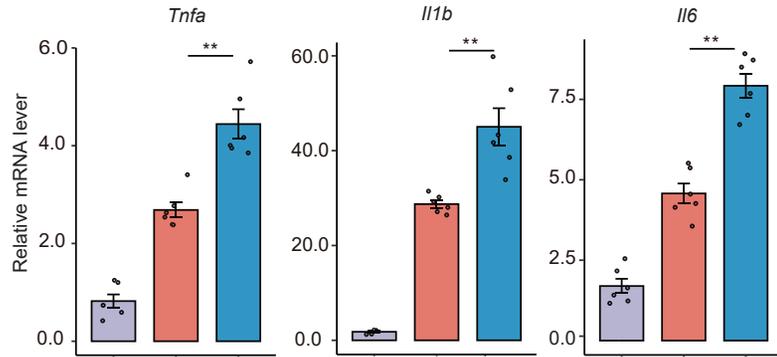
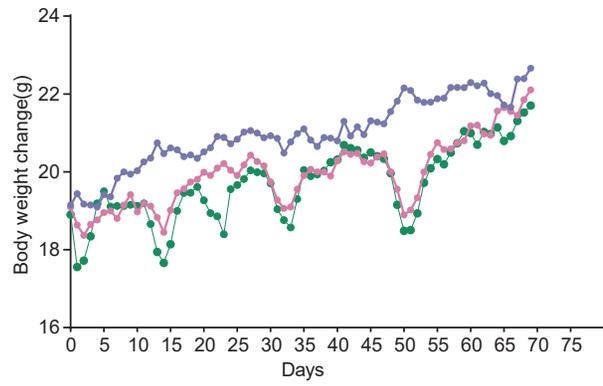
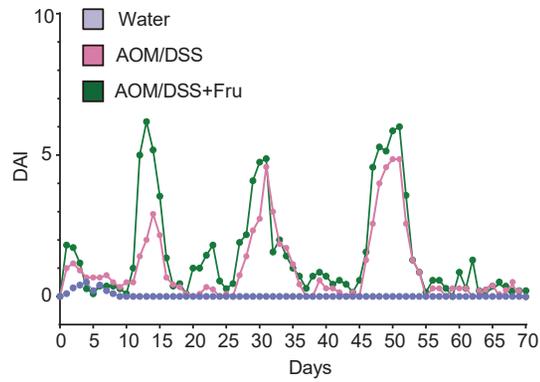
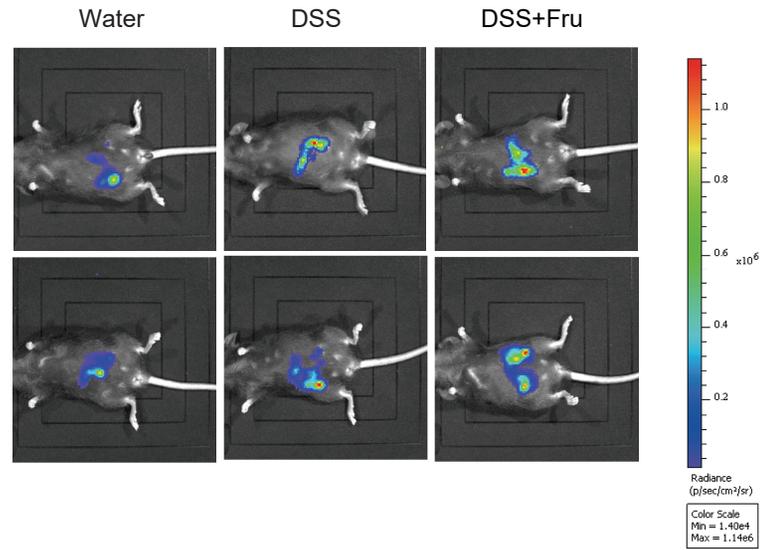
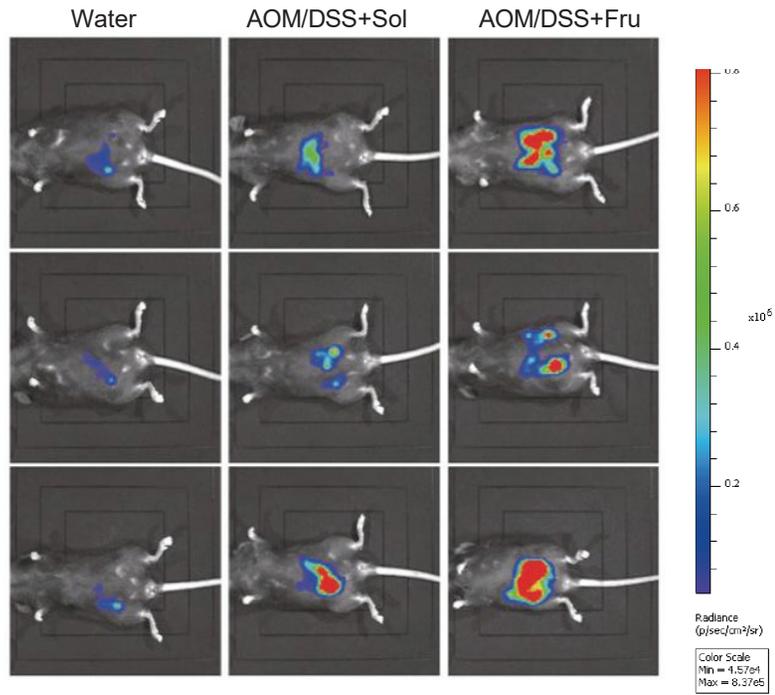
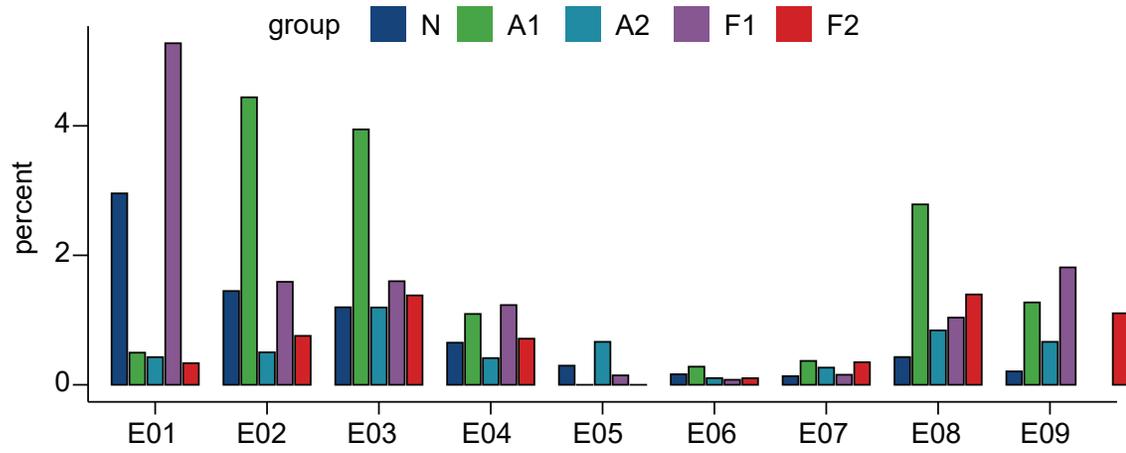
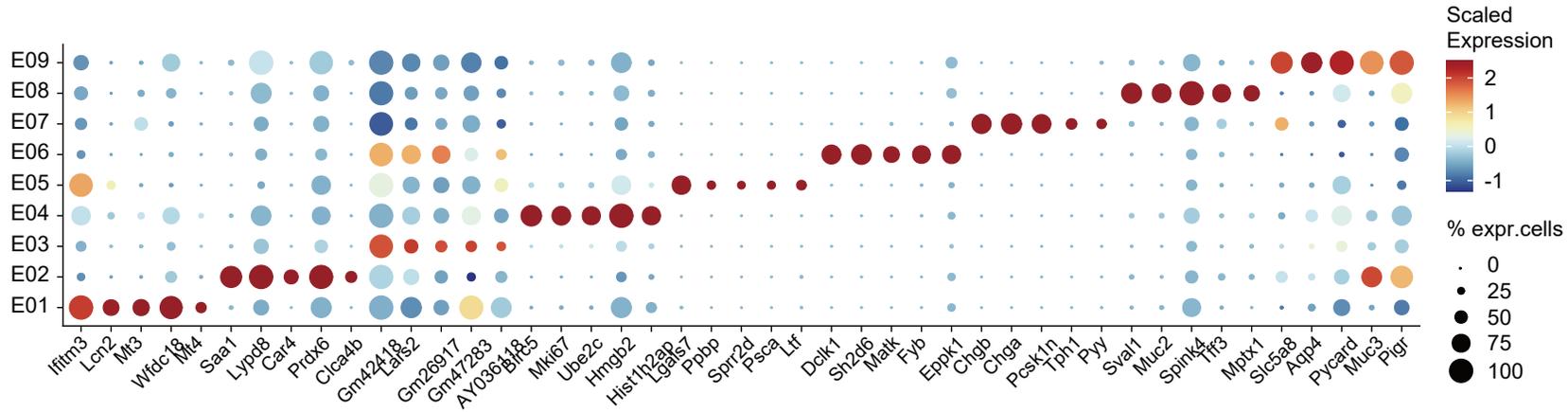
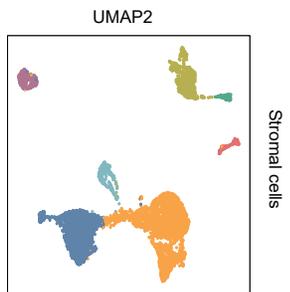
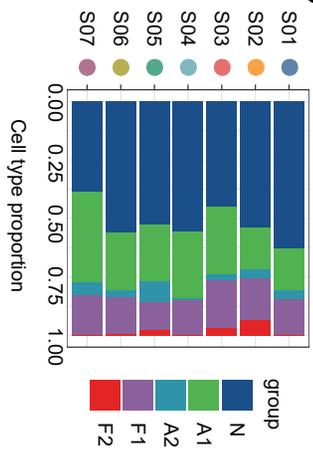
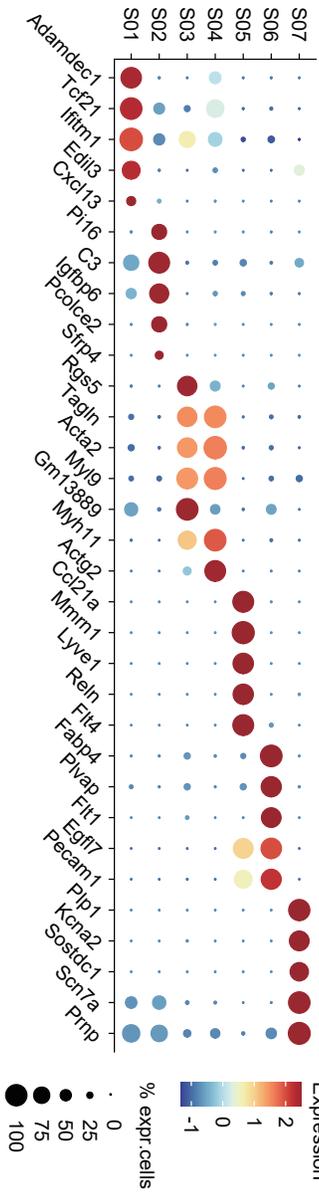


