

An Integrated Network Toxicology and Machine Learning Framework for Deciphering the Gastrointestinal Toxicity of Ginkgo biloba Seeds

Yu Qi

qy13871199637@163.com

Hubei University of Chinese Medicine

Jinrong He

The Central Hospital of Wuhan

Chengjie Xu

Sichuan Academy of Medical Sciences & Sichuan Provincial People's Hospital

Pei Li

Wuhan Hospital of Traditional Chinese Medicine

Shang Huang

Wuhan Hospital of Traditional Chinese Medicine

Yong Li

Wuhan Hospital of Traditional Chinese Medicine

Chengkai Tan

Wuhan Hospital of Traditional Chinese Medicine

Mingfeng Xia

Shouzheng Pharma Biotechnology Co., Ltd

Yaohui Song

Wuhan Eighth Hospital

Saili Chen

Hubei University of Science and Technology PhD Start-up Fund Project Support

Xiaohong Guo

Hubei University of Chinese Medicine

Article

Keywords: Ginkgo biloba seeds (GBS), Gastrointestinal toxicity, Network toxicology, Machine learning, MMP9

Posted Date: April 17th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-9164926/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Background: *Ginkgo biloba* seeds (GBS), a dual-use food and medicine, can cause severe gastrointestinal toxicity, but the underlying mechanisms remain unclear.

Purpose: This study aims to elucidate the integrated mechanisms of GBS toxicity by combining network toxicology, machine learning, and experimental validation.

Methods: A network toxicology approach was employed to map the interaction between GBS toxicants and gastrointestinal toxicity-related targets. The presence of the three major toxic constituents—ginkgolic acids (GAs), 4'-O-methylpyridoxine (MPN), and 4'-O-methylpyridoxine-5'-glucoside (MPNG)—was experimentally confirmed in both the raw GBS powder and the GBS-containing serum using UHPLC-Q Exactive HFX-MS, verifying their systemic bioavailability. An integrated machine learning framework, combining LASSO, Random Forest, and SVM-RFE, was used to prioritize key genes, with SHAP analysis enhancing model interpretability. Molecular docking and molecular dynamics (MD) simulations were performed to assess the potential direct interaction between GBS constituents and the MMP9 protein. *In vitro* assays (CCK-8, EdU, LDH, ROS, apoptosis) evaluated cytotoxicity, while qPCR, WB, IF, and CETSA assessed MMP9 expression and target engagement.

Results: The integrated analysis, initiated by UHPLC-Q Exactive HFX confirmation of GAs, MPN, and MPNG in both raw GBS and GBS-containing serum, identified MMP9 as the top key gene associated with toxicity through a combined network toxicology and machine learning approach. MD revealed stable binding conformations between these toxicants and the MMP9 protein. GBS-containing serum significantly inhibited cell proliferation, induced oxidative stress, promoted apoptosis, and triggered an inflammatory response. MMP9 expression was markedly elevated at both the mRNA and protein levels, with CETSA confirming direct target engagement.

Conclusion: This study establishes a novel framework for investigating herbal toxicity and identifies MMP9 as a central mediator of GBS-induced gastrointestinal injury. The direct interaction between GBS constituents and MMP9, along with the activation of the IL-17/NF- κ B/IL-6 inflammatory axis and the Bax/Bcl-2/Caspase-3 apoptotic pathway, provides a systems-level understanding of GBS toxicity and highlights MMP9 as a potential biomarker for safety assessment.

Introduction

Ginkgo biloba L., one of the oldest living gymnosperm species, is often regarded as a “living fossil” with an evolutionary history spanning over 200 million years (Hassan et al., 2020). Widely cultivated for its medicinal and ornamental value, this species yields edible

seeds—known as Bai Guo (百果) in Traditional Chinese Medicine (TCM)—which have been employed in East Asia for hundreds of years, serving both as functional foods and therapeutic agents (Liu et al., 2022). Historically, *Ginkgo biloba seeds* (GBS) were prescribed in classical pharmacopoeias such as the Compendium of Materia Medica (Bencao Gangmu) to treat respiratory disorders (e.g., asthma, cough) and urinary ailments. Modern pharmacological studies have identified bioactive constituents in GBS—including flavonoids, terpene lactones, and polysaccharides—that are associated with a range of pharmacological activities (Fang et al., 2020). These include antimicrobial effects, neuroprotective properties for neurodegenerative disorders such as Alzheimer's disease, and cardiovascular protective effects by improving cerebral and coronary blood flow (He et al., 2023). However, accumulating clinical and epidemiological evidence suggests that excessive intake may induce gastrointestinal toxicity, characterized by symptoms such as nausea, vomiting, and abdominal pain (Kajiyama et al., 2002). These findings highlight the urgent need to elucidate the underlying toxicological mechanisms of gastrointestinal toxicity induced by GBS, particularly in relation to its major toxic constituents.

The toxicity of GBS is primarily attributed to several well-characterized constituents, including ginkgolic acids (GAs), 4'-O-methylpyridoxine (MPN), and 4'-O-methylpyridoxine-5-glucoside (MPNG) (Boateng, 2022). GAs are long-chain alkylphenolic compounds characterized by variable side chain lengths and degrees of unsaturation, and they exhibit genotoxic and cytotoxic properties by inducing DNA damage, cell cycle arrest, and mitochondrial dysfunction, ultimately leading to apoptosis in a variety of cell types. These effects are associated not only with systemic toxicity, such as immunotoxicity and hepatotoxicity, but also contribute to gastrointestinal symptoms including abdominal pain, diarrhea, nausea, and vomiting, particularly following excessive oral intake (Shao et al., 2025). In addition to GAs, MPN and MPNG are two major endotoxins found in GBS, both capable of inducing gastrointestinal and neurological toxicity. MPN, as a vitamin B6 antagonist, inhibits glutamate decarboxylase and reduces GABA levels, leading to convulsions and gastrointestinal distress (Sasaki et al., 2000). It has been shown to be rapidly absorbed in humans, with serum concentrations closely correlating with the severity of poisoning symptoms (Hori et al., 2004). In contrast, MPNG, the glucosylated form of MPN, is more abundant in raw seeds but exhibits lower acute toxicity, with a reported lethal dose of 0.8 mmol/kg in mice compared to 0.2 mmol/kg for MPN (Zhang et al., 2021). However, MPNG can be hydrolyzed by β -glucosidase in the gastrointestinal tract, releasing MPN and amplifying the overall toxicity. Moreover, heat treatment significantly reduces MPN levels while increasing MPNG content, suggesting a dynamic interconversion between the two compounds (Gong et al., 2018).

Although increasing evidence has linked the consumption of GBS to gastrointestinal toxicity, the underlying toxicological mechanisms remain incompletely understood. Traditional toxicological studies have primarily focused on single components or isolated

targets, thus failing to capture the complex interactions between the toxic constituents of GBS and biological systems. In this context, network toxicology provides a promising framework for elucidating the intricate relationships between toxic ingredients and their biological targets. Furthermore, machine learning techniques offer powerful tools for toxicity prediction, identification of key molecular features, and analysis of immunotoxicological relevance (Bai et al., 2025). Integrating these two approaches enables a more comprehensive, systematic, and predictive understanding of the mechanisms underlying GBS-induced gastrointestinal toxicity. In this study, we innovatively applied UHPLC-Q Exactive HFX analysis to identify the toxic components of GBS that enter the bloodstream. We further validated the binding affinity between these toxic constituents and key targets using molecular docking and molecular dynamics simulations. *In vitro* experiments with drug-containing serum were conducted to confirm the gastrointestinal toxicity of GBS and to further elucidate the underlying mechanisms. To our knowledge, this study represents the first comprehensive investigation into GBS toxicity through an integrated approach combining network toxicology, machine learning, and experimental validation. Our findings provide novel insights into the safety evaluation of GBS and offer a scientific basis for understanding the mechanisms of GBS toxicity and identifying potential biomarkers and therapeutic targets. The detailed process of this study is shown in Fig. 1.

Materials and Methods

Identification of Toxicity-Related Targets for GBS Constituents

To explore the molecular mechanisms underlying the gastrointestinal toxicity induced by GBS, we first identified potential molecular targets for its major toxic constituents: GAs, MPN, and MPNG. Potential molecular targets of the selected compounds were retrieved from multiple bioinformatics platforms, including TCMSP (<https://old.tcmsp-e.com/>), ChEMBL (<https://www.ebi.ac.uk/chembl/>), SwissTargetPrediction (<http://swisstargetprediction.ch/>), PharmMapper (<http://www.lilab-ecust.cn/pharmmapper/>), and the Similarity Ensemble Approach (<https://sea.bkslab.org/>). All predicted targets were standardized by mapping to human (*Homo sapiens*) protein entries in the UniProt database (<https://www.uniprot.org/>) using gene symbols as primary identifiers. Finally, duplicate targets were eliminated using the Jvenn online tool (<http://www.bioinformatics.com.cn/static/others/jvenn/index.html>), resulting in a non-redundant target set for subsequent network analysis.

Identification of Gastrointestinal Toxicity-Related Targets and Intersection Analysis

To identify potential targets associated with gastrointestinal toxicity, we conducted a systematic search in public databases including OMIM (<https://omim.org/>), GeneCards (<https://www.genecards.org/>), PharmGKB (<https://www.pharmgkb.org/>), and TTD

(Therapeutic Target Database, <https://db.idrblab.net/ttd/>), using the keywords “Gastrointestinal Toxicity”, “Inflammatory Bowel Disease (IBD)”, and “Gastritis”. These three diseases were selected as representative models of intestinal and gastric toxicity, respectively, due to their relevance to inflammation, epithelial barrier dysfunction, and immune-mediated injury—mechanisms commonly associated with herbal-induced gastrointestinal damage. The resulting targets were standardized and merged to generate a comprehensive list of disease-associated proteins.

An intersection analysis using the Jvenn online tool identified common targets between GBS constituents and gastrointestinal toxicity, which were used to construct a component-target-disease network in (v3.8.2, Cytoscape Consortium, Seattle, WA, USA). This network illustrated the complex interactions among the toxic constituents, their molecular targets, and the associated gastrointestinal diseases, providing a system-level understanding of the underlying mechanisms of GBS-induced gastrointestinal toxicity.

To further explore the functional interactions among the intersected targets and elucidate their potential roles in the mechanisms of GBS-induced gastrointestinal toxicity, a protein-protein interaction (PPI) network was generated via the STRING database (<https://string-db.org/>) by submitting the overlapping gene set under high-confidence settings (minimum interaction score > 0.7). The resulting network was then imported into Cytoscape for visualization.

Identification of Gastrointestinal Toxicity-Related Differentially Expressed Genes (DEGs)

To further explore the molecular signatures associated with gastrointestinal toxicity, we retrieved publicly available gene expression datasets from the Gene Expression Omnibus (GEO) database (<https://www.ncbi.nlm.nih.gov/geo/>) that were relevant to IBD and Gastritis. The selected datasets included GSE3365 and GSE36807 for IBD, and GSE233973, GSE5081 for Gastritis.

Raw microarray data were downloaded and normalized using the R/Bioconductor package limma, with background correction and quantile normalization applied to reduce technical variability. Batch effects across datasets were corrected using the ComBat algorithm from the sva package. DEGs were identified using the following thresholds: $|\log_2FC| > 0.415$ and $P\text{-value} < 0.05$. The resulting DEGs were merged and filtered to generate a comprehensive list of gastrointestinal toxicity-related genes. A volcano plot was generated to visualize the DEGs, highlighting the most significantly upregulated and downregulated genes associated with gastrointestinal toxicity.

Weighted Gene Co-Expression Network Analysis (WGCNA)

WGCNA was performed using the WGCNA R package to identify gene co-expression modules associated with gastrointestinal toxicity. The batch-corrected gene expression matrix was normalized using the `normalizeBetweenArrays` function in the `limma` package, and modules were detected based on a soft-thresholding power of 5, which yielded a scale-free topology fit index $R^2 > 0.85$. Hierarchical clustering was applied to group genes with similar expression patterns, and modules with similar eigengenes were merged using a height cut-off of 0.25. Module-trait relationships were assessed by correlating module eigengenes with clinical traits, and modules significantly associated with gastrointestinal toxicity (Pearson's $|r| > 0.5$, $P < 0.05$) were selected for further analysis. Hub genes within the most relevant module were identified based on module membership (MM) and gene significance (GS), and used for downstream intersection analysis with predicted targets of GBS constituents.

Intersection Analysis and Functional Enrichment of Key Targets

To identify the targets associated with GBS-induced gastrointestinal toxicity, we performed an intersection analysis among predicted toxic constituent targets, DEGs from GEO datasets, and WGCNA-identified hub genes. The overlapping gene set was subjected to functional enrichment analysis.

Functional enrichment analysis of the overlapping targets was performed using the R package `clusterProfiler`. Specifically, GO analysis revealed enriched categories within BP, MF, and CC (biological processes, molecular functions, cellular components), and KEGG pathway analysis delineated relevant signaling cascades—all to explore the biological events and pathways potentially implicated in GBS-induced toxicity. Significantly enriched terms were selected using a P -value < 0.05 and FDR < 0.05 as thresholds, and the results were illustrated through bar plots and bubble plots.

Machine learning-driven identification of key targets in GBS-induced gastrointestinal toxicity

To identify key molecular targets associated with GBS-induced gastrointestinal toxicity, a machine learning-based approach was employed to refine the intersected gene set obtained from network toxicology analysis. Three widely used feature selection algorithms—LASSO regression, Random Forest (RF), and Support Vector Machine with Recursive Feature Elimination (SVM-RFE)—were utilized to pinpoint genes most critically associated with the toxic phenotype.

LASSO regression was carried out using R's `glmnet` package, where the optimal λ was identified through 10-fold cross-validation; only genes with non-zero coefficients were kept as putative toxicity-related targets. For RF analysis, the `randomForest` package was employed, with feature importance assessed by the Mean Decrease Gini index.

Additionally, SVM-RFE was executed via the e1071 package, combining recursive feature elimination and 10-fold cross-validation to minimize overfitting and prioritize robust, discriminative features.

The core genes identified by all three machine learning methods were determined through Venn diagram intersection analysis using the Jvenn online tool, yielding a high-confidence gene signature for downstream validation.

Validation of Key Gene Discriminatory Power Using Violin Plots and ROC Analysis

To validate the discriminatory power of the intersected key genes between control and gastrointestinal toxicity samples, two visualization approaches were employed.

First, violin plots with embedded boxplots were generated using the ggpubr package in R. Gene expression data were normalized via the limma package and averaged across replicates. Statistical significance was assessed using the Wilcoxon rank-sum test, with P -values annotated directly on the plots. Additionally, volcano plots were constructed to visualize the $\log_2\text{FC}$ and statistical significance ($-\log_{10} P\text{-value}$) of these genes; genes satisfying $|\log_2\text{FC}| > 0.585$ and $P\text{-value} < 0.05$ were designated as significantly differentially expressed, thereby distinguishing their expression profiles across the two groups.

Second, receiver operating characteristic (ROC) curves were constructed using the pROC package to evaluate the diagnostic performance of the selected genes. The area under the curve (AUC) and 95% confidence interval (CI) were computed, with the optimal cutoff value identified based on the Youden index.

SHAP Analysis for Interpretation of Machine Learning Predictions

To enhance the interpretability of the machine learning models and identify the most influential features contributing to the prediction of GBS-induced gastrointestinal toxicity, we performed SHapley Additive exPlanations (SHAP) analysis. SHAP values were calculated via the "shap" package in R, based on the trained models (LASSO, Random Forest, and SVM-RFE) from the previous step.

