Cirsium japonicum mitigates allergic nasal inflammation by regulating NRF2-mediated mucin production: Systematic transcriptome and molecular docking analyses

Bo-Jeong Pyun [a, 1], Su-Jin Baek [b, 1], Kyuhyung Jo [a], Ik Soo Lee [a], Musun Park [b], Hye Jin Kim [a], Joo Young Lee [a], Susanna Choi [a], Yun Hee Kim [a, **] and Taesoo Kim [a, *]

[a] KM Convergence Research Division, Korea Institute of Oriental Medicine, 1672 Yuseong-daero Yuseong-gu, Daejeon 34054, Republic of Korea

[b] KM Data Division, Korea Institute of Oriental Medicine, 1672 Yuseong-daero, Yuseong-gu, Daejeon 34054, Republic of Korea

[1] These authors contributed equally to this work.

* Corresponding author.

** Corresponding author.

E-mail addresses: xotn91@kiom.re.kr (T.K.), ddyunee@kiom.re.kr (Y.K.).

1. Supplementary Figures (S1–S11)

Fig. S1. Effects of CSJ on allergy nasal symptoms and NALF in the OVA-induced AR mouse model.

Fig. S2. Gene expression profiling of CSJ in the OVA-induced AR mouse model.

Fig. S3. Heatmap of the gene set enrichment results for CSJ.

Fig. S4. GSEA plot of gene expression signatures.

Fig. S5. Integrated gene network showing the expression changes in enriched pathways and genes after CSJ treatment in the OVA-induced AR mouse model.

Fig. S6. Identification of key genes and major transcription factors (TFs) related to nasal inflammation.

Fig. S7. Expression profiling of NRF2-associated pathways and NRF2 downstream genes included in MSigDB.

Fig. S8. Effect of CSJ on oxidative stress damage and the NRF2/KEAP1/HO-1 signaling pathway in the OVA-induced AR mouse model.

Fig. S9. HPLC analysis of CSJ.

Fig. S10. Three-dimensional molecular docking structure of KEAP1 with five CSJ compounds and ginnalin A showing the interactions between compounds and the KEAP1 Kelch domain pocket.

Fig. S11. Three-dimensional molecular docking structure of MUC5AC and five CSJ compounds showing the interactions between compounds and MUC5AC.

2. Supplementary Materials and Methods

3. Supplementary Tables (S1–S10)

* Tables S1–S7. Data are provided as Excel files (Supplementary Tables: S1–S7_JPA).

Table S1. GSEA results for each condition (Excel).

Table S2. Integrated network genes for each condition (Excel).

Table S3. List of genes in the “O-linked glycosylation of mucins” pathway (Excel).

Table S4. List of genes ranked by gene expression changes (Excel).

Table S5. List of genes significantly correlated with Muc5ac (Excel).

Table S6. Enriched TFs for Muc5ac-associated genes (Excel).

Table S7. Results of functional enrichment analysis of Muc5ac-associated genes (Excel).

Table S8. Calibration curves, linearity, LOD, and LOQ for the five reference standards (n = 3).

Table S9. Precision (intra- and inter-day) and recovery of the five reference standards (n = 6).

Table S10. Contents of the five marker compounds in the *Cirsium japonicum* herb extract.

1. Supplementary Figures

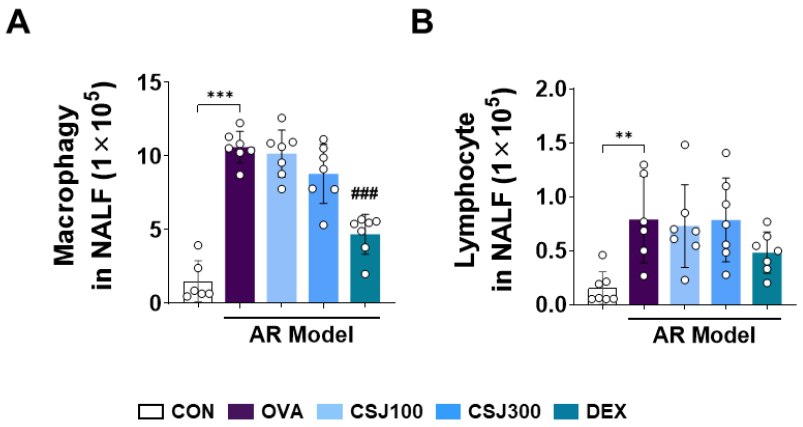


Fig. S1. Effects of CSJ on allergy nasal symptoms and nasal lavage fluid (NALF) in the OVA-induced AR mouse model. Numbers of infiltrated (A) macrophages and (B) lymphocytes in NALF were calculated. $^{**}P < 0.01$ and $^{***}P < 0.001$ compared with the CON group; $^{###}P < 0.001$ compared with the OVA group.

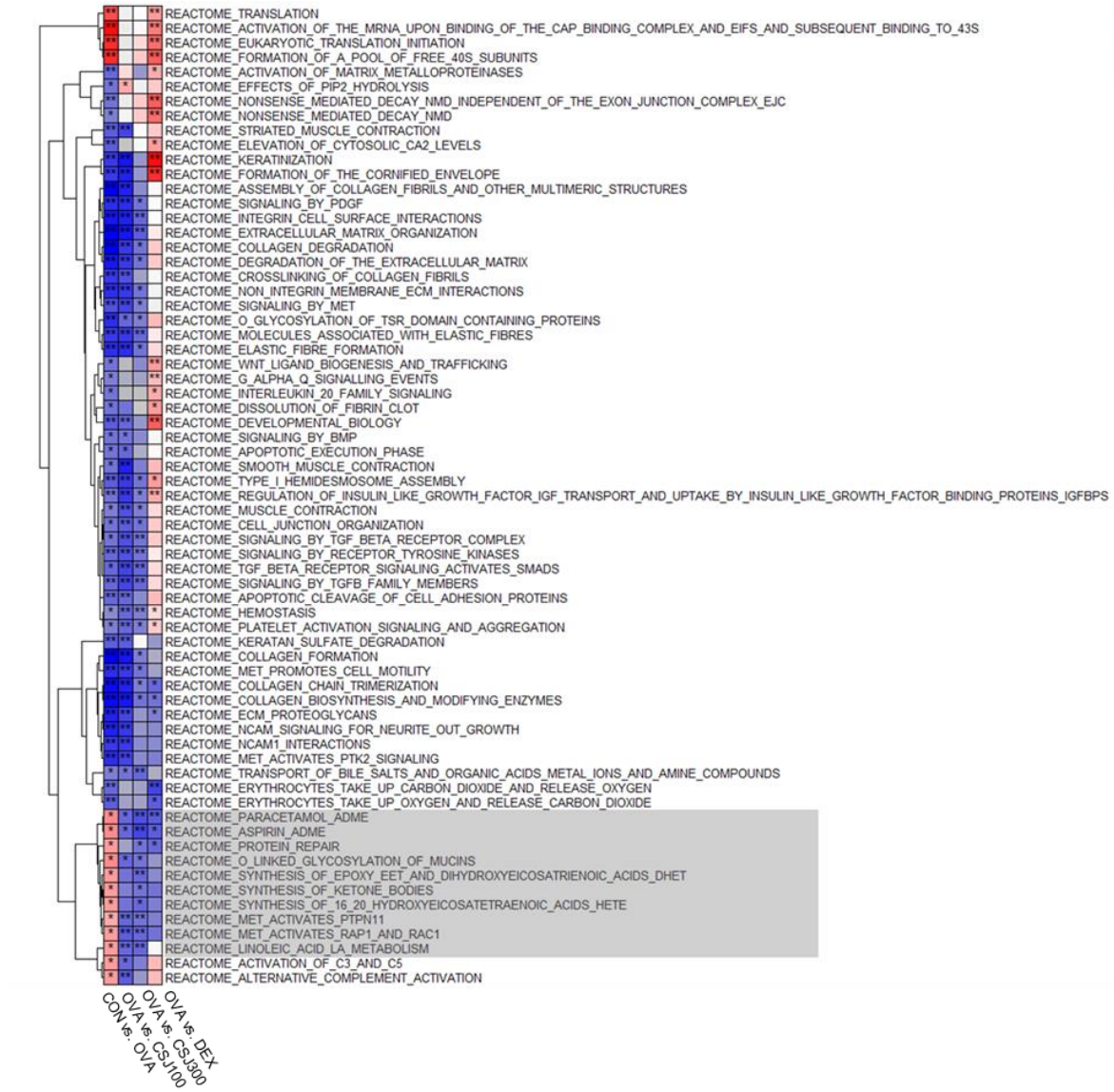


Fig. S2. Gene expression profiling of CSJ in the OVA-induced AR mouse model. Heat map of gene set enrichment analysis (GSEA) results for CSJ treatment using Reactome gene sets. Red/blue colors represent positive/negative normalized enrichment scores. * $P < 0.05$, ** $P < 0.01$. CSJ: *Cirsium japonicum*, AR: allergic rhinitis, CON: control, OVA: ovalbumin-induced AR mouse model, CSJ100: OVA-induced AR mouse model treated with CSJ (100 mg/kg/mouse), CSJ300: OVA-induced AR mouse model treated with CSJ (300 mg/kg/mouse), DEX: OVA-induced AR mouse model treated dexamethasone (1 mg/kg/mouse).