**A****C****D****B****E**

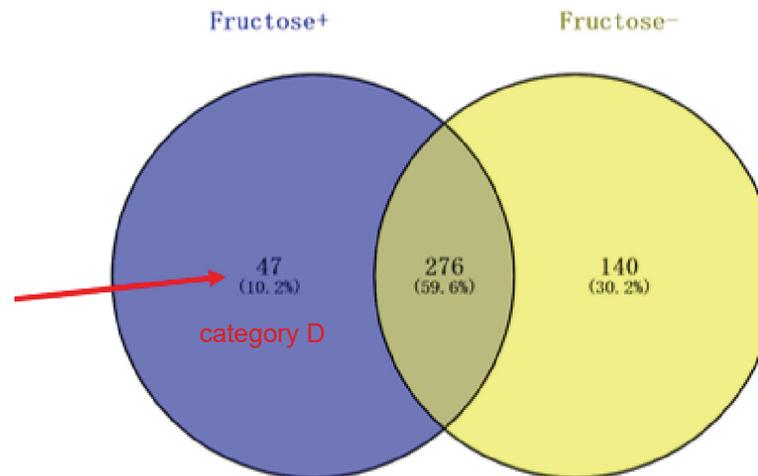
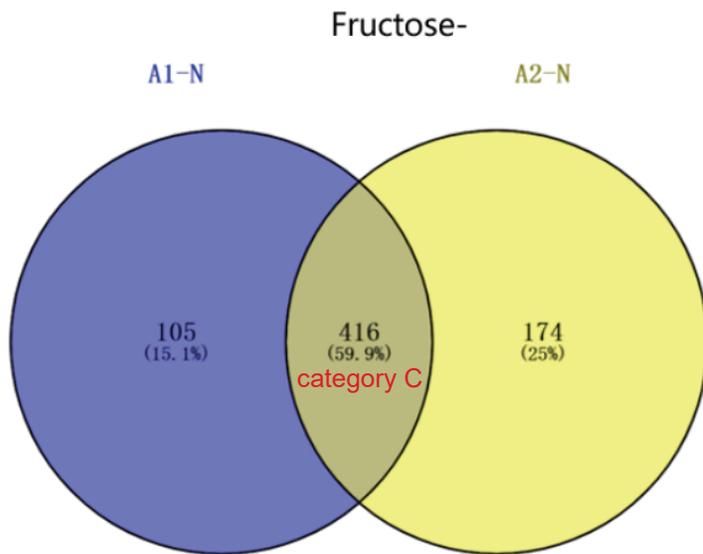
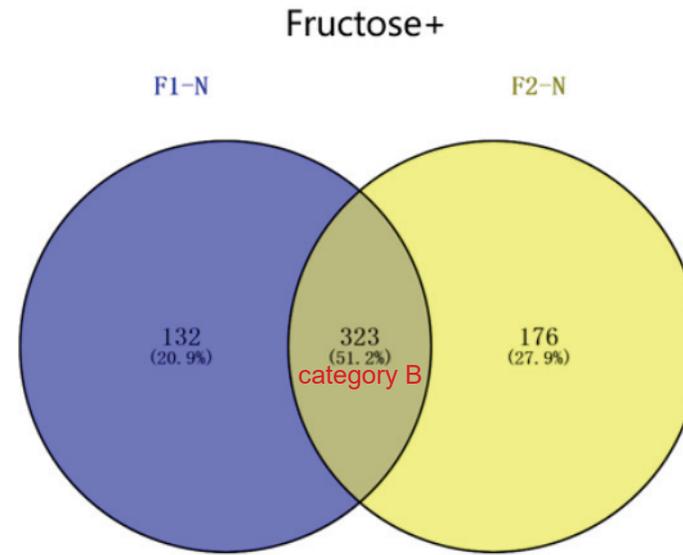
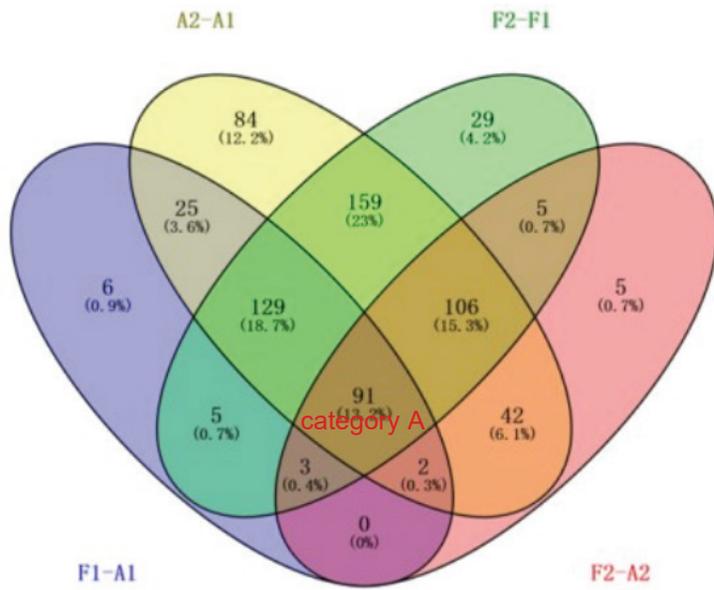
**A****B**

