathy, their association with the immune microenvironment was further investigated and validated at the protein level: The gene - immune cell correlation heatmap (Fig. 5A) revealed that both FGR and LDHB were significantly negatively correlated with the infiltration levels of immune cells such as activated dendritic cells

and plasmacytoid dendritic cells. Combined with the background information from the inter-immune cell correlation heatmap (Fig. 5B), this suggests that both may be involved in the inflammatory process by regulating immune cell infiltration.

Compared with the normal control group, FGR expression was significantly upregulated in the IgAN patient group, while the ssGSEA infiltration scores for activated DCs/plasmacytoid DCs showed a downregulated trend. This suggests that the IgAN samples may predominantly consist of patients in the chronic stage, where prolonged inflammation leads to DC functional exhaustion, and the body initiates excessive compensatory mechanisms to suppress uncontrolled inflammation. Within the IgAN patient group, Spearman correlation analysis demonstrated a significant negative correlation between FGR expression and the infiltration scores of activated DCs/plasmacytoid DCs ($r=-0.38/-0.30$, $p < 0.05$), further validating this compensatory regulatory feature (Fig. 5C, D).

Compared with the normal control group, LDHB expression was significantly downregulated in the IgAN patient group. This result is consistent with the metabolic reprogramming characteristics under the chronic inflammatory state of IgAN and represents an active mechanism by which the body adapts to immune dysregulation through the regulation of lactate metabolism(Luo et al. 2026).Within the IgAN patient group, Spearman correlation analysis showed that LDHB expression was strongly and significantly negatively correlated with the ssGSEA infiltration scores of activated DCs, plasmacytoid DCs, and regulatory T cells ($r = -0.58, -0.63, -0.56$, all $p < 0.001$). The strength of these associations far exceeded that observed for FGR, suggesting that

LDHB is a central hub for immune regulation in IgAN (Fig. 5E-G).

Utilizing immunohistochemistry data (HPA: <https://www.proteinatlas.org/>) from the database (Fig. 5H-I), (antibody IDs: HPA002024, CAB004641), we analyzed two independent sets of LDHB immunohistochemical validation data from the HPA database (HPA002024, CAB004641). The results corresponding to HPA002024 showed weak positive staining for FGR in normal kidney tissue, whereas moderate to strong positive staining was observed in renal carcinoma tissue. In the results corresponding to CAB004641, LDHB exhibited strong positive staining (dark brown) in normal kidney tissue, which was significantly attenuated (light brown/pale staining) in renal carcinoma tissue (Fig. 5J-K). Both independent validation results consistently confirmed a significant upregulation of FGR protein expression in renal lesion tissues. This finding not only corroborates the conclusions from the transcriptomic data at the protein level but also provides reliable multi-dimensional evidence supporting the mechanism by which FRLE targets and regulates LDHB to intervene in renal inflammatory injury.

472 with FGR revealed distinct differences in their binding potential to the FGR kinase.
473 Ellagic acid (Fig. 6A). Rutin (Fig. 6C) binds through its sugar part working together
474 with its main flavonoid part. This combined action allows it to bind in a stable way,
475 and it has the second strongest binding potential. Hyperoside (Fig. 6B) binds using
476 only hydrogen bonds and van der Waals forces. It has the least stable binding among
477 them.

478 Analysis of the molecular docking results for ellagic acid, hyperoside, and rutin
479 with LDHB also indicated significant differences in binding stability. Hyperoside (Fig.
480 6E) made the most stable binding network with LDHB. It uses hydrogen bonds, π
481 -cation interactions, and van der Waals forces. No bad interactions were seen. Rutin
482 (Fig. 6F) has the second most stable binding. Its two-sugar structure gives many polar
483 spots for binding. It uses hydrogen bonds and van der Waals forces. But there is one
484 weak bad acceptor-acceptor interaction. Ellagic acid (Fig. 6D) binds with the least
485 stability. It forms hydrogen bonds and van der Waals interactions. But two bad
486 acceptor-acceptor interactions happen with it.