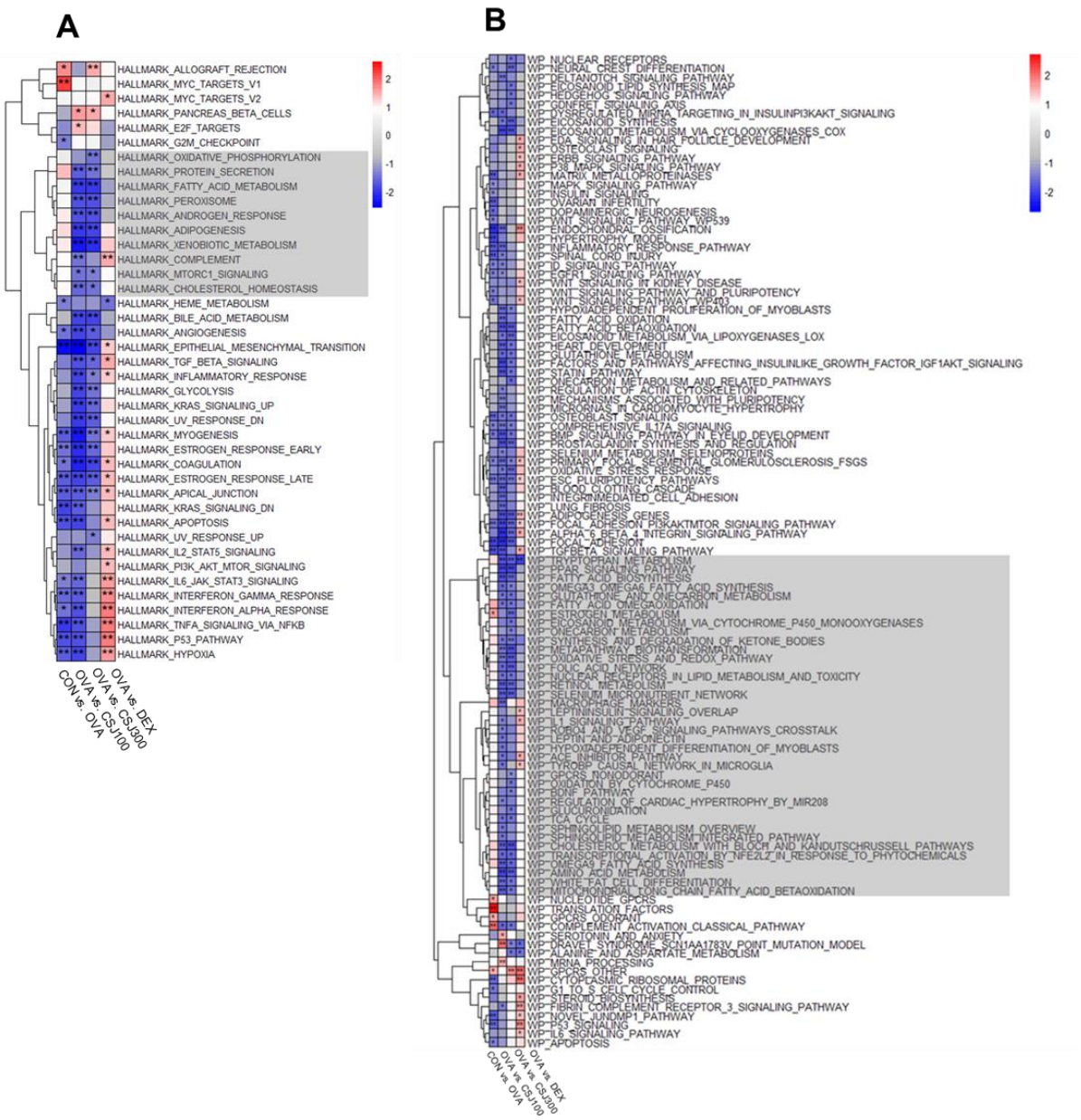


Fig. S3. Heatmap of the gene set enrichment results for CSJ. (A) Hallmark; (B) WikiPathways. Red and blue represent positive and negative normalized enrichment score (NES) values, respectively. * $P < 0.05$, ** $P < 0.01$.

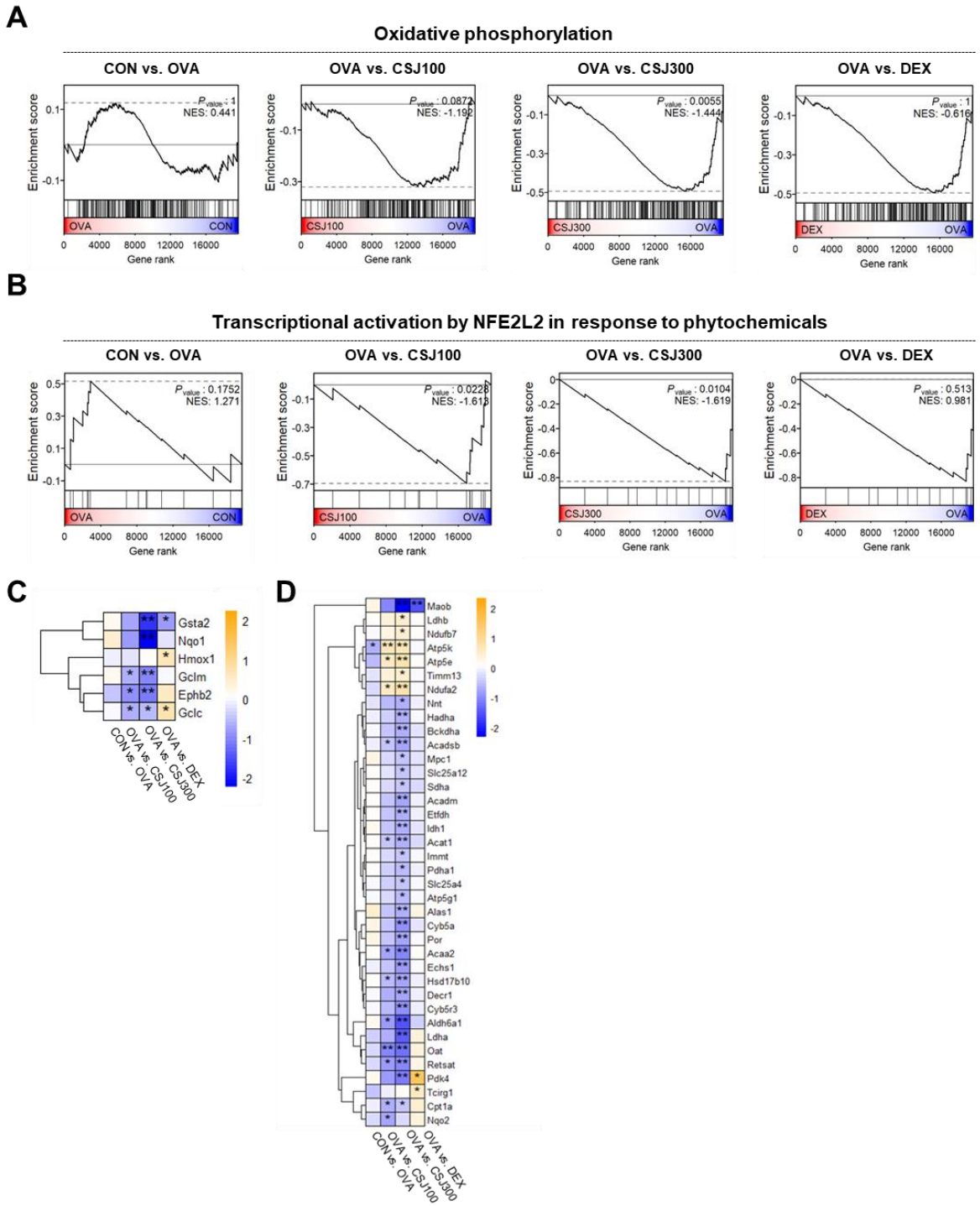
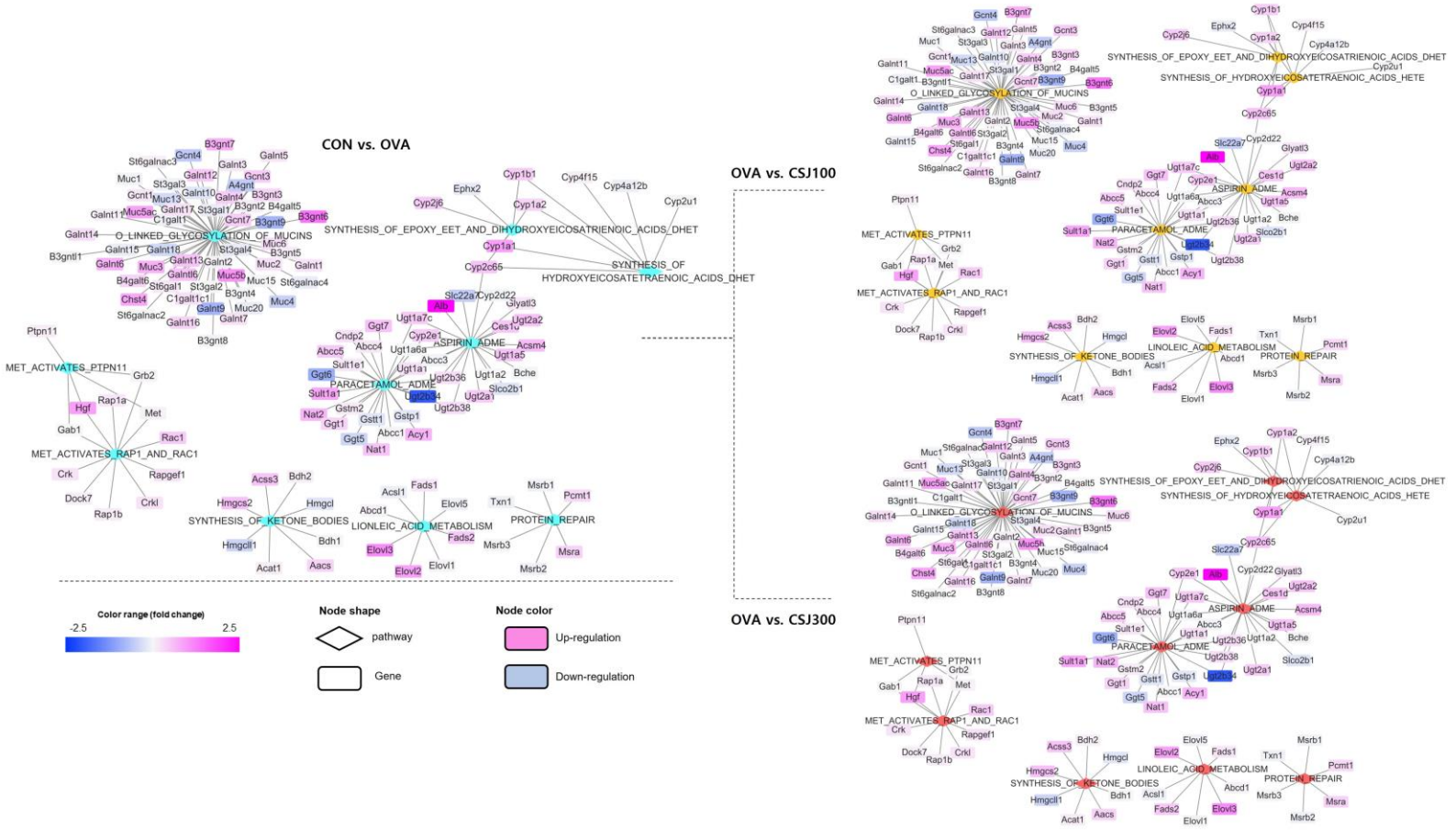


