SUPPLEMENTARY INFORMATION

MATERIALS AND METHODS

Histological diagnosis

Histological diagnoses were established through liver biopsy specimens processed using standard procedures, embedded in paraffin blocks, and stained with H&E and Masson's trichrome stains. A seasoned hepatopathologist evaluated all liver biopsies at a single instance using the MASH Clinical Research Network histological scoring system. Hepatic fibrosis was graded on a 5-point scale (0–4) according to the Kleiner classification, with advanced fibrosis defined as a score of 3–4. Hepatic steatosis and lobular inflammation were assessed on a 4-point scale (0–3), while hepatic ballooning was scored on a 3-point scale (0–2). Finally, we categorized samples into steatosis, characterized by steatosis (>= 5%), with or without lobular and portal inflammation, and MASH, which met all diagnostic criteria, including steatosis, hepatocellular ballooning, and lobular inflammation.[1]

Genomic and epigenomic analysis

After trimming adapter sequences and removing low-quality reads using Trim Galore (version 0.6.6), the remaining clean read pairs were aligned to the human reference genome (version hg19) using theBurrows-Wheeler Aligner MEM (version 0.7.17).[2] The mapped files were then transformed into pileupfiles using SAMtools (version 1.10).[3] After eliminating duplicate reads using Picard tools, we performeda somatic variation calling analysis using the Genome Analysis Tool Kit (GATK) Mutect2 tool (version 4.1.9) with the default options.[4,5] We applied Panel of Normal (PoN) to eliminate common variations ascalled by Mutect2 (tumor-only mode). For this step, we used additional WES and WGS data from blood.SNVs and indels from the Korean Reference Genome Database (KRGDB) were regarded as commonpopular variations.[6] We called germline variations using HaplotypeCaller (GATK, version 4.1.2.0).[5] To annotate these variations with gene names and other information, we used table_annovar from the Annovar toolkit.[7] Then maftools used to further analyze and visualize the variation calling result.[8] ForWGBS analysis, we used Bismark (version 0.20.0) to align reads with Bowtie2 (version 2.3.5.1) and eliminated duplicate reads using sambamba (version 0.6.8).[9-11] Then, we used the bismark methylationextractor to calculate the ratio of methylated cytosine residues. The results were analyzed with the R packages BSseq for smoothing and the DMRseq library for calling DMRs.[12,13]

Transcriptomic data analysis

We used RSEM (version 1.3.1) with the STAR aligner (version 2.6.1c) to map total RNA-seq data to thehuman reference genome (version hg19) and profile gene expression levels.[14,15] Then, we used the Bioconductor package DESeq2 to normalize the data set and identify DEGs.[16] DEGs were defined asgenes with over 1.3-fold expression change (FC), a *P*-value of less than 0.05, an average of value over1 in at least one group, and an average in each group less than the standard deviation of each group. We downloaded MASLD (GSE135251) and liver cancer (GSE77314) RNA-seq data sets for a more comprehensive analysis.[17,18] We processed the raw data via the aforementioned methods and removedunwanted variations using the Bioconductor package RUVseq, which is designed to correct for batch effects.[19] This allowed us to avoid artifacts arising from our use of different sized sequencing data sets.Then, we used DESeq2 to normalize gene expression levels by disease stage. Gene ontology (GO) analyses and protein-protein interaction (PPI) analyses were conducted using Metascape.[20]

Machine learning

We used a support vector machine (SVM) algorithm to perform feature selection from the gene expression data. For this, we used the normalized read counts from the RNA-seq data as inputs. First,we filtered out genes with expression levels under 1 and then scaled the expression levels of the remaining genes with robust scaler. After scaling, absolute values over 1.1 were regarded as extreme outliers and assigned to either 1.1 or –1.1 depending on their direction, to exclude their influence. Afterthat, we performed a feature elimination step. The worst features had the lowest correlations in the linear SVM model. Features were eliminated repetitively until less than 1,500 remained. Then, the remaining features were used to find the best parameters

 https://doi.org/10.3350/cmh.2023.0449

with the grid search method for a radical basisfunction (RBF) with the SVM model through cross-over validation. For this step, we modified python code published by Vabalas and Gowen.[21] Finally, we performed a principal component analysis (PCA)using R and the prcomp library. The results were visualized in a 3D plot using the R scatterplot3d library.