SHAP analysis provided a model-agnostic interpretation of feature importance across the trained models, enabling us to rank genes according to their contribution to toxicity prediction. Dependence plots were used to explore the interaction effects between gene expression levels and SHAP values for the top-ranked genes. SHAP bar plots were generated to visualize the overall importance of each gene, while beeswarm plots illustrated the direction and magnitude of SHAP values across samples. These

visualizations collectively offer a comprehensive and interpretable view of the molecular mechanisms underlying GBS-induced gastrointestinal toxicity.

Selection and Functional Analysis of MMP9 as a Key Toxicity-Associated Gene

MMP9 was selected as a key gene for further analysis based on its high predictive performance in ROC analysis and feature importance in SHAP analysis. To explore its functional role in the context of GBS-induced gastrointestinal toxicity, samples were divided into low- and high-MMP9 expression groups based on the median expression value as the threshold. Differential gene expression analysis was conducted using the limma package in R, and the expression profiles of significantly dysregulated genes were visualized through a clustering heatmap generated with the pheatmap package. Additionally, Pearson correlation analysis was conducted between MMP9 and other toxicity-related genes to assess co-expression patterns.

GSEA Analysis of MMP9 in Gastrointestinal Toxicity

To further explore the functional implications of MMP9 in GBS-induced gastrointestinal toxicity, we conducted Gene Set Enrichment Analysis (GSEA) with the clusterProfiler package in R, aiming to uncover pathways significantly linked to MMP9 expression levels.

For GSEA, the gene expression matrix was first normalized using the limma package and filtered for quality control. Based on the median expression level, samples were categorized into low- and high-MMP9 expression subgroups. A ranked gene list was generated by calculating the $\log_2FC$ between the two groups and sorting genes by their expression differences. This ranked gene list was used to perform GSEA using the c2.cp.kegg.Hs.symbols.gmt gene set, with the following parameters: $pvalueCutoff = 1$, and $P.adjustCutoff = 0.05$. Enriched pathways with normalized enrichment score (NES) > 1 or < -1 and $P < 0.05$ were considered significant. GSEA plots were generated for the top enriched pathways in the high- and low-MMP9 expression groups.

Immune Cell Infiltration Analysis Using CIBERSORT

To investigate the immunological profile of GBS-induced gastrointestinal toxicity, we applied the CIBERSORT algorithm to estimate the relative proportions of 22 immune cell types based on the LM22 signature matrix. Gene expression data were normalized using the preprocessCore package and submitted to CIBERSORT for deconvolution analysis. Only samples with $P < 0.05$ were retained for downstream analysis. The resulting immune infiltration matrix was then used to generate grouped barplots and boxplots for visual comparison between control and treatment groups. These visualizations offered insights into the immunotoxicological impact of GBS and aided in pinpointing immune cell populations potentially implicated in gastrointestinal toxicity.

Immune Function Enrichment Analysis

Following CIBERSORT-based immune cell infiltration analysis, we further explored the functional implications of immune responses in GBS-induced gastrointestinal toxicity using Gene set variation analysis (GSVA). The “immune. Gmt” gene set was used to define immune-related functional modules, and single-sample GSEA (ssGSEA) was carried out. Pathway scores were normalized using a min-max scaling function to allow for inter-pathway comparisons. Samples were classified into low- and high-MMP9 expression groups according to the median MMP9 expression level. The relationship between immune function scores and MMP9 expression was visualized and evaluated using boxplots.

Immune Cell Correlation Analysis

To investigate the relationship between MMP9 expression and immune cell infiltration in the context of GBS-induced gastrointestinal toxicity, we performed immuneCor analysis by integrating gene expression data with immune infiltration results from CIBERSORT. Gene expression data were first normalized using the limma package and filtered to include only treatment samples. The expression of MMP9 was then extracted and merged with the immune infiltration matrix from CIBERSORT to generate a unified dataset for correlation analysis.

Spearman's rank correlation analysis was performed to evaluate the association between MMP9 expression and the relative proportions of 22 immune cell types. Correlation coefficients and corresponding P-values were calculated, with statistical significance set at $P < 0.05$. For significant associations, a lollipop plot was generated using the R packages ggpubr and graphics to simultaneously visualize the correlation coefficients and significance levels across all immune cell types. This integrative visualization provides an intuitive overview of the immunomodulatory role of MMP9 in GBS-induced gastrointestinal toxicity.

Molecular Docking and Molecular Dynamics Simulation

To investigate the molecular interactions between the major toxic constituents of GBS and the key target MMP9, we adopted the computational protocol established in our prior work (Xu et al., 2025).

Briefly, the 3D structures of GAs and MPN were retrieved from TCMSP and PubChem (SDF format). Ligands were converted to the PDBQT format using Open Babel, while the MMP9 crystal structure (PDB ID: 1L6J) was prepared via AutoDock Tools (water removed, polar hydrogens added, Gasteiger charges assigned). Docking was executed with AutoDock Vina (exhaustiveness = 10); poses exhibiting binding energy ≤ -6 kcal mol⁻¹ were retained.

MD simulations (100 ns) for GA-MMP9, MPN-MMP9 complexes were run with GROMACS 2022 under the identical AMBER14SB/GAFF/TIP3P setting described in the above-cited paper (150 mM NaCl, 100 ps NVT + 100 ps NPT, 298 K, 1 bar, LINCS, 2 fs step). Trajectory analyses (RMSD, RMSF, Rg, H-bonds) and MM-PBSA free-energy calculations were likewise performed as reported therein. All visualizations were generated with VMD and PyMOL.

Preparation of GBS Powder and Drug-Containing Serum

GBS were purchased from the Department of Pharmacy, Wuhan Hospital of TCM, and authenticated by Professor Mei, a certified pharmacognosist. The seeds were lyophilized for 48 h and then ground into a fine powder using a mechanical grinder. The powder was sieved through a 60-mesh screen to ensure uniform particle size and stored at -20°C until use. This preparation method was performed according to previously established protocols (Yu, 2017).

Drug-containing serum was prepared as follows: Sprague-Dawley (SD) rats (8 weeks old, 300 ± 20 g, $n = 20$) were obtained from Beijing Charles River Laboratory Animal Technology Co., Ltd. and maintained in a specific pathogen-free (SPF) facility at 25°C with a 12-h light/dark cycle and 55% humidity (three rats of the same sex per cage). After one week of acclimatization, the rats were randomly divided into two groups ($n = 10$ per group): the GBS-treated group and the control group. The dosing regimen was designed based on human clinical dosage and body surface area scaling. According to the Chinese Pharmacopoeia (2020 edition), the daily dose for a 60 kg adult is 3-9 g (150 mg/kg). Using a human-to-rat interspecies scaling factor of 6.17 (Zou et al., 2012), the equivalent high-dose for rats was calculated as 900 mg/kg/day, which was considered the upper limit of the safe dose. To investigate the toxicological effects under high-exposure conditions while ensuring animal survival, a dose of 2000 mg/kg/day was selected for the GBS group, based on preliminary experimental results showing no mortality and observable gastrointestinal toxicity at this dose level. This dose was administered via oral gavage twice daily for 3 days, with the control group receiving an equal volume of distilled water.

On the fourth day, 1 h after the last gavage, all rats were anesthetized by intraperitoneal injection of 2% sodium pentobarbital. Blood was collected via abdominal aorta puncture, incubated at room temperature for 1 h, and then centrifuged at 1500g for 15 min at 4°C. The supernatant serum was harvested, heat-inactivated at 56°C for 30 min, filtered through a 0.22- μ m filter for subsequent use.

All experimental procedures were reviewed and approved by the Institutional Animal Care and Use Committee of Shouzheng Pharma (Wuhan) Biotechnology Co., Ltd. (Approval No. 2025080801, 1 August 2025), following the guidelines outlined in GB/T 35892-2018.

Profiling of Bioactive and Blood-Entering Constituents in GBS-Containing Serum

The presence and systemic exposure of ginkgolic acids (GAs), 4'-O-methylpyridoxine (MPN), and 4'-O-methylpyridoxine-5'-glucoside (MPNG) in GBS-containing serum were confirmed using UHPLC-Q Exactive HFX-MS. All experimental procedures were conducted according to the previously published protocol (Qi et al., 2025). This analysis verified the bioavailability of these compounds, supporting the use of GBS-containing serum for subsequent experiments.

Cell Culture and Viability Assay

The human gastric epithelial cell line GES-1 and the colorectal adenocarcinoma cell line Caco-2 were purchased from Pricella (Wuhan, Hubei Province, China). Cells were cultured in Dulbecco's Modified Eagle Medium (DMEM) (Servicebio, Wuhan, China) supplemented with 10% fetal bovine serum (FBS) (MeiSenCTCC, Shanghai, China) and maintained at 37°C in a humidified incubator with 5% CO₂. Cells were passaged using 0.25% trypsin-EDTA, transferred to 96-well plates (3.5 × 10⁴ cells/well), and allowed to adhere overnight. On the following day, the medium was aspirated and replaced with fresh medium containing serum derived from GBS-dosed or vehicle-treated rats.

The final serum concentrations used were 0% (control), 5%, 10%, and 20%. Cells were exposed to the serum for 24, 48, and 72 h. After treatment, 10 µL of Cell Counting Kit-8 (CCK-8) reagent (Beyotime, Shanghai, China) was introduced into each well, followed by an additional incubation period of 1–2 h at 37°C. Absorbance was then recorded at 450 nm with a microplate reader (PerkinElmer, Massachusetts, USA). Cell viability was presented as a percentage relative to the untreated control group, calculated using the formula: Cell Viability (%) = $(A_{\text{treatment}} - A_{\text{blank}}) / (A_{\text{control}} - A_{\text{blank}}) \times 100$.

Cell Proliferation Assay by EdU Staining

To evaluate the effect of GBS-containing serum on cell proliferation, a 5-ethynyl-2'-deoxyuridine (EdU) incorporation assay was performed using the EdU Cell Proliferation Kit (Beyotime, Shanghai, China). Briefly, GES-1 and Caco-2 cells were seeded into 48-well plates at a density of 1.5 × 10⁴ cells per well, followed by overnight culture to promote cell attachment. Cells were then exposed to GBS-containing serum at final concentrations of 0%, 5%, 10%, and 20% for a 48 h incubation period.

Following treatment, 50 µM EdU was introduced into each well, followed by a 2h incubation at 37 °C. Cells were subsequently fixed in 4% paraformaldehyde for 15 min, permeabilized with 0.3% Triton X-100 for 10 min, and rinsed three times with PBS supplemented with 3% bovine serum albumin (BSA). Click-iT reaction cocktail (containing Alexa Fluor 594 azide, CuSO₄, reaction buffer, and additive) was added to each well and

incubated for 30 min at room temperature in the dark. Nuclei were counterstained with Hoechst 33342 (1:1000 dilution) for 10 min. Images were captured using an inverted fluorescence microscope (Olympus Corporation, Tokyo, Japan).

Lactate Dehydrogenase (LDH) Release and Total LDH Assay

To evaluate the cytotoxic effects of GBS-containing serum on GES-1 and Caco-2 cells, both total LDH activity and LDH release was quantified using the LDH Cytotoxicity Assay Kit (C0016, Beyotime, Shanghai, China) .

For the total LDH activity assay, after treatment with GBS-containing serum (0%, 5%, 10%, 20%) for 48 h, the cells in the original culture plate were lysed with 1% Triton X-100 to release all intracellular LDH. After 1 h of incubation at 37°C, 100 µL of the cell lysate was transferred to a new 96-well plate. An equal volume of LDH detection working solution was added to each well, and the plate was incubated at 37°C for 30 min in the dark. The reaction was terminated by adding 50 µL of stop solution, and absorbance was read at 490 nm using a microplate reader.

For the LDH release assay, following 48 h of treatment with GBS-containing serum at concentrations of 0%, 5%, 10%, and 20%, 100 µL of the cell culture supernatant was transferred to a fresh 96-well plate. The same detection procedure as the total LDH assay was followed.

Prior to any calculations, the absorbance of the background blank control wells (containing medium and reagents without cells) was subtracted from all sample absorbance values. The percentage of cytotoxicity was then calculated using the following formula: Cytotoxicity (%) = $(A_{\text{treatment}} - A_{\text{control}}) / (A_{\text{maximum}} - A_{\text{control}}) \times 100$.

Glutathione (GSH) Assay

To evaluate the impact of GBS-containing serum on the intracellular antioxidant defense system, reduced GSH levels in GES-1 and Caco-2 cells were measured using the Total Glutathione Assay Kit (Beyotime, Shanghai, China), according to the manufacturer's instructions. Briefly, cells were seeded in 12-well plates at a density of 1×10^5 cells/well and incubated with GBS-containing serum at final concentrations of 0%, 5%, 10%, and 20% for 48 h, with three replicates per condition. After treatment, cells were washed twice with ice-cold PBS, collected by scraping, and lysed with 5% protein removal reagent M solution freshly prepared in total glutathione detection buffer. The lysate was incubated on ice for 10 min and centrifuged at $10,000 \times g$ for 10 min at 4°C. The supernatant was used immediately for GSH measurement. Total glutathione and oxidized glutathione (GSSG) were quantified using a 96-well plate format: 10 µL of sample or standard was combined with 150 µL of total glutathione detection working solution (containing glutathione

reductase diluted 1:5 and DTNB) and incubated at 25°C for 5 min, and then 50 µL of 0.5 mg/mL NADPH was added to initiate the reaction. The absorbance at 412 nm was measured every 5 min for 25 min using a microplate reader. A standard curve was generated using GSSG standards (0.5-15 µM), and total glutathione concentration was calculated as twice the GSSG-equivalent concentration. GSH concentration was calculated by subtracting twice the GSSG concentration from the total glutathione level. All values were normalized to the total protein content.

Intracellular Reactive Oxygen Species (ROS) Detection by Fluorescence Microscopy

To investigate whether GBS-induced cytotoxicity is mediated through oxidative stress, ROS levels were measured using the fluorescent probe 2',7'-dichlorofluorescein diacetate (DCFH-DA) (Beyotime, Shanghai, China) and fluorescence microscopy. Briefly, GES-1 and Caco-2 cells were seeded into 48-well plates at 1.5×10^4 cells per well and cultured overnight to facilitate attachment, followed by exposure to GBS-containing serum at distinct final concentrations (0%, 5%, 10%, and 20%) and cultured for 48 h.

After treatment, cells were rinsed with PBS, followed by incubation in serum-free DMEM containing 10 µM DCFH-DA at 37°C for 20 min in the dark. For positive control, a separate set of cells was treated with 10 µM Rosup (Beyotime, Shanghai, China) for 30 min after DCFH-DA loading. Fluorescence images were immediately acquired using a fluorescence microscope with excitation at 488 nm and emission at 525 nm. The intensity of green fluorescence, indicative of ROS levels, was analyzed using ImageJ software.

Cell Apoptosis Assay by Flow Cytometry

To evaluate the pro-apoptotic effects of GBS-containing serum on GES-1 and Caco-2 cells, annexin V-FITC/PI dual staining was performed using the Annexin V-FITC Apoptosis Detection Kit (Apexbio, Texas, USA), according to the manufacturer's instructions. Briefly, cells were seeded into 6-well plates and treated with GBS-containing serum-containing medium at final serum concentrations for 48 h.

After treatment, cells were harvested using EDTA-free trypsin (Servicebio, Wuhan, China), washed twice with cold PBS, and resuspended in 500 µL of 1×binding buffer. Then, 5 µL of annexin V-FITC and 5 µL of propidium iodide (PI) were added to each sample, followed by 15 min incubation at room temperature in the dark. Apoptosis was immediately analyzed using a CytoFLEX flow cytometer (FACSAria™ II, New Jersey, USA), with data acquisition and analysis performed using FlowJo software (v10.8.1, BD Biosciences, San Jose, CA, USA).

