



OPEN ACCESS

EDITED BY

Deborah Giordano,
National Research Council (CNR), Italy

REVIEWED BY

Jiachen Liu,
Washington University in St. Louis, United States
Uzma Bashir,
University of Poonch Rawalakot, Pakistan

*CORRESPONDENCE

Zhaohui Fang,
✉ fzh2024@yeah.net.com

RECEIVED 13 January 2025

ACCEPTED 25 March 2025

PUBLISHED 14 April 2025

CITATION

Hu T and Fang Z (2025) Explore potential immune-related targets of leeches in the treatment of type 2 diabetes based on network pharmacology and machine learning. *Front. Genet.* 16:1554622. doi: 10.3389/fgene.2025.1554622

COPYRIGHT

© 2025 Hu and Fang. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Explore potential immune-related targets of leeches in the treatment of type 2 diabetes based on network pharmacology and machine learning

Tairan Hu¹ and Zhaohui Fang^{2*}

¹Anhui University of Traditional Chinese Medicine, Hefei, China, ²Department of Cardiology, First Hospital Affiliated to Anhui University of Traditional Chinese Medicine, Branch of National Clinical Research Center for Chinese Medicine Cardiology, Institution of Cardiovascular Disease, Anhui Academy of Chinese Medicine Sciences, Hefei, China

Introduction: Type 2 diabetes mellitus (T2DM) is a chronic metabolic disorder that poses a significant global health burden due to its profound effects on systemic physiological homeostasis. Without timely intervention, the disease can progress insidiously, leading to multisystem complications such as cardiovascular, renal, and neuropathic pathologies. Consequently, pharmacological intervention becomes crucial in managing the condition. Leeches have been traditionally used in Chinese medicine for their potential to inhibit the progression of T2DM and its associated complications; however, the specific mechanisms underlying their action and target pathways remain poorly understood. The objective of this study was to predict potential therapeutic targets of leeches in the treatment of T2DM.

Methods: We collected active components and targets associated with leeches from four online databases, while disease-related targets were sourced from the GeneCards and OMIM databases. Following this, we performed Gene Ontology (GO) analysis and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis. Gene expression data were obtained from the GSE184050 dataset. Important immune cell types were identified through immunoinfiltration analysis in conjunction with single sample enrichment analysis (ssGSEA). Additionally, weighted co-expression network analysis (WGCNA) was utilized to identify significantly associated genes. Finally, we employed LASSO regression, SVM-RFE, XGBoost, and random forest algorithms to further predict potential targets, followed by validation through molecular docking.

Results: Leeches may influence cellular immunity by modulating immune receptor activity, particularly through the activation of RGS10, CAPS2, and OPA1, thereby impacting the pathology of Type 2 Diabetes Mellitus (T2DM).

Discussion: However, it is important to note that our results lack experimental validation; therefore, further research is warranted to substantiate these findings.

KEYWORDS

leech, type 2 diabetes, network pharmacology, immune infiltrate, machine learning

1 Introduction

Type 2 diabetes is a chronic disease that seriously threatens human health (Kautzky-Willer et al., 2016). Current studies suggest that the pathogenesis of type 2 diabetes mellitus is primarily attributed to both insulin resistance and impaired systemic capacity to metabolize surplus glucose (Lee et al., 2022). Diminished cellular glucose oxidation (DeFronzo, 2004) and dysregulated mitochondrial bioenergetics (Rovira-Llopis et al., 2017) constitute significant contributors to the pathogenesis of type 2 diabetes mellitus (T2DM). Pharmacotherapy remains the primary therapeutic approach for type 2 diabetes mellitus (T2DM); however, adverse drug reactions persist as a significant clinical challenge affecting a substantial patient population.

Leech-derived bioactive constituents demonstrate targeted anti-hyperglycemic efficacy, while concomitant administration with exogenous insulin achieves synergistic glycemic regulation, thereby overcoming the therapeutic constraints inherent to conventional monotherapeutic approaches targeting isolated pathways (Mohammed et al., 2013). Concurrently, experimental studies have revealed that leech therapy ameliorates streptozotocin (STZ)-induced nephrotic microangiopathy in diabetic rats by suppressing endothelial cell migration and angiogenesis (Pang et al., 2020). Hirudin ameliorates kidney injury in diabetic nephropathy by suppressing GSDMD-mediated pyroptosis, thereby reducing the release of renal inflammatory factors (e.g., IL-1 β and IL-18) (Han et al., 2023). The bioactive components of leeches ameliorate diabetic nephropathy and preserve renal function through anticoagulant, antifibrotic, antithrombotic, and anti-inflammatory mechanisms (Tian et al., 2024). These results suggest that leeches may play a role in the treatment of T2DM, however, the specific mechanisms require further study.

Network pharmacology is an effective method to explore the underlying mechanisms of drugs and diseases, and the mechanism of action and potential targets of leech therapy for T2DM are explored in combination with immunoinfiltration analysis, and verified by molecular docking.

2 Materials and methods

2.1 Network pharmacological analysis

To investigate the mechanisms underlying the therapeutic effects of leech therapy in diabetes, we performed a network pharmacology-based analysis. First, the active ingredients of leeches were identified using the Herb database (<http://herb.ac.cn/>) (Fang et al., 2021), with data retrieval performed through the “Related Ingredients” section. The SMILES notation of the ingredients, which represents their structural formulas, was obtained from the PubChem database. These notations were subsequently utilized for target prediction analyses. Target data through SwissTargetPrediction (<https://www.swisstargetprediction.ch/>) (Daina et al., 2019), CTD (<https://ctdbase.org/>) (Davis et al., 2023), SuperPred (Database (<http://prediction.charite.de/>)) (Gallo et al., 2022), disease targets from Genecards (www.genecards.org), OMIM (<https://omim.org/>) (Stelzer et al., 2016) database access, Keyword “Type 2 diabetes” was set, and disease targets with

Relevance score > 10 were screened. The leech-active component-target network was generated using Cytoscape (version 3.8.2). Node degree centrality was calculated to quantify connection patterns, with higher values indicating greater nodal influence within the network architecture. Using the “cytohubb plugin” in Cytoscape software, select “Degree” to obtain the main active ingredients of leeches. Finally, microscopic letter platform (<http://www.bioinformatics.com.cn/>) to further optimize the “leech - active ingredient - targets” network diagram. R software “clusterProfiler” “pathview” package for enrichment analysis.

2.2 Obtain GEO datasets and GSEA analysis for T2DM

The GSE184050 dataset was retrieved from the Gene Expression Omnibus (GEO) database (<http://www.ncbi.nlm.nih.gov/geo>) (Barrett et al., 2013). This dataset comprises 116 whole blood transcriptome samples, including fifty patients with type 2 diabetes and sixty-six healthy controls. The Series Matrix File and SOFT formatted files were obtained from the database to facilitate data normalization processes. Differential expression analysis was performed using GEO2R, with the disease and control groups defined to generate differentially expressed genes (DEGs). These DEGs were subsequently processed in R (version 4.2.1) for downstream analyses. Finally, differential analysis was conducted using the limma package (Ritchie et al., 2015) in R software, with significance thresholds defined as $p < 0.05$ and absolute fold change ($|FC|$) > 1. Finally, a differential expression heatmap was generated from the expression profile dataset. Gene Set Enrichment Analysis (GSEA) (Canzler and Hackermüller, 2020) was subsequently conducted to analyze enriched gene sets. Through GSEA analysis, the gene ontology (GO) and associated signaling pathway (KEGG) of DEGs were determined.

2.3 Single sample enrichment analysis (ssGSEA)

Immune-related gene set enrichment in the study samples was assessed using single-sample gene set enrichment analysis (ssGSEA) (Jin et al., 2021), a modified version of traditional GSEA specifically designed for evaluating predefined gene set enrichment at the individual sample level. This method enables independent evaluation of immune pathway and related process activity within individual samples, eliminating the need for comparative data from other specimens. Transcriptome data was annotated by GRCh38 genome annotation file, and mRNA was extracted and normalized into tpm for subsequent analysis (Yi et al., 2020).

2.4 Weighted gene coexpression network analysis (WGCNA)

WGCNA is designed to identify co-expression patterns of gene modules associated with specific biological processes, diseases, or other phenotypes (Xu et al., 2023; Feng et al., 2022). In this study, we constructed a gene co-expression network using all expressed genes

(DEGs) and activated CD4 cell scores, activated dendritic cell scores, central memory CD4 T cell scores, and type 1 T helper cell scores in the dataset. Firstly, the power scatter plot was established and the optimal soft threshold ($\beta = 12$) was selected to obtain the similarity matrix between the genes. Then, hierarchical clustering of genes was performed, and dynamic module identification and clipping were performed with the minimum module number of 80. The resulting gene modules were visualized by co-expression patterns, and similar modules were combined. Correlations between activated DC scores and each module were assessed by correlation tests, and heat maps were drawn to show the correlation results.