**A**

● S01\_Fibroblast(pro-inflammatory)  
 ● S02\_Fibroblast\_P116  
 ● S03\_Pericytes\_Rgs5  
 ● S04\_Myofibroblasts\_Acta2  
 ● S05\_LECs\_Lyve1  
 ● S06\_Endothelial cells  
 ● S07\_Schwann cells

**B****C**





**A**

# Food $\cap$ TCM

For a longer, better life.

Herb



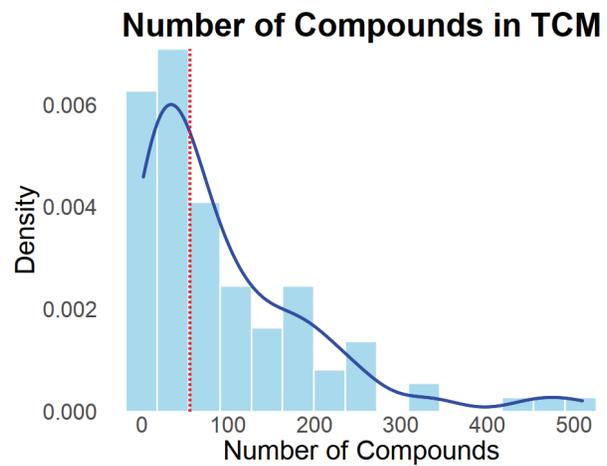
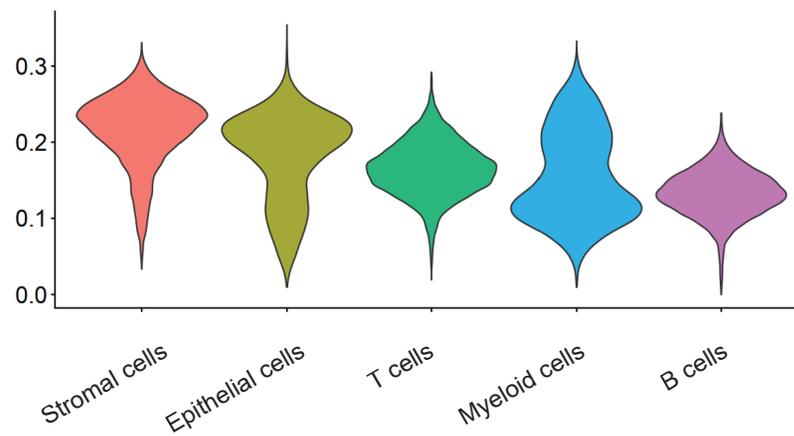
112 Herbs

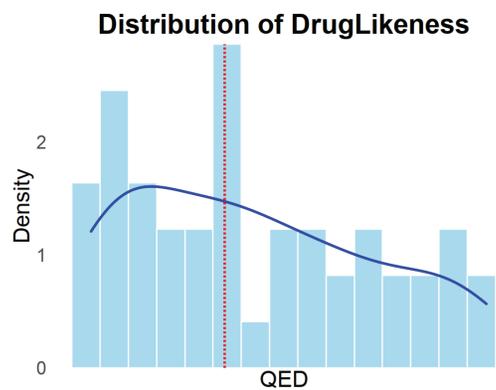
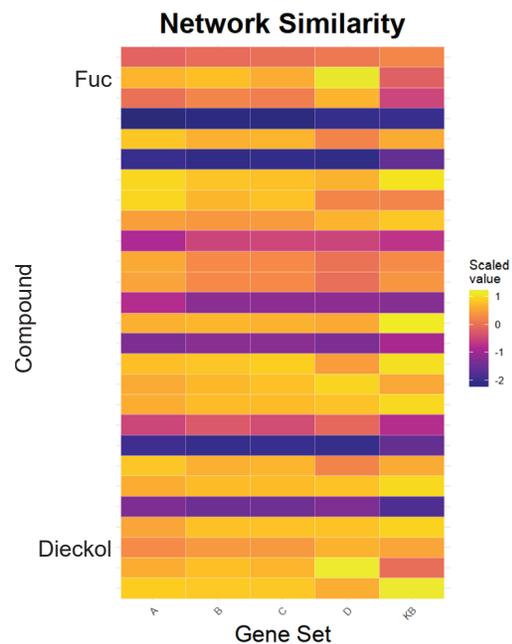
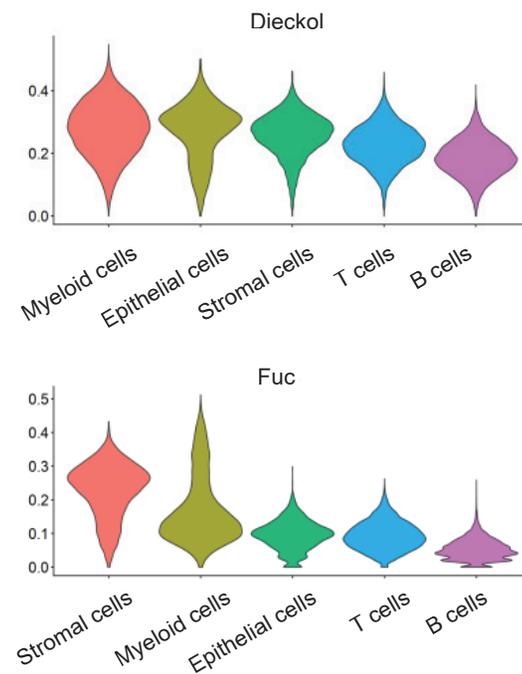
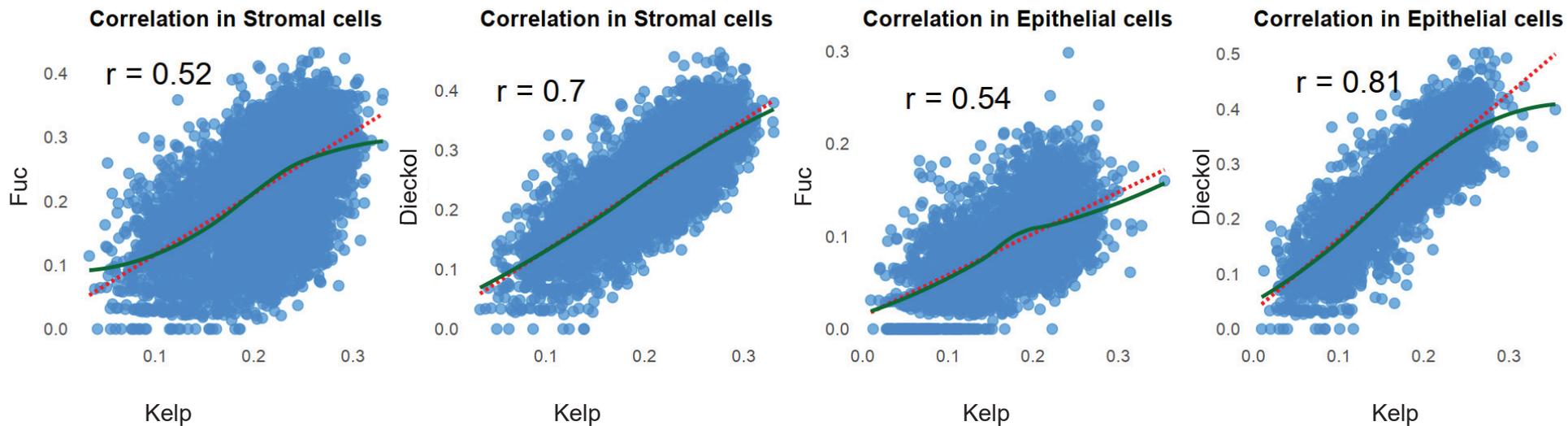