Hoechst/PI staining assay

To further evaluate the effects of GBS-containing serum on nuclear morphology and membrane integrity, a dual fluorescence staining assay using Hoechst 33342 and PI was performed. Briefly, GES-1 and Caco-2 cells were seeded into 48-well plates at a density of 1.5×10^4 cells per well and cultured overnight to allow for attachment. Cells were then treated with GBS-containing serum at final serum concentrations for 48 h.

After treatment, cells were washed with PBS, followed by staining with a mixture of 10 $\mu\text{g/mL}$ Hoechst 33342 and 10 $\mu\text{g/mL}$ PI in the dark at room temperature for 15 min. The cells were then washed twice with PBS to remove excess dye and imaged using a fluorescence microscope.

RNA Extraction and RT-qPCR Analysis

To evaluate the mRNA expression levels of key genes involved in GBS-induced gastrointestinal toxicity, quantitative real-time PCR (qPCR) was performed on GES-1 and Caco-2 cells treated with GBS-containing serum. Cells were seeded into 6-well plates and incubated with GBS-containing serum at varying final concentrations (0%, 5%, 10%, and 20%) and incubated for 24 h. Total RNA was extracted using TRIzol® reagent (Servicebio, Wuhan, China) according to the manufacturer's instructions. RNA concentration and purity were determined using a NanoDrop One spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA), with A260/A280 ratios between 1.8 and 2.0 considered acceptable.

Complementary DNA (cDNA) was synthesized from 1 μg of total RNA using the PrimeScript™ RT Master Mix (Takara, Kyoto, Japan) in a 20 μL reaction volume. qPCR was performed using the TB Green® Premix Ex Taq™ II (Takara, Kyoto, Japan) on a StepOnePlus™ Real-Time PCR System (Applied Biosystems, Waltham, MA, USA). The thermal cycling conditions were as follows: initial denaturation at 95°C for 30 s, followed by 40 cycles of 95°C for 5 s and 60°C for 30 s. The relative mRNA expression of the target gene MMP9 was determined using the $2^{-\Delta\Delta C_q}$ method, with β -actin serving as the reference gene for normalization. Primer sequences were designed as follows and synthesized by Sangon Biotech (Shanghai, China): MMP9- Forward: 5'-GGACCACAACCTCGTCATCGT-3', Reverse: 5'-ACCTTCACTCGCGTGTACAG-3'; β -actin - Forward: 5'-CATGTACGTTGCTATCCAGGC-3', Reverse: 5'-CTCCTTAATGTCACGCACGAT-3'.

Western Blot Analysis

To evaluate the protein expression levels of key targets involved in GBS-induced toxicity, Western blot (WB) analysis was performed on GES-1 and Caco-2 cells treated with GBS-containing serum. Total protein was isolated using RIPA lysis buffer (Servicebio, Wuhan,

China) containing phenylmethanesulfonyl fluoride (PMSF) to suppress protease activity. Protein concentration was determined using the Bicinchoninic Acid (BCA) Assay Kit (Beyotime, Shanghai, China), according to the manufacturer's instructions. Equal amounts of protein were separated by 10% Sodium Dodecyl Sulfate-Polyacrylamide Gel Electrophoresis (SDS-PAGE) (Epizyme, Shanghai, China) and transferred onto nitrocellulose (NC) membranes (MilliporeSigma, St. Louis, USA) via wet transfer. The membranes were blocked with QuickBlock™ Blocking Buffer (Beyotime, Shanghai, China) for 15 min at room temperature to reduce non-specific binding. After blocking, membranes were incubated overnight at 4°C with primary antibodies against Ki67, MMP9, Caspase-3, Bcl-2, Bax, NF-κB p65, Phospho-NF-κB p65, IL-6 and β-Actin (see Table 1 for antibody details). Following three washes with Tris-Buffered Saline with Tween 20 (TBST), membranes were incubated with horseradish peroxidase (HRP)-conjugated goat anti-rabbit IgG (H+L) secondary antibody (Abways, Shanghai, China) for 1 h at room temperature. After three additional TBST washes, protein bands were visualized using ECL chemiluminescent reagent (Vazyme, Nanjing, China) and imaged with a ChemiScope 6200 gel imaging system (Clinx Science Instruments, Shanghai, China). The gray values of the protein bands were quantified using ImageJ software (v1.54, National Institutes of Health, USA). Protein expression levels were normalized to β-Actin as the internal loading control, and relative expression changes were calculated using the control group as a reference.

Immunofluorescence (IF) Staining

To further validate the upregulation of MMP9 protein expression and visualize its subcellular localization, IF staining was performed on GES-1 and Caco-2 cells treated with GBS-containing serum. Cells were plated on glass coverslips in 24-well plates and incubated with GBS-containing serum at concentrations of 0% and 20% for 48 h. After treatment, cells were fixed with 4% paraformaldehyde for 15 min, followed by permeabilization with 0.3% Triton X-100 in PBS for 15 min. The cells were then blocked with 5% BSA in PBS for 2 h to minimize non-specific binding. Subsequently, the cells were incubated with a primary antibody against MMP9 overnight at 4°C. After three washes with PBS, the cells were incubated with a Goat Anti-Rabbit IgG (Cy3) secondary antibody for 1 h at room temperature in the dark. After 15 min DAPI staining (Beyotime, Shanghai, China), nuclei were visualized by fluorescence microscopy, and MMP9 fluorescence intensity was quantified using ImageJ.

Cellular Thermal Shift Assay (CETSA)

To validate the direct interaction between GBS-containing serum and the key target MMP9 in a cellular context, a CETSA was performed using GES-1 and Caco-2 cells. The assay was conducted based on the principle that ligand binding stabilizes proteins against

thermal denaturation (Tu et al., 2023). In brief, cells were exposed to either 20% GBS-containing serum or an equivalent volume of dimethyl sulfoxide (DMSO) and incubated on a horizontal shaker for 1 h at room temperature. After treatment, cells were harvested, washed with PBS, and divided into eight aliquots, which were heated at different temperatures (37, 44, 51, 58, 65, 72°C) for 3 min in a thermal cycler. Following heat treatment, cells were immediately chilled on ice for 3 min and subjected to three freeze-thaw cycles (-80°C to 37°C) to lyse the cells. After centrifugation at 12,000 × g for 15 min at 4°C, the supernatant was harvested and processed for WB.

Statistical Analysis

All data are presented as mean ± standard deviation from at least three independent experiments. Statistical analyses were performed via GraphPad Prism 9.0 (GraphPad Software, San Diego, CA, USA) and R software (v4.2.0). The Shapiro-Wilk test was employed to evaluate the normality of data distribution, and an unpaired Student's t-test was used for comparing two groups with normally distributed data, while the Mann-Whitney U test was applied for non-normally distributed data. For comparisons among multiple groups, one-way analysis of variance (ANOVA) followed by Tukey's post hoc test was used for normally distributed data, and the Kruskal-Wallis test followed by Dunn's post hoc test was used for non-normally distributed data. For dose-response and time-course experiments, two-way ANOVA with Sidak's or Tukey's multiple comparisons test was performed. A *P*-value of less than 0.05 was considered a statistically significant difference (**P* < 0.05, ***P* < 0.01, ****P* < 0.001, ns indicates not significant).

Results

Network Toxicology Analysis of GBS-Induced Gastrointestinal Toxicity

To identify the potential molecular targets underlying GBS-induced gastrointestinal toxicity, a comprehensive network toxicology approach was employed. Putative targets for the major toxic constituents—GAs, MPN, MPNG—were predicted by integrating data from multiple bioinformatics databases, including TCMSP, ChEMBL, SwissTargetPrediction, PharmMapper, and SEA. After removing duplicates, these 524 targets were considered as the drug-target set (Figure 2A).

Concurrently, 11,717 disease-associated targets related to gastrointestinal toxicity were retrieved from public databases such as OMIM, GeneCards, PharmGKB, and TTD (Figure 2B). These targets were associated with conditions including IBD, gastritis, and intestinal injury, which were used as phenotypic models of gastrointestinal toxicity.

The intersection of the drug-target set and disease-target set yielded 439 common targets, which are potentially involved in mediating the toxic effects of GBS on the gastrointestinal

system (Figure 2C). A component-target-disease network was constructed using Cytoscape 3.8.2 to visualize the complex interactions among the three toxicity-related components, their shared targets, and the gastrointestinal toxicity (Figure 2D).

The 439 overlapping targets were subjected to PPI network analysis to elucidate their functional interrelationships. The gene list was analyzed using the STRING database, yielding a PPI network composed of 437 nodes and 5,649 edges, was imported into Cytoscape 3.8.2 for visualization and topological assessment. To identify the most central genes within the network, the cytoHubba plugin was applied using the Maximal Clique Centrality (MCC) algorithm, which ranked the top 10 hub genes as BCL2L1, STAT3, EGFR, MMP9, CASP3, HSP90AA1, ESR1, SRC, ALB, and AKT1 (Figure 2E). These hub genes are implicated in key biological processes such as apoptosis, inflammation, cell proliferation, and signal transduction, suggesting their critical roles in the molecular mechanisms underlying GBS-induced gastrointestinal toxicity. Notably, MMP9 emerged as a central hub node, highlighting its potential as a key regulator in the toxicity network.

Identification of Gastrointestinal Toxicity-Related Genes from GEO Datasets

To identify molecular signatures associated with gastrointestinal toxicity, gene expression profiles from publicly available GEO datasets (including GSE3365, GSE36807, GSE233973, and GSE5081) were analyzed. Raw data were first normalized and batch effects adjusted with the ComBat method, prior to initiating differential expression analysis. Principal component analysis (PCA) demonstrated effective data normalization, with samples clearly clustering by disease status before and after batch correction (Figure 3A-B). DEGs were identified by comparing gastrointestinal toxicity samples with healthy controls, using the thresholds $|\log_2FC| > 0.415$ and $P < 0.05$. This analysis yielded a total of 758 DEGs, comprising 550 upregulated and 208 downregulated genes (Figure 3C). Concurrently, a WGCNA was performed to identify gene modules associated with gastrointestinal toxicity. Based on a soft-thresholding power of 5, a scale-free topology ($R^2 > 0.85$) was achieved, and genes were clustered into distinct modules. The module-trait relationship analysis revealed that the MEbrown module showed the highest positive correlation with gastrointestinal toxicity ($r = 0.58$, $P = 5.81e-21$) (Figure 3D-F). Therefore, the genes within the MEbrown module were selected for further analysis.

To prioritize high-confidence candidate genes, an intersection analysis was conducted among three gene sets: (1) the 439 overlapping targets identified from network toxicity, (2) the 758 DEGs from GEO datasets, and (3) the genes in the MEbrown module from WGCNA. This stringent integration yielded 13 key genes that are potentially central to GBS-induced gastrointestinal toxicity (Figure 3G). These 13 key genes represent a high-confidence target set for subsequent machine learning and functional validation.

Functional Enrichment Analysis Reveals Key Biological Processes and Pathways

To gain insights into the biological functions and signaling pathways associated with the 13 overlapping targets identified in the bioinformatics analysis, GO and KEGG enrichment analyses were performed.

For GO analysis, the top enriched terms were categorized into three domains. In the BP category, the most significantly enriched terms included collagen catabolic process, extracellular matrix disassembly, cellular response to UV-A, response to UV-A, collagen metabolic process, and multiple purine nucleotide catabolic processes, suggesting active extracellular matrix remodeling and nucleotide metabolism in Ginkgo biloba seed-induced gastrointestinal toxicity. The CC analysis revealed enrichment in serine-type peptidase complex, ficolin-1-rich granule lumen, specific granule, tertiary granule, peptidase inhibitor complex, and protein complex involved in cell-matrix adhesion, indicating the involvement of specialized secretory granules and protease complexes in the toxic response. For MF, the top terms were serine-type endopeptidase activity, metallopeptidase activity, phosphoric ester hydrolase activity, and protein tyrosine/threonine phosphatase activity, highlighting the critical roles of proteolytic enzymes and phosphatases in mediating the toxic effects (Figure 3H).

KEGG pathway analysis further revealed that the target genes were significantly enriched in several key signaling pathways. The most significantly enriched pathways included IL-17 signaling pathway and Prostate cancer, both of which are closely associated with inflammation and tissue damage. Additionally, Complement and coagulation cascades and Transcriptional misregulation in cancer were also significantly enriched, underscoring the roles of immune system activation, hemostatic dysregulation, and transcriptional control in the pathogenesis of gastrointestinal toxicity (Figure 3I).

Machine Learning-Based Identification of Key Genes for Gastrointestinal Toxicity

To prioritize the most discriminative genes from the 13 intersected targets, a multi-algorithm machine learning approach was employed, integrating LASSO regression, SVM-RFE, and RF.

First, LASSO logistic regression was performed to identify a minimal gene signature with high predictive power. The LASSO coefficient profiles showed a progressive reduction in the number of non-zero coefficients as the penalty parameter (λ) increased. Ten-fold cross-validation was used to determine the optimal λ value ($\log(\lambda) = -4$), which minimized the binomial deviance and yielded a 11-gene signature (Figure 4A). These eleven genes were selected as potential biomarkers for GBS-induced gastrointestinal toxicity.

Next, SVM-RFE was applied to identify the most stable and discriminative gene signature. The algorithm iteratively eliminated less informative features and ranked genes based on their contribution to classification performance, with lower ranks indicating higher

importance. The cross-validation error curve exhibited a characteristic U-shape, with the minimum error rate (13.1%) achieved when 6 genes were retained. The corresponding accuracy curve peaked at 86.9% with this 6-gene set, outperforming the baseline no-information rate (Figure 4B).

Finally, Random Forest analysis was conducted using 500 trees. The out-of-bag (OOB) error progressively decreased and stabilized as the number of trees increased, indicating model convergence. Gene importance was ordered according to the Mean Decrease Gini index, and a gradient bubble plot was created to visualize the relative contributions of each gene. Notably, MMP9, TREM1, and LCN2 exhibited the highest importance scores, suggesting their critical roles in distinguishing toxicity samples from controls (Figure 4C).

The overlapping genes identified by all three methods were considered high-confidence key genes for subsequent validation (Figure 4D).

Validation of Key Genes by Expression and Diagnostic Performance Analysis

To validate the discriminatory power of the six key genes identified through integrated machine learning (MMP9, TREM1, PLAU, DUSP6, F3, and LCN2), their expression profiles and diagnostic performance were further evaluated. Violin plots with embedded boxplots revealed significant differences in the expression levels of all six genes between gastrointestinal toxicity samples and controls (Figure 4E). Specifically, MMP9, TREM1, PLAU, DUSP6, F3, and LCN2 were all significantly upregulated in the toxicity group (all $P < 0.001$, Wilcoxon rank-sum test), indicating their strong association with the disease state.

Additionally, volcano plots were constructed to visualize the fold changes ($\log_2FC$) and statistical significance ($-\log_{10} P$ -value) of these genes; genes meeting the criteria of $|\log_2FC| > 0.585$ and P -value < 0.05 were designated as significantly differentially expressed, thereby distinguishing their expression profiles between the two groups (Figure 4F).

To assess their diagnostic accuracy, ROC curves were constructed (Figure 4G). Among the six genes, MMP9 exhibited the highest predictive performance, with an AUC value of 0.873 (95% CI: 0.921–0.917). These results demonstrate that the six-gene signature, particularly the top-ranking gene MMP9, has excellent sensitivity and specificity for distinguishing GBS-induced gastrointestinal toxicity from normal conditions.