2.5 Machine learning analysis of intersection targets

Immune-related targets for leeches to treat diabetes were screened by machine learning algorithms. First, the intersection targets of leeches and WGCNA core modules were selected as objects, and the targets were further screened by LASSO regression (Hohberg and Groll, 2024), SVM-REF (Yang et al., 2024), xgboost algorithm (Hou et al., 2020) and random forest algorithm (Hu and Szymczak, 2023). LASSO regression is a feature screening algorithm. The intersection targets are set as high data, and the ten-fold crossover algorithm is used to avoid overfitting and maintain the accuracy of the algorithm. SVM-REF algorithm is a classification algorithm, which reduces the error through classification and screening method. Xgboost algorithm and random forest algorithm can process data more efficiently, and process variables through cross-analysis method to realize high-latitude data sorting.

2.6 Molecular docking

The macromolecular structures were downloaded from the PDB website (Prestegard, 2021), specifically for RGS10, CAPS2, and OPA1, with PDB numbers 2DLR, Q86UW7, and 6JT, respectively. The small molecule structure compositions were sourced from PubChem (<https://pubchem.ncbi.nlm.nih.gov/>) (Kim, 2016), while the CAPS2 molecular structure was obtained from AlphaFold (<https://alphafold.ebi.ac.uk>). AutoDockTools-1.5.7 was used to process both small and large molecules. This involved “dehydrating and hydrogenating,” setting the docking box, and calculating the binding energy based on the number of hydrogen bonds and interactions between the small and large molecules. For visualization, we utilized PyMOL 3.1, presenting the final results with both overall and partial views of the docking.

2.7 Statistical analysis

All statistical analyses were conducted using R software (version 4.3.1), with a p-value of less than 0.05 deemed statistically significant. For network pharmacology investigations, Cytoscape (version 3.8.2) and its plugins were utilized to construct and analyze networks. The topological assessment of the “drug-component-target” network was primarily based on node degree

centrality. Protein-protein interaction (PPI) network clusters were identified using Cytoscape’s MCODE plugin with the following parameters: degree cut-off = 2, node score cut-off = 0.2, K-core = 2, and maximum depth = 100. Core targets were subsequently determined through node degree ranking. Molecular docking simulations were performed with an energy range of 5 kcal/mol and exhaustiveness set to 400.

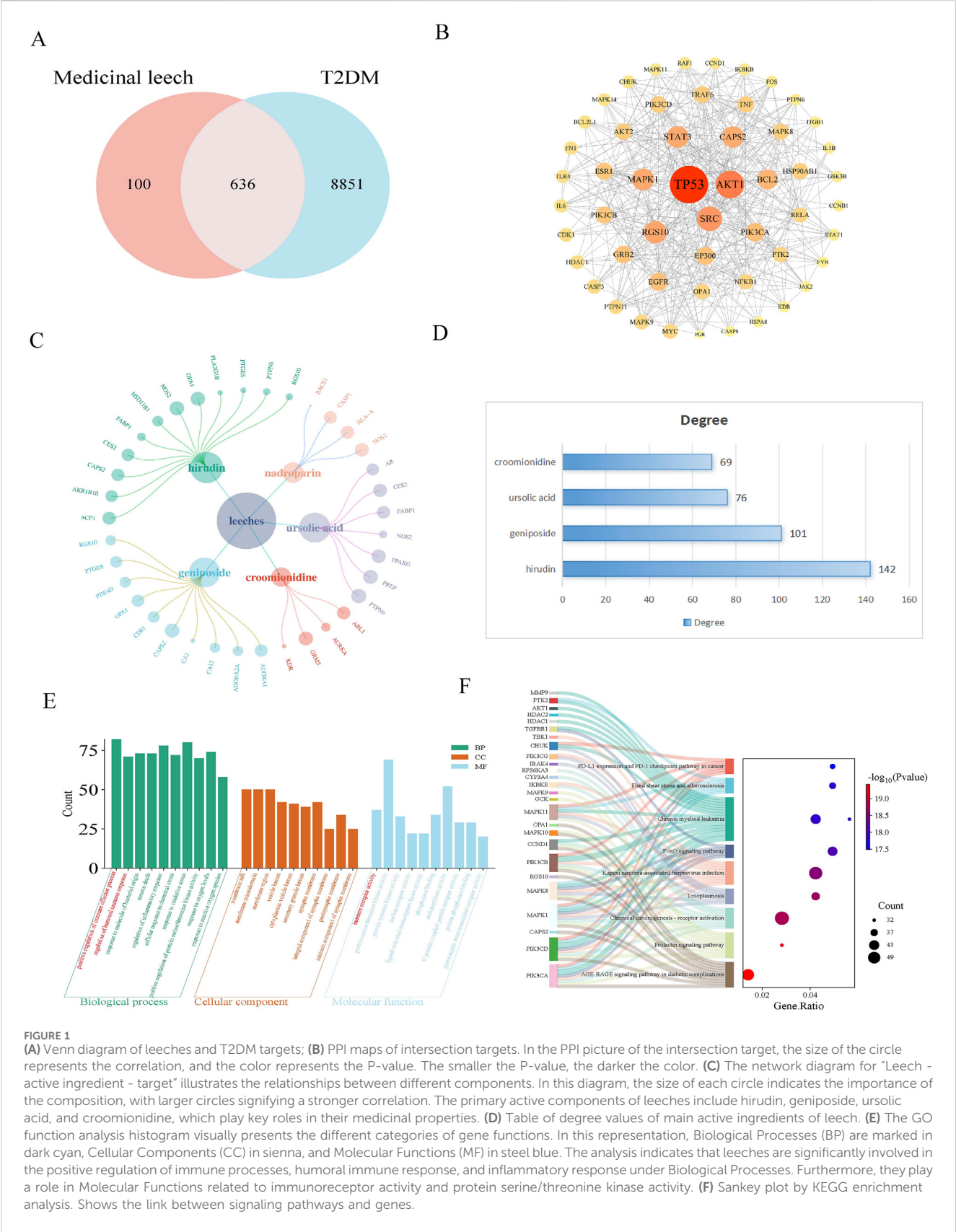
3 Results

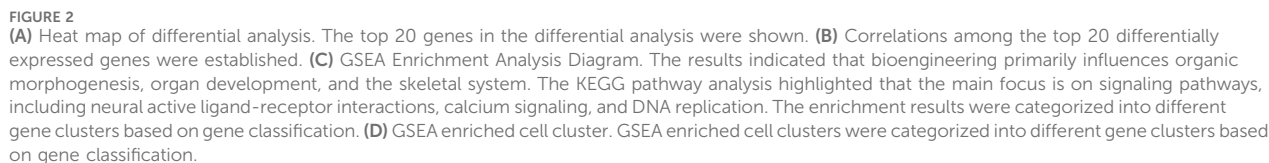
3.1 Network pharmacological analysis

A total of 19 active components from leeches were obtained through the HERB database. Subsequently, 461 component targets were identified using SWI, while 96 drug targets were retrieved from the CTD, and 258 drug targets were sourced from SuperPred. After removing duplicates, a total of 736 unique drug targets were identified. A total of 9,487 disease targets were identified through screenings from GeneCards and OMIM. The results revealed that there are 636 common targets associated with both leeches and diabetes mellitus (Figure 1A). Protein interaction analysis was also performed (Figure 1B). The component-target network illustrated the relationship between the effective components and targets of leeches (Figure 1C). The results indicate that the main active ingredient of leeches is hirudin, geniposide, ursolic acid, croomionidine (Figure 1D). The results of the GO enrichment analysis for biological processes indicated that leeches are involved in the positive regulation of immune response, humoral immune response, oxidative stress response, and inflammatory response. The molecular function analysis revealed that *Hirudo* is involved in immune receptor activity, protein serine/threonine kinase activity, and protein tyrosine kinase activity (Figure 1E); KEGG results showed involvement in AGE-RAGE signaling pathway and chemical carcinogenic-receptor activation signaling pathway (Figure 1F).