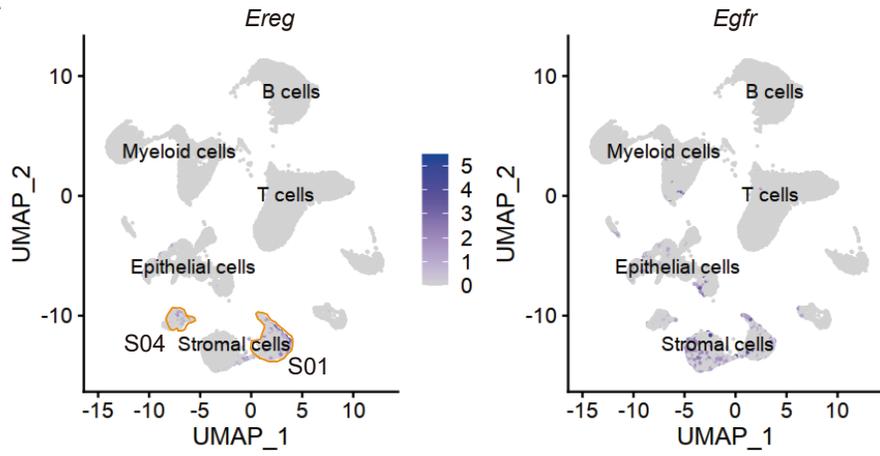
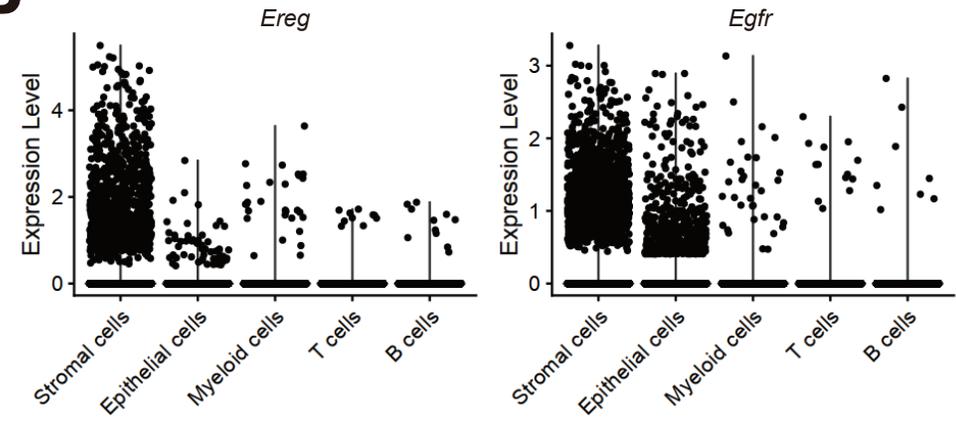
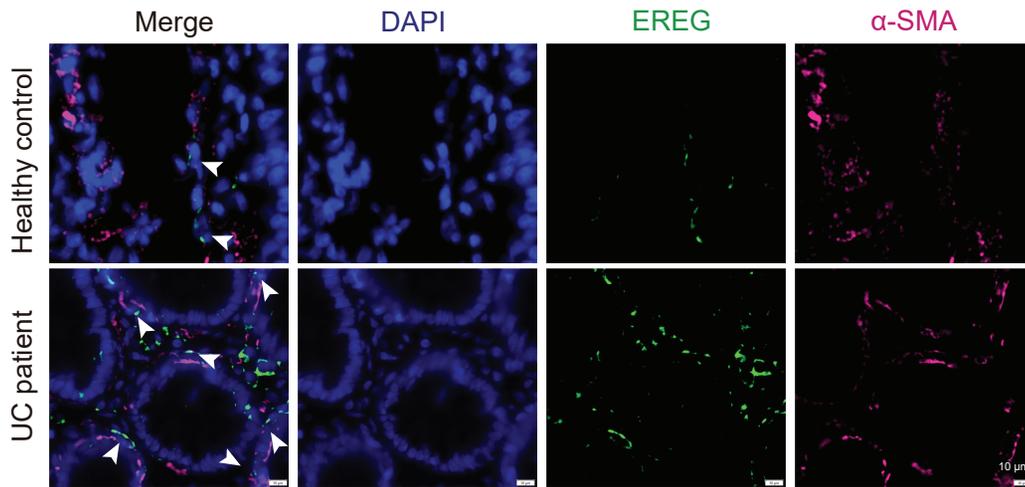
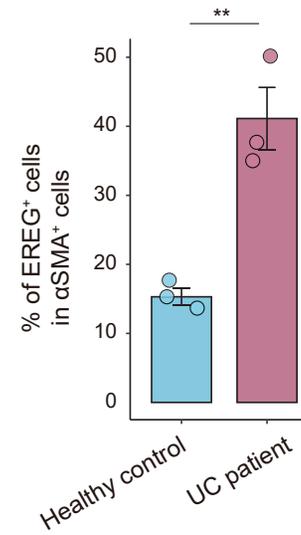
14,988 Compounds

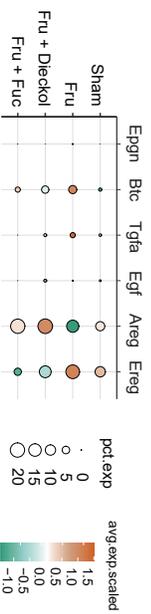
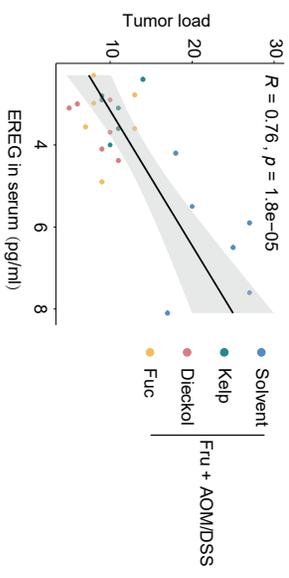
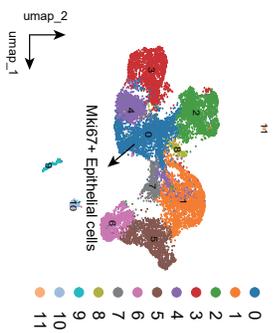
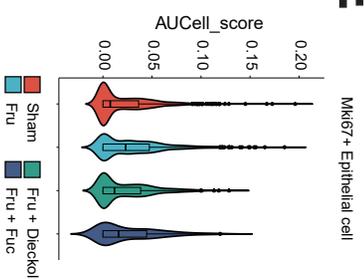