SHAP Analysis Reveals Key Drivers of Model Prediction

To enhance the interpretability of the gene signature identified by machine learning and identify the most influential genes in predicting GBS-induced gastrointestinal toxicity, SHAP analysis was performed on the six key genes.

To provide a clear quantitative ranking, the mean absolute SHAP values were also visualized as a bar plot (Figure 4H-I). This plot confirmed that MMP9 has the highest contribution to the model (0.160), followed by TREM1 (0.092), PLAU (0.079), DUSP6 (0.078), LCN2 (0.033), and F3 (0.017). This quantitative ranking further validates the central role of this six-gene signature in the toxicity prediction.

The SHAP beeswarm plot (Figure 4J) provides a global view of feature importance by ranking genes based on their mean absolute SHAP value. This plot not only shows the relative importance of each gene but also reveals the direction of their impact on the model's prediction. MMP9 exhibited the highest importance, followed by TREM1, PLAU, DUSP6, LCN2, and F3. The color of the points represents the expression level of each gene (red = high, orange = low). For MMP9, TREM1, and PLAU, high expression (red points) is predominantly associated with positive SHAP values, indicating that elevated levels of these genes strongly push the model's prediction towards the “toxicity” class. This pattern confirms their role as pro-toxic drivers.

To further explore the complex interplay within the gene signature, we analyzed the SHAP dependence patterns. This analysis revealed critical non-linear relationships: TREM1, DUSP6, LCN2, and F3 exhibited expression patterns that were coordinately upregulated with MMP9, suggesting a synergistic network where high expression of MMP9 is associated with high expression of these pro-inflammatory and stress-response genes (Figure 4K). In contrast, PLAU displayed a distinct pattern, with its expression showing a non-linear inverse relationship with MMP9, indicating a potentially antagonistic or independently regulated role within the toxicity network.

Force-directed analysis (Figure 4L) illustrates how individual gene features collectively shift the model output from the baseline expectation ($E[f(x)] = 0.699$) to the final high-risk prediction ($f(x) = 0.986$) in sample GSM7440039. In this sample, MMP9 emerged as the dominant positive driver, contributing the largest incremental increase in the predicted probability ($\Delta = +0.203$), indicating that its elevated expression is a primary determinant of toxicity classification. TREM1 ($\Delta = +0.0606$) and DUSP6 ($\Delta = +0.0523$) also acted as consistent positive contributors, reinforcing the pro-toxicity signal. In contrast, PLAU exhibited a slight negative contribution ($\Delta = -0.0451$), suggesting that its expression level in this instance was insufficient to promote toxicity and may have exerted a modest dampening effect on the prediction. These results highlight MMP9 as the pivotal regulator in driving the high-risk prediction for this GBS-exposed sample.

Functional Role of MMP9 in GBS-Induced Gastrointestinal Toxicity

To elucidate the functional role of MMP9—the top-ranked gene identified through integrated bioinformatics and machine learning analysis—we performed a comprehensive

investigation of its co-expression network, pathway enrichment, and biological implications.

Clustering heatmap analysis revealed a distinct expression pattern in gastrointestinal samples, with MMP9 exhibiting significant upregulation in toxicity samples. This high expression was strongly correlated with a cluster of pro-inflammatory genes, including CXCL6, CXCL8, and CXCR2, forming a tightly co-expressed network (Figure 5A). In contrast, genes associated with mucosal function, such as SLC3A1 and SI, showed low expression, suggesting a coordinated impairment of epithelial barrier integrity in the high-MMP9 state.

Further Pearson correlation analysis indicated a notably positive correlation between MMP9 expression and core mediators involved in inflammation as well as immune cell recruitment (Figure 5B). Strong correlations were observed with BCL2A1 ($r > 0.6$, apoptosis inhibition), VNN2 ($r > 0.6$, oxidative stress), and FPR1 ($r > 0.6$, chemotaxis), as well as with CXCR2 (neutrophil chemotaxis), suggesting that MMP9 may drive immune cell infiltration into the gastrointestinal tract.

To investigate the pathway-level implications of elevated MMP9 expression, GSEA was conducted. The high-MMP9 group exhibited significant enrichment in central immune and inflammatory pathways, such as the Toll-like receptor signaling pathway, chemokine signaling pathway, and cytokine-cytokine receptor interaction (Figure 5C), indicating a pro-inflammatory transcriptional program.

MMP9 Is Associated with a Pro-Inflammatory Immune Microenvironment in Gastrointestinal Toxicity

To investigate the potential role of MMP9 in modulating the immune microenvironment of GBS-induced gastrointestinal toxicity, we performed CIBERSORT analysis to deconvolute the proportions of 22 immune cell types from bulk gene expression data (Figure 5D). The analysis revealed a significant shift in immune cell composition between the control and toxicity groups. In the control group, T cells CD4 memory resting and mast cells resting were significantly elevated, suggesting a state of immune quiescence and surveillance. In contrast, the toxicity group exhibited a marked increase in neutrophil infiltration, indicating a robust activation of the innate immune response and acute inflammatory state.

Further correlation analysis demonstrated a significant positive correlation between MMP9 expression and multiple immune cell populations, functional pathways (Figure 5E), as well as various pro-inflammatory and immune-activating modules—including those related to inflammation promotion and parainflammation; antigen presentation pathways (APC co-stimulation, MHC class I); chemokine receptor (CCR) signaling; Type I and Type

II interferon responses; T cell co-stimulation signals and T follicular helper (Tfh) and Th1 cell scores. Notably, MMP9 was also positively correlated with immune checkpoint signaling and regulatory T (Treg) cell scores, suggesting a complex interplay between immune activation and potential feedback regulation in the inflamed tissue. Additionally, tumor-infiltrating lymphocyte (TIL) score was strongly associated with high MMP9 expression, reflecting an overall increase in immune cell infiltration.

To investigate the association between MMP9 expression and specific immune cell subsets, Spearman's rank correlation analysis was conducted, and a lollipop plot was constructed to visualize the correlation profile (Figure 5F). MMP9 expression was significantly positively correlated with several key innate immune cells, most notably M0 macrophages and neutrophils, as well as activated mast cells. In contrast, MMP9 expression showed a negative correlation with CD4⁺ resting memory T cells and CD8⁺ T cells, suggesting a potential suppression of adaptive immune memory in the context of high MMP9-driven inflammation.

Metabolite Identification by UHPLC-Q Exactive HFX-MS

To confirm the presence and systemic bioavailability of key toxic constituents in GBS, an untargeted metabolomics approach was employed using UHPLC-Q Exactive HFX-MS. The analysis revealed the characteristic chromatographic peaks of three major toxicants: GAs, MPN, and MPNG. These compounds were successfully identified based on their retention times and MS/MS fragmentation patterns, which were matched against a self-built TCM secondary mass spectrometry database and reference standards. The retention times for GAs, MPN, and MPNG were 6.88 min, 7.52 min, and 9.35 min, respectively (Figure 6A-D). This result verifies that these toxic constituents are not only present in raw GBS but are also absorbed into the bloodstream, as evidenced by their detection in the GBS-containing serum. Their systemic bioavailability provides a direct link between oral consumption of GBS and the subsequent exposure of gastrointestinal epithelial cells to these toxins, thereby enabling their direct contribution to gastrointestinal toxicity.

Molecular Dynamics Simulations Reveal Stable Interactions between GBS Toxicants and MMP9

To further evaluate the binding stability and dynamic interactions between the major toxic constituents of GBS and the key target MMP9, molecular docking and MD simulations were performed for three complexes: GAs, MPN bound to MMP9.

Molecular docking analysis revealed favorable binding energies for all three compounds, with GAs-MMP9 exhibiting the strongest binding affinity (-8.4 kcal/mol), followed by MPN-MMP9 (-7.6 kcal/mol), suggesting a high potential for stable complex formation (Figure 6E).

The subsequent 100-nanosecond MD simulations demonstrated that all two complexes reached a state of equilibrium with stable RMSD values. The GAs-MMP9 complex showed a stable RMSD plateau around 2.0 Å, suggesting a high degree of structural stability. Similarly, the MPN-MMP9 complex exhibited a stable RMSD of approximately 1.8 Å, indicating a rigid and well-maintained protein-ligand conformation. (Figure 6F).

Furthermore, the RMSF analysis revealed that the binding of all two ligands significantly reduced the flexibility of the MMP9 active site, particularly in the catalytic domain (residues 110-120 and 240-250). The GAs-MMP9 complex showed the most pronounced reduction in fluctuation, highlighting its strong interaction with the protein (Supplementary 2).

The Rg values remained constant throughout the simulation for all complexes, indicating no significant change in the overall compactness of the protein. The SASA of the protein also remained stable, suggesting no major conformational changes upon ligand binding (Supplementary 1 and 4). The center-of-mass distance between each ligand and the MMP9 protein rapidly decreased and stabilized within the first 10 ns, indicating the formation of a stable complex (Supplementary 3). The binding free energy decomposition analysis revealed that van der Waals and electrostatic interactions were the dominant forces driving the binding, with the GAs-MMP9 complex showing the most favorable energy contributions. Additionally, the number of hydrogen bonds between the ligands and MMP9 remained stable over time, with the GAs-MMP9 complex forming the highest average number of hydrogen bonds (Supplementary 5-6 and Figure 6G).

Collectively, these results provide strong evidence that GAs, MPN can form stable and persistent complexes with MMP9, with GAs exhibiting the most robust binding characteristics. This supports the hypothesis that these toxic constituents may directly interact with and modulate the activity of MMP9, contributing to GBS-induced gastrointestinal toxicity.

GBS-Containing Serum Inhibits Cell Proliferation in GES-1 and Caco-2 Cells

To evaluate the cytotoxic effects of GBS-containing serum on gastrointestinal epithelial cells, a CCK-8 assay was performed. The findings revealed that GBS-positive serum exerted a notable inhibitory effect on the proliferation of both GES-1 and Caco-2 cells, with the effect dependent on time and concentration. Exposure to 10% and 20% serum for 48 h led to a marked reduction in cell viability, with the effect intensifying over time and at higher concentrations. The 20% serum group exhibited the strongest inhibition, indicating a robust dose-response relationship (Figure 7A).

To further validate these findings, an EdU incorporation assay was performed. Consistent with the CCK-8 results, treatment with 10% and 20% GBS-containing serum for 48 h

significantly suppressed DNA synthesis ($P < 0.05$), as evidenced by a marked reduction in EdU-positive cells (Figure 7B).

GBS-Containing Serum Induces Cytotoxicity in GES-1 and Caco-2 Cells

To further evaluate the cytotoxic effects of GBS-containing serum, both total intracellular LDH activity (a surrogate for viable cell number) and extracellular LDH release (an indicator of membrane integrity) were measured in GES-1 and Caco-2 cells. After 48 h of treatment, the total intracellular LDH activity was significantly reduced in both cell lines treated with 5%, 10%, and 20% GBS-containing serum ($P < 0.05$), with the 20% group exhibiting the most pronounced decrease, suggesting a significant loss of viable cells. Concomitantly, the release of LDH into the culture medium was significantly increased in a concentration-dependent manner ($P < 0.01$), confirming membrane damage and cytotoxicity (Figure 7C).

These findings, combined with the CCK-8 and EdU results, demonstrate that GBS-containing serum exerts significant cytotoxic effects on gastrointestinal epithelial cells through multiple mechanisms, including inhibition of proliferation and induction of cell membrane damage.

GBS-Containing Serum Induces Oxidative Stress in GES-1 and Caco-2 Cells

To evaluate the impact of GBS-containing serum on the cellular redox balance, intracellular reduced GSH levels and the GSH/GSSG ratio were measured in GES-1 and Caco-2 cells after 48 h of treatment. Results indicated that serum containing GBS exerted a significant depleting effect on intracellular GSH, with the effect being concentration-dependent. In comparison to the control (0% serum), exposure to 10% and 20% GBS-containing serum caused a notable reduction in GSH levels across both cell lines ($P < 0.05$). Furthermore, the GSH/GSSG ratio, a key indicator of cellular oxidative stress, was markedly decreased in cells treated with 10% and 20% GBS-containing serum ($P < 0.01$) (Figure 7D).

Consistent with these findings, fluorescence microscopy analysis using the DCFH-DA probe demonstrated a marked elevation in intracellular ROS levels. Cells treated with 10% and 20% GBS-containing serum exhibited markedly enhanced green fluorescence intensity compared to the control, with the 20% treatment group showing the most pronounced effect ($P < 0.001$). These findings imply that GBS-containing serum triggers severe oxidative stress in gastrointestinal epithelial cells (Figure 7E).

GBS-Containing Serum Induces Apoptosis in GES-1 and Caco-2 Cells

To investigate the pro-apoptotic effects of GBS-containing serum, both Hoechst 33342/PI dual staining and annexin V-FITC/PI flow cytometry were performed on GES-1 and Caco-2

cells after 48 h of treatment.

Hoechst-PI staining revealed characteristic morphological changes associated with apoptosis. With increasing concentrations of GBS-containing serum, the number of PI-positive cells (indicating late apoptosis or necrosis, red fluorescence) increased significantly. Concurrently, Hoechst 33342-stained nuclei exhibited clear signs of chromatin condensation and nuclear fragmentation (bright, fragmented blue fluorescence), which are hallmarks of apoptotic cell death. The most pronounced apoptotic effects were observed in the 20% serum group ($P < 0.001$), indicating a strong concentration-dependent induction of apoptosis (Figure 7F).

These findings were further quantitatively confirmed by flow cytometry analysis. The results showed that treatment with 10% and 20% GBS-containing serum notably elevated the proportion of annexin V-positive cells in both GES-1 and Caco-2 cell lines ($P < 0.001$), compared to the control group (Figure 7G).

Collectively, these results demonstrate that GBS-containing serum induces significant apoptosis in gastrointestinal epithelial cells in a concentration-dependent manner.

Molecular Mechanisms Underlying GBS-Induced Gastrointestinal Toxicity

To elucidate the molecular mechanisms of GBS-induced gastrointestinal toxicity, we investigated the expression of key target proteins and pathways in GES-1 and Caco-2 cells treated with GBS-containing serum.

First, to validate MMP9 as a central target, qPCR, WB, and IF analyses were performed. All three methods consistently demonstrated that MMP9 expression was significantly upregulated at both the mRNA and protein levels in a concentration-dependent manner, with the most pronounced increase observed in the 20% serum group (Figure 8A-C).

Furthermore, CETSA was conducted to assess the direct interaction between GBS components and the MMP9 protein. The CETSA results showed that the thermal stability of MMP9 was significantly enhanced in cells treated with 20% GBS-containing serum compared to the DMSO control, particularly at higher temperatures (e.g., 65°C and 72°C). This thermal stabilization provides direct biophysical evidence that GBS constituents bind to and stabilize the MMP9 protein, confirming its role as a functional target (Figure 8D).

Given that KEGG enrichment analysis highlighted the IL-17 signaling pathway as a key player in GBS-induced toxicity, we further investigated the activation of its downstream inflammatory and apoptotic cascades. WB analysis revealed significant changes in cellular signaling pathways. The expression of Ki67, a marker of cell proliferation, was markedly downregulated, while the levels of cleaved caspase-3, a key executor of apoptosis, were significantly upregulated, accompanied by a decreased Bcl-2/Bax ratio, collectively

indicating the induction of apoptosis. Moreover, the phosphorylation of NF- κ B p65 (p-p65) was significantly increased, leading to a higher p-p65/p65 ratio, and the expression of the pro-inflammatory cytokine IL-6 was also upregulated—both of which are well-established downstream effectors of IL-17 signaling (Figure 8C).

Collectively, these results delineate the core molecular cascade underlying GBS-induced gastrointestinal injury, as illustrated in Figure 9.