497 conformational stabilizing effects of the ligands on the protein complexes (Fig. 7A-N).
498 After the binding of rutin with LDHB, the RMSD curve exhibited a stable trend, and
499 the PCA conformation showed a highly aggregated state, indicating excellent
500 conformational stability. The binding of hyperoside with LDHB also effectively
501 reduced conformational fluctuations, demonstrating comparable binding stability. For
502 FGR, after binding with hyperoside or rutin, the RMSD fluctuations are significantly
503 reduced, and the clustering of PCA conformations is markedly enhanced; both ligands
504 show outstanding stabilizing effects on the FGR complex. In summary, rutin exhibits
505 the optimal binding stability with LDHB, while hyperoside and rutin show similar
506 binding stability with FGR. Both ligands can efficiently enhance the conformational
507 stability of the LDHB and FGR complexes, demonstrating excellent binding
508 performance.

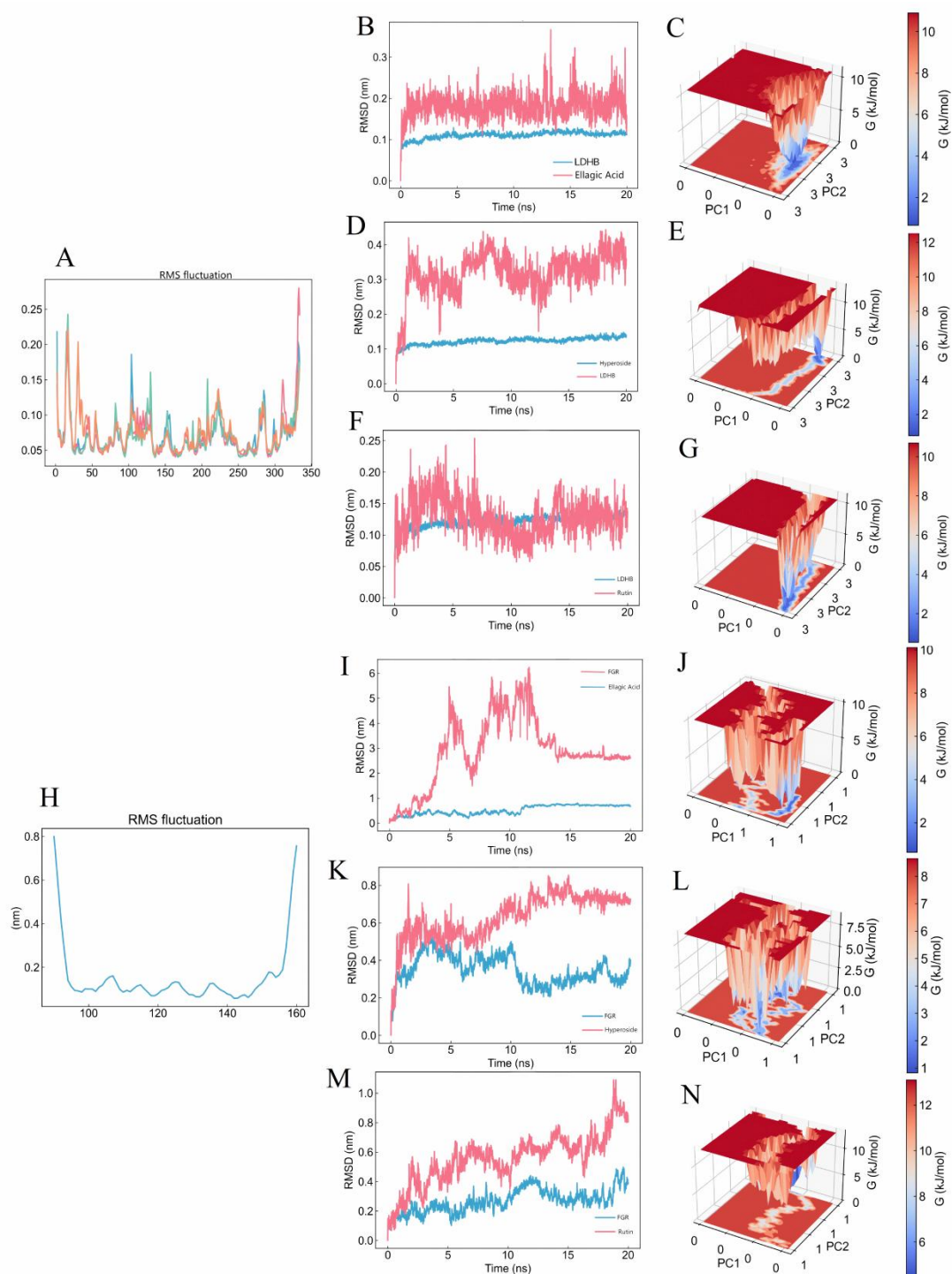


Fig. 7 Stability analysis of molecular dynamics simulations for the complexes formed between FRLE core components and FGR or LDHB. (A: RMS fluctuation curve of LDHB protein residues; B: RMSD curve of the LDHB–ellagic acid complex; C: PCA conformational distribution plot of the LDHB–ellagic acid complex; D: RMSD curve of the LDHB–hyperoside complex; E: PCA conformational distribution plot of the LDHB–hyperoside complex; F: RMSD curve of the LDHB–rutin complex; G: PCA conformational distribution plot of the LDHB–rutin complex; H: RMS fluctuation

curve of FGR protein residues; I: RMSD curve of the FGR–ellagic acid complex; J: PCA conformational distribution plot of the FGR–ellagic acid complex; K: RMSD curve of the FGR–hyperoside complex; L: PCA conformational distribution plot of the FGR–hyperoside complex; M: RMSD curve of the FGR–rutin complex; N: PCA conformational distribution plot of the FGR–rutin complex)

Conclusion