Fig. S4. GSEA plot of gene expression signatures. Plot for "Oxidative phosphorylation" (A) and "Transcriptional activation by Nfe2l2 in response to phytochemicals" (B) under each condition. (C) Heatmap of gene expression related to "Oxidative phosphorylation". (D) Heatmap of gene expression related to "Transcriptional activation by NFE2L2 (NRF2) in response to phytochemicals". * $P < 0.05$, ** $P < 0.01$.



92

93 **Fig. S5.** Integrated gene network showing the expression changes in enriched pathways and genes after CSJ treatment in the OVA-induced AR mouse model.

94 Blue and pink squares represent downregulation and upregulation of expression, respectively. CSJ: *Cirsium japonicum*, CON: control, OVA: ovalbumin-

95 induced AR mouse model, CSJ100: OVA-induced AR mouse model treated with CSJ (100 mg/kg/mouse), CSJ300: OVA-induced AR mouse model treated

96 with CSJ (300 mg/kg/mouse).

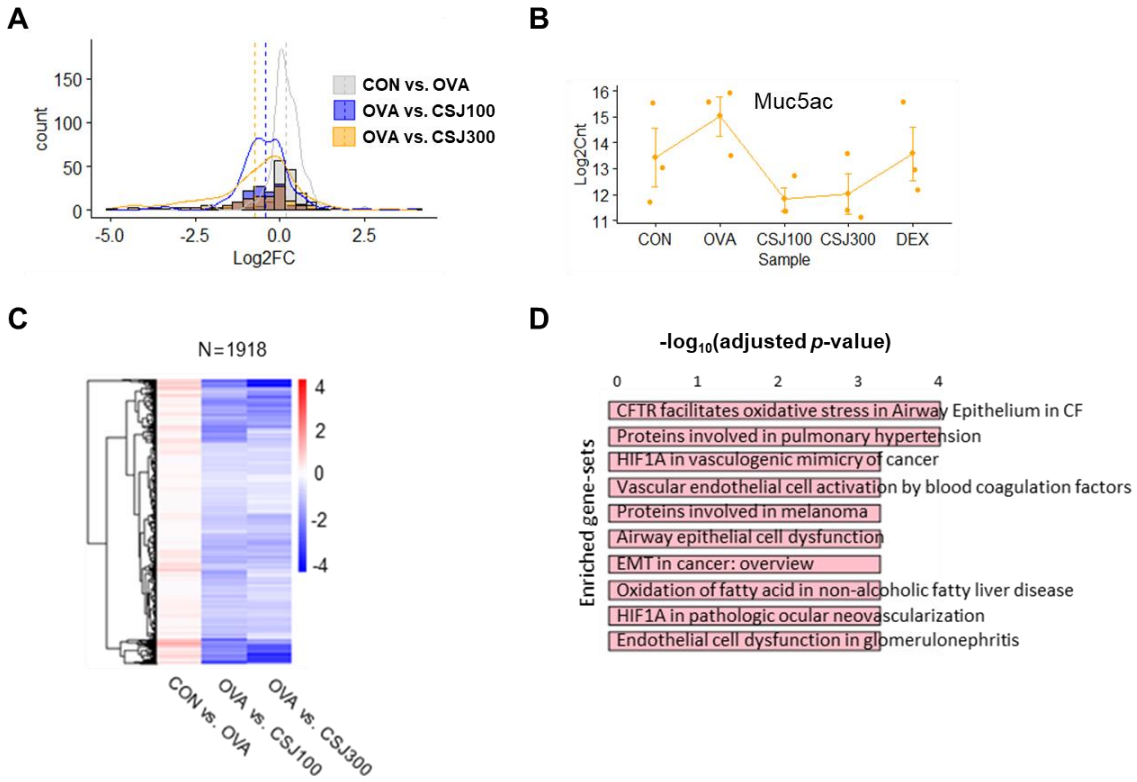


Fig. S6. Identification of key genes and major transcription factors (TFs) related to nasal inflammation. (A) DEG distribution under the three conditions. Log2FC indicates the fold change of expression differences for each gene under the three conditions (B) Gene expression abundance for *Muc5ac*. (C) Heatmap for genes significantly correlated with *Muc5ac* expression. (D) Bar plot of enriched functional gene sets significantly correlated with *Muc5ac* expression. CSJ: *Cirsium japonicum*, AR: allergic rhinitis, CON: control, OVA: ovalbumin-induced AR mouse model, CSJ100: OVA-induced AR mouse model treated with CSJ (100 mg/kg/mouse), CSJ300: OVA-induced AR mouse model treated with CSJ (300 mg/kg/mouse).