63 Macro-nutrients

**B****C**

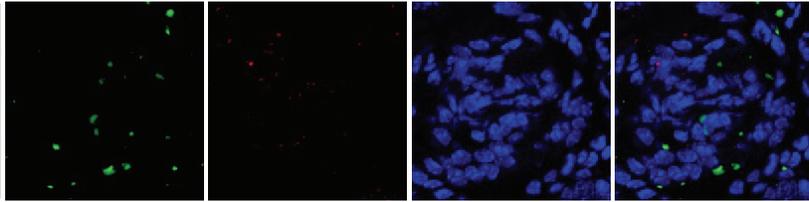
**A****B****C****D**

**A****B****C****D**

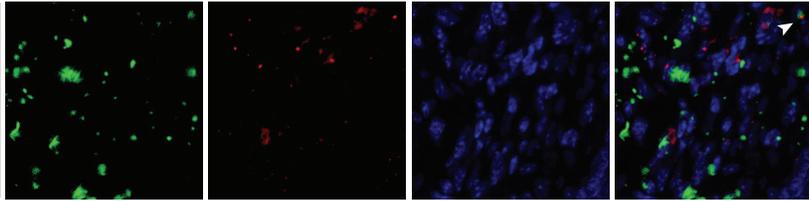
**A****B****C****D****E**

EREG OGT DAPI

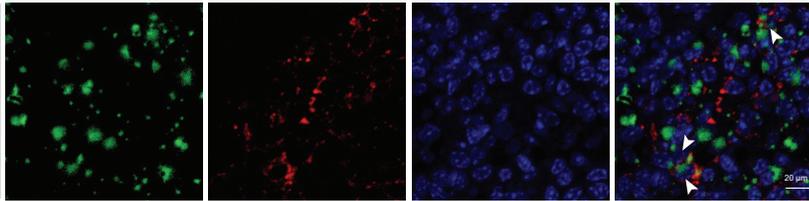
Normal mucosa

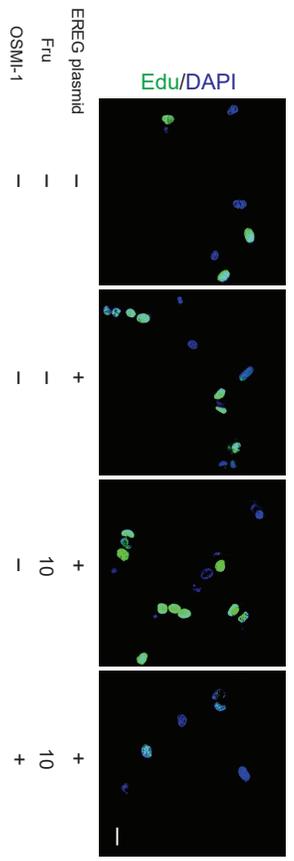


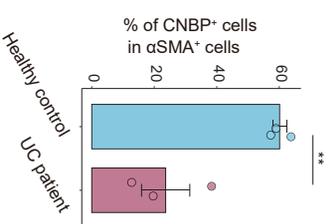
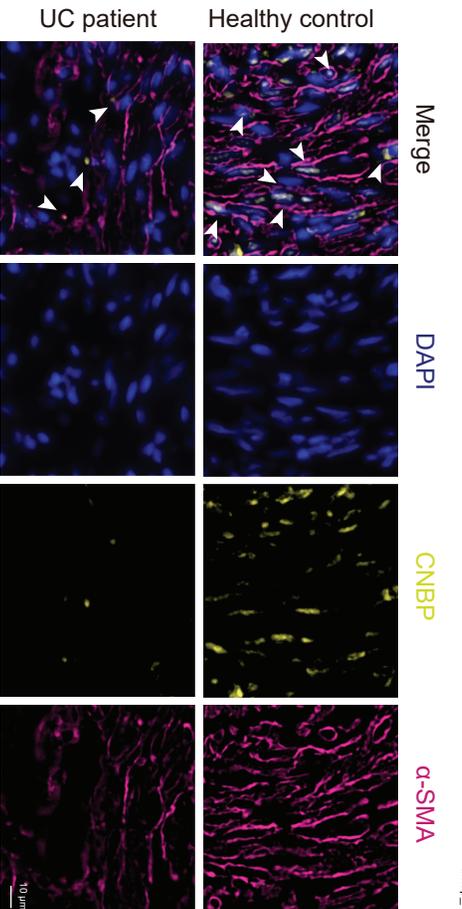
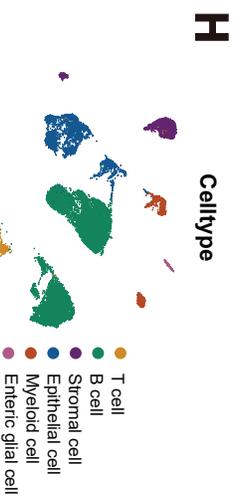
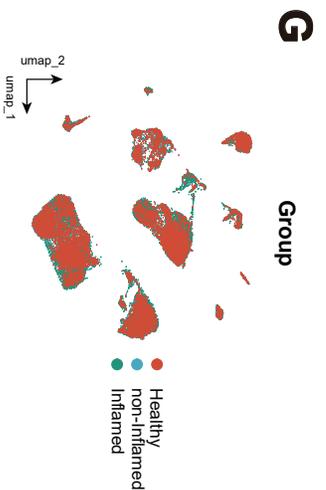
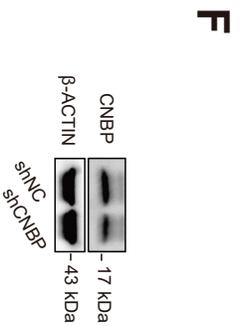
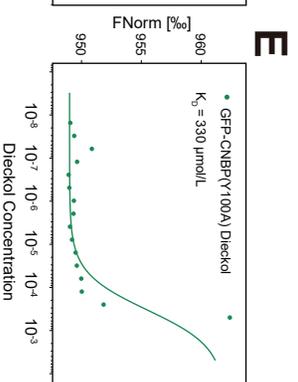
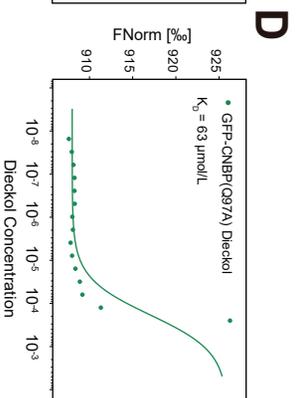
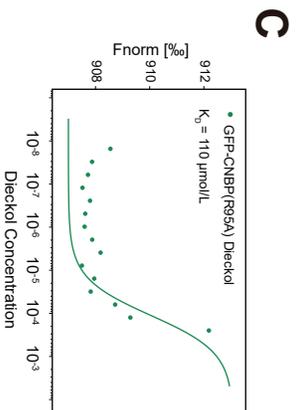
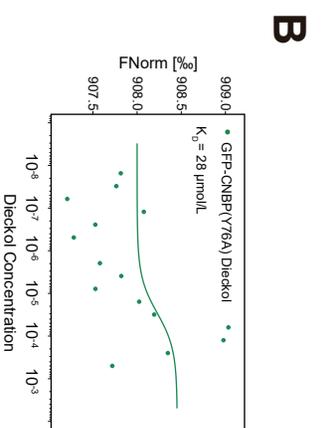
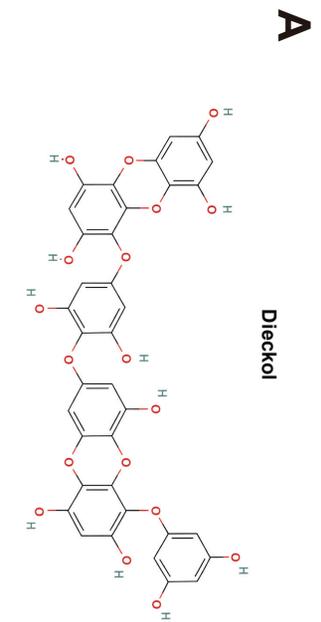
AOM/DSS Day 75