Open chromatin accessibility analysis

We used published ATAC-seq data from the Gene Expression Omnibus database (PRJNA725028).[22] To analyze ATAC-seq data, we used the Bowtie2 (version 2.3.5.1) aligner to map the sequencing reads to the human reference genome (version hg19) and eliminated duplicated reads with Picard tools (version 0.6.8).[4,10] Genome-wide chromatin accessibility regions in steatosis and MASH samples werecalled using MACS2 (version 2.2.7.1).[23] Then, we calculated the density of open chromatin regions insteatosis and MASH samples using the Homer package annotatePeaks and visualized the results withIGV (version 2.9.4).[24,25]

Statistics

The independent t-test was conducted to compare the means of two different groups such as steatosisand MASH. Statistical analyses were conducted using R (Welch's *t*-test, Student's *t*-test) and Prism (Student's *t*-test). R packages include a *t*-test function, this function takes such as normalized expression counts and calculate *P*-value according to the significant difference between two groups. The variance of the samples in each group was considered when the *P*-value was calculated by *t*-test analysis in R. If the *P*-value was less than 0.05, it was considered as a significantly difference.

HFD mouse model and H&E, PAS staining

Animal studies were approved by the IACUC at Sookmyung women's university. Mice were randomly assigned into the following two groups: 1) Low fat diet (LFD, n=3); 2) High fat diet (HFD, n=3). Fatty liver disease was induced by feeding male C57/BL6J mice with HFD. The HFD contains 60% kcal fromfat (D12492, Research Diet, Inc.) while the LFD contained 18% kcal from fat (2018S, Envigo, Inc.). Micewere fed *ad libitum* for 10weeks. MASLD was characterized by the existence of hepatic steatosis in liver tissues. After obtaining liver tissues from mice fed a HFD or LFD, we fixed the tissues in 10% formalin and dehydrated them in ethanol. After cutting paraffin sections, we stained them with H&E and with PAS.

Hepatocyte organoid culture

Using the method of Huch et al, we isolated primary hepatocytes from wild type liver tissue.[26] Then, we filtered the isolated hepatocytes through a 70 μm filter, washed them twice with cold AdDMEM/F12,counted them, and then mixed them with matrigel in 24-well plates. Finally, we established organoids in expansion medium consisting of AdDMEM/F12 plus 15% RSPO1 conditioned medium, B27 (minus vitamin A), 50 ng/mL EGF, 1.25 mM N-acetylcysteine, 10 nM gastrin, 3 μm CHIR99021, 25 ng/mL HGF,50 ng/mL FGF7, 50 ng/mL FGF10, 1 μM A83-01, 10 mM Nicotinamide, and 10 μM Rho Inhibitor γ-27632.

Free Fatty Acid (FFA) treatment and Oil Red O staining

We treated hepatic organoids with 2 mM oleic acid and 1 mM palmitic acid. After a 48-hour incubation,we washed the cells two times with PBS and fixed them with 4% paraformaldehyde for 15 minutes. After fixation, we again washed the cells two times with PBS and stained them with Oil Red O solutionfor 20 min at room temperature. Finally, we washed the stained cells with distilled water to remove anyunbound Oil Red O. Staining area of the cells were measured and quantified using Image J software version 1.53k (National Institute of Health, USA).

Triglyceride assay

For organoid TG analysis, cells were washed with cold PBS, homogenized in 5% NP-40 in DW. The TGcontents in each sample were estimated using TG assay kit (Abcam) in accordance with the manufacturer's instructions.

Quantitative real time PCR (qRT-PCR)

We extracted RNA from the livers of mice fed on an HFD or LFD using a NucleoSpin RNA/Protein kit (MACHEREY-NAGEL) and from organoids using a Hybrid-R kit (GeneAll Biotechnology) according to the manufacturer's instructions. 500 ng–1 μg of RNAs were reverse transcribed using M-MLV ReverseTranscriptase (Promega). We performed a qRT-PCR on a LightCycler 96 Real-Time PCR System (Roche Life Science) using qPCRBIO SyGreen Blue Mix (PCR Biosystems). The cycling conditions included an preincubation at 95℃ for 600 s followed by 45 cycles at 95℃ for 10 s, 60℃ for 10 s and 72℃ for 10 s, with a melting stage of 95℃ for 10 s, 65℃ for 60 s and 97℃ for 1 s. Expression levels were determined using the $2^{-\Delta\Delta Ct}$ method for a relative quantification of the results. We then normalized theserelative expression results to a reference gene (i.e., the 18s rRNA internal control). The relevant primersare listed in Supplementary Table 8.

 https://doi.org/10.3350/cmh.2023.0449