The pathological core of IgAN is a complex network interwoven with immune dysregulation and metabolic imbalance. Nowadays, main clinical treatment plans first pay attention to symptom easing and functional support, and they do not directly aim at the disease’ s root reason or occurrence mechanism.(Liu et al. 2026). Natural products have intrinsic advantages of “multi-component, multi-target, and synergistic regulation.” These advantages make them an important research field for IgAN therapy. In detail, this study displayed how FRLE acts at molecular level to interfere with IgAN. Therefore, thus, the research findings prove FRLE ’ s various characteristics; these characteristics indicate its special potential as a natural product for IgAN treatment.

The “multi-component, multi-target” action mode of FRLE matches with the complex pathological mechanisms of gAN. Previous studies (by relevant researchers, year unknown) have identified 12 key active components inside FRLE, which are capable to retrieve 817 potential proteins. Therefore, by making intersection between these 817 potential proteins and the differentially expressed genes in IgAN, we can obtain 118 overlapping proteins. Hence, analysis result shows that these overlapping proteins are enriched in core inflammatory and immune pathways such as integrins, NF- κ B, and MAPK. These results are consistent with the pathological process of

IgAN, which involves "immune activation, adhesion abnormalities, and inflammatory damage." Multiple components in FRLE can synergistically regulate the excessive growth of immune cells and mesenchymal cells, as well as the release of inflammatory factors, thereby fundamentally correcting the pathway abnormalities in IgAN. This "multi-component, multi-target" action mode represents a unique advantage of natural products and provides a new direction to address the challenge of "limited efficacy of single-intervention approaches" in the treatment of IgAN.