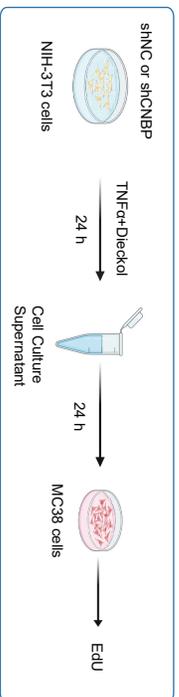
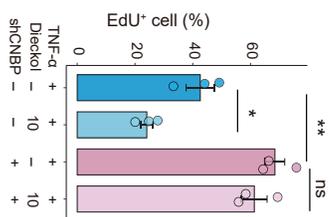
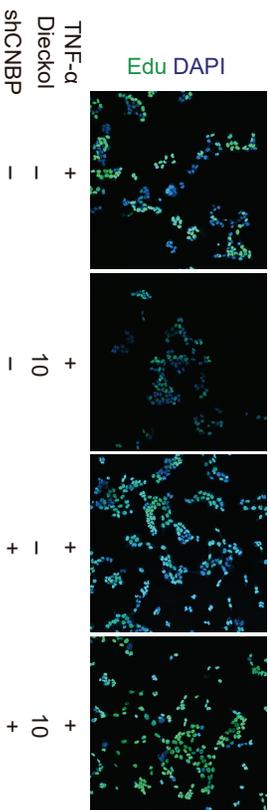
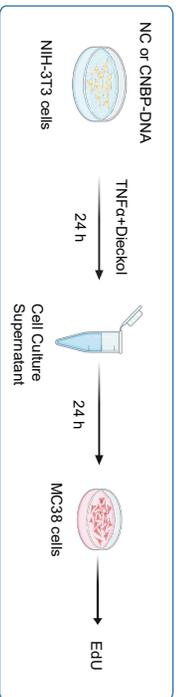
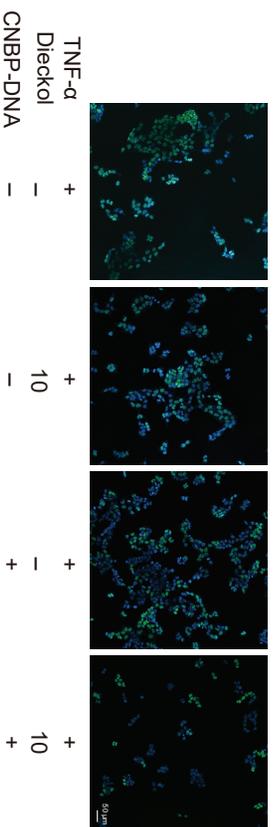
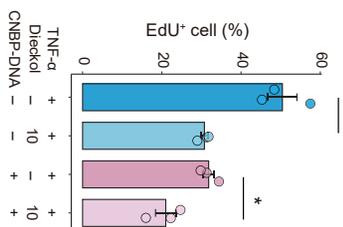


Fru + AOM/DSS Day 75







**A****C****B****D****E****F**

**Supplementary Table 1. Antibodies used in the study**

Antibodies	Source	Identifier
EREG	Santa Cruz	sc-376284
p-ERK	CST	4370
ERK	CST	4695
p-EGFR	CST	11862
EGFR	Santa Cruz	sc-81488
OGT	Santa Cruz	sc-74546
O-GlcNAc	Abcam	ab2739
Ki67	Abcam	ab15580
HA-Tag	CST	3724
$\alpha$ -SMA	CST	19245
GAPDH	Affinity	T0004
$\beta$ -ACTIN	Affinity	AF7018
$\beta$ -Tublin	Affinity	T0023
GFP-Tag	Proteintech	50430-2-AP
CNBP	Santa Cruz	sc-515387

**Supplementary Table 2. Compound information**

Name	Source	Identifier
DSS	MP Biomedicals	0216011090
D-(-)-fructose	Sigma	F0127
AOM	Sigma	A5486
Fucoidan	Yuanye	S11142
Dieckol	MCE	HY-147059
CCK-8	MCE	HY-K0301
EdU	Beyotime	C0071S
HiScript III RT SuperMix for qPCR	Vazyme	R323-01
SYBR Green Mix	Vazyme	Q712-02
Trizol	Takara	9109
L-012	Wako	120-04891
DAPI	Thermo Fisher Scientific	D3751
Alexa Fluor™594	Thermo Fisher Scientific	A11005
Alexa Fluor™488	Thermo Fisher Scientific	A11008
Multiple immunofluorescence kit	Shang Hai GuduBio Technology Co. Ltd	HYDS0045
IHC Detect Kit	Proteintech Group Inc	PK10006
Protein A/G	Thermo Fisher Scientific	88802
ChIP assay kit	Beyotime	P2078
Dual Luciferase Reporter Gene Assay Kit	Beyotime	RG027

Mouse EREG(Epiregulin) ELISA Kit	ELK Biotechnology	ELK4009
TNF- $\alpha$	Peptotech	PMC3016
OSM1-1	MCE	HY-119738
FITC-Dextran	MCE	HY-128868G
Lipofectamine™ 3000	Thermo Fisher Scientific	L3000015
Cycloheximide	MCE	HY-12320

### Supplementary Table 3. Oligonucleotides used in the study information

qPCR Primers	Sequence
Mouse- <i>Ereg</i>	5'-ACGGGGATCGTCTTCCATCT-3' 5'-CGCTGCTTGTGTAGGTTCCC-3'
Mouse- <i>Ereg</i> promoter (ChIP)	5'-AAGGCATGCTGAAGGACTC-3' 5'-GATTCCAGTGCCGAACCAGA-3'
Mouse- <i>Ereg</i> Peak1 (ChIP)	5'-CGACAGCAGCTCAGACTCTCC-3' 5'-CCATGAGTCAGGACCCCAACC-3'
Mouse- <i>Ereg</i> Peak2 (ChIP)	5'-GGATCAGCTCCGTTCCAACT-3' 5'-AAAGCAGCGTCAAGACCCAA-3'
Mouse- <i>Tnfa</i>	5'-ATGAGCACAGAAAGCATGATC-3' 5'-TACAGGCTTGTCACTCGAATT-3'
Mouse- <i>Il1b</i>	5'-GAAATGCCACCCTTTGACACAGTG-3' 5'-TGGATGCCTCTCATCAGGACAG-3'
Mouse- <i>Il6</i>	5'-TAGTCCTTCCTACCCCAATTTCC-3' 5'-TTGGTCCTTAGCCCACTCCTTC-3'

## Methods

### High-fructose exacerbated DSS-induced colitis model and colitis-associated

#### cancer (CAC) model

Eight-week-old male C57BL/6 mice were obtained from GemPharmatech Co. Ltd. and housed under controlled conditions (21±3°C, 12-hour light/dark cycle). All experiments were approved by an Institutional Animal Care and Use Committee

(IACUC) at the Model Animal Research Center of Nanjing University

(IACUC2506005). For the high-fructose exacerbated colitis model, thirty 8-week-old

male C57BL/6 mice were randomly divided into three groups. The Fructose+DSS group received drinking water supplemented with 30% fructose, which was refreshed every 48 hours for 4 consecutive weeks. The remaining groups consumed standard drinking water. At week 5, 2.5% DSS was administered in the drinking water of both DSS and Fructose+DSS groups, with the latter maintaining 30% fructose supplementation. On day 6 post-DSS administration, the DSS solution was replaced with either standard water or 30% fructose water. Body weight, fecal consistency, and anal hygiene status were monitored throughout the experiment. All mice were euthanized on day 8 post-DSS initiation following a 12-hour fasting period.