3.2 GSEA analysis of differential genes

Whole blood transcriptome data were obtained from the GSE184050 dataset, which included 116 patients—50 with type 2 diabetes and 66 healthy controls. By standardizing the sample data and using gene IDs instead of probe IDs, expression profiles for both the normal and disease groups were constructed. A total of 17,732 gene data were collected, resulting in the identification of 3,939 differentially expressed genes that were upregulated. The heat map shows the significantly different genes (Figure 2A). Correlations among the top 20 differentially expressed genes were established (Figure 2B). GSEA analyzed the Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) related to disease genes. The results indicated that bioengineering primarily influences organic morphogenesis, organ development, and the skeletal system (Figure 2C). In terms of molecular functions, it is involved in transmembrane transport and organ development. The KEGG pathway analysis highlighted that the main focus is on signaling pathways, including neural active ligand-receptor interactions, calcium signaling, and DNA

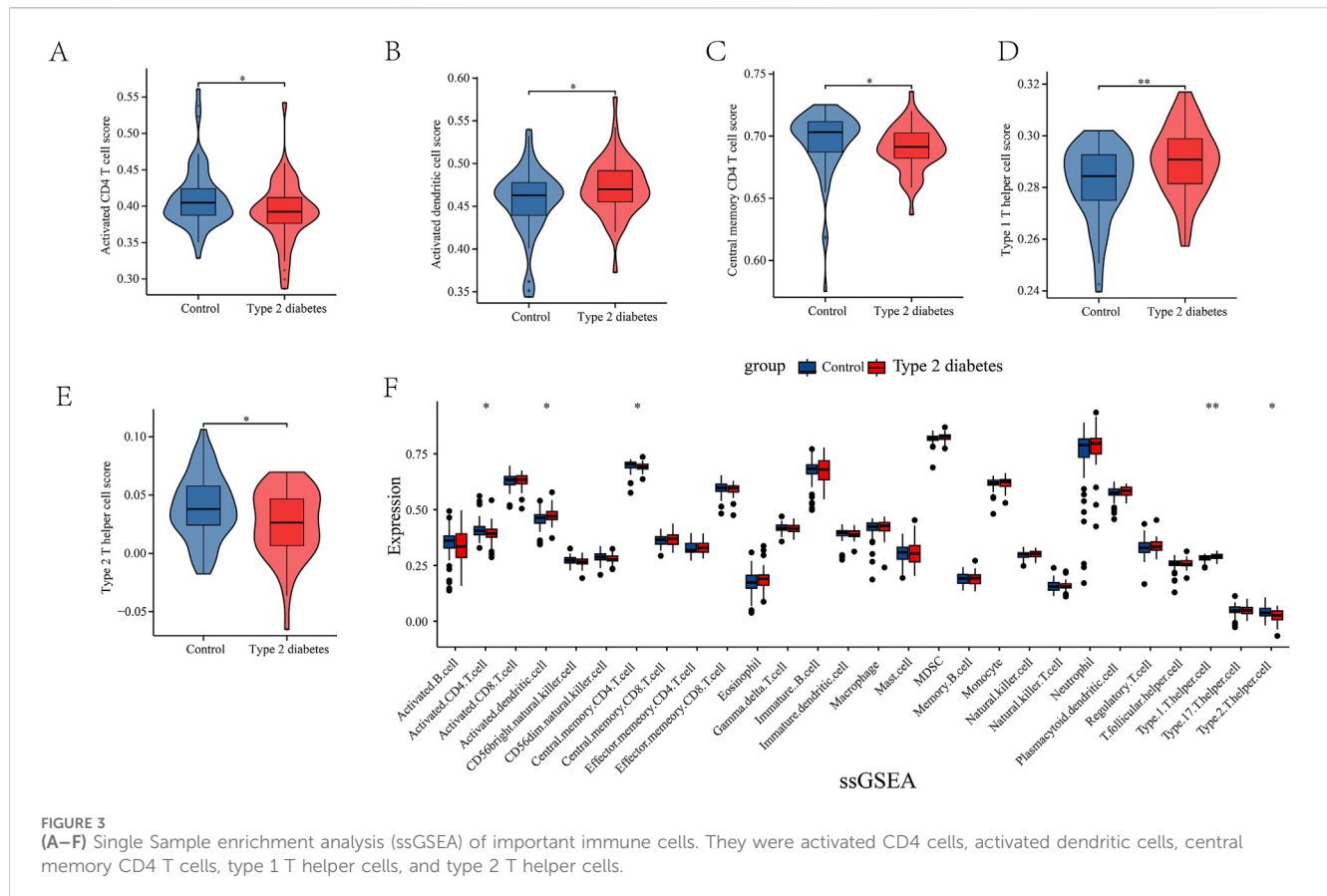




3.3 Immunoinfiltration and single sample enrichment analysis (ssGSEA)

Frontiers in Genetics

To further identify immune-related targets for T2DM treatment in leeches, we conducted WGCNA analysis. Gene co-expression networks were constructed for activated CD4 cells, activated dendritic cells, central memory CD4 T cells, type 1 T helper cells, and type 2 T helper cells using the GSE184050 dataset. Assessment of outliers No significant outliers were detected in the data (Figure 4A). A soft threshold power assessment of 10, with a scale-free index of 0.9, indicates that the connectivity within the network is reasonable (Figure 4B). The correlation trend and correlation heatmap among different samples were generated, and a multi-level modular clustering tree was created using a weighted co-representation network (Figure 4C). The final results indicated that all samples were categorized into a normal group and a disease group, which were further divided into 26 distinct modules. The correlation between these modules was represented by P-values (Figure 4D). The correlation coefficients between phenotypes and modules were calculated. The results revealed that the MEDarkturquoise module ($\text{cor} = 0.5$, $p = 5e-04$) and the MELightgreen module ($\text{cor} = 0.54$, $p = 1e-04$) displayed the most significant correlations with Type 2 Diabetes Mellitus (T2DM). Module membership (MM) and gene significance (GS) reflect the



correlation of genes within a particular module. The scatter plot shows that the MM and GS coefficients of MEDarkturquoise module are positively correlated ($\text{cor} = 0.075$, $p = 0.32$), MM and GS coefficients of MELightgreen module were positively correlated ($\text{cor} = 0.23$, $p = 5.2 \times 10^{-5}$). The MEDarkturquoise module and the MELightgreen module may be critical modules that play the most significant roles in Type 2 Diabetes Mellitus (T2DM), and they are also associated with activated dendritic cells. (Figures 4E,F).

3.5 Machine learning

By intersecting the leech targets with those analyzed through WGCNA, a total of 122 overlapping targets were identified, which served as the focus for machine learning (Figure 5A). The correlation among the top 20 targets is illustrated in a heatmap (Figures 5B,C). In the LASSO regression analysis, we performed 10-fold cross-validation (10-fold CV) using `cve.glmnet` to optimize the hyperparameter λ , and selected the optimal λ value based on binomial deviance: $\lambda_{\text{min}} = 0.02197$ (corresponding to 13 features). $\lambda_{1\text{se}} = 0.11727$ (all features are compressed to 0, the regularization intensity is high). And 13 key genes were selected (Figure 6A). During the training process of the random forest model, we determined the optimal `mtry` value using `tuneRF` and selected the corresponding `n`tree that yielded the minimum out-of-bag (OOB) error for hyperparameter optimization. Finally, we selected the top 20 key genes based on the feature importance ranking (Figure 6B). During the XGBoost training, we used `eval_metric = "error"` to monitor model errors and selected the best model based on the final training

errors. To optimize model performance, we adjusted hyperparameters such as `max_depth`, `eta`, and `gamma` through Grid Search (Figure 6C). SVM-RFE employed 10-fold cross-validation to assess model performance and ensure the generalization capability of feature selection. This method divided the data into training and test sets, resulting in the identification of 39 variables to build the model (Figure 6D). Ultimately, we identified three common targets: RGS10, CAPS2, and OPA1 (Figure 6E).

3.6 Molecular docking

Due to the large molecular weight of hirudin, the error in the molecular docking process is significantly affected; therefore, the hirudin component was excluded from the docking analysis. To better validate our results, geniposide, ursolic acid, and croomionidine were selected along with CAPS2, RGS10, and OPA1 as the top three components for docking analysis, respectively. The binding heat energy was less than -5 kcal/mol, indicating good bonding results. The results indicated that the binding affinity of CAPS2-ursolic acid is -8.9 kcal/mol, with a docking score of -8.93 kcal/mol. Additionally, the binding affinity of OPA1-croomionidine is -7.37 kcal/mol. The docking score is -7.68 kcal/mol, while the minimum binding affinity of RGS10-ursolic acid is -8.69 kcal/mol, and the docking score is -8.7 kcal/mol. The binding affinities and docking scores are largely consistent, indicating a strong binding strength between the ligand and the receptor (Figures 7B–D). All docking scores are displayed in Figure 7A.

