Identification of Alcoholic Hepatitis-related and Mesenchymal Stem Cell Therapy Target Genes

Seul Ki Han  
Yonsei University Wonju College of Medicine

Taesic Lee  
Yonsei University Wonju College of Medicine

Jisun Lim  
Pharmicell Co., Ltd

Hyunsoo Kim  
Pharmicell Co., Ltd

Young Uh  
u931018@yonsei.ac.kr

Yonsei University Wonju College of Medicine

Moon Young Kim  
Yonsei University Wonju College of Medicine

Research Article

Keywords:

Posted Date: April 19th, 2024

DOI: https://doi.org/10.21203/rs.3.rs-4245638/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

Alcoholic hepatitis (AH) is a widespread and life-threatening chronic liver condition that poses a risk of short-term mortality if not properly managed. Clinicians often encounter challenges due to insufficient knowledge about the underlying mechanisms of AH. This study employs a meta-analysis to identify the molecular mechanisms and potential cell therapy targets for AH.

Methods

We collected four gene expression datasets, three from liver tissues and one from blood tissues, to identify genes associated with AH. Two liver datasets that had data on deaths after steroid treatment in patients with alcoholic hepatitis were also examined to uncover signatures associated with poor prognosis. Additionally, we curated three cohorts, including a mesenchymal stem cell (MSC) intervention group, to identify genes responsive to stem cell interventions. Candidate genes were selected using the inverse weighted variance-based method implemented in the METAL software. We utilized prior knowledge to narrow down potential upstream genes, including a transcription factor (TF) catalog, protein-protein interaction (PPI) networks, disease-gene association databases, and summary statistics for single nucleotide polymorphisms (SNP) linked to disease and expression.

Results

Through four stepwise meta-analyses of nine gene expression datasets, we identified genes targeted by MSC therapy. In detail, the first, second, third, and fourth steps of meta-analysis provided the liver-specific, liver-blood, severe-mortality, and MSC-Tx meta genes linked to AH condition, respectively. Multiple lines of evidence (TF, PPI, and SNP databases) were used to identify 47 AH-related upstream genes.

Conclusions

This study presented critical genes involved in the progression of AH and the therapeutic effects of MSC through meta-analysis. Utilizing these genes, we can confirm genetic changes induced by stem cell treatment, providing a foundation for targeted cell or function-enhanced genetic therapies.

Introduction

Alcoholic hepatitis (AH) is a life-threatening condition characterized by a necroinflammatory process leading to cirrhosis and liver failure induced by alcohol use disorder [1]. In high-risk patients, the 90-day mortality rate can reach up to 50% (as observed in the STOPHA trial) [2]. Therefore, urgent treatment decisions, such as steroids or liver transplantation, are often necessary. For decades, steroid treatment has been the primary treatment for severe AH, demonstrating improved 28-day survival in carefully indicated patients [3]. However, steroid therapy was related to a high incidence of infection and adverse
events [4]. Moreover, patients not eligible for steroid therapy have limited alternatives, primarily liver transplantation [4].

Chronic alcohol consumption induces intestinal dysbiosis, leading to an increase in pathogen-associated molecular patterns (PAMP) [5]. These, in turn, activate the recruitment of inflammatory cells—Kupffer cells