Discussion

Ginkgo biloba L., the sole surviving species of the Ginkgoaceae family, is a widely used medicinal and edible plant. Its dried ripe seeds, GBS, were classified as a “dual-use” resource (for both food and medicine) by the Chinese Ministry of Health in 2002 (Sun et al., 2025). While traditionally valued for their putative health benefits, excessive or improper consumption of raw GBS is associated with a range of adverse effects. Clinical reports have documented that GBS poisoning, particularly in children and the elderly, can lead to severe symptoms such as nausea, vomiting, abdominal pain, diarrhea, convulsions, loss of consciousness, and even multi-organ damage to the gastrointestinal tract, heart, liver, and kidneys (Mei et al., 2017). The toxicity of GBS is primarily attributed to three classes of compounds: GAs, MPN, and MPNG. Numerous studies have demonstrated the toxic effects of these individual constituents at both cellular and animal levels (Al-Yahya et al., 2006). In this study, we first confirmed the presence of these three key toxic constituents in both raw GBS and the GBS-containing serum using UHPLC-Q Exactive HFX-MS, verifying their systemic bioavailability. While a large portion of existing research has centered on the isolated toxicity of individual compounds, TCM emphasizes the multi-component, multi-target nature of herbal medicines. To achieve a comprehensive understanding of the integrated toxicological mechanisms underlying GBS, we applied an integrated approach combining network toxicology, machine learning, and experimental validation to pinpoint critical targets and pathways, providing a systems-level perspective on GBS-induced toxicity. This strategy establishes a novel and systematic framework for elucidating the complex toxicological mechanisms of TCM, offering new insights for the safety assessment of herbal products.

The convergence of network pharmacology and systems biology approaches in this study has unveiled a complex molecular landscape underlying GBS-induced gastrointestinal toxicity. The identification of 439 overlapping targets provides a comprehensive map linking the major toxic constituents—GAs, MPN, and MPNG—to a broad network of biological processes associated with gastrointestinal injury. The construction of the PPI network further revealed a highly interconnected topology, with TNF, IL-6, AKT1 and MMP9 emerging as central hub genes. The centrality of these nodes is biologically meaningful: TNF and IL-6 are master regulators of inflammation (Holtmann et al., 2002),

AKT1 is a key node in cell survival and proliferation signaling(Li et al., 2025), and MMP9 is a potent effector of extracellular matrix degradation(Wang et al., 2023). The central roles of these hub genes suggest that GBS exposure may concurrently induce inflammatory activation, disrupt cell fate decisions, and compromise tissue barrier integrity, collectively driving gastrointestinal injury. Functional enrichment analysis corroborated these findings, showing significant enrichment in extracellular matrix remodeling, inflammatory response, and cellular stress pathways, as well as key signaling cascades such as IL-17 signaling and complement activation, all of which are hallmarks of gastrointestinal toxicity(Wang et al., 2025).

To prioritize the most discriminative genes from the network-derived targets and enhance the interpretability of our predictive model, we employed an integrated machine learning framework, combining LASSO regression, Random Forest, and SVM-RFE. This multi-algorithm approach identified a robust six-gene signature (MMP9, TREM1, PLA2, DUSP6, F3, LCN2) with consistently high predictive performance. To move beyond conventional feature importance and gain deeper biological insights, we applied SHAP analysis, a model-agnostic method that quantifies the contribution of each gene to the prediction for every individual sample(Hajjar et al., 2025). SHAP analysis confirmed MMP9 as the top contributor, with high expression levels strongly pushing the model output toward the “toxicity” class, while beeswarm and dependence plots revealed the directionality of gene effects and non-linear relationships—such as a threshold effect in MMP9 expression—beyond which its contribution to toxicity prediction sharply increased.

The central role of MMP9 in our computational models is strongly supported by its well-established and functionally validated involvement in human gastrointestinal inflammatory diseases. MMP9, a zinc-dependent endopeptidase, exerts a pivotal function in degrading extracellular matrix components—most notably type IV collagen, a key constituent of the basement membrane (Vandooren et al., 2013). Numerous clinical and experimental studies have demonstrated that MMP9 is significantly upregulated in the mucosal tissues of patients with IBD, including both Crohn's disease and ulcerative colitis, where its expression correlates with disease severity, endoscopic activity, and tissue destruction(de Bruyn et al., 2015). As a key member of the matrix metalloproteinase family, MMP9's ability to degrade the basement membrane directly mediates epithelial barrier breakdown—a hallmark of IBD pathogenesis. Evidence from DSS-induced colitis models shows that MMP9 deficiency significantly ameliorates inflammatory responses and tissue damage, confirming its causal role in disease progression(O'Sullivan et al., 2017). Furthermore, the MMP9-p53 axis has been implicated in the regulation of intestinal epithelial apoptosis, linking extracellular matrix degradation to programmed cell death in inflammatory settings(Wang et al., 2023). Similarly, in gastritis and *Helicobacter pylori* infection, elevated MMP9 expression is consistently observed: studies on human gastric tissues have shown that MMP9 is either undetectable or lowly expressed in normal mucosa, but

exhibits moderate to high expression in gastritis and gastric cancer samples, indicating its progressive upregulation with disease severity and association with mucosal injury, inflammation, and pathological remodeling (Sampieri et al., 2010). Concurrently, elevated MMP9 activity triggers mitochondrial release of cytochrome-c and elevates cleaved caspase-3, thereby potentiating gastric epithelial apoptosis in *H. pylori*-associated gastritis (Al-Sadi et al., 2021).

The pivotal role of MMP9 was further substantiated by functional analysis, which demonstrated that high MMP9 expression is associated with a profound reprogramming of the immune microenvironment. Concurrently, CIBERSORT analysis revealed a notable elevation in the infiltration of innate immune cells—particularly neutrophils and M0 macrophages—which are key mediators of acute inflammation and tissue damage. Conversely, adaptive immune components, such as T cells CD4 memory resting, were significantly reduced, suggesting a potential suppression of immune surveillance.

To validate the computational predictions and elucidate the underlying mechanisms of GBS-containing serum-induced gastrointestinal toxicity, we employed a drug-containing serum strategy to simulate the *in vivo* exposure of gastrointestinal cells to orally administered GBS. Functional assays demonstrated that GBS-containing serum exerted significant cytotoxic effects on gastrointestinal epithelial cells. These effects directly point to the early damage and dysfunction of the gastrointestinal mucosal barrier. First, CCK-8 and EdU assays demonstrated that GBS-containing serum remarkably suppressed cell proliferation in a time- and concentration-dependent manner. The continuous proliferation of epithelial cells is fundamental for maintaining mucosal integrity and repairing daily damage; thus, its inhibition indicates impaired mucosal self-renewal capacity. Second, the LDH release assay showed that cell membrane integrity was compromised, directly leading to the loss of the epithelial layer's physical barrier function. More importantly, we observed that GBS-containing serum induced significant oxidative stress (elevated intracellular ROS levels and depletion of the GSH antioxidant system) and apoptosis (confirmed by Hoechst-PI staining and flow cytometry). These processes are core pathological events in mucosal injury: oxidative stress can directly damage epithelial cells and activate pro-inflammatory signaling, while excessive apoptosis creates “gaps” in the epithelial layer, disrupting its continuity. These cellular-level pathological changes constitute the microscopic foundation of the “mucosal barrier disruption,” a macroscopic phenomenon observed clinically.

To investigate the mechanistic link between GBS-containing serum and the central hub gene MMP9, we performed a series of molecular biology experiments. qPCR, WB, and IF analyses consistently demonstrated that GBS-containing serum significantly upregulated MMP9 expression at both the mRNA and protein levels in a concentration-dependent manner. Furthermore, CETSA provided direct biophysical evidence of target engagement,

showing significant thermal stabilization of the MMP9 protein upon treatment with GBS-containing serum. This finding, combined with molecular docking and MD simulations that revealed stable binding conformations among key GBS constituents (GAs, MPN, MPNG) and the MMP9 protein, solidifies MMP9 as a direct functional target of GBS-induced toxicity.

This experimental confirmation of MMP9 activation provides a crucial foundation for understanding its downstream effects. Notably, a well-established positive feedback loop exists between NF- κ B and MMP9 in inflammatory contexts: IL-17-activated NF- κ B p65 binds to the κ B motif in the MMP9 promoter to drive its transcription, while secreted MMP9, in turn, proteolytically activates pro-inflammatory cytokines like IL-1 β and TNF- α and facilitates IL-17 release, thereby re-activating the IKK/NF- κ B pathway (Chen et al., 2018).

Critically, our experimental data confirm the operation of this axis. WB analysis of GBS-containing-serum-treated cells showed sharply elevated levels of phosphorylated NF- κ B p65 and its downstream cytokine IL-6, coincident with decreased Ki67, increased cleaved caspase-3, and a reduced Bcl-2/Bax ratio—providing a coherent molecular explanation for the observed suppression of proliferation and enhancement of apoptosis. These findings confirm the operation of the “inflammation-NF- κ B-MMP9” cascade, demonstrating how persistent MMP9 activity is translated into mucosal barrier failure through proliferation, oxidative stress, and apoptotic signaling.

In summary, our integrated data highlight MMP9 as the pivotal mediator of GBS-induced gastrointestinal toxicity. The convergence of direct target engagement by GBS constituents and upstream pro-inflammatory signaling leads to a significant and sustained upregulation of MMP9 expression and activity. This dual regulation results in a sustained surge of MMP9 activity, which orchestrates a cascade of events—including the suppression of proliferation, induction of oxidative stress, and activation of apoptosis—ultimately leading to mucosal barrier disruption.

Despite the comprehensive nature of our study, several limitations should be acknowledged. First, our experimental validation was conducted primarily *in vitro* using a drug-containing serum model. While this approach effectively simulates the systemic exposure of gastrointestinal cells to orally administered GBS-containing serum, it cannot fully recapitulate the complex *in vivo* pathophysiology, including the dynamic interactions between the gut, liver, and immune system. The absence of an animal model of GBS-containing serum-induced gastrointestinal toxicity limits our ability to validate the role of key targets like MMP9 in an animal model. Second, although our bioinformatics analyses strongly suggest that high MMP9 expression is associated with a pro-inflammatory immune microenvironment, these findings are computational predictions based on publicly

available GEO datasets, which lack direct experimental validation of immune cell infiltration or cytokine profiles in the tissue microenvironment. This represents a critical gap between our computational insights and functional confirmation. Future studies should therefore establish a robust *in vivo* model and integrate spatial transcriptomics or multiplex immunofluorescence to directly characterize the immune landscape, which will be essential for translating our findings into a deeper understanding of the clinical implications of GBS consumption.

Conclusion

In conclusion, our integrated framework identifies MMP9 as a central mediator of GBS-induced gastrointestinal toxicity. GBS exposure activates the IL-17/NF- κ B signaling axis, leading to significant upregulation of MMP9 expression. This upregulation initiates a self-amplifying loop that integrates suppressed proliferation, oxidative stress, and apoptosis, ultimately disrupting the gastrointestinal epithelial barrier. This study provides a systems-level understanding of GBS toxicity and highlights MMP9 as a potential biomarker and therapeutic target.

Declarations

Author contributions

Jinrong He, Saili Chen, Pei Li: Conceptualization, Supervision, Project administration, Writing-review & editing. **Yu Qi, Jinrong He, Chengjie Xu, Xiaohong Guo:** Investigation, Methodology, Writing the original draft. **Jinrong He:** Funding acquisition. **Shang Huang, Yong Li, Chengkai Tan:** Visualization, Writing the original draft. **Mingfeng Xia, Yaohui Song:** Data curation, Methodology. The final manuscript has been approved by all authors. All authors agree to be accountable for all aspects of work ensuring integrity and accuracy.

Acknowledgements

This research received funding from the Natural Science Foundation of Hubei Province (No.2025AFB647), the Scientific Research Projects from Wuhan Municipal Health Commission (No. WX23Z16), the Hubei University of Science and Technology PhD Start-up Fund Project Support (No. BK202422), and the Exploring the Mechanism by which Xiaoyao Sanjie Decoction Induces Ferroptosis in Triple-Negative Breast Cancer via the AKT1/NRF2/GPX4 Signaling Pathway (D20242803).

Declaration of competing interest

The authors declare no conflict of interests.

Supplementary materials

Supplementary material associated with this article can be found in the online version.

References

1. Al-Sadi, R., Engers, J., Haque, M., King, S., Al-Omari, D., Ma, T.Y., 2021. Matrix Metalloproteinase-9 (MMP-9) induced disruption of intestinal epithelial tight junction barrier is mediated by NF- κ B activation. *PLoS One*, 16, e0249544.
2. Al-Yahya, A.A., Al-Majed, A.A., Al-Bekairi, A.M., Al-Shabanah, O.A., Qureshi, S., 2006. Studies on the reproductive, cytological and biochemical toxicity of Ginkgo Biloba in Swiss albino mice. *J Ethnopharmacol*, 107, 222-8.
3. Bai, C., Wu, L., Li, R., Cao, Y., He, S., Bo, X., 2025. Machine Learning-Enabled Drug-Induced Toxicity Prediction. *Adv Sci (Weinh)*, 12, e2413405.
4. Boateng, I.D., 2022. A critical review of current technologies used to reduce ginkgotoxin, ginkgotoxin-5'-glucoside, ginkgolic acid, allergic glycoprotein, and cyanide in Ginkgo biloba L. seed. *Food Chem*, 382, 132408.
5. Chen, X., Chang, L., Li, X., Huang, J., Yang, L., Lai, X., Huang, Z., Wang, Z., Wu, X., Zhao, J., Bellanti, J.A., Zheng, S.G., Zhang, G., 2018. Tc17/IL-17A Up-Regulated the Expression of MMP-9 via NF- κ B Pathway in Nasal Epithelial Cells of Patients With Chronic Rhinosinusitis. *Front Immunol*, 9, 2121.
6. de Bruyn, M., Arijs, I., De Hertogh, G., Ferrante, M., Van Assche, G., Rutgeerts, P., Vermeire, S., Opdenakker, G., 2015. Serum Neutrophil Gelatinase B-associated Lipocalin and Matrix Metalloproteinase-9 Complex as a Surrogate Marker for Mucosal Healing in Patients with Crohn's Disease. *J Crohns Colitis*, 9, 1079-87.
7. Fang, J., Wang, Z., Wang, P., Wang, M., 2020. Extraction, structure and bioactivities of the polysaccharides from Ginkgo biloba: A review. *Int J Biol Macromol*, 162, 1897-1905.
8. Gong, H., Wu, C.E., Fan, G.J., Li, T.T., Wang, J.H., Wang, T., 2018. Determination and Comparison of 4'-O-Methylpyridoxine Analogues in Ginkgo biloba Seeds at Different Growth Stages. *J Agric Food Chem*, 66, 7916-7922.
9. Hajjar, M., Aldabbagh, G., Albaradei, S., 2025. Interpretable Multi-Cancer Early Detection Using SHAP-Based Machine Learning on Tumor-Educated Platelet RNA. *Diagnostics (Basel)*, 15.
10. Hassan, I., Wan Ibrahim, W.N., Yusuf, F.M., Ahmad, S.A., Ahmad, S., 2020. Biochemical Constituent of Ginkgo biloba (Seed) 80% Methanol Extract Inhibits Cholinesterase Enzymes in Javanese Medaka (*Oryzias javanicus*) Model. *J Toxicol*, 2020, 8815313.