For the CAC model, the mice received a single intraperitoneal injection of 7.5 mg/kg AOM solution on day 0. The AOM/DSS+Fructose group was provided with 30% fructose-supplemented drinking water continuously throughout the entire modeling period. On day 5, both AOM/DSS and AOM/DSS+Fructose groups were administered 2.5% DSS in drinking water for 5 consecutive days. This was followed by a 14-day recovery period starting during which mice received either standard water or 30% fructose water. The DSS challenge-recovery cycle was repeated twice. Body weight, fecal consistency, and anal health were monitored throughout the experiment. Tissues were harvested for analysis at the inflammatory phase (tumor initiation, day 30) and adenoma phase (day 75). At the experimental endpoint, colons were opened longitudinally and visible tumors ( $\geq 1$  mm in diameter) were counted. Tumor load was determined by summing the calculated areas of all identified tumors for each mouse. Tumor length and width were measured using an electronic caliper, and tumor area was calculated accordingly. All measurements were performed in a blinded manner.

To evaluate the ameliorative effects of the selected medicinal-food homologous compounds on high-fructose diet-aggravated colitis-associated colorectal cancer, mice were administered daily via intragastric gavage with Kelp extract (200 mg/kg), dieckol (100 mg/kg), and fucoidan (100 mg/kg). According to previous reports, the disease activity index (DAI) was determined based on three parameters: weight loss, stool consistency, and the severity of intestinal bleeding.

### **Kelp extract**

First, thoroughly wash dried Kelp (*Laminaria japonica*) with abundant distilled water to remove surface dust, salts, and soluble impurities. Kelp fragments were mixed with distilled water, followed by the addition of enzymatic agents (pectinase and cellulase).

The suspension was stirred at 50°C for 24 hours, subsequently centrifuged at 3000 g for 20 minutes at 4°C, and vacuum-filtered. Three volumes of 60% ethanol were then added for precipitation. After 18 hours of extraction, the solution was filtered, concentrated using a rotary evaporator, and finally spray-dried to obtain the powdered extract<sup>21</sup>.

### **Cell culture**

L929 cells were cultured in DMEM (11965092, Gibco) supplemented with 10% fetal

bovine serum (10100147, Gibco) and 1% penicillin-streptomycin (C0222, Beyotime); NIH-3T3 cells were maintained in DMEM containing 10% newborn calf serum (26010074, Gibco). Both cell lines were incubated at 37°C with 5% CO<sub>2</sub>.

### **Single-cell RNA sequencing**

Murine colon tissues were collected and generate single-cell gel bead in-emulsions (GEMs) using the 10× Chromium platform according to the manufacturer's protocols. Subsequent library construction and sequencing were performed. All libraries underwent paired-end sequencing with dual indexing on the NextSeq 500 platform (Illumina).

### **Data preprocessing**

Raw scRNA-seq data were processed by Cell Ranger (v.3.1.0) software (10× Genomics) for read alignment, barcode assignment and unique molecular identifier (UMI) counting, based on genomes (mm10-3.0.0). Cells meeting criteria of UMI >200 and mitochondrial RNA <10% were retained. Subsequent analyses were performed using the Seurat package in R. Quality-controlled single-cell data were first integrated and subjected to principal component analysis (PCA) for linear dimensionality reduction. A K-nearest neighbor (KNN) graph was then constructed to refine cellular neighborhood relationships based on edge weights. Proximity distances were calculated using FindNeighbors, followed by iterative clustering with

FindClusters. Finally, the dataset was visualized through Uniform Manifold Approximation and Projection (UMAP) for non-linear dimensionality reduction. Cell clustering was conducted via FindClusters function using the top 30 principal components (PCs) to resolve cell identities. Cluster-specific marker genes were identified using the Wilcoxon rank-sum test (FindAllMarkers, default parameters) and annotated based on known markers and top-ranked genes

### **Pseudotime analysis**

Monocle2 was applied to illustrate the developmental epithelial cells trajectories using the UMI count matrix. Cluster-specific variable genes were employed to order cells along the pseudotemporal axis. The epithelial cells cluster 3 and 8 (E03 and E08) was defined as the root state.

### **CellPhoneDB analysis**

The analysis starts with input preparation, a single-cell expression matrix with cell-type annotations and gene symbols. Next, mean expression levels of ligands and receptors are calculated for each cell type. Finally, interaction scores are computed by multiplying the average ligand expression in a source cell type with the average receptor expression in a target cell type, yielding a quantitative measure of potential communication strength between cell populations.

## **Medicine and food homologous TCM screening for high-fructose exacerbated**

### **CAC**

FoodNTCM is the first and most comprehensive database in China that integrates fundamental information on the medicinal sources and uses of medicine–food homologous substances with their chemical and nutritional compositions (<http://101.200.131.49:8000/>). Users can browse or search the database based on personal preferences, either by the name of the medicinal substance (hereafter referred to as “herb”) or by the compound name. Future versions will support additional search options, including the source of the herb as well as its descriptions and functional annotations.

The current version of FoodNTCM (v1.0) integrates data from multiple authoritative sources, including FoodB, FoodDisNET, TASYL-TCM, PubChen, PubMed, the Chinese Pharmacopoeia (2020 edition), and the Chinese Food Composition Table (Standard Edition, 6th Edition). This database serves as a valuable reference and guiding resource for researchers studying medicine–food homologous substances in Traditional Chinese Medicine and for the development of related health products.

To identify targeted food with potential therapeutic effects against the key mechanisms of high-fructose exacerbated CAC, we systematically compiled data on medicine–food homologous TCM and their compound compositions from the literature and various databases. Additionally, we established an online database, FoodNTCM, to facilitate further exploration and application of these findings. In

total, 112 herbs and their related compound were documented. Before screening, the data was further processed by being annotated to PubChem database and TCM with at least one valid compound in PubChem were kept for further analysis. The compounds in these homologous were then transferred into chemical similarities between reference compounds in a network-based whole-genome drug-target prediction algorithm<sup>17</sup>. And the biological effect profile of every compound in each medicine-food homologous TCM was calculated by this algorithm. Further, according to our previous statistical strategy<sup>22</sup>, the holistic targets of each medicine-food homologous TCM were obtained (adjust P value < 0.05, Benjamini-Hochberg procedure).

With the four gene sets obtained from scRNA-seq of high-fructose exacerbated CAC and predicted holistic targets of 101 medicine and food homologous, the closeness between each gene set and each medicine-food homologous TCM was estimated on the hypergeometric distribution. The candidate medicine and food homologous for high-fructose-induced CAC was screened based on the significance and mapping count. The expression pattern of Kelp in scRNA-seq was measured by the mean expression of the holistic targets of Kelp in different cell types and visualized by VlnPlot function from Seurat package in R.

### **Key Compound Identification from Kelp**

The biological effect profiles of compounds in Kelp were kept for further analysis. In

total, 50 compounds were identified in Kelp according to our collection. And the druglikeness of these compounds were calculated by the implementation based on Python rdkit package<sup>23</sup>. 27 compounds with quantitative estimate of drug-likeness (QED) score no less than 0.3 were kept for further analysis. Additionally, the same as screening TCM for high-fructose exacerbated CAC, the closeness between each gene set and every compound in Kelp was also estimated on the hypergeometric distribution. The expression pattern correlations of these compounds and Kelp were estimated on the Spearman correlation based on the corresponding expression on each cell belonging to stromal cells or epithelial cells, visualized by ggplot2 package in R. The network similarity of two gene sets was implemented by GOSemSim package in R, with methods set as “Wang”.