Machine learning algorithms were used together to find core targets. FRLE pulled out FGR and LDHB as these core targets. In IgAN, FGR shows high expression and LDHB shows low expression. This is an unusual condition of one rising and one falling. Both targets can very well distinguish IgAN samples from normal samples; their AUC values reach 0.907 and 0.938. Therefore, they are key markers of the disease and key action points for treatment. The main components of FRLE are hyperoside, rutin, and ellagic acid. These components can lock onto FGR and LDHB and bind with them tightly. Among them, ellagic acid has the best binding effect with FGR. Hyperoside and rutin can greatly increase the shape stability of the FGR-LDHB complex. This kind of binding mode, in which specific components attach to specific targets, maintains the exact effect of FRLE. Hence, it prevents possible side effects caused by binding to incorrect targets. FRLE can correct the abnormal expression of FGR and LDHB, reducing the overly high FGR level and increasing the overly low LDHB level. Moreover, FRLE is consistent with FGR's function in immune regulation, where FGR has an interactive relationship with other

factors. Therefore, FRLE interferes to inhibit inflammation by blocking FGR ' s main function of promoting inflammation; at the same time, it maintains immune system balance through LDHB, which connects metabolism and immune function.

The active components with medicinal value in natural products are mostly extracted from plants that have undergone prolonged artificial selection and practical validation. These plants are typically those that have been used and screened over time in traditional medicine or folk practices, demonstrating certain efficacy and high safety profiles for human use.(Xie et al. 2025). As a natural product, FRLE holds promising prospects in clinical applications and other fields.In comparison with single-drug therapy, natural products hold greater suitability for long-term pharmacological interference in chronic kidney disease patients (CKD). Based on the research results of this study, rosehip fruit shows pharmacological capabilities including anti-inflammation, anti-oxidation, and kidney protection effects. FRLE can aim at key targets FGR and LDHB to interfere in the pathological progress of immune dysregulation and metabolic disorder in IgA nephropathy (IgAN); therefore, thus it realizes a triple function of “anti-inflammation, immune regulation, and metabolic control.”

This research used a mixed investigation method that combines high-throughput sequence screening, machine learning verification, molecular docking, and kinetic simulation to explain FRLE ' s treatment mechanism in IgAN. Therefore, the research found FRLE ' s core active components and key targets that affect IgAN, meanwhile, it also provides a method framework for mechanism researches of other natural

products. Hence, the built standards for FRLE' s targeted regulation of core targets
FGR and LDHB supply possible diagnostic chances for IgAN, thus showing clinical
application value to realize precise treatment in this disease.

This study also has certain limitations. The direct treatment effects of FRLE on
IgAN and the cooperative effects of its core parts have not received validation via in
vitro and in vivo tests. Therefore, future research must center on performing related
cell experiments to verify FRLE' s influence on FGR and LDHB expression, as well
as its adjustment effects on inflammatory pathways. Hence, in addition, animal
experiments should be implemented to build an IgAN animal model for assessing its
kidney protection effects; after that, preclinical safety evaluations must be done, thus
laying a more firm basis for clinical translation. In summary, as a natural product,
FRLE demonstrates significant potential value in IgAN treatment by virtue of its
unique advantages: "multi-component synergy, multi-target precision, adaptation to
pathological complexity, and good biocompatibility." This study systematically
elucidated its mechanism of action, providing a new theoretical basis for the
development and utilization of natural products and opening a new direction for the
precise treatment of IgAN.

Abbreviations

Rosa laxa Retz.Fruit Extract	RLFE
IgA nephropathy	IgAN
Glomerulonephritis	GN
end-stage renal disease	ESRD

607	immunoglobulin A	IgA
608	Support Vector Machine Recursive Feature Elimination	SVM-RFE
609	Least Absolute Shrinkage and Selection Operator	LASSO
610	Random Forest	RF
611	Root Mean Square Error	RMSE
612	Differentially expressed genes	DEGs
613	variance-stabilizing transformation	VST
614	receiver operating characteristic	ROC
615	area under the curve	AUC
616	Single-sample gene set enrichment analysis	ssGSEA
617	Human Protein Atlas	HPA

618 **Acknowledgements**

619 Not applicable.

620 **Author contributions**

621 All authors participated in the design of the experiment. FQQ: Writing—original draft, Writing—review & editing.

622 HY: Formal Analysis, Validation. GY: Data curation. LAQ: Writing—original draft. DLN: Validation, Supervision.

623 TL: Funding acquisition, Conceptualization.

624 **Funding**

625 This work was supported by the Natural Science Foundation of Xinjiang Uygur Autonomous Region

626 (2025D01D27 and the National Natural Science Foundation of China (82460821).