11. He, B., Qian, K., Han, X., Li, J., Zhou, Q., Xu, L.A., Liu, H., Cui, P., 2023. Novel mechanisms for the synthesis of important secondary metabolites in *Ginkgo biloba* seed revealed by multi-omics data. *Front Plant Sci*, 14, 1196609.
12. Holtmann, M.H., Schütz, M., Galle, P.R., Neurath, M.F., 2002. Functional relevance of soluble TNF- α , transmembrane TNF- α and TNF-signal transduction in gastrointestinal diseases with special reference to inflammatory bowel diseases. *Z Gastroenterol*, 40, 587-600.
13. Hori, Y., Fujisawa, M., Shimada, K., Oda, A., Katsuyama, S., Wada, K., 2004. Rapid analysis of 4-O-methylpyridoxine in the serum of patients with *Ginkgo biloba* seed poisoning by ion-pair high-performance liquid chromatography. *Biol Pharm Bull*, 27, 486-491.
14. Kajiyama, Y., Fujii, K., Takeuchi, H., Manabe, Y., 2002. *Ginkgo* seed poisoning. *Pediatrics*, 109, 325-7.
15. Li, X., Wang, J., Zhou, J., Xiao, H., Liu, L., Zhang, Z., Si, J., Yang, C., Wang, M., Ye, J., Sun, G., 2025. *Myristica fragrans* water extract modulates multiple biological processes to pre-protect anhydrous ethanol-induced gastric ulcers via Akt/JNK/Nrf2 pathway activation. *J Ethnopharmacol*, 340, 119258.
16. Liu, Y., Xin, H., Zhang, Y., Che, F., Shen, N., Cui, Y., 2022. Leaves, seeds and exocarp of *Ginkgo biloba* L. (*Ginkgoaceae*): A Comprehensive Review of Traditional Uses, phytochemistry, pharmacology, resource utilization and toxicity. *J Ethnopharmacol*, 298, 115645.
17. Mei, N., Guo, X., Ren, Z., Kobayashi, D., Wada, K., Guo, L., 2017. Review of *Ginkgo biloba*-induced toxicity, from experimental studies to human case reports. *Journal of environmental science and health. J Environ Sci Health C Environ Carcinog Ecotoxicol Rev*, 35, 1-28.
18. O'Sullivan, S., Wang, J., Pigott, M.T., Docherty, N., Boyle, N., Lis, S.K., Gilmer, J.F., Medina, C., 2017. Inhibition of matrix metalloproteinase-9 by a barbiturate-nitrate hybrid ameliorates dextran sulphate sodium-induced colitis: effect on inflammation-related genes. *Br J Pharmacol*, 174, 512-524.
19. Qi, Y., Li, P., Li, B., Lv, P., Hu, H., Tan, C., Wang, W., Jiang, Y., He, J., 2025. Integrating of serum pharmacochimistry, metabolomics, and experimental validation to explore the mechanism of Erzhi Pill against Triple-negative breast cancer. *J Ethnopharmacol*, 353, 120287.
20. Sampieri, C.L., de la Peña, S., Ochoa-Lara, M., Zenteno-Cuevas, R., León-Córdoba, K., 2010. Expression of matrix metalloproteinases 2 and 9 in human gastric cancer and superficial gastritis. *World J Gastroenterol*, 16, 1500-5.
21. Sasaki, K., Hatta, S., Wada, K., Ohshika, H., Haga, M., 2000. Bilobalide prevents reduction of gamma-aminobutyric acid levels and glutamic acid decarboxylase activity

- induced by 4-O-methylpyridoxine in mouse hippocampus. *Life Sci*, 67, 709-15.
22. Shao, Y., Chen, Y., Zhu, Q., Yi, L., Ma, Y., Zang, X., Yao, W., 2025. The Pharmacology and Toxicology of Ginkgolic Acids: Secondary Metabolites from *Ginkgo biloba*. *Am J Chin Med*, 53, 147-177.
 23. Sun, X., Zheng, Y., Sun, K., He, C., Xiao, P., 2025. Review, revision, and prospect of list of substances with both edible and medicinal values in China. *China J. Chin. Mater. Med.*, 50, 346-355.
 24. Tu, Y., Tan, L., Tao, H., Li, Y., Liu, H., 2023. CETSA and thermal proteome profiling strategies for target identification and drug discovery of natural products. *Phytomedicine*, 116, 154862.
 25. Vandooren, J., Van den Steen, P.E., Opdenakker, G., 2013. Biochemistry and molecular biology of gelatinase B or matrix metalloproteinase-9 (MMP-9): the next decade. *Crit Rev Biochem Mol Biol*, 48, 222-72.
 26. Wang, J., Niu, G., Mai, H., Zhang, X., Zhu, J., Li, B., Gao, Y., Huang, T., Meng, Q., Chen, R., 2025. The protective role of 3-Indoleglyoxylic acid in Bisphenol S-induced intestinal barrier dysfunction via mitochondrial ROS-Mediated IL-17/CXCL10/TNF- α signaling. *Environment international* 199, 109477.
 27. Wang, X., Chen, S., Wang, J., Chen, Y., Guo, Y., Wang, Q., Liu, Z., Zeng, H., Xu, C., 2023. Olfactomedin-4 deletion exacerbates DSS-induced colitis through a matrix metalloproteinase-9-dependent mechanism. *Environ Int*, 199, 109477.
 28. Xu, B., Yu, Y., Zhang, J., Jiang, B., Yan, L., Chen, S., Hu, L., Miao, Q., Qi, Y., 2025. Investigating the Mechanism of Jiawei Weijin Decoction in Treating Non-Small Cell Lung Cancer Using Network Pharmacology, Bioinformatics Analysis and Experimental Validation. *Drug Des Devel Ther*, 19, 8711-8743.
 29. Yu, Y., 2017. Toxic and Active Compositions and The Intervention Effecton Alzheimer's disease of Ginkgo seeds. Master's thesis. Jiangsu University.
 30. Zhang, W., Shi, M., Zhang, F., Cao, F., Su, E., 2021. A Facile Method to Determine the Native Contents of 4'-O-Methylpyridoxine and 4'-O-Methylpyridoxine-5'-glucoside in *Ginkgo biloba* Seeds. *J Agric Food Chem*, 69, 14270-14277.
 31. Zou, P., Yu, Y., Zheng, N., Yang, Y., Paholak, H.J., Yu, L.X., Sun, D., 2012. Applications of human pharmacokinetic prediction in first-in-human dose estimation. *Aaps j*, 14, 262-81.

Table

Table 1. Information on antibodies involved.

Antibody	Application	Dilution	Company	Catalog Number
MMP9	IF	1:50	Abclonal	A25299
Goat Anti-Rabbit IgG (Cy3)	IF	1:500	Abways	AB0133
Ki67	WB	1:1000	Abclonal	A20018
MMP9	WB	1:1000	Abways	CY1226
Caspase-3	WB	1:1500	Abclonal	A19654
Bcl-2	WB	1:1000	Abclonal	A19693
Bax	WB	1:1000	Abclonal	A19684
NF-κB p65	WB	1:1000	Abclonal	A19653
Phospho-NF-κB p65	WB	1:1000	Abclonal	AP1294
IL6	WB	1:1000	Abclonal	A22222
β-Actin	WB	1:1000	Abclonal	AC026
Rabbit IgG (H + L) HRP	WB	1:10000	Abways	AB0101

Figures

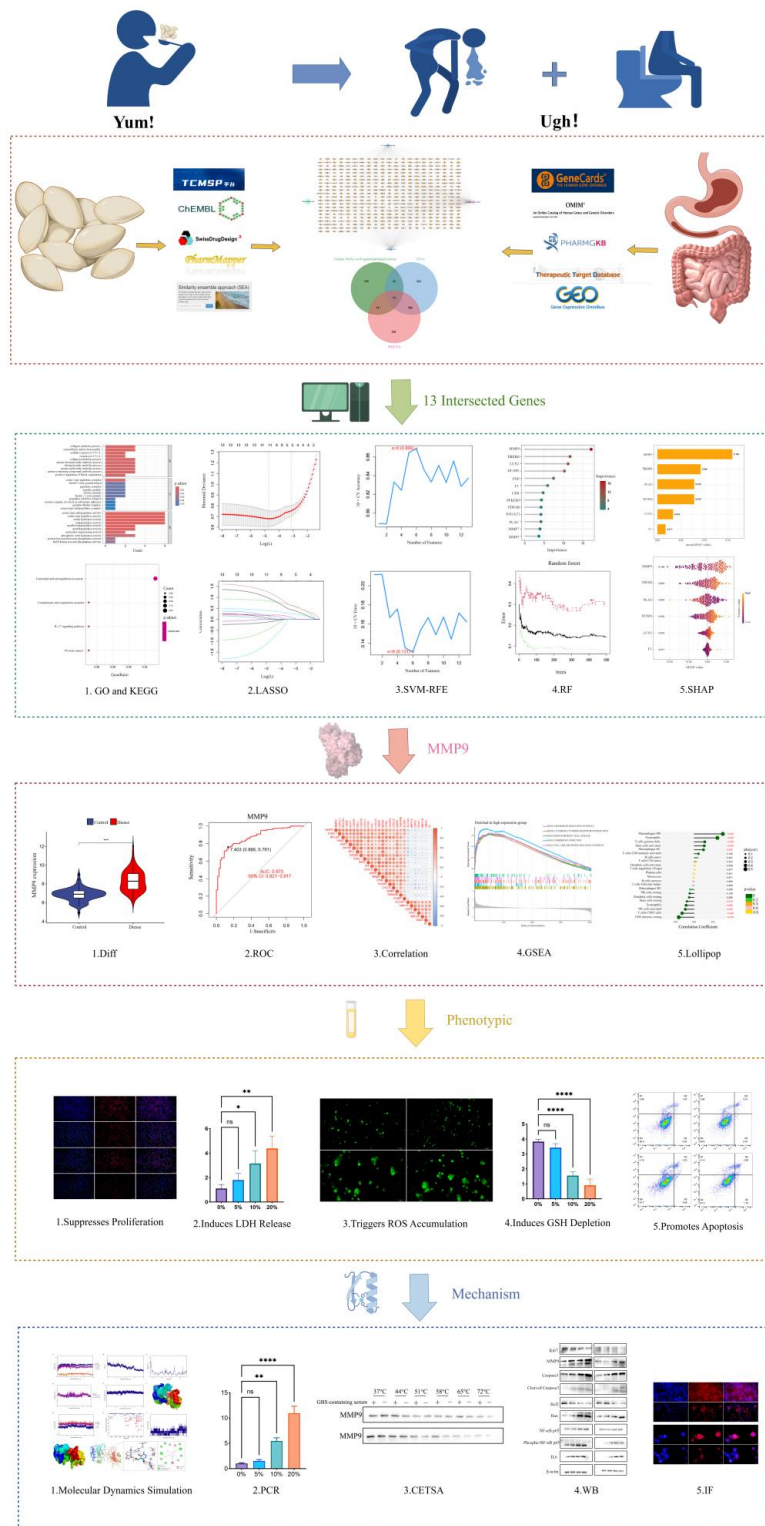


Figure 1

Flow chart of the present study.

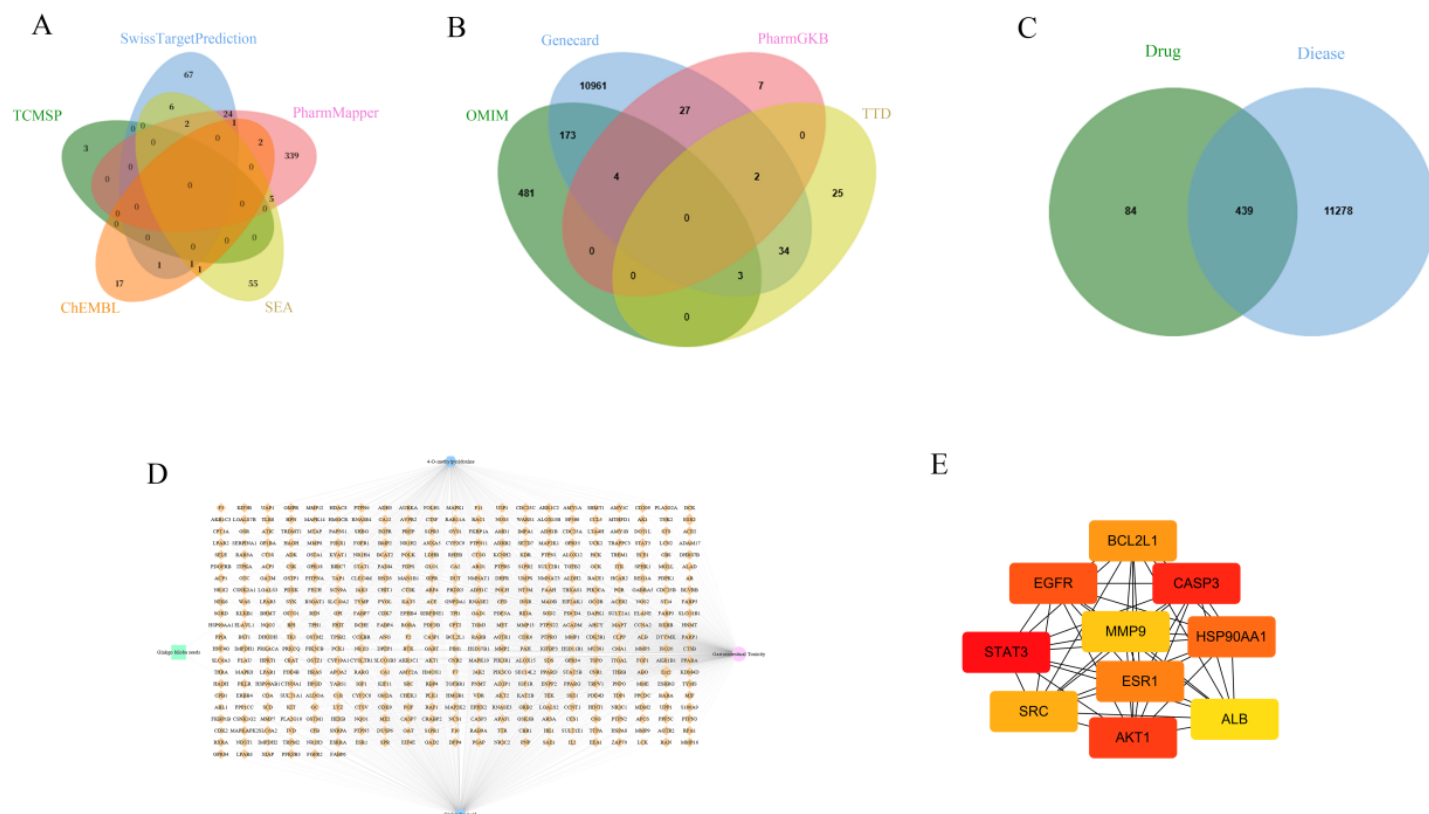


Figure 2

Network Toxicology Analysis of GBS-Induced Gastrointestinal Toxicity. (A) Venn diagrams illustrate the overlap of predicted targets from TCMSP, ChEMBL, SwissTargetPrediction, PharmMapper, and SEA, yielding 524 unique drug targets. (B) Venn diagrams show the overlap of disease-associated targets retrieved from OMIM, GeneCards, PharmGKB, and TTD, identifying 11,717 targets related to gastrointestinal toxicity. (C) The intersection of the drug-target set and disease-target set yielded 439 common targets, potentially mediating GBS-induced gastrointestinal toxicity. (D) A component-target-disease network was constructed using Cytoscape 3.8.2, visualizing interactions among toxic constituents, shared targets, and gastrointestinal toxicity.