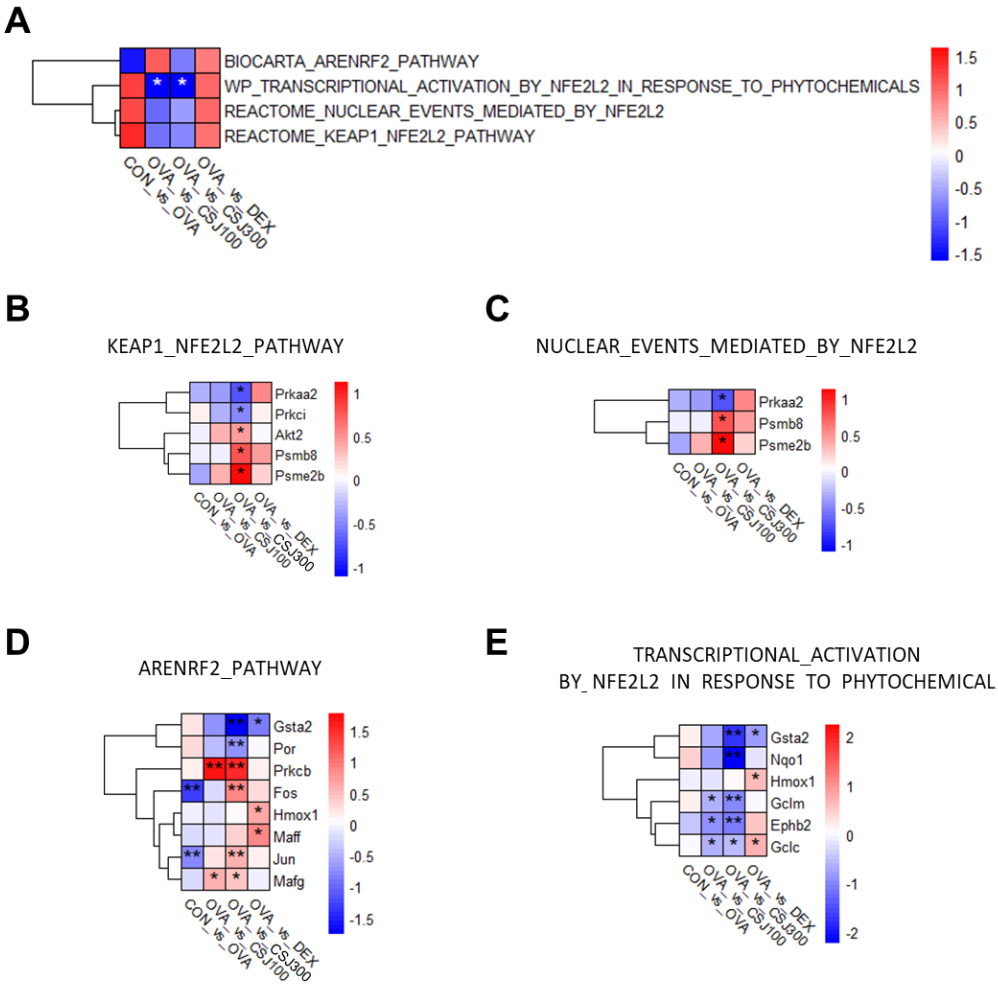
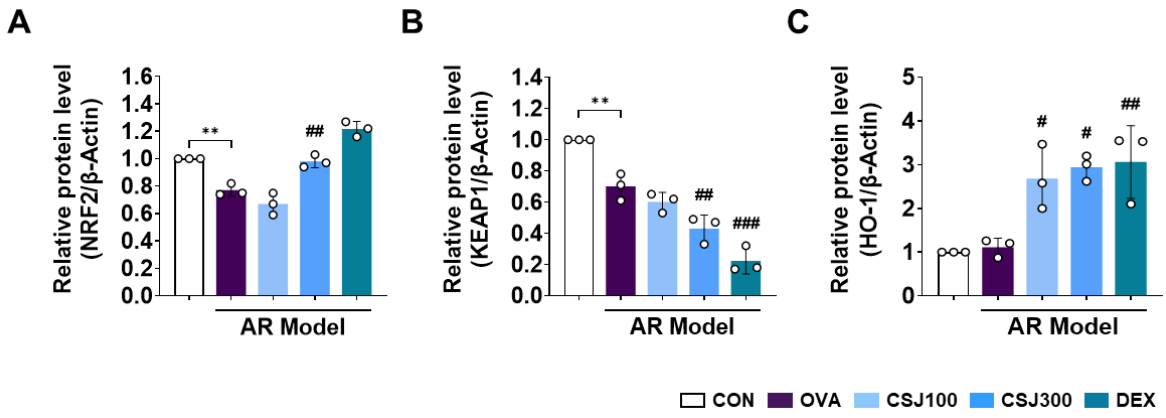


Fig. S7. Expression profiling of NRF2-associated pathways and NRF2 downstream genes included in MSigDB. (A) GSEA results of the NRF2 (NFE2L2)-associated pathway included in MSigDB. Gene expression abundance for downstream genes in Reactome (B and C), Biocarta (D), and WikiPathways (E). * $P < 0.05$, ** $P < 0.01$.



126

127

128 **Fig. S8.** Effect of CSJ on oxidative stress damage and the NRF2/KEAP1/HO-1 signaling pathway in
129 the OVA-induced AR mouse model. (A–C) Levels of NRF2, KEAP1, and HO-1 protein in
130 homogenized nasal tissues were evaluated using western blotting analysis. Protein levels were
131 normalized to the total β-actin level, and relative band intensities in western blots compared with those
132 in the CON group were calculated using ImageJ software. Results are presented as the means ± standard
133 deviation (SD, n = 3–4). $^{**}P < 0.01$, significantly different from the CON group, $^{\#}P < 0.05$, $^{\#\#}P < 0.01$,
134 and $^{\#\#\#}P < 0.001$ significantly different from the group treated with OVA. CSJ: *Cirsium japonicum*, AR:
135 allergic rhinitis, CON: control, OVA: ovalbumin-induced AR mouse model, CSJ100: OVA-induced
136 AR mouse model treated with CSJ (100 mg/kg/mouse), CSJ300: OVA-induced AR mouse model
137 treated with CSJ (300 mg/kg/mouse).