627 **Data availability**

628 The datasets used and/or analyzed during the current study are available from the corresponding author on

reasonable request.

Declarations

Ethics approval and consent to participate

No humans or animals were used in this study.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests

References

Adediwura VA,Koirala K,Do HN,Wang J, Miao Y (2024) Understanding the impact of binding free energy and kinetics calculations in modern drug discovery. *Expert Opin Drug Discov.* 19(6) pp 671–682.

Ashburner M,Ball CA,Blake JA,Botstein D,Butler H,Cherry JM,Davis AP,Dolinski K,Dwight SS,Eppig JT,Harris MA,Hill DP,Issel-Tarver L,Kasarskis A,Lewis S,Matese JC,Richardson JE,Ringwald M,Rubin GM, Sherlock G (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet.* 25(1) pp 25–9.

Che H,Zhang R,Wang X,Yu H,Shi X,Yi J,Li J,Qi Q,Dong R, Li Q (2024) Ultrasound-assisted extraction of polyphenols from *Phyllanthi Fructus*: Comprehensive insights from extraction optimization and antioxidant activity. *Ultrason Sonochem.* 111(107083).

Dennis JM, Witting PK (2017) Protective Role for Antioxidants in Acute Kidney Disease. *Nutrients.* 9(7) pp.

Fang Y,Li Z,Yang L,Li W,Wang Y,Kong Z,Miao J,Chen Y,Bian Y, Zeng L (2024) Emerging roles of lactate in acute and chronic inflammation. *Cell Commun Signal.* 22(1) pp 276.

Friedman J,Hastie T, Tibshirani R (2010) Regularization Paths for Generalized Linear Models via Coordinate Descent. *J Stat Softw.* 33(1) pp 1–22.

Ghaddar M,Canney M, Barbour SJ (2024) IgA Nephropathy: Epidemiology and Disease Risk Across the World. *Semin Nephrol.* 44(5) pp 151564.

Hänzelmann S,Castelo R, Guinney J (2013) GSEA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics.* 14(7).