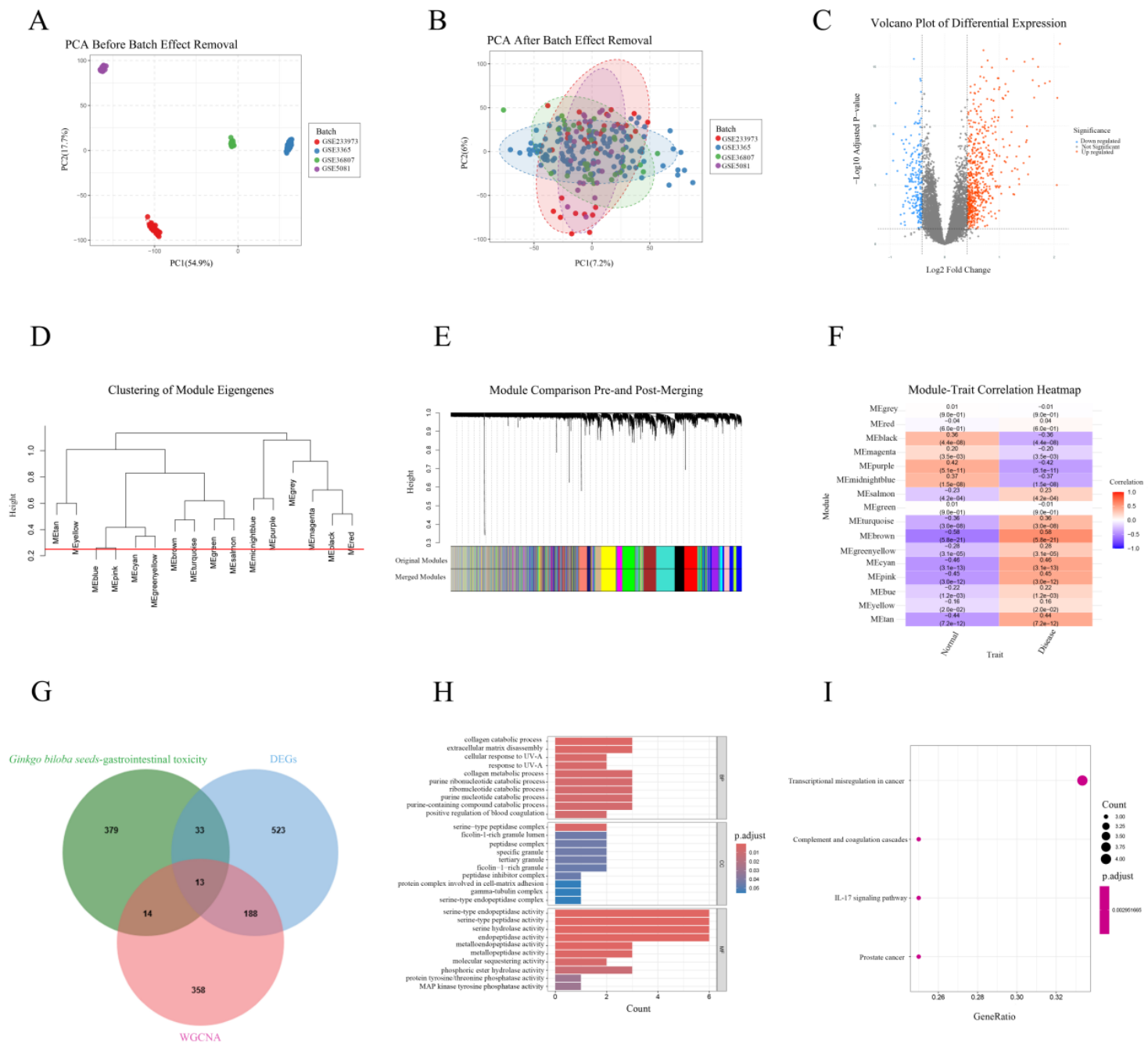


Figure 3

Identification and Functional Analysis of Gastrointestinal Toxicity-Related Genes. (A-B) PCA plots demonstrate effective data normalization and batch effect removal, with samples clustering by disease status before and after correction. (C) A volcano plot visualizes the distribution of 758 DEGs identified by comparing gastrointestinal toxicity samples with healthy controls. (D-F) WGCNA identifies gene modules associated with gastrointestinal toxicity, with the MEbrown module showing the highest positive correlation. (G) A Venn diagram illustrates the intersection of three gene sets: 439 overlapping targets from network toxicology, 758 DEGs from GEO datasets, and 573 genes in the MEbrown module from WGCNA, yielding 13 key genes. (H) GO enrichment analysis

reveals top enriched terms in BP, CC, and MF categories. (I) KEGG pathway enrichment analysis showing significant enrichment in the IL-17 signaling pathway and other pathways.

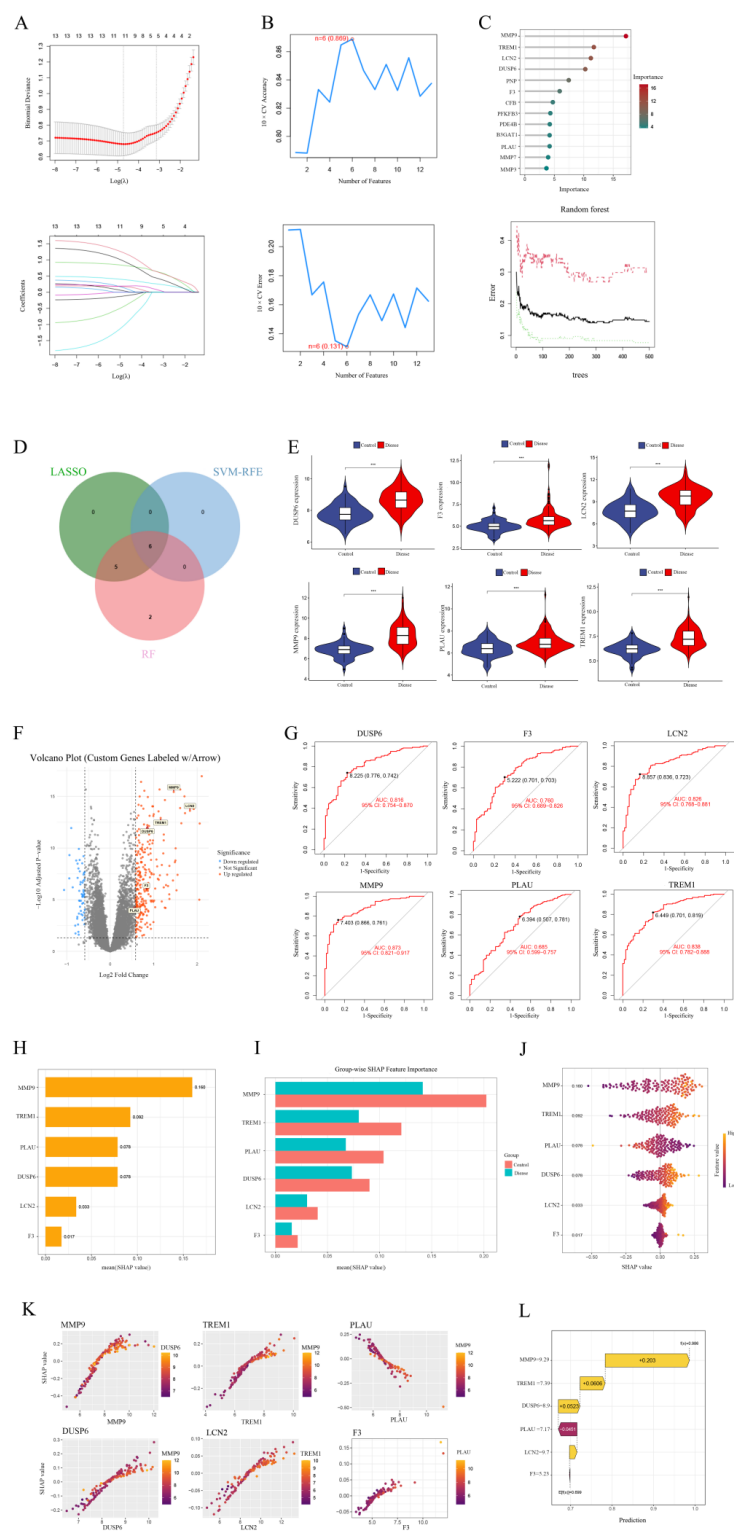


Figure 4

Machine Learning-Based Identification and Validation of Key Genes for Gastrointestinal Toxicity. (A) LASSO regression coefficient profiles and cross-validation results identify the

optimal λ value, yielding a 9-gene signature. (B) SVM-RFE selects a 6-gene signature with the lowest cross-validation error (13.1%) and highest accuracy (86.9%), outperforming the baseline. (C) Random Forest analysis reveals model convergence and ranks MMP9, TREM1, and DUSP6 as top contributors based on Mean Decrease Gini index. (D) A Venn diagram shows that six genes are common to all three machine learning methods. (E) Violin plots demonstrate significant upregulation of all six key genes in gastrointestinal toxicity samples compared to controls. (F) A volcano plot visualizes the $\log_2FC$ and statistical significance of the six key genes. (G) ROC curves show high diagnostic performance for all six genes, with MMP9 exhibiting the highest AUC (0.986). (H) A bar plot ranks genes by mean absolute SHAP values, confirming MMP9 as the most important feature (0.160). (I) A beeswarm plot reveals that MMP9 is the most important gene. (J) Dependence plots show coordinated upregulation of TREM1, DUSP6, LCN2, and F3 with MMP9, while PLAU exhibits an inverse relationship. (K) Force-directed analysis illustrates how MMP9 is the dominant positive driver ($\Delta = +0.203$) in a high-risk prediction, with TREM1 and DUSP6 reinforcing the signal, and PLAU exerting a slight negative effect ($\Delta = -0.0451$). * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$ vs. the control group.

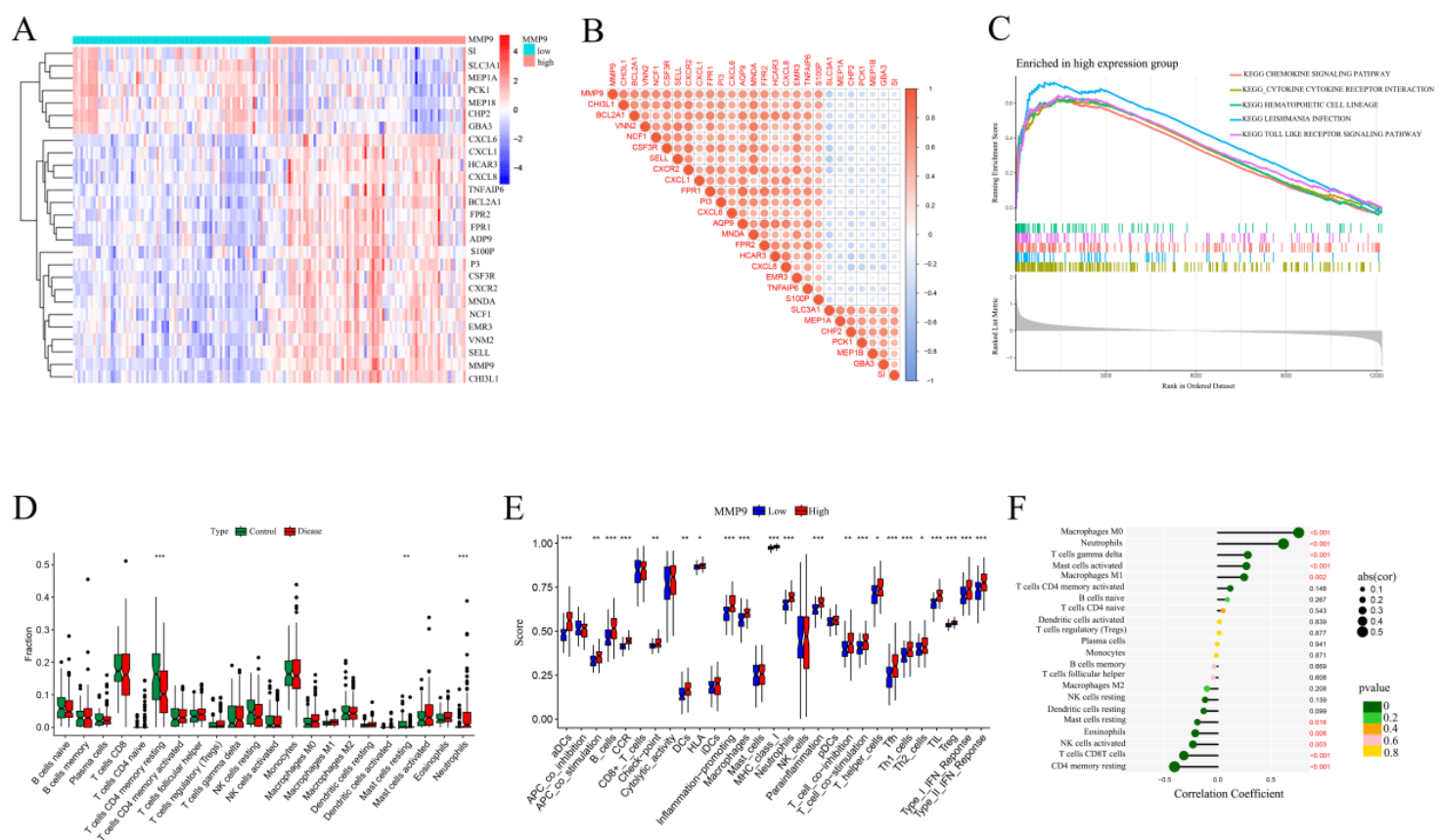


Figure 5

Functional Role of MMP9 in GBS-Induced Gastrointestinal Toxicity. (A) A clustering heatmap reveals that high MMP9 expression is associated with upregulation of pro-inflammatory genes and downregulation of mucosal protective genes. (B) A correlation

matrix shows strong positive correlations between MMP9 and key mediators of inflammation and immune cell recruitment. (C) GSEA identifies significant enrichment of core immune-inflammatory pathways in the high-MMP9 group, reflecting a pro-inflammatory transcriptional program. (D) CIBERSORT analysis reveals a shift from immune quiescence in controls to acute inflammation in toxicity samples. (E) Boxplots show that MMP9 expression is positively correlated with multiple pro-inflammatory and immune-activating modules. (F) A lollipop plot indicates that MMP9 is strongly positively correlated with M0 macrophages and neutrophils, but negatively correlated with CD4+ memory resting T cells and CD8+ T cells. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$ vs. the control group.