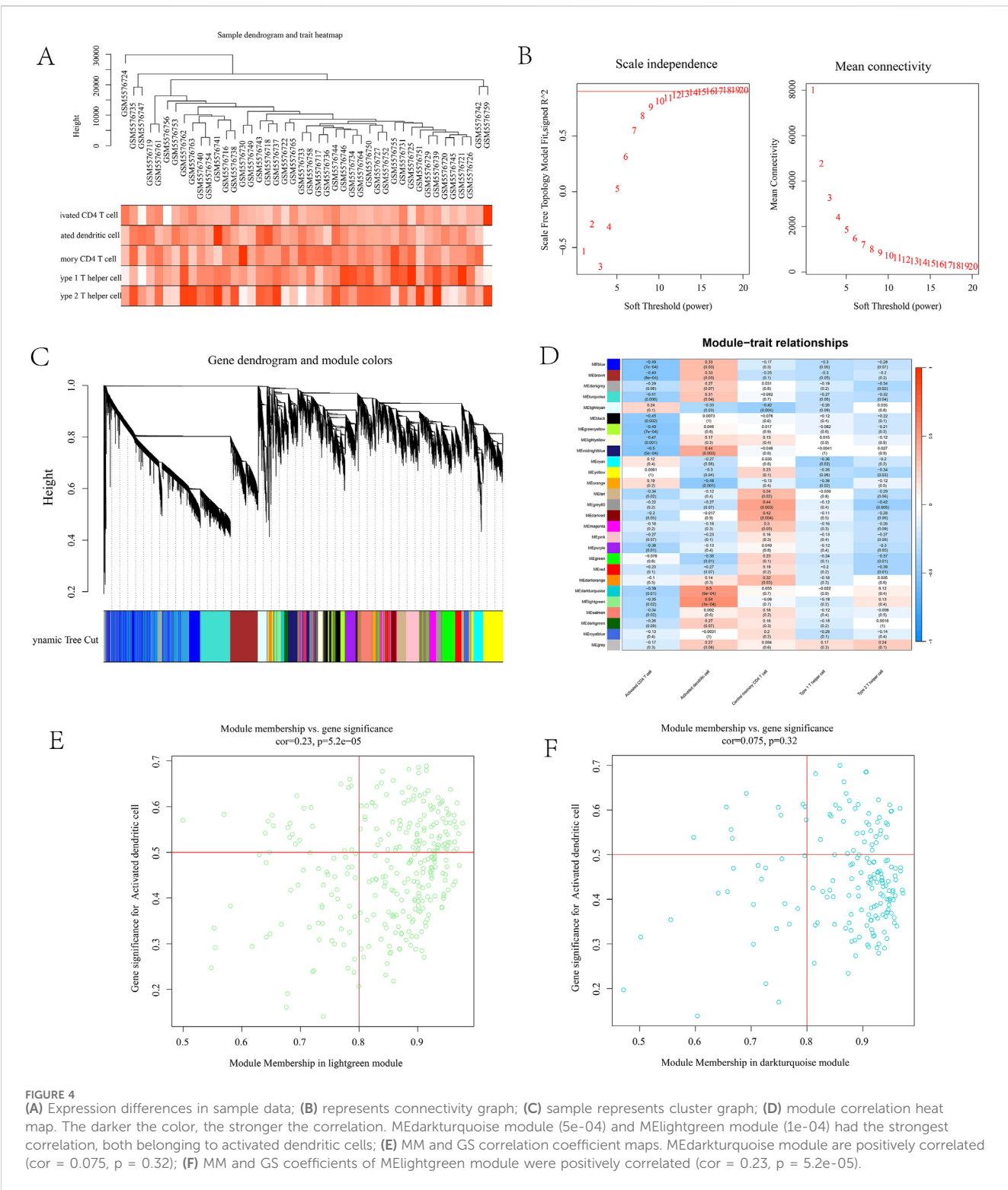
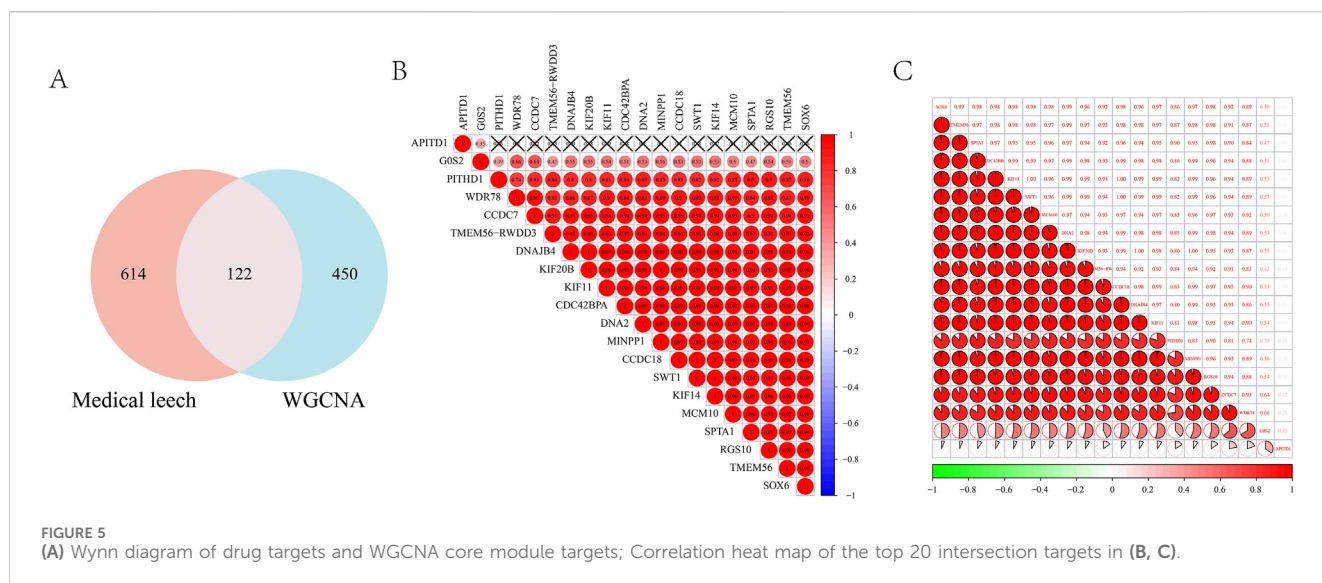


FIGURE 4 (A) Expression differences in sample data; (B) represents connectivity graph; (C) sample represents cluster graph; (D) module correlation heat map. The darker the color, the stronger the correlation. MEdarkturquoise module ($5e-04$) and MELightgreen module ($1e-04$) had the strongest correlation, both belonging to activated dendritic cells; (E) MM and GS correlation coefficient maps. MEdarkturquoise module are positively correlated ($cor = 0.075$, $p = 0.32$); (F) MM and GS coefficients of MELightgreen module were positively correlated ($cor = 0.23$, $p = 5.2e-05$).

4 Discussion

Type 2 diabetes mellitus (T2DM) constitutes a major global health burden, with its associated complications significantly compromising patient quality of life. Current clinical management primarily relies on pharmacological interventions; however, therapeutic outcomes demonstrate significant

interindividual variability (Kautzky-Willer et al., 2016). To improve disease control, optimize treatment efficacy, and mitigate complication risks, the discovery of novel therapeutic agents and mechanisms remains critically imperative. Emerging evidence demonstrates the therapeutic potential of Traditional Chinese Medicine (TCM) in T2DM management, particularly in ameliorating diabetic nephropathy (Cao et al., 2025).