### **Network target of Kelp in the treatment of high-fructose exacerbated CAC**

In order to find the key mechanism of Kelp in treating high-fructose-induced CAC, network target analysis was performed based on the holistic targets of Kelp and the four gene sets. The biological network composed of certain molecules in gene set A, B, C, D, as well as the holistic targets of Kelp and CRC-related biomolecules were constructed based on protein-protein interaction data, and were then visualized by Cytoscape. KEGG enrichment was conducted to find the significantly enriched pathways in either Kelp or the four gene sets, implemented by clusterProfiler package in R. Pathways significantly enriched by the holistic targets of Kelp were then

categorized into 4 modules, including Metabolism, Environmental Information Processing, Cellular Processes and Organismal Systems. On the other hand, 10 pathways were found to be significantly enriched by the molecules in all four gene sets, and 9 of them were potentially targeted by Kelp intervention (adjust P value < 0.05, Benjamini-Hochberg procedure). The pathways were visualized using the R package ggplot2, and the  $-\log_{10}$  of the enrichment significance (P value) was standardized for each gene set to observe more obvious differences.

The correlations of the expression of EGFR and EREG between Kelp, Fuc and dieckol were estimated on the Spearman correlation on each cell belonging to the whole detected cells, stromal cells or epithelial cells, respectively. And the expression levels of EREG and EGFR in scRNA-seq were measured and visualized by VlnPlot function from Seurat package in R, in different cell types.

### **H&E staining**

Colon tissues were fixed in 4% paraformaldehyde for 24 hours, followed by paraffin embedding and sectioning. Paraffin sections were deparaffinized in xylene after 90-minute incubation at 65°C, then rehydrated through descending concentrations of ethanol. Sections were rinsed in distilled water before staining with hematoxylin for 1 minute. Differentiation was performed using 1% hydrochloric acid-ethanol, followed by bluing in 1% ammonia water. After rinsing, sections were counterstained with 1%

eosin and mounted with neutral balsam for microscopic examination.

### **Immunohistochemistry**

Following deparaffinization and antigen retrieval, sections were permeabilized with 0.5% Triton X-100. Endogenous peroxidase activity was blocked, and non-specific binding was prevented using 3% goat serum. Primary antibodies were incubated at 4°C overnight, followed by species-specific secondary antibodies at room temperature for 2 hours. After PBS washes, DAB was applied, and the reaction was stopped with double-distilled water. Hematoxylin counterstaining was performed prior to mounting. All steps were optimized to ensure specificity and reproducibility.

### **Immunofluorescence staining**

Following antigen retrieval, tissue sections or cells were permeabilized with 0.5% Triton X-100 for 1 minute, then washed with PBS on a rocking platform. After blocking non-specific binding sites, species-specific primary antibodies were applied and incubated overnight at 4°C. On the following day, sections were allowed to equilibrate to room temperature for 30 minutes. Corresponding fluorescent secondary antibodies (selected based on host species of primary antibodies) were subsequently incubated in the dark for 2 h, followed by three PBS washes. DAPI working solution (1:5000 dilution) was applied for nuclear counterstaining at room temperature for 30 minutes, then removed by washing. Finally, sections were mounted with antifade

mounting medium to preserve fluorescence signals.

### **Microscale thermophoresis (MST)**

NIH-3T3 cells were transfected with GFP-CNBP plasmids or mutant plasmids and lysed 48 hours post-transfection. Dieckol, utilized as a ligand, was diluted twofold and combined with an equal volume of cell lysate to yield a final fluorescence intensity of 400. Binding assays were conducted using the Monolith NT.115 instrument, followed by quantitative analysis of protein-ligand interactions via MO.Affinity Analysis software.

### **ChIP assay**

NIH-3T3 cells were transfected with GFP-CNBP plasmids and, 48 hours post-transfection, were subjected to appropriate stimulatory conditions. Following transfection, chromatin immunoprecipitation was performed. Briefly, target proteins and DNA were crosslinked using 1% formaldehyde. Cells were collected, lysed, and genomic DNA was fragmented via sonication. Reverse crosslinking was then conducted to assess the size distribution of sheared DNA fragments. After optimizing sonication conditions, chromatin immunoprecipitation was performed using Protein A+G Agarose/Salmon Sperm DNA beads. Enriched DNA fragments were subsequently quantified by qPCR. Primers were designed based on the promoter sequence of the *Ereg* gene and chromatin accessible regions identified using the

Cistrome Data Browser, ensuring targeted amplification of regulatory elements.

### **Cellular thermal shift assay**

L929 cells were equally divided into two 10 cm dishes and cultured overnight. One dish was treated with dieckol (10 mM), while the other received 0.1% DMSO as a control. After 2 hours, cells were harvested, aliquotted into ten 200  $\mu$ L PCR tubes, and subjected to heat treatment in a thermocycler. Samples were centrifuged at 13,000  $\times$  g for 20 minutes, and analyzed by Western blotting to detect target proteins.

### **Plasmid DNA transfection**

NIH-3T3 cells were seeded into culture plates and transfected with GFP-CNBP, EREG plasmids, or corresponding vectors. All plasmids were sourced from Corruces Biotechnology. Lipofectamine™ 3000 were purchased from Thermo Fisher Scientific.

### **ELISA assay**

The protein level in mice serum was detected by an ELISA kit (ELK Biotechnology). Following initial incubation for 1.5 hours, the samples were washed, and the supernatant was discarded. The detection antibody was then added, followed by a secondary 1.5-hour incubation. After incubation with TMB substrate and stop solution, absorbance at 450 nm was measured to quantify EREG protein levels.

### **Cell counting kit-8 assay**

Cell viability was assessed using the Cell Counting Kit-8 (CCK-8, MCE) following the manufacturer's protocol. NIH-3T3 cells were seeded in culture plates for 12 hours to transfection with the indicated EREG plasmid. After 48 hours, varying concentrations of fructose or OSMI-1 were added to the culture medium and incubated for an additional 24 hours. The supernatant was collected, centrifuged at  $1000 \times g$  to remove cellular debris, and then incubated with MCC38 cells for 24 hours. CCK-8 solution was added to the reaction mixture, and the absorbance was measured at 450 nm using a microplate reader.

### **Dual Luciferase Reporter Gene Assay**

NIH-3T3 cells were seeded in a 96-well plate and transfected with the indicated plasmids or vectors. After 48 hours, 100  $\mu$ L of cell lysis buffer was added to each well. Following supernatant collection, 100  $\mu$ L of firefly Luciferase substrate was added to measure relative light units (RLU). Subsequently, 100  $\mu$ L of Renilla Luciferase detection reagent was added to determine RLU. Using Renilla Luciferase activity as control, the RLU value from firefly Luciferase was normalized to that of Renilla Luciferase. The resulting ratio was used to compare the activation levels of the reporter gene across different samples.

## **TRAP analysis**

To identify dieckol-binding proteins in L929 cells, we implemented the Target-Responsive Accessibility Profiling (TRAP) approach. Two independent cultures of L929 cells were treated with either 50 nM TNF $\alpha$  alone or 10  $\mu$ M dieckol combined with 50 nM TNF $\alpha$ . Following a 2-hour incubation period, cellular membranes were permeabilized using M-PER buffer. The resulting lysates were subjected to covalent labeling via the addition of formaldehyde and borane pyridine complex, which specifically target lysine residues in proteins under ambient conditions for accessibility profiling. Detailed methodological protocols are provided in previously published literature.

## **EREG degradation analysis**

L929 cells were cultured in serum-free medium supplemented with 50  $\mu$ g/mL cycloheximide for 1 hour. Subsequently, cells were treated with 10 mM fructose and incubated for 3, 6, or 9 hours. Following incubation, cells were lysed, and total protein extracts were subjected to Western blotting analysis to assess target protein expression levels.

## **Statistical analysis**

The data represent the mean  $\pm$  SEM, Unpaired two-tailed Student's t tests and One-

way analysis of variance (ANOVA) with Tukey multiple comparison tests were used to assess statistical significance. Statistical significance was set at \* $P < 0.05$ , \*\* $P < 0.01$ , ns, not significant.