138

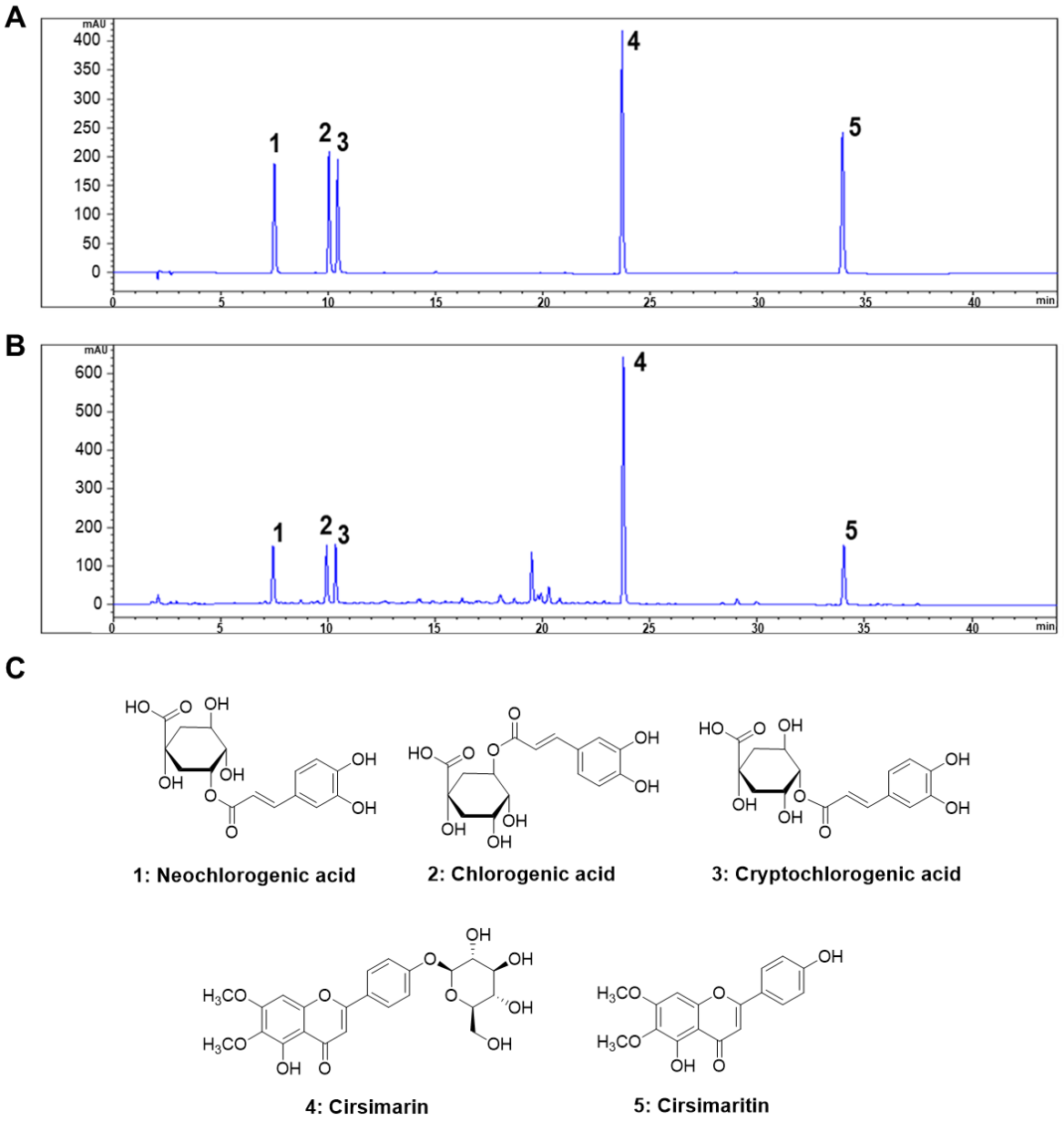
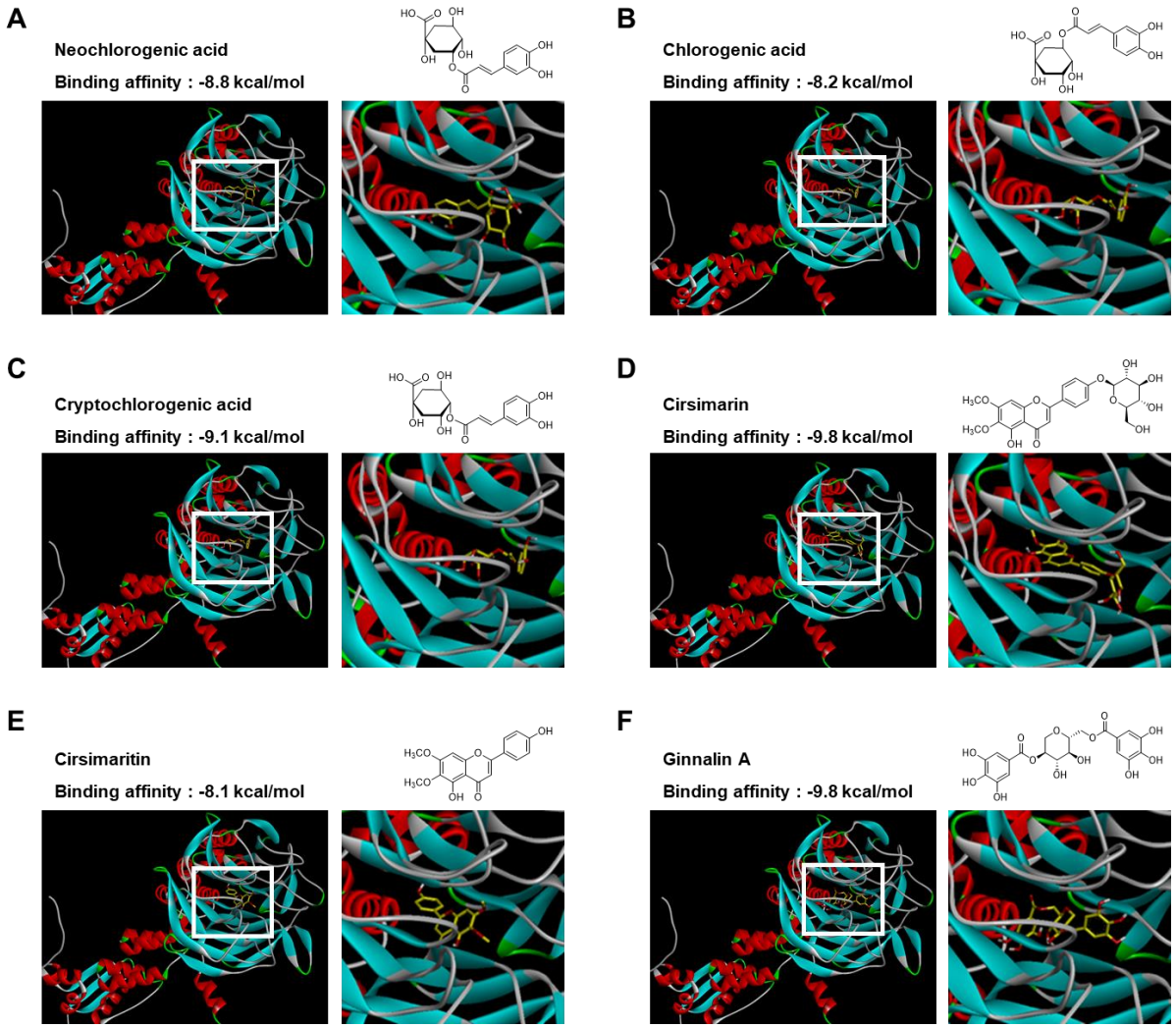
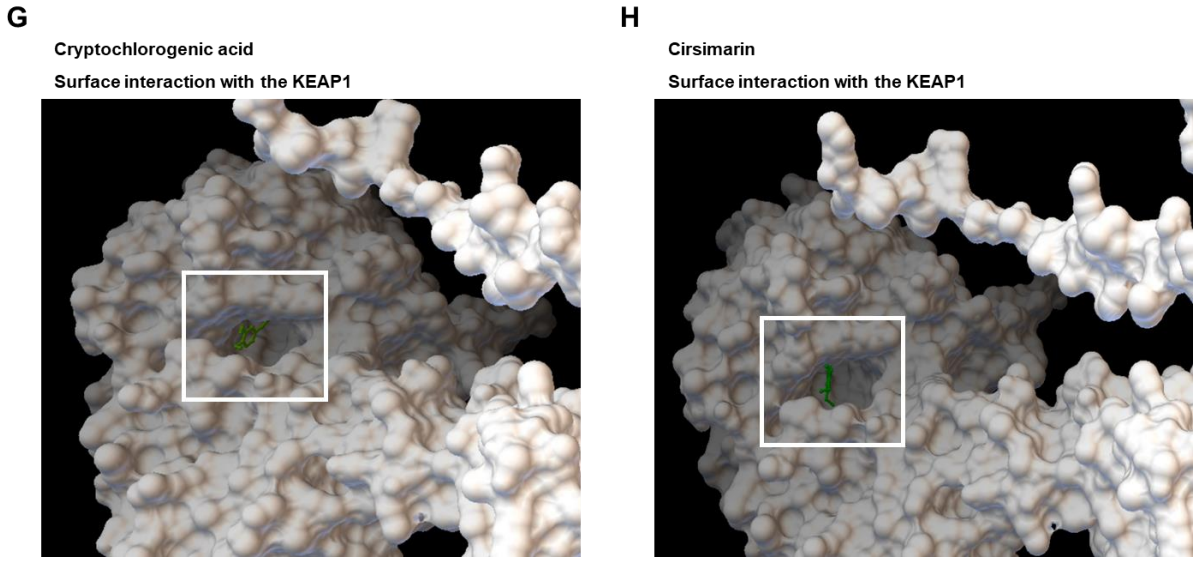


Fig. S9. HPLC analysis of CSJ. HPLC chromatograms of five reference standard mixtures (A) and CSJ (B). Chemical structures of five compounds (C). Peak identification: 1, neochlorogenic acid; 2, chlorogenic acid; 3, cryptochlorogenic acid; 4, cirsimarin; 5, cirsimaritin. Chromatographic conditions are described in the text. Detection wavelength was 340 nm.

146


A
Neochlorogenic acid
Binding affinity : -8.8 kcal/mol
B
Chlorogenic acid
Binding affinity : -8.2 kcal/mol
C
Cryptochlorogenic acid
Binding affinity : -9.1 kcal/mol
D
Cirsimarin
Binding affinity : -9.8 kcal/mol
E
Cirsimaritin
Binding affinity : -8.1 kcal/mol
F
Ginnalin A
Binding affinity : -9.8 kcal/mol