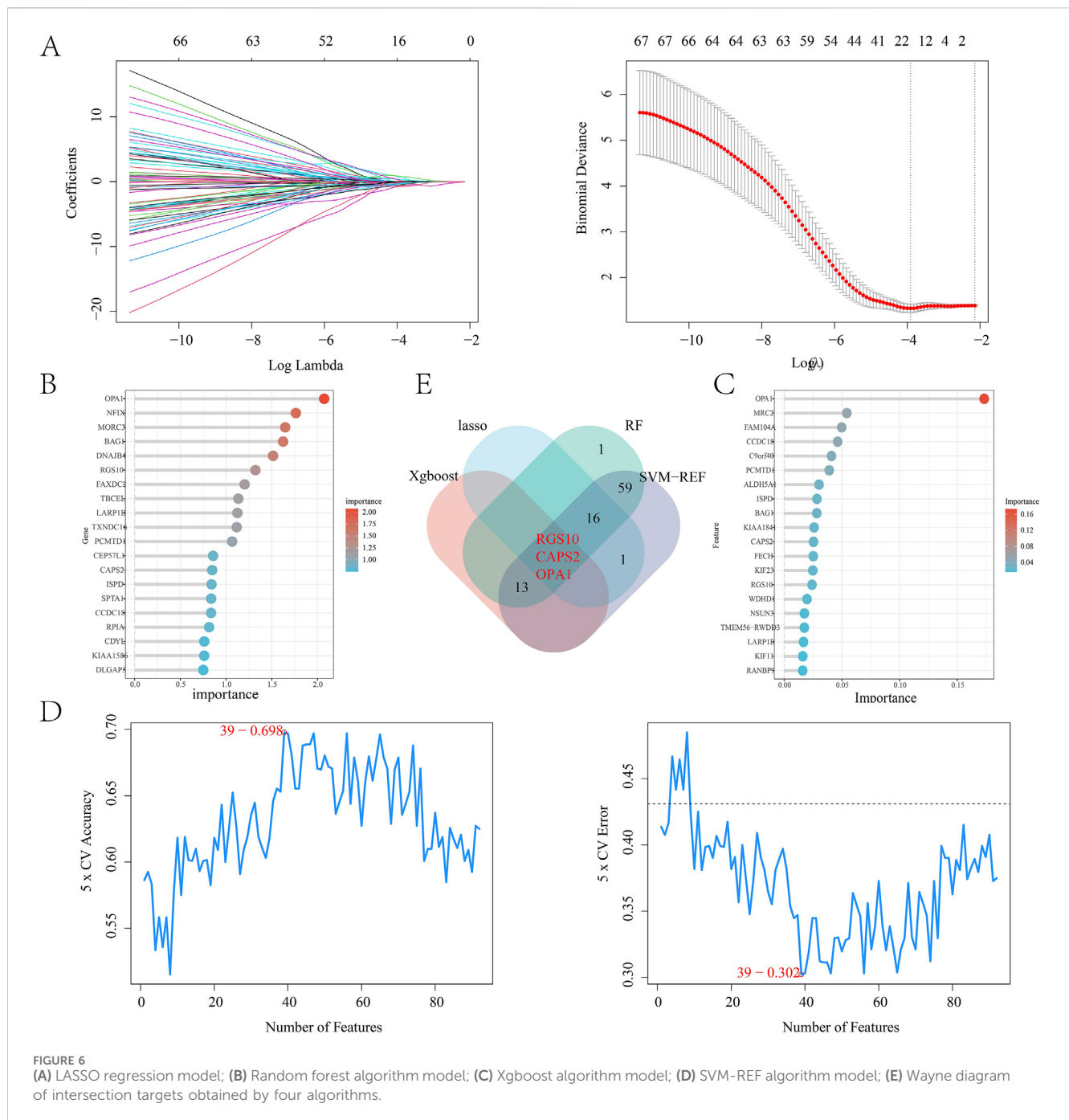
Furthermore, multimodal therapeutic strategies for T2DM include metabolic regulation through reduction of serum triglyceride and cholesterol levels (Tong et al., 2018), as well as pharmacological modulation of α -glucosidase activity (Zhou et al., 2023).

Hirudo (leech), an animal-derived traditional Chinese medicine, has been historically employed for its therapeutic properties in activating blood circulation and resolving stasis. Empirical studies have validated its antidiabetic efficacy, particularly through its principal bioactive component - hirudin. This naturally occurring thrombin inhibitor (Greinacher and Lubenow, 2001), identified as a core bioactive constituent in our network pharmacological analysis, exhibits dual therapeutic mechanisms: α -thrombin inhibition and anti-thrombotic activity. The bioactive constituents of Hirudo (leech) are primarily administered through dual delivery pathways: enteral absorption via oral intake and transdermal permeation through topical application. Hirudin, a thrombin-specific polypeptide inhibitor, demonstrates bioavailability contingent upon intestinal mucosal permeability and resistance to proteolytic degradation. Pharmacokinetic investigations reveal that while partial enzymatic cleavage in the gastrointestinal tract converts hirudin into low-molecular-weight peptides, these derivatives maintain substantial bioactivity (Stepensky, 2018). Hepatic and renal systems predominantly mediate hirudin metabolism, yielding amino acid derivatives and bioactive peptide fragments that may modulate endogenous biosynthetic pathways (Yuan et al., 2013). Notably, hirudin's potent thrombin-inhibiting capacity necessitates cautious clinical application due to dose-dependent hemorrhagic risks and potential toxicological manifestations in coagulopathic patients.

Modern pharmacological studies have demonstrated that blood sugar and lipids in the blood vessels of diabetic patients are prone to deposition. Prolonged deposition can induce inflammation and damage to blood vessels (Lee et al., 2020), as well as result in blood vessel blockage, giving rise to microvascular complications involving the heart, brain, and kidneys (Opazo-Ríos et al., 2020). Hirudin has the ability to regulate blood lipids and exerts a preventive effect against blood

vessel blockage induced by diabetes (Cheng et al., 2022). Simultaneously, hirudin can inhibit the inflammatory response via the P38 MAPK/NF - κ B pathway and treat kidney injury in diabetic rats (Han et al., 2020). In particular, it can regulate inflammatory factors such as TNF - α , IL - 1, and IL - 6, thereby preventing the occurrence of diabetic nephropathy (Vaidya et al., 2011). In addition to hirudin, the main active components of hirudin analyzed in this network pharmacological study were geniposide, ursolic acid, croomionidine, nadroparin, etc.,. Geniposide has also been confirmed to have anti-blood sugar effects. *In vitro* studies indicate that geniposide can help lower blood sugar by regulating α -glucosidase. This was confirmed by the analysis conducted by Zhou H et al. using PLC-ESI-QTOF-MS/MS. Additionally, geniposide can regulate the AKT-FOXO1 pathway to inhibit liver glucose production (Yang et al., 2018). Ursolic acid has been shown to reduce postprandial blood glucose levels in patients. Additionally, enzyme kinetics experiments have demonstrated that ursolic acid can inhibit the activities of α -amylase and α -glucosidase, making it an important hypoglycemic component (Wang et al., 2020; Camer et al., 2014).

The results of network pharmacology indicate that the active targets of leeches in diabetes treatment primarily revolve around immune regulation. Studies have found that the main active components of leeches are linked to immune regulatory functions. Hirudin, a natural anticoagulant protein, has demonstrated an anti-thrombotic effect in recent years. This effect is achieved not only through its inhibition of thrombin but also by regulating immune receptors and pathways, which may influence inflammatory responses and immune cell function (Liu et al., 1991). It also functions as a natural antibody, reducing the activity of human T cells (Asgari et al., 2017). Geniposide exhibits a variety of biological activities, including anti-inflammatory, antioxidant, and immunomodulatory effects (Zhou et al., 2019). *In vitro* experiments show that geniposide can enhance the immune response by regulating immune receptors such as STAT and JAK (Gong et al., 2024). It has also been shown to inhibit the expression of TLR4 and MMP9 proteins, thereby influencing the MyD88/NF-



κ B p65 signaling pathway involved in immune regulation (Liu et al., 2024).

If the active ingredients of leeches and their targets can treat diabetes through immune regulation, what are the specific targets of their effects? To investigate this, we combined the GSE184050 dataset to search for relevant immune cells using single-sample enrichment analysis. Ultimately, five important types of immune cells were identified: activated CD4 cells, activated dendritic cells, central memory CD4 T cells, type 1 T helper cells, and type 2 T helper cells. These five cell types may play an important role in the treatment of diabetes in leeches.

To further investigate the correlation between the active components of leeches and immune cells, we conducted WGCNA analysis and identified 26 analysis modules based on the phenotyping of the five immune cells. Among these, the MEdarkturquoise and MELightgreen modules were the most significant, exhibiting the strongest immune correlations in this study. By combining the genes from both modules, a total of 572 module genes were obtained. We intersected the drug genes with the core module genes and identified 122 common targets.

Machine learning can be utilized to identify potential targets by developing predictive models. In this study, to further explore the