- Huang PY,Hsieh YH,Ting YH, Lee CC, Tsai JP (2024) Ellagic acid ameliorates renal fibrogenesis by blocking epithelial-to-mesenchymal transition. *Tzu Chi Med J.* 36(1) pp 59–66.
- Jing X,Ren D,Gao F,Chen Y,Wu X,Han Y,Han Q,Li L,Wang X,Tang W, Zhang Y (2021) Gene deficiency or pharmacological inhibition of PDCD4-mediated FGR signaling protects against acute kidney injury. *Acta Pharm Sin B.* 11(2) pp 394–405.
- Kanehisa M, Goto S (2000) KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 28(1) pp 27–30.
- Liberzon A,Birger C,Thorvaldsdóttir H,Ghandi M,Mesirov JP, Tamayo P (2015) The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst.* 1(6) pp 417–425.
- Liu H,Yao Y,Zhang C, Xiong J (2026) Metabolic Checkpoints in IgA Nephropathy: From Pathogenesis to Precision Medicine. *Drugs.* 86(3) pp 301–319.
- Luo Z,Li K,Yu Y,Liu Y,Weng H,Zeng XT,Zheng Y, Li W (2026) LDHB Deficiency in Fibroblasts Induces Lactate-Mediated Inflammatory Reprogramming that Promotes Breast Cancer Metastasis. *Cancer Res.*
- Qing J,Zhao Y,Cheungpasitporn W,Miao J,Thongprayoon C,Wu H, Wu J (2025) Single-cell transcriptomic analysis reveals T cell heterogeneity and metabolic reprogramming in human immune-mediated glomerulonephritis. *Autoimmunity.* 58(1) pp 2582720.
- Ren Y,Li Y,Wang Z,Cui Y,Xing Z,Zhang Y,Cao N,Yu Y,Guo Y, Li X (2025) Integrative bioinformatics and experimental validation unveil CRISP3 as a hypoxia-, epithelial mesenchymal transition-, and immune-related prognostic biomarker and therapeutic target in breast cancer. *Front Immunol.* 16(1634399).
- Sanches PHG,de Melo NC,Porcari AM, de Carvalho LM (2024) Integrating Molecular Perspectives: Strategies for Comprehensive Multi-Omics Integrative Data Analysis and Machine Learning Applications in Transcriptomics, Proteomics, and Metabolomics. *Biology (Basel).* 13(11) pp.
- Shankar M, Lamech TM (2025) Treatment burden in glomerular diseases: advances and challenges in immunosuppressive therapy. *Front Nephrol.* 5(1545373).
- Tang D,Chen M,Huang X,Zhang G,Zeng L,Zhang G,Wu S, Wang Y (2023) SRplot: A free online platform for data visualization and graphing. *PLoS One.* 18(11) pp e0294236.
- Tonelli M, Dickinson JA (2020) Early Detection of CKD: Implications for Low-Income, Middle-Income, and High-Income Countries. *J Am Soc Nephrol.* 31(9) pp 1931–1940.
- Wang N, Tian L (2025) Study on the absorption characteristics of euscaphic acid and tiliroside in fruits of *Rosa laxa* Retz. *PeerJ.* 13(e18638).

- Wheeler DC,Toto RD,Stefánsson BV,Jongs N,Chertow GM,Green T,Hou FF,McMurray JJV,Pecoits-Filho R,Correa-Rotter R,Rossing P,Sjöström CD,Umanath K,Langkilde AM, Heerspink HJL (2021) A pre-specified analysis of the DAPA-CKD trial demonstrates the effects of dapagliflozin on major adverse kidney events in patients with IgA nephropathy. *Kidney Int.* 100(1) pp 215–224.
- Xie S,Zhan F,Zhu J,Xu S, Xu J (2025) The latest advances with natural products in drug discovery and opportunities for the future: a 2025 update. *Expert Opin Drug Discov.* 20(7) pp 827–843.
- Yu G,Wang LG,Han Y, He QY (2012) clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics.* 16(5) pp 284–7.
- Yuan Y,Liu L,Chen N,Tang S, Qi Y (2025) Shenqi compound ameliorates renal fibrosis in diabetic kidney disease by inhibiting LRG1/TGF- β /Smad signaling pathway. *Phytomedicine.* 148(157355).
- Zhao H,Li Y,Sun J,Xu G,Wang C,Zhou S,Nie S,Li Y,Su L,Chen R,Cao Y,Gao Q,Hu Y,Li H,Yang Q,Wan Q,Liu B,Xu H,Li G,Weng J,Chen C,Liu H,Shi Y,Zha Y,Kong Y,Su G,Tang Y,Zhou Y,Guo A,Gong M,Xu X, Hou FF (2023) Immunosuppression versus Supportive Care on Kidney Outcomes in IgA Nephropathy in the Real-World Setting. *Clin J Am Soc Nephrol.* 18(9) pp 1186–1194.
- Zhou H,Li X,Zhang Y,Wei F,Liu Z,Zhao Y,Zhuang X,Liu X, Zhou H (2025) Machine learning combined multi-omics analysis to explore key oxidative stress features in systemic lupus erythematosus. *Front Immunol.* 16(1567466).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [GraphicalAbstract.png](#)