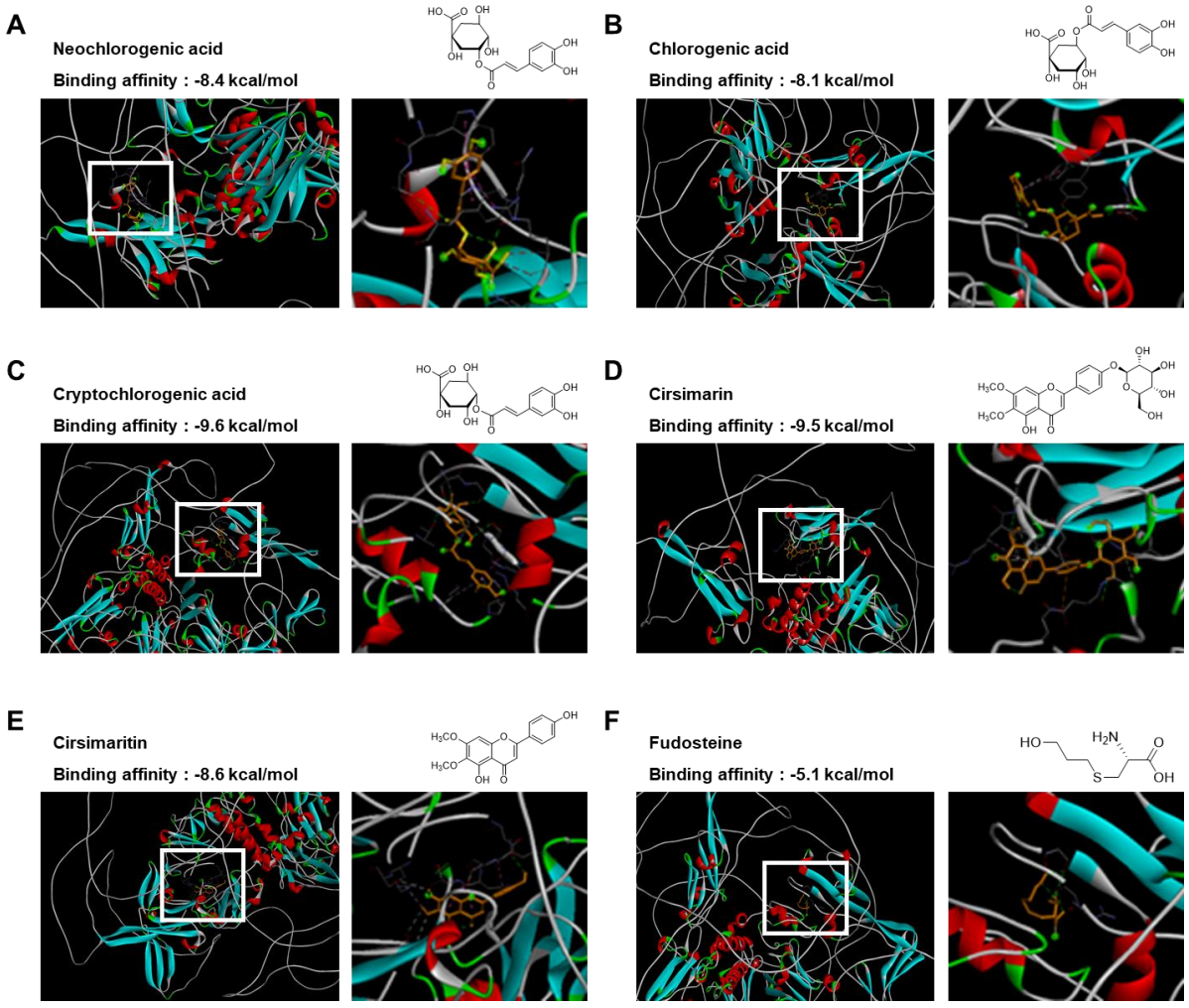
147


G
Cryptochlorogenic acid
Surface interaction with the KEAP1
H
Cirsimarin
Surface interaction with the KEAP1

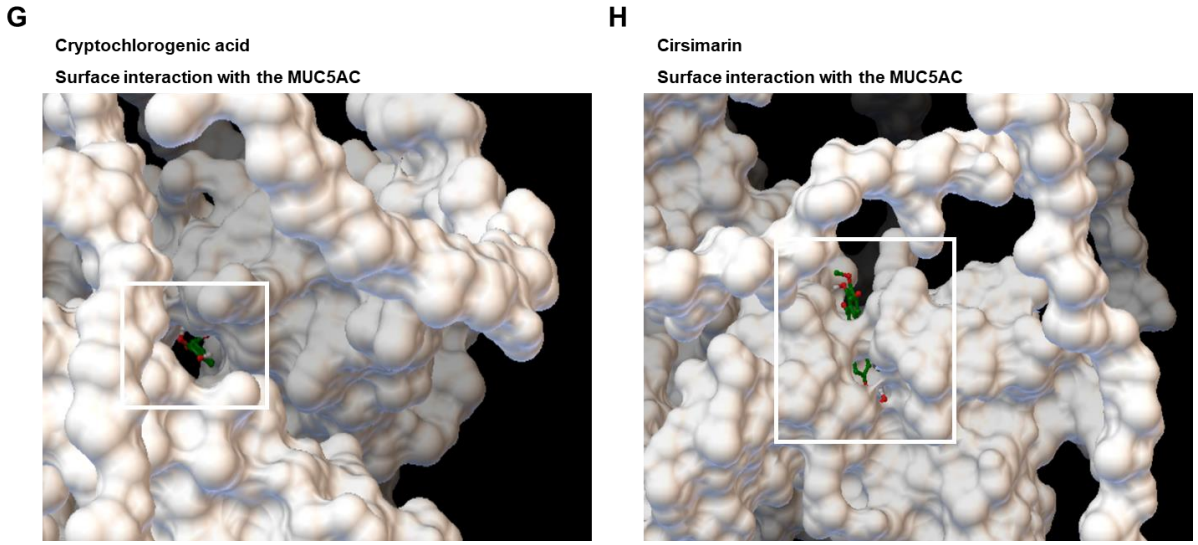
148

Fig. S10. Three-dimensional molecular docking structure of KEAP1 with five CSJ compounds and ginnalin A showing the interactions between compounds and the Kelch domain of KEAP1. The compounds directly bind to KEAP1 with a binding energy of (A) −8.8 kcal/mol (neochlorogenic acid), (B) −8.2 kcal/mol (chlorogenic acid), (C) −9.1 kcal/mol. (cryptochlorogenic acid), (D) − 9.8 kcal/mol (cirsimarin), (E) − 8.1 kcal/mol (cirsimaritin), and (F) −9.8 kcal/mol (ginnalin A). (G) Surface interaction representation of cryptochlorogenic acid binding to the Kelch domain of KEAP1. (H) Surface interaction representation of cirsimarin binding to the Kelch domain of KEAP1. The binding site of each major CSJ compound docked to the Kelch domain of KEAP1 is indicated by a white square.

158



159



160

161

Fig. S11. Three-dimensional molecular docking structure of MUC5AC and five CSJ compounds showing the interactions between compounds and the cysteine-rich domain (CysD) of MUC5AC. The compounds directly bind to MUC5AC with a binding energy of (A) −8.4 kcal/mol (neochlorogenic acid), (B) −8.1 kcal/mol (chlorogenic acid), (C) −9.6 kcal/mol. (cryptochlorogenic acid), (D) −9.5 kcal/mol (cirsimarin), (E) −8.6 kcal/mol (cirsimaritin), and (F) −5.1 kcal/mol (fudosteine). (G) Surface interaction representation of cryptochlorogenic acid binding to the CysD of MUC5AC. (H) Surface interaction representation of cirsimarin binding to the CysD of MUC5AC. The binding site of each major CSJ compound docked to the CysD of MUC5AC is indicated by a white square.

2. Supplementary Materials and Methods

2.1. Preparation of CSJ

CSJ plants were obtained from Omni Herb Co., Ltd. (Kyungpook, Republic of Korea). A voucher specimen (#JW207) and the herbal components were deposited in the herbarium of the Korea Institute of Oriental Medicine (Daejeon, Republic of Korea). CSJ (100 g) was extracted with distilled water (1 L) at 100 °C for 3 h using a heat-reflux system (MS-DM609, Misung Scientific, Yangu, Republic of Korea). The extract solution was filtered and concentrated under reduced pressure using a rotary evaporator (Ev-1020t, SciLab, Seoul, Republic of Korea) at below 60 °C. The concentrated extract was then freeze-dried using freeze-drying equipment (LP20; Ilshin Biobase, Dongduchen, Republic of Korea) at −80 °C for 96 h to obtain 12 g of a CSJ extract powder. An aliquot (20 mg) of the extract powder was dissolved in 60% methanol (10 mL), and the solution was filtered using a 0.45-μm syringe filter (Whatman, Clifton, NJ, USA) before high-performance liquid chromatography-grade (HPLC) analysis.

2.2. Animals and ethics approval

Six-week-old female BALB/c mice (weight: 18–20 g) were obtained from Samtako Bio (Osan, Republic of Korea). During the study, the mice were maintained in an animal room at 22 °C ± 2 °C and 55% ± 15% humidity under a 12 h light/dark cycle and provided ad libitum access to food and water. Animal experiments were performed according to the ARRIVE guidelines and were approved by the Institutional Animal Care and Use Committee of the Chonnam National University (approval no. CNU IACUC-YB-2022-83).

2.3. OVA-induced AR mouse model and treatment regimen

After a week of adaptation, the mice were sensitized by intraperitoneal injection of OVA (50 μg; Sigma-Aldrich, St. Louis, MO, USA) dissolved in aluminum hydroxide (2 mg; Sigma-Aldrich) on days 0, 7, and 14. Control (non-sensitized) mice were administered saline only. On day 21, the mice were

randomly divided into five groups (n = 6–8 mice per group): CON group (control), OVA group (AR model treated with OVA), CSJ100 group (AR model treated with CSJ100 mg/kg/mice), CSJ300 group (AR model treated with CSJ300 mg/kg/mice), and DEX group (AR model treated with dexamethasone 1 mg/kg mice). During 7 consecutive days (from day 21 to 27), CSJ or DEX was orally administered to mice in the treatment groups once daily, whereas mice in the CON and OVA groups were only treated with saline. On day 21, 23, and 27, the mice were intranasally challenged with 400 μg OVA solubilized in 20 μL saline. The mice were then sacrificed using alfaxalone (Jurox Pty Ltd., Rutherford, Australia) 28 h and 24 h after the last nasal challenge, and the nasal mucosa and blood were collected for subsequent experiments.

2.4. Evaluation of allergic nasal symptoms

After the final intranasal OVA challenge, allergic nasal symptoms were evaluated in a blinded manner by counting the frequencies and scores of sneezing and nasal rubbing behaviors, respectively, in the first 5 min.