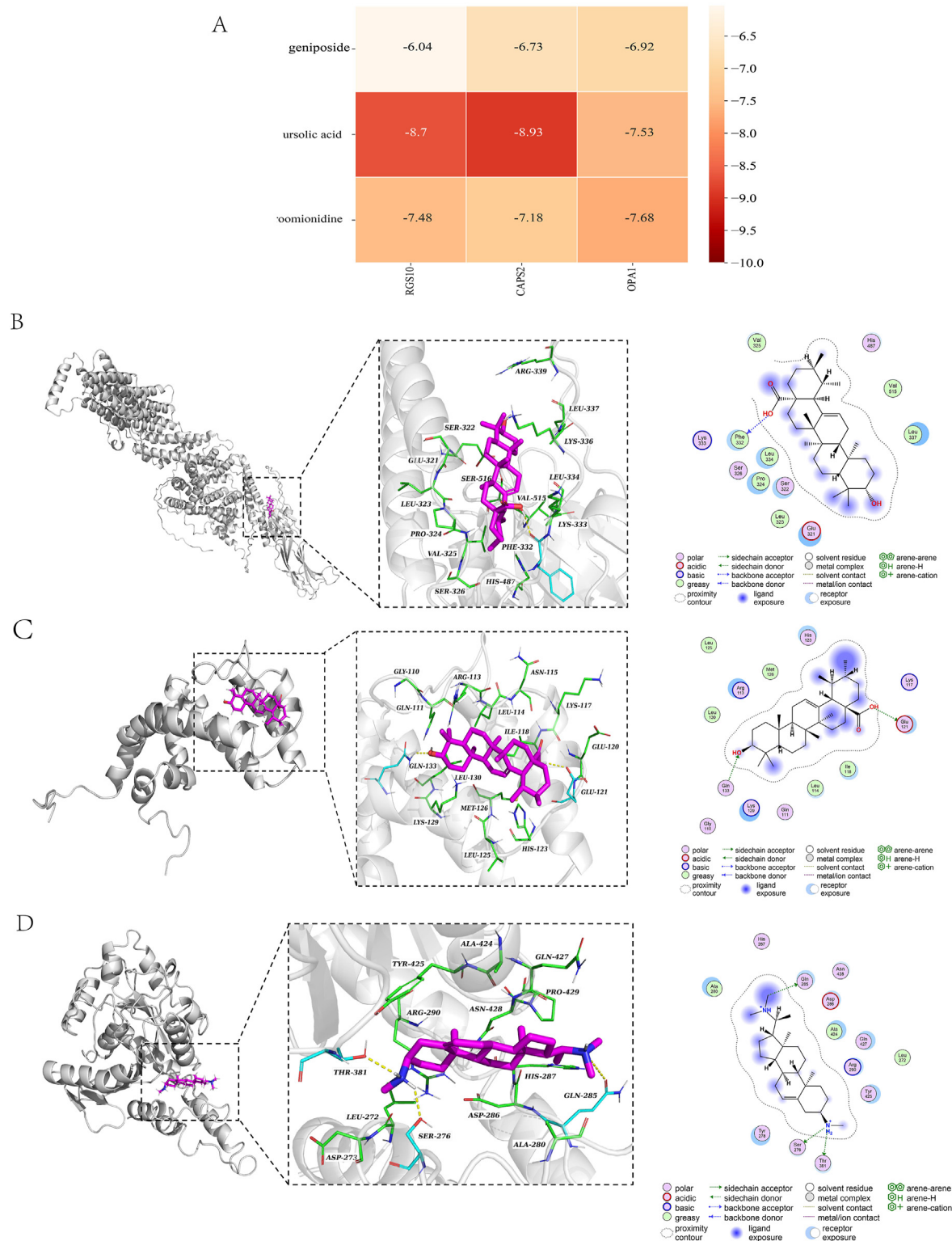


FIGURE 7

(A) Table of docking scores; (B) CAPS2-ursolic acid molecular docking diagram; (C) RGS10-ursolic acid molecular docking diagram; (D) OPA1-croomionidine molecular docking diagram.

potential immune targets of leeches in the treatment of diabetes, we constructed a prediction model using 122 intersection targets. This was achieved through LASSO regression, random forest algorithms,

SVM-REF, and XGBoost algorithms from machine learning. LASSO regression is an effective analytical method that allows for feature selection in high-dimensional data. It helps prevent overfitting and

enhances the predictive performance of the model (Yan et al., 2023). By incorporating L1 regularization terms, LASSO regression compresses the regression coefficients and automatically selects the features that are most relevant to the target variable. To prevent overfitting, we selected the optimal regularization parameter (λ) using 10-fold cross-validation. Ultimately, we identified 16 important variables to construct the model. To further identify genes associated with the disease, we trained the data using a random forest algorithm. This involved constructing multiple decision trees and aggregating the predictions from each tree. Compared to traditional single decision tree models, random forest can more efficiently handle high-dimensional data and prevent overfitting. During the model-building process, we set the number of trees to 56 ($n_{\text{tree}} = 56$), optimizing this parameter through cross-validation to ensure minimal error. Additionally, we calculated the importance of each variable and selected the top 20 variables as disease-related genes. The SVM-REF algorithm combines support vector machine (SVM) and recursive feature elimination (RFE) methods to facilitate effective feature selection (Zhao et al., 2022). In the SVM model, recursive feature elimination optimizes the model by incrementally removing the features that contribute the least to classification, based on the importance of weights or support vectors. This process ultimately retains the features with the greatest classification power. When evaluating the SVM-REF model, we used accuracy and error rate as evaluation metrics. We selected the top 39 variables, which exhibited the highest accuracy and the lowest error rate, to construct the model. The XGBoost algorithm is an effective machine learning model, so we incorporated it into our research. Ultimately, by building this machine learning model, we identified the core targets RGS10, CAPS2, and OPA1.

RGS10 is a member of the RGS protein family, primarily functioning to enhance the GTPase activity of G proteins. This action modulates the signaling pathways associated with G protein-coupled receptors (GPCRs). These pathways are essential for various cellular processes, including metabolism and hormone secretion. Animal studies have shown that RGS10 can help maintain metabolic homeostasis by mitigating inflammatory responses, which in turn reduces insulin resistance (Fang et al., 2019). Additionally, RGS10 is recognized as a significant immune-related target in diabetic retinopathy (Xia et al., 2024). CAPS2 (also known as NLRP3) is an inflammatory body protein that plays a crucial role in the innate immune system. The precise relationship between CAPS2 (NLRP3) and diabetes is not yet fully understood; however, recent investigations suggest that NLRP3 and its associated inflammatory pathways may play a role in the pathogenesis of diabetes (Duan et al., 2020; Jiang et al., 2017). CAPS2 may influence insulin resistance and beta cell dysfunction by regulating the release of inflammasomes and pro-inflammatory cytokines, such as IL-1 β (Kelley et al., 2019). Mitochondria are essential energy and metabolic organelles. Diabetes can cause mitochondrial dysfunction, leading to a range of metabolic abnormalities. OPA1 plays a crucial role in regulating mitochondrial fusion. It can mitigate the impairment of mitochondrial fusion caused by diabetes, potentially reducing the development of complications such as diabetic cardiomyopathy (Wang et al., 2023). Research has indicated that OPA1-mediated mitochondrial fusion exerts both regulatory and protective effects

against liver and myocardial damage induced by diabetes, primarily through the PKG-STAT3 pathway (Wang et al., 2022; Chang et al., 2023). Existing studies have shown that OPA1 can influence the energy acquisition of dendritic cells by regulating mitochondrial morphology (Virga et al., 2024). We can speculate that a similar effect may also be present in diabetic patients. CAPS2 promotes differentiation of Activated CD4 T cells by initiating Caspase-1-dependent interleukin-1 β secretion (Arbore et al., 2016).

To summarize, this study predicts immune targets and regulations using network pharmacology and bioinformatics methods, which could represent novel molecular mechanisms for leeches in treating T2DM. The coordinated multi-target interaction of RGS10, CAPS2, and OPA1 may play a significant role in regulating metabolism, secretion, and mitochondrial function in the treatment of T2DM. However, our study has some limitations. The database analysis is constrained by sample size, which restricts the scope of our research. Additionally, further experimental studies are needed to explore how RGS10, CAPS2, and OPA1 targets dynamically regulate T2DM, particularly how these dynamic changes occur during the early, middle, and late stages of disease development. Current research agrees that leeches possess anticoagulant properties, and anticoagulant treatment proves to be highly beneficial for managing diabetic complications and cardiomyopathy (Bhatt et al., 2020). Therefore, we can anticipate that future studies will further explore the potential of leeches in integrating Chinese and Western medicine through their unique anticoagulant and immune-regulating properties, ultimately benefiting a greater number of diabetic patients.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary Material.

Author contributions

TH: Software, Visualization, Writing – original draft. ZF: Funding acquisition, Writing – review and editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This study was supported by the National Natural Science Foundation of China (grant numbers 81774286 and 81603574) as well as the ZF Anhui Famous Chinese Medicine Studio (2019-8-515). This study was supported by ZF's research team.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or

those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2025.1554622/full#supplementary-material>