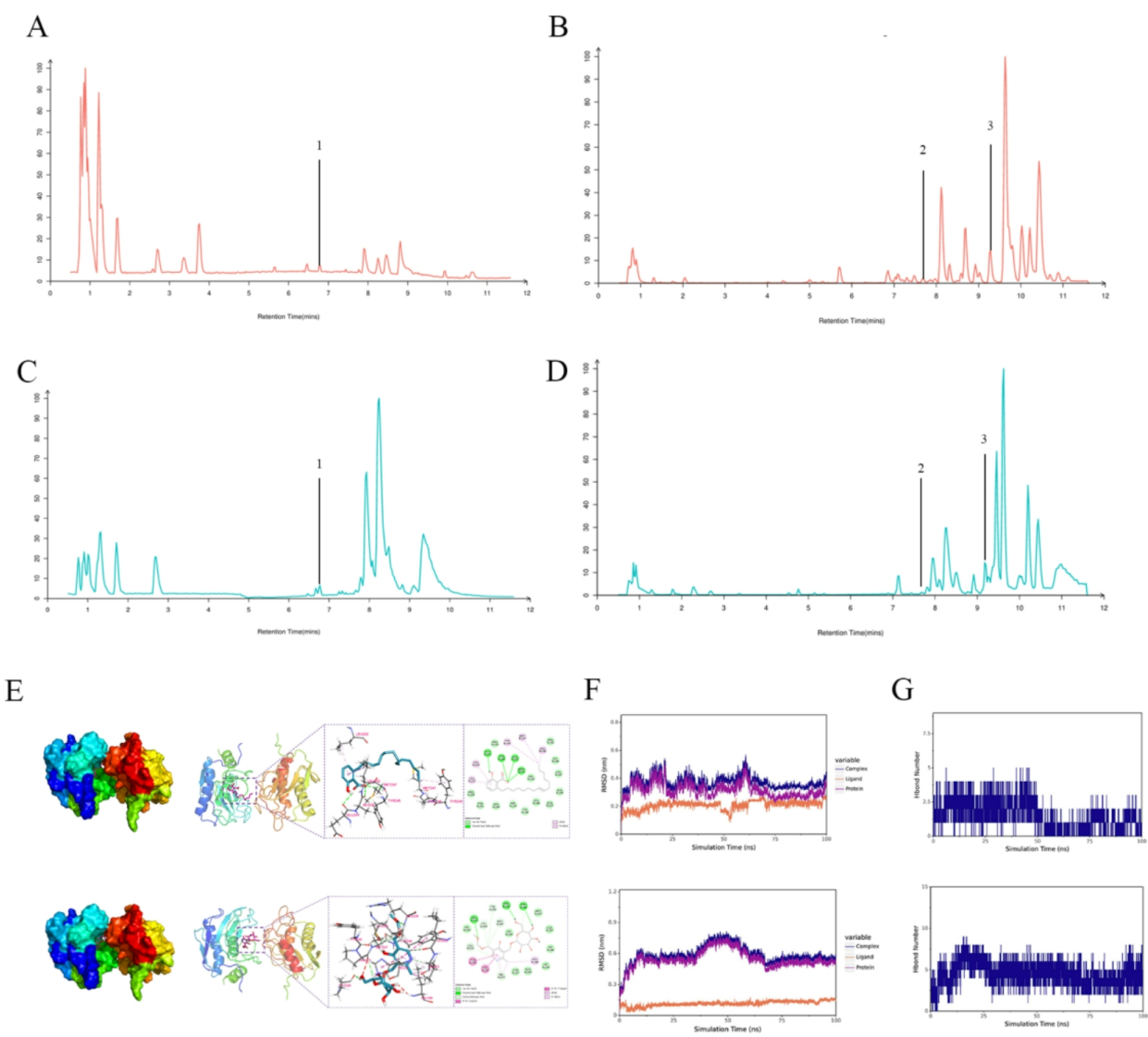


Figure 6

Metabolite Identification and Molecular Dynamics Simulations of GBS Toxicants. (A-B) UHPLC-Q Exactive HFX-MS chromatograms (positive and negative ion modes) of raw GBS water extract, showing characteristic peaks for GAs, MPN, and MPNG. (C-D) UHPLC-Q Exactive HFX-MS chromatograms (positive and negative ion modes) of GBS-containing serum, confirming the presence of GAs, MPN, and MPNG in the bloodstream, demonstrating their systemic bioavailability. (E) Molecular docking models illustrate the binding poses of GAs, MPN within the active site of MMP9, with key interactions such as hydrogen bonds and hydrophobic contacts highlighted. (F) RMSD plots show stable conformational dynamics for two complexes over a 100-nanosecond simulation, with GAs-MMP9 exhibiting an RMSD plateau around 2.0 Å, MPN-MMP9 around 1.8 Å. (G) Hydrogen bond analysis reveals that the GAs-MMP9 complex forms the highest average number of hydrogen bonds, indicating strong and persistent interactions.

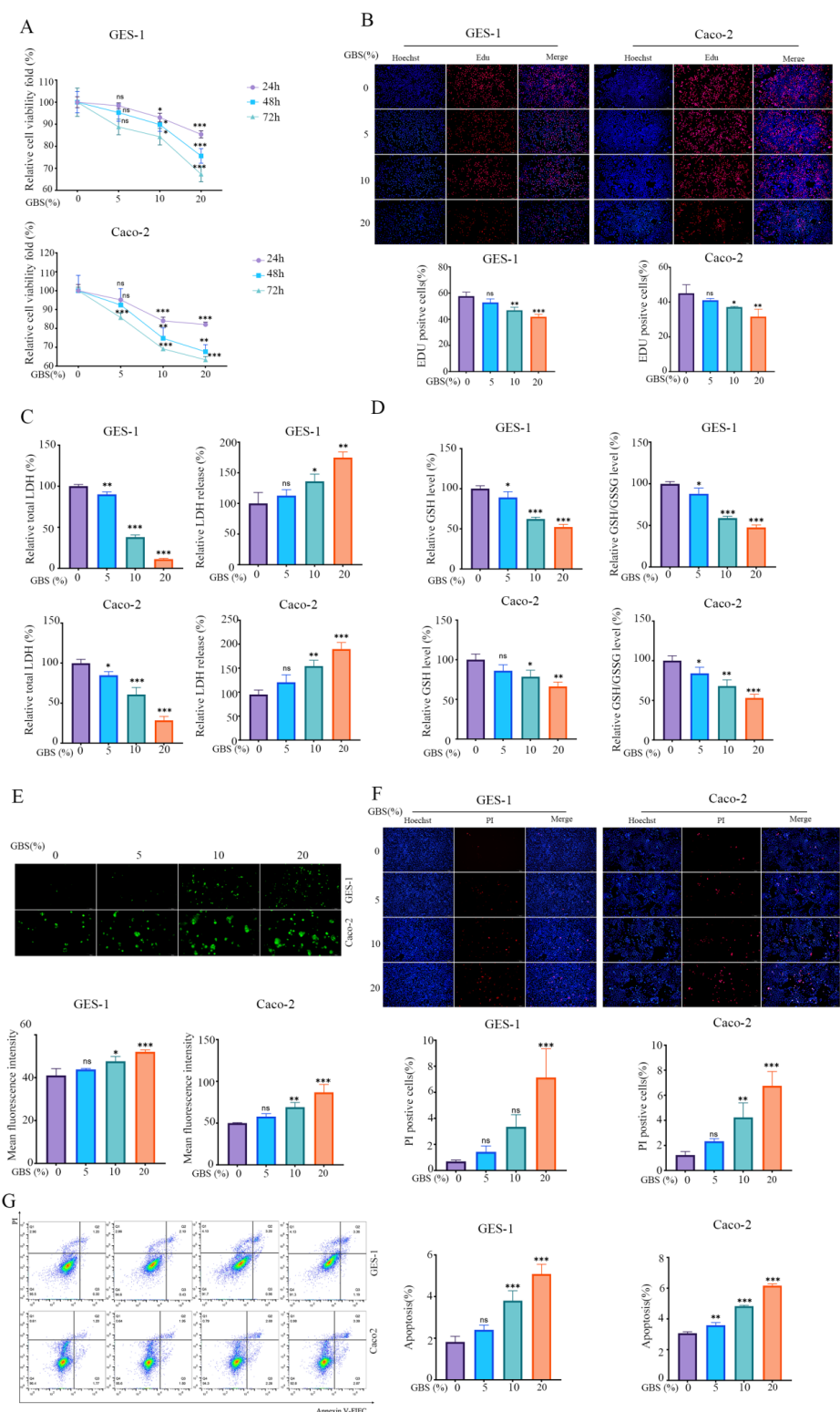


Figure 7

GBS-Containing Serum Induces Cytotoxicity, Oxidative Stress, and Apoptosis in GES-1 and Caco-2 Cells. (A) CCK-8 assay shows a significant concentration- and time-dependent reduction in cell viability treated with GBS-containing serum. (B) EdU incorporation assay reveals that GBS-containing serum significantly suppresses DNA synthesis, Scale bar: 200 μ M. (C) LDH assay demonstrates that GBS-containing serum reduces total intracellular

LDH activity and increases extracellular LDH release in a concentration-dependent manner. (D) GSH levels and the GSH/GSSG ratio are significantly decreased in both cell lines after treatment with 10% and 20% GBS-containing serum. (E) DCFH-DA staining shows a significant increase in intracellular ROS levels with increasing concentrations of GBS-containing serum, Scale bar: 200 μ M. (F) Hoechst/PI dual staining reveals characteristic apoptotic morphology, including chromatin condensation and nuclear fragmentation, Scale bar: 200 μ M. (G) Annexin V-FITC/PI flow cytometry quantifies a significant increase in the proportion of annexin V-positive cells in both cell lines treated with GBS-containing serum. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$ vs. the control group. All data are shown as the mean $\pm$ SD from three independent experiments.

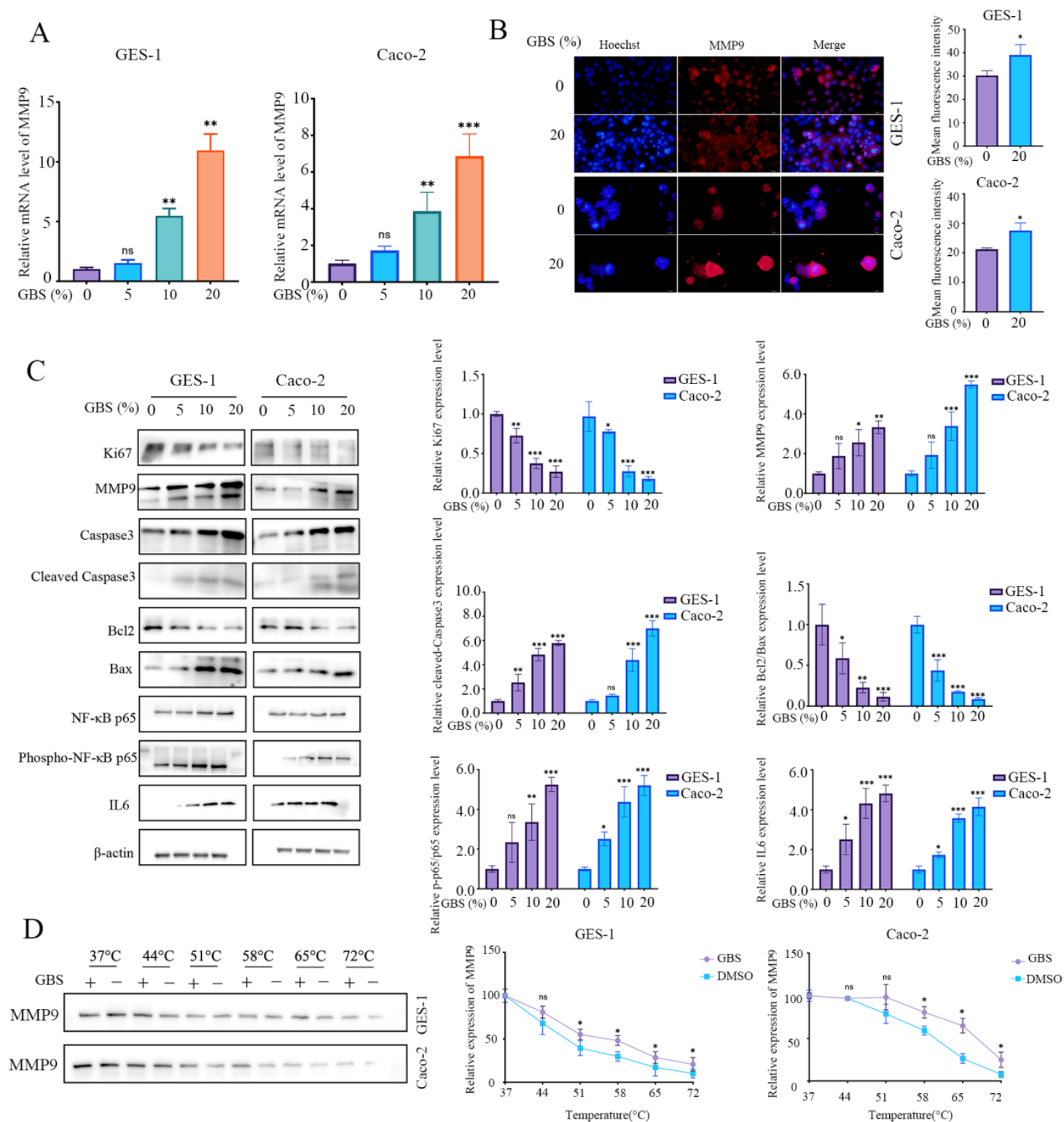


Figure 8

Molecular Mechanisms of GBS-Induced Gastrointestinal Toxicity. (A) qPCR analysis shows a significant concentration-dependent upregulation of MMP9 mRNA in GES-1 and Caco-2 cells treated with GBS-containing serum. (B) Immunofluorescence staining reveals strong MMP9 protein expression and subcellular localization in GES-1 and Caco-2 cells treated with 20% GBS-containing serum, Scale bar: 50 μ M. (C) WB analysis demonstrates that

GBS-containing serum significantly downregulates Ki67, upregulates cleaved caspase-3 and p-p65, decreases the Bcl-2/Bax ratio, and increases IL-6 expression in a concentration-dependent manner. (D) CETSA shows enhanced thermal stability of MMP9 in GES-1 and Caco-2 cells treated with 20% GBS-containing serum compared to DMSO controls. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$ vs. the control group. All data are shown as the mean $\pm$ SD from three independent experiments.

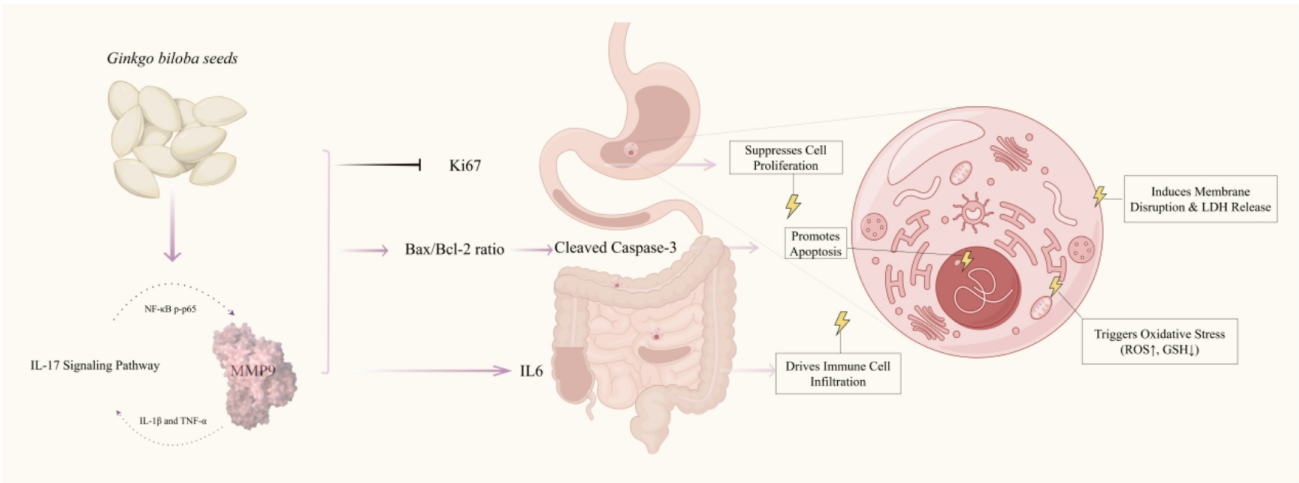


Figure 9

Proposed mechanism underlying the gastrointestinal toxicity of GBS. Illustration created with FigDraw (<https://www.figdraw.com/>).

Supplementary Files

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

- [WB.zip](#)
- [Supplement.tif](#)