2.5. Measurement of serum and nasal lavage fluid

Blood samples were centrifuged at 1,000 × *g* and 4 °C for 15 min, and the serum layer was collected. Serum OVA-specific IgE (500840, Cayman Chemical, Ann Arbor, MI, USA), histamine (ENZ-KIT140–0001, Enzo Life Science, Farmingdale, NY, USA), and IL-13 (DY413, R&D Systems, Minneapolis, MN, UK) levels were quantified using appropriate enzyme-linked immunosorbent assay (ELISA) kits according to the manufacturer's protocols. Nasal lavage fluid (NALF) analysis was performed by partially resecting the trachea, inserting a catheter into the nasopharynx, and gently administering 1 mL ice-cold saline. NALF was centrifuged for 10 min at 2,000 × *g* and 4 °C. To determine the differential immune cell counts in the NALF, the pellet was centrifuged onto the sliders using a cytospin device (CellSpin Clinical Centrifuge; Hanil Scientific, Incheon, Republic of Korea) for 10 min at 1,000 × *g* and 4 °C. The slides were stained using the Diff-Quik stain kit (38721, Sysmex Co., Kobe, Japan) for cell staining according to the corresponding protocol and observed under a light

microscope (400× magnification; Olympus Corporation, Tokyo, Japan). The total cell number in NALF was calculated using a cell counter (Thermo Fisher Scientific, Waltham, MA, USA), and the differential eosinophil and neutrophil counts in each group were assessed. The separated supernatant was stored at –80 °C until further analysis.

2.6. Histopathological and immunohistochemical analysis of nasal mucosa tissues

Mouse nasal tissues of mice were fixed in 10% formalin solution (Sigma-Aldrich) then decalcified in EDTA buffer (0.1 M; Bio-solution Co. Ltd., Seoul, Republic of Korea), processed using an alcohol-xylene series, and further embedded in paraffin. Subsequently, the paraffin-embedded nasal tissues were cut into 5-μm-thick sections and stained with Giemsa staining solution (BBC Biochemical, Mount Vernon, WA, USA) to analyze the degree of eosinophil infiltration. The thickness of the nasal mucosa was examined using hematoxylin and eosin (H&E; Sigma-Aldrich) staining, and the degree of mucus secretion was analyzed for goblet cell hyperplasia using a periodic acid–Schiff (PAS) staining kit (Sigma-Aldrich). Immunohistochemical staining was performed to assess mucin expression in the nasal mucosa. Specifically, an anti-MUC5AC antibody (dilution, 1:1,000; NBP2-15196, Novus Biologicals, Littleton, CO, USA) was diluted in antibody diluent (S0809; Dako, Agilent Technologies, Inc., Santa Clara, CA, USA) and incubated with nasal tissue sections overnight at 4 °C. After three PBS washes, tissue slides were incubated with a secondary antibody (MP-7801-15, Vector Labs, Burlingame, CA, USA) for 30 min at 24 °C, then stained with 3,3′-diaminobenzidine solution (Vector Labs). Stained tissue slides were scanned and digitalized using a Pannoramic DESK (3DHISTECH, Budapest, Hungary) digital slide scanner. Histopathological changes were assessed using Pannoramic CaseViewer software (3DHISTECH). Finally, we quantified the data using ImageJ software (version 1.52).

2.7. Library preparation and RNA-sequencing

Total RNA (1 μg) was isolated from nasal tissues of OVA-induced AR mice treated with two doses of CSJ (100 and 300 mg/kg/mice) using the RNeasy Mini Kit (Qiagen, Valencia, CA, USA). RNA integrity number (RIN) values were determined on a 2100 Bioanalyzer Instrument (Agilent, Santa Clara,

CA, USA), and samples with RIN values > 7 were used for sequencing. The sequencing library was prepared using the MGI Easy RNA Directional Library Prep Kit, and paired -end reads (100bp × 2) were generated for high-throughput sequencing via MGISEQ-2000 (MGI Tech, Shenzhen, China).

2.8. Estimate expression abundance

To obtain pure sequencing reads, adapter sequences and low-quality reads below Q30 were removed using Cutadapt [1]. Pure and high-quality sequence reads were mapped to the mouse genome (mm10), and RNA expression abundances were quantified using DESeq2 [2]. Differentially expressed genes (DEGs) between CSJ and OVA groups were determined by the log-transformed change (LogFC) and statistical significance ($p < 0.05$) using edgeR [3].

2.9. Pathway analysis

Gene set enrichment analysis (GSEA) of pre-ranked expression values was performed according to the DEG analysis using fgsea (v.1.12.0) in the Bioconductor package [4]. Gene sets (Hallmark, Reactome, KEGG, and WikiPathways) used in pathway analysis were obtained from the Molecular Signature Database (MSigDB; https://www.gsea-msigdb.org/gsea/msigdb). When identifying the enriched pathways, 67/41/101 Reactome/Hallmark/WikiPathways were selected according to their normalized enrichment scores and at least one condition showing a change with $p < 0.05$.

2.10. Analysis of the NRF2/KEAP1/HO-1 signaling pathway

The expression levels of nuclear factor erythroid 2-related factor 2 (NRF2; CSB-E16188m, Cusabio, Wuhan, China), Kelch-like ECH-associated protein 1 (KEAP1; CSB-EL012147MO, Cusabio), and heme oxygenase (HO-1; CSB-E08268m, Cusabio) in NALF were analyzed using an ELISA kits according to the manufacturer's instructions. Additionally, lipid peroxidation was examined by measuring the concentration of 4-hydroxynonenal (4-HNE; MBS7606509, MyBioSource, San Diego, CA, USA), which is expressed in high quantities in NALF during oxidative stress. Data were normalized to the control data.

2.11. Western blotting

Protein (20–30 μg) from homogenized nasal tissues was separated on 4–20% Mini-PROTEAN TGX Precast Protein Gels and transferred onto PVDF (0.2 μm) membranes using the Trans-Blot transfer system (Bio-Rad, Hercules, CA, USA). The membranes were then blocked with a blocking solution (Thermo Fisher Scientific) at 24 °C for 1 h to inhibit non-specific binding and probed with primary antibodies against NRF2 (ab137550; Abcam, Cambridge, UK; dilution, 1:1,000), KEAP1 (8047; Cell Signaling Technology, Danvers, MA, USA; dilution, 1:1,000), HO-1 (70081; Cell Signaling Technology; dilution, 1:1,000), and β-actin (4970; Cell Signaling Technology; dilution, 1:1,000) overnight at 4 °C. After membranes were washed with a TBS-T, they were incubated at 24 °C for 1 h with the relevant secondary antibodies conjugated with horseradish peroxidase (Santa Cruz Biotechnology, Dallas, TX, USA), and the immunoreactivity was detected using an enhanced chemiluminescence reagent (EzWestLumiOne, Atto Corporation, Tokyo, Japan). Protein bands were visualized using the ChemiDoc Imaging System (Bio-Rad) and quantified using ImageJ software (version 1.52a).

2.12. Chemicals and reagents for standard solution preparation

Reference standards for neochlorogenic acid (PubChem CID: 5280633, CFN97472), chlorogenic acid (PubChem CID: 1794427, CFN99116), cryptochlorogenic acid (PubChem CID: 9798666, CFN99117), cirsimarin (PubChem CID: 159460, CFN96507), and cirsimaritin (CID: 188323, CFN97126) were purchased from ChemFaces (Wuhan, China). The purity of all reference standards was >98%. Acetonitrile, methanol, and water were of HPLC grade and purchased from J. T. Baker (Phillipsburg, NJ, USA). Analytical-grade formic acid was purchased from Merck (Darmstadt, Germany). Standard stock solutions of five reference standards (all at 2 mg/mL) were prepared in HPLC-grade methanol, stored at <4 °C, and used for HPLC analysis after serial dilution in methanol.

2.13. Instrumentation and optimum chromatographic conditions

HPLC analyses were conducted on an Agilent 1200 HPLC instrument (Agilent Technologies) equipped with a vacuum degasser, binary pump, column compartment, autosampler, and diode array detector (DAD). Agilent ChemStation software was used for data collection and analysis. HPLC conditions were optimized for the column, mobile phase, flow rate, and detection wavelength to identify the bioactive compounds in CSJ. Specifically, a Zorbax Eclipse Plus C18 column (150 × 4.6 mm, 3 μm; Agilent Technologies) was used for chromatographic separation, and the column temperature was maintained at 35 °C. The mobile phase consisted of 0.1% formic acid in water (A) and acetonitrile (B), with gradient elution. The gradient solvent system was as follows: 95–60% A (0–30 min), 60–30% A (30–40 min), and 30–0% A (40–45 min). The column was re-equilibrated with 95% A for 10 min prior to each analysis and the flow rate was set at 0.8 mL/min. A wavelength of 340 nm yielded the highest S/N ratio for the five reference standards; therefore, detection was conducted at 340 nm. The injection volume of each sample was 5 μL. Using the optimized chromatographic conditions, the five reference standards were successfully separated and eluted within 45 min (Fig. S9A). The presence of the five markers in CSJ was confirmed by comparing their UV spectra and retention times with those of the corresponding reference standards (Fig. S9B). The chemical structures of the five reference standards (neochlorogenic acid, chlorogenic acid, cryptochlorogenic acid, cirsimarin, and cirsimaritin) are shown in Fig. S9C.