References

- Arbore, G., West, E. E., Spolski, R., Robertson, A. A. B., Klos, A., Rheinheimer, C., et al. (2016). T helper 1 immunity requires complement-driven NLRP3 inflammasome activity in CD4⁺ T cells. *Science* 352 (6292), aad1210. doi:10.1126/science.aad1210
- Asgari, S., Ebrahim-Habibi, A., Mahdavi, M., Choopani, M., and Mirzahoseini, H. (2017). Therapeutic protein deimmunization by T-cell epitope removal: antigen-specific immune responses *in vitro* and *in vivo*. *APMIS* 125 (6), 544–552. doi:10.1111/apm.12682
- Barrett, T., Wilhite, S. E., Ledoux, P., Evangelista, C., Kim, I. F., Tomashevsky, M., et al. (2013). Ncbi geo: archive for functional genomics data sets--update. *Nucleic Acids Res.* 41 (Database issue), D991–D995. Epub 2012/11/30. doi:10.1093/nar/gks1193
- Bhatt, D. L., Eikelboom, J. W., Connolly, S. J., Steg, P. G., Anand, S. S., Verma, S., et al. (2020). Role of combination antiplatelet and anticoagulation therapy in diabetes mellitus and cardiovascular disease: insights from the COMPASS trial. *Circulation* 141 (23), 1841–1854. doi:10.1161/CIRCULATIONAHA.120.046448
- Camer, D., Yu, Y., Szabo, A., and Huang, X. F. (2014). The molecular mechanisms underpinning the therapeutic properties of oleanolic acid, its isomer and derivatives for type 2 diabetes and associated complications. *Mol. Nutr. Food Res.* 58 (8), 1750–1759. Epub 2014/04/18. doi:10.1002/mnfr.201300861
- Canzler, S., and Hackermüller, J. (2020). Multigsea: a gsea-based pathway enrichment analysis for multi-omics data. *BMC Bioinforma.* 21 (1), 561. Epub 2020/12/09. doi:10.1186/s12859-020-03910-x
- Cao, M., Yang, J., Ye, H., Cui, G., Li, W., and Zhang, X. (2025). Leech granules inhibit ferroptosis and alleviate renal injury in mice with diabetic kidney disease. *J. Ethnopharmacol.* 337 (Pt 3), 118995. doi:10.1016/j.jep.2024.118995
- Chang, P., Zhang, X., Zhang, J., Wang, J., Wang, X., Li, M., et al. (2023). BNP protects against diabetic cardiomyopathy by promoting Opa1-mediated mitochondrial fusion via activating the PKG-STAT3 pathway. *Redox Biol.* 62, 102702. doi:10.1016/j.redox.2023.102702
- Cheng, S. H., Wang, S., Chen, H., Wang, C. A., Li, S. J., Shi, J. B., et al. (2022). Effects of hirudin on lipid metabolism and hemorheology in rats with diabetic nephropathy. *Clin. J. Chin. Med.* 14 (22), 37–40. doi:10.3969/j.issn.1674-7860.2022.22.009
- Daina, A., Michielin, O., and Zoete, V. (2019). SwissTargetPrediction: updated data and new features for efficient prediction of protein targets of small molecules. *Nucleic Acids Res.* 47 (W1), W357–W364. Epub 2019/05/21. doi:10.1093/nar/gkz382
- Davis, A. P., Wieggers, T. C., Johnson, R. J., Sciaky, D., Wieggers, J., and Mattingly, C. J. (2023). Comparative toxicogenomics database (ctd): update 2023. *Nucleic Acids Res.* 51 (D1), D1257–D1262. Epub 2022/09/29. doi:10.1093/nar/gkac833
- DeFronzo, R. A. (2004). Pathogenesis of type 2 diabetes mellitus. *Med. Clin. North Am.* 88 (4), 787–835. doi:10.1016/j.mcna.2004.04.013
- Duan, Y., Zhang, L., Angosto-Bazarra, D., Pelegrín, P., Núñez, G., and He, Y. (2020). RACK1 mediates NLRP3 inflammasome activation by promoting NLRP3 active conformation and inflammasome assembly. *Cell Rep.* 33 (7), 108405. doi:10.1016/j.celrep.2020.108405
- Fang, S., Dong, L., Liu, L., Guo, J., Zhao, L., Zhang, J., et al. (2021). Herb: a high-throughput experiment- and reference-guided database of traditional Chinese medicine. *Nucleic Acids Res.* 49 (D1), D1197–D1206. Epub 2020/12/03. doi:10.1093/nar/gkaa1063
- Fang, X., Chung, J., Olsen, E., Snider, I., Earls, R. H., Jeon, J., et al. (2019). Depletion of regulator-of-G-protein signaling-10 in mice exaggerates high-fat diet-induced insulin resistance and inflammation, and this effect is mitigated by dietary green tea extract. *Nutr. Res.* 70, 50–59. doi:10.1016/j.nutres.2018.06.004
- Feng, S., Xu, Y., Dai, Z., Yin, H., Zhang, K., and Shen, Y. (2022). Integrative analysis from multicenter studies identifies a wgcna-derived cancer-associated fibroblast signature for ovarian cancer. *Front. Immunol.* 13, 951582. Epub 2022/07/26. doi:10.3389/fimmu.2022.951582
- Gallo, K., Goede, A., Preissner, R., and Gohlke, B. O. (2022). Superpred 3.0: drug classification and target prediction-a machine learning approach. *Nucleic Acids Res.* 50 (W1), W726–W731. Epub 2022/05/08. doi:10.1093/nar/gkac297
- Gong, J., Jin, Q., and Zhu, F. (2024). Effects of geniposide on innate immunity and antiviral activity of Scyllaparamamosain. *Fish. Shellfish Immunol.* 145, 109303. doi:10.1016/j.fsi.2023.109303
- Greiner, A., and Lubenow, N. (2001). Recombinant hirudin in clinical practice: focus on lepirudin. *Circulation* 103 (10), 1479–1484. doi:10.1161/01.cir.103.10.1479
- Han, J., Pang, X., Zhang, Y., Peng, Z., Shi, X., and Xing, Y. (2020). Hirudin protects against kidney damage in streptozotocin-induced diabetic nephropathy rats by inhibiting inflammation via P38 MAPK/NF- κ B pathway. *Drug Des. Devel. Ther.* 14, 3223–3234. Published 2020 Aug 7. doi:10.2147/DDDT.S257613
- Han, J., Zuo, Z., Shi, X., Zhang, Y., Peng, Z., Xing, Y., et al. (2023). Hirudin ameliorates diabetic nephropathy by inhibiting gsdmd-mediated pyroptosis. *Cell Biol. Toxicol.* 39 (3), 573–589. Epub 2021/07/03. doi:10.1007/s10565-021-09622-z
- Hohberg, M., and Groll, A. (2024). A flexible adaptive lasso cox frailty model based on the full likelihood. *Biom. J.* 66 (7), e202300020. doi:10.1002/bimj.202300020
- Hou, N., Li, M., He, L., Xie, B., Wang, L., Zhang, R., et al. (2020). Predicting 30-days mortality for mimic-iii patients with sepsis-3: a machine learning approach using xgboost. *J. Transl. Med.* 18 (1), 462. Epub 2020/12/09. doi:10.1186/s12967-020-02620-5
- Hu, J., and Szymczak, S. (2023). A review on longitudinal data analysis with random forest. *Brief. Bioinform.* 24 (2), bbad002. Epub 2023/01/19. doi:10.1093/bib/bbad002
- Jiang, H., He, H., Chen, Y., Huang, W., Cheng, J., Ye, J., et al. (2017). Identification of a selective and direct NLRP3 inhibitor to treat inflammatory disorders. *J. Exp. Med.* 214 (11), 3219–3238. doi:10.1084/jem.20171419
- Jin, Y., Wang, Z., He, D., Zhu, Y., Chen, X., and Cao, K. (2021). Identification of novel subtypes based on ssgea in immune-related prognostic signature for tongue squamous cell carcinoma. *Cancer Med.* 10 (23), 8693–8707. Epub 2021/10/21. doi:10.1002/cam4.4341
- Kautzky-Willer, A., Harreiter, J., and Pacini, G. (2016). Sex and gender differences in risk, pathophysiology and complications of type 2 diabetes mellitus. *Endocr. Rev.* 37 (3), 278–316. doi:10.1210/er.2015-1137
- Kelley, N., Jeltama, D., Duan, Y., and He, Y. (2019). The NLRP3 inflammasome: an overview of mechanisms of activation and regulation. *Int. J. Mol. Sci.* 20 (13), 3328. Published 2019 Jul 6. doi:10.3390/ijms20133328
- Kim, S. (2016). Getting the most out of pubchem for virtual screening. *Expert Opin. Drug Discov.* 11 (9), 843–855. Epub 2016/07/28. doi:10.1080/17460444.2016.1216967
- Lee, M., Zhao, H., Liu, X., Liu, D., Chen, J., Li, Z., et al. (2020). Protective effect of hydroxysafflower yellow a on nephropathy by attenuating oxidative stress and inhibiting apoptosis in induced type 2 diabetes in rat. *Oxid. Med. Cell Longev.* 2020, 7805393. Epub 2020/04/08. doi:10.1155/2020/7805393
- Lee, S. H., Park, S. Y., and Choi, C. S. (2022). Insulin resistance: from mechanisms to therapeutic strategies. *Diabetes Metab. J.* 46 (1), 15–37. doi:10.4093/dmj.2021.0280
- Liu, L., Zhang, H., Tang, X., Zhang, M., Wu, Y., Zhao, Y., et al. (2024). Geniposide ameliorates psoriatic skin inflammation by inhibiting the TLR4/MyD88/NF- κ B p65 signaling pathway and MMP9. *Int. Immunopharmacol.* 133, 112082. doi:10.1016/j.intimp.2024.112082
- Liu, L. W., Vu, T. K., Esmon, C. T., and Coughlin, S. R. (1991). The region of the thrombin receptor resembling hirudin binds to thrombin and alters enzyme specificity. *J. Biol. Chem.* 266 (26), 16977–16980. doi:10.1016/s0021-9258(19)47326-2
- Mohammed, A. A., Mohammad, G. A., Mohamed, A., Mohamed, A., and Ahmed, M. (2013). *In vivo* anti-hyperglycemic activity of saliva extract from the tropical leech *Hirudinaria manillensis*. *Chin. J. Nat. Med.* 11 (5), 488–493. doi:10.1016/S1875-5364(13)60089-8
- Opazo-Rios, L., Mas, S., Marín-Royo, G., Mezzano, S., Gómez-Guerrero, C., Moreno, J. A., et al. (2020). Lipotoxicity and diabetic nephropathy: novel mechanistic insights and therapeutic opportunities. *Int. J. Mol. Sci.* 21 (7), 2632. Epub 2020/04/16. doi:10.3390/ijms21072632
- Pang, X., Zhang, Y., Peng, Z., Shi, X., Han, J., and Xing, Y. (2020). Hirudin reduces nephropathy microangiopathy in stz-induced diabetes rats by inhibiting endothelial cell

- migration and angiogenesis. *Life Sci.* 255, 117779. Epub 2020/05/18. doi:10.1016/j.lfs.2020.117779
- Prestegard, J. H. (2021). A perspective on the pdb's impact on the field of glycobiology. *J. Biol. Chem.* 296, 100556. Epub 2021/03/22. doi:10.1016/j.jbc.2021.100556
- Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., et al. (2015). Limma powers differential expression analyses for rna-sequencing and microarray studies. *Nucleic Acids Res.* 43 (7), e47. Epub 2015/01/22. doi:10.1093/nar/gkv007
- Rovira-Llopis, S., Bañuls, C., Diaz-Morales, N., Hernandez-Mijares, A., Rocha, M., and Victor, V. M. (2017). Mitochondrial dynamics in type 2 diabetes: pathophysiological implications. *Redox Biol.* 11, 637–645. doi:10.1016/j.redox.2017.01.013
- Stelzer, G., Rosen, N., Plaschkes, I., Zimmerman, S., Twik, M., Fishilevich, S., et al. (2016). The Genecards suite: from gene data mining to disease genome sequence analyses. *Curr. Protoc. Bioinforma.* 54 (3), 1. Epub 2016/06/21. doi:10.1002/cpbi.5
- Stepensky, D. (2018). Pharmacokinetics of toxin-derived peptide drugs. *Toxins (Basel)* 10 (11), 483. Published 2018 Nov 20. doi:10.3390/toxins10110483
- Tian, F., Yi, X., Yang, F., Chen, Y., Zhu, W., Liu, P., et al. (2024). Research progress on the treatment of diabetic nephropathy with leech and its active ingredients. *Front. Endocrinol. (Lausanne)* 15, 1296843. Published 2024 Jan 26. doi:10.3389/fendo.2024.1296843
- Tong, X., Xu, J., Lian, F., Yu, X., Zhao, Y., Xu, L., et al. (2018). Structural alteration of gut microbiota during the amelioration of human type 2 diabetes with hyperlipidemia by metformin and a traditional Chinese herbal formula: a multicenter, randomized, open label clinical trial. *mBio* 9 (3), e02392. Published 2018 May 22. doi:10.1128/mBio.02392-17
- Vaidya, V. S., Niewczas, M. A., Ficociello, L. H., Johnson, A. C., Collings, F. B., Warram, J. H., et al. (2011). Regression of microalbuminuria in type 1 diabetes is associated with lower levels of urinary tubular injury biomarkers, kidney injury molecule-1, and N-Acetyl-B-D-Glucosaminidase. *Kidney Int.* 79 (4), 464–470. Epub 2010/10/29. doi:10.1038/ki.2010.404
- Virga, D. M., Hamilton, S., Osei, B., Morgan, A., Kneis, P., Zamponi, E., et al. (2024). Activity-dependent compartmentalization of dendritic mitochondria morphology through local regulation of fusion-fission balance in neurons *in vivo*. *Nat. Commun.* 15 (1), 2142. Published 2024 Mar 8. doi:10.1038/s41467-024-46463-w
- Wang, J., Zhao, J., Yan, Y., Liu, D., Wang, C., and Wang, H. (2020). Inhibition of glycosidase by ursolic acid: *in vitro*, *in vivo* and *in silico* study. *J. Sci. Food Agric.* 100 (3), 986–994. Epub 2019/10/28. doi:10.1002/jsfa.10098
- Wang, S., Zhao, H., Lin, S., Lv, Y., Lin, Y., Liu, Y., et al. (2023). New therapeutic directions in type II diabetes and its complications: mitochondrial dynamics. *Front. Endocrinol. (Lausanne)* 14, 1230168. Published 2023 Aug 21. doi:10.3389/fendo.2023.1230168
- Wang, T., Chen, Y., Li, Y., Wang, Z., Qiu, C., Yang, D., et al. (2022). TRPV1 protect against hyperglycemia and hyperlipidemia induced liver injury via OPA1 in diabetes. *Tohoku J. Exp. Med.* 256 (2), 131–139. doi:10.1620/tjem.256.131
- Xia, N., Zhao, Q., Xu, J., and Cheng, Z. (2024). Deciphering immune-related gene signatures in diabetic retinopathy: insights from *in silico* analysis and *in vitro* experiment. *Curr. Pharm. Biotechnol.* 25 (15), 2032–2045. doi:10.2174/0113892010276304240103084841
- Xu, M., Zhou, H., Hu, P., Pan, Y., Wang, S., Liu, L., et al. (2023). Identification and validation of immune and oxidative stress-related diagnostic markers for diabetic nephropathy by wgcna and machine learning. *Front. Immunol.* 14, 1084531. Epub 2023/03/14. doi:10.3389/fimmu.2023.1084531
- Yan, M., Li, W., Wei, R., Li, S., Liu, Y., Huang, Y., et al. (2023). Identification of pyroptosis-related genes and potential drugs in diabetic nephropathy. *J. Transl. Med.* 21 (1), 490. Published 2023 Jul 21. doi:10.1186/s12967-023-04350-w
- Yang, S. Q., Chen, Y. D., Li, H., Hui, X., and Gao, W. Y. (2018). Geniposide and gentiopicroside suppress hepatic gluconeogenesis via regulation of akt-foxo1 pathway. *Arch. Med. Res.* 49 (5), 314–322. Epub 2018/11/10. doi:10.1016/j.arcmed.2018.10.005
- Yang, Z., Kao, X., Huang, N., Yuan, K., Chen, J., and He, M. (2024). Identification and analysis of PANoptosis-related genes in sepsis-induced lung injury by bioinformatics and experimental verification. *J. Inflamm. Res.* 17, 1941–1956. Published 2024 Mar 28. doi:10.2147/JIR.S452608
- Yi, M., Nissley, D. V., McCormick, F., and Stephens, R. M. (2020). Ssgsea score-based ras dependency indexes derived from gene expression data reveal potential ras addiction mechanisms with possible clinical implications. *Sci. Rep.* 10 (1), 10258. Epub 2020/06/26. doi:10.1038/s41598-020-66986-8
- Yuan, H., Zhang, L., Jin, M., Tang, W., and Liao, S. (2013). Leech medicinal composition and main pharmacological efficacy research progress [in Chinese]. *J. gansu Med.* 32 (4), 270–273. doi:10.15975/j.carolcarrollinkigsyy.2013.04.004
- Zhao, Z., He, S., Yu, X., Lai, X., Tang, S., Mariya M, E. A., et al. (2022). Analysis and experimental validation of rheumatoid arthritis innate immunity gene CYFIP2 and pancreatic. *Front. Immunol.* 13, 954848. Published 2022 Jul 11. doi:10.3389/fimmu.2022.954848
- Zhou, H., Zhang, S., Chen, L., Liu, Y., Shen, L., and Zhang, J. (2023). Effective therapeutic verification of crocin I, geniposide, and gardenia (gardenia jasminoides ellis) on type 2 diabetes mellitus *in vivo* and *in vitro*. *Foods* 12 (8), 1668. Published 2023 Apr 17. doi:10.3390/foods12081668
- Zhou, Y. X., Zhang, R. Q., Rahman, K., Cao, Z. X., Zhang, H., and Peng, C. (2019). Diverse pharmacological activities and potential medicinal benefits of geniposide. *Evid. Based Complement. Altern. Med.* 2019, 4925682. Published 2019 Apr 16. doi:10.1155/2019/4925682