2.14. Validation of the HPLC method and sample analysis results

The HPLC method and analysis results were validated by determining the linearity, limit of detection (LOD), limit of quantification (LOQ), precision, and accuracy. Linearity was assessed from calibration curves generated from three replicate injections of standard solutions at five levels. The LOD and LOQ of each standard solution were determined using signal-to-noise (S/N) ratios of 3 and 10, respectively. The recovery test was performed to determine the accuracy of the method. The regression equations, linear ranges, correlation coefficients, LOD, and LOQ values of the reference standards are listed in Table S8. All calibration curves showed good linearity ($r^2 \geq 0.999$) within the tested concentration ranges. The relative standard deviation (RSD) values of the intra- and inter-day

precisions for the five reference standards were in the ranges 0.09–0.76% and 0.15–1.23%, respectively (Table S9). The recovery test was performed by adding three known concentrations (80%, 100%, and 120%) of the five reference standards into the CSJ extract. As shown in Table S9, the recovery rate of each reference standard was in the range 99.24–100.88%, and the RSD values were less than 4%. These data indicated that the developed HPLC/DAD method is reliable and highly accurate. The developed analytical method was applied for simultaneous quantitative analysis of the five marker compounds in the CSJ extract. The samples were analyzed in triplicate, and the results are summarized in Table S10. The contents of the five markers in the CSJ extract were in the range of 1.64–7.93 mg/g; the most abundant component in the extract was cirsimarin (7.93 ± 0.04 mg/g) followed by cirsimaritin (3.54 ± 0.04 mg/g).

2.15. Collection of major CSJ compounds and KEAP1 protein structures for docking analysis

Structural information was collected for ginnalin A, a positive control known to bind to KEAP1 [5], and five compounds derived from CSJ (neochlorogenic acid, chlorogenic acid, cryptochlorogenic acid, cirsimarin, and cirsimaritin). 3D sdf files were downloaded from the PubChem database and used to obtain structural information of the compounds [6]. Compound structure files were converted into pdbqt files using OpenBabel software [7]. KEAP1 protein structural information was obtained using human-derived data provided by the AlphaFold 2.0 (AF) database [8]. As AF predicts the structure of a protein using artificial intelligence and provides the prediction result, it has the advantage of confirming all sequence structures of a protein. The KEAP1 (AF entry: Q14145) structure in pdb format was downloaded from AF and converted to pdbqt format using OpenBabel software.

2.16. Molecular docking of major CSJ compounds with KEAP1 and statistical processing

Molecular docking analysis was performed between the six compounds collected from PubChem and the KEAP1 protein collected from AF. Docking analysis was performed using AutoDock vina software [9]. The exhaustiveness of the docking analysis parameters was set to 100, the center coordinates were set to (0, 0, 0), and the grid box size was set to (126, 126, 126). In each docking

analysis, the interaction with the lowest binding score was selected based on the interaction between the compound and KEAP1. Additionally, docking analysis using AF tends generates better binding scores than that using the more common Protein Data Bank database [10] because it excludes specific compounds, such as water molecules present around the protein, and uses only the protein sequence [11]. To eliminate this bias, a significant binding-affinity score was derived by calculating the *p*-value using a permutation test [12]. The 108,625 pairs of docking affinity scores obtained using the same software and parameters as a previous study [11] were used as permutation sets. For the docking score belonging to the permutation set, the value corresponding to the top 10% was selected as the threshold ($p < 0.1$). Compounds with binding-affinity scores below this threshold were selected as significant compounds for the docking assays.

2.17. Collection of major CSJ compounds and MUC5AC protein structures for docking analysis

Structural information was collected for fudosteine, a positive control known to interact with the MUC5AC protein [13], and the five CSJ compounds. 3D sdf files were downloaded from the PubChem database and used to obtain structural information on the compounds. The compound structure files were converted into pdbqt files using OpenBabel software. The MUC5AC protein structure was based on the human protein structure provided in the AF database. AF provides data by cutting proteins with more than 2,700 amino acids every 1,400 amino acids, with a difference of 200 for the analysis of overlapping parts. For example, first-fragment data are given in the form of sequence numbers 1–1,400, and second-fragment data are given in the form of sequence data numbers 200–1,600. Therefore, as MUC5AC contains 5,654 amino acids, it was provided in 23 fragments (AF entry: P98088). After all 23 fragment structures were downloaded, they were preprocessed in pdbqt form using OpenBabel software for subsequent analysis.

2.18. Molecular docking of CSJ compounds with MUC5AC and statistical processing

Molecular docking was performed in the same manner as that for KEAP1 protein docking analysis (Section 2.16). The analysis generated 23 docking results for each compound, and the interaction with

the lowest binding score was selected as the interaction between the compound and MUC5AC. The statistical processing method was also the same as that for KEAP1, and a value of $p < 0.1$ was selected as the threshold value using the data set obtained in a previous study [11].

2.19. Statistical analysis

All statistical analyses were performed using Prism version 9 (GraphPad Software, Inc., La Jolla, CA, USA). Data are expressed as means ± standard deviation (SD). Data between groups were analyzed using one-way analysis of variance (ANOVA) followed by Tukey's multiple comparison test. Statistical significance was set to $p < 0.05$.

3. Supplementary Tables

Table S8. Calibration curves, linearity, LOD, and LOQ for the five reference standards (n = 3)

Compound	Regression equation[a]	Linear range (μg/mL)	Correlation coefficient (r^2)	LOD[b] (μg/mL)	LOQ[c] (μg/mL)
Neochlorogenic acid	y = 15.598x + 17.597	20–100	0.9996	0.05	0.15
Chlorogenic acid	y = 17.011x + 19.165	20–100	0.9993	0.02	0.07
Cryptochlorogenic acid	y = 15.879x + 15.690	20–100	0.9991	0.02	0.07
Cirsimarin	y = 13.934x + 65.830	60–300	0.9990	0.12	0.38
Cirsimaritin	y = 8.808x + 11.190	20–100	0.9990	0.03	0.09

[a]y, Peak area of the compound; x, concentration (μg/mL) of the compound

[b]S/N = 3

[c]S/N = 10

Abbreviations: LOD, limit of detection; LOQ, limit of quantification

Table S9. Precision (intra- and inter-day) and recovery of the five reference standards (n = 6)

Compound	Intra-day precision RSD (%)	Inter-day precision RSD (%)	Recovery	
			Recovery (%)	RSD (%)
Neochlorogenic acid	0.71	1.21	99.24	3.04
Chlorogenic acid	0.72	1.23	99.83	1.64
Cryptochlorogenic acid	0.76	1.18	100.88	1.72
Cirsimarin	0.24	0.34	100.82	1.64
Cirsimaritin	0.09	0.15	99.63	2.79

Abbreviation: RSD, relative standard deviation

Table S10. Contents of the five marker compounds in the *Cirsium japonicum* herb extract

Compound	Content (n = 3)	
	mg/g (mean ± SD)	%
Neochlorogenic acid	1.92 ± 0.01	0.19
Chlorogenic acid	1.64 ± 0.02	0.16
Cryptochlorogenic acid	1.73 ± 0.02	0.17
Cirsimarin	7.93 ± 0.04	0.79
Cirsimaritin	3.54 ± 0.04	0.35

413 Abbreviation: SD, standard deviation

414

415

References

[1] M. Martin, Cutadapt removes adapter sequences from high-throughput sequencing reads, EMBnet.journal 17 (2011) 1.

[2] M.I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2, Genome Biol. 15 (2014) 550.

[3] M.D. Robinson, D.J. McCarthy, G.K. Smyth, edgeR: a Bioconductor package for differential expression analysis of digital gene expression data, Bioinformatics 26 (2010) 139–140.

[4] S.V. Korotkevich G, Sergushichev A, Fast gene set enrichment analysis, bioRxiv (2019). https://doi.org/doi:10.1101/060012

[5] Z. Zhang, L. Peng, Y. Fu, et al., Ginnalin A binds to the subpockets of Keap1 Kelch domain to activate the Nrf2-regulated antioxidant defense system in SH-SY5Y cells, ACS Chem. Neurosci. 12 (2021) 872–882.

[6] S. Kim, J. Chen, T. Cheng, et al., PubChem 2019 update: improved access to chemical data, Nucleic Acids Res. 47 (2019) D1102–D1109.

[7] N.M. O'Boyle, M. Banck, C.A. James, et al., Open Babel: An open chemical toolbox, J. Cheminform. 3 (2011) 33.

[8] J. Jumper, R. Evans, A. Pritzel, et al., Highly accurate protein structure prediction with AlphaFold, Nature 596 (2021) 583–589.

[9] O. Trott, A.J. Olson, AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading, J. Comput. Chem. 31 (2010) 455–461.

[10] H.M. Berman, J. Westbrook, Z. Feng, et al., The Protein Data Bank, Nucleic Acids Res. 28 (2000) 235–242.

[11] M. Park, S.-Y. Lee, H. Lee, J.-M. Yi, Effect of terpenes from *Poria cocos*: Verifying modes of action against Alzheimer's disease using molecular docking, drug-induced transcriptomes and diffusion network, biorxiv (2023). https://doi.org/10.1101/2023.06.01.543358

[12] F. Pesarin, L. Salmaso, The permutation testing approach a review, Statistica 70 (2010) 481–509.

442 [13] C.K. Rhee, C.M. Kang, M.B. You, et al., Effect of fudosteine on mucin production, Eur. Respir. J.
443 32 (2008) 1195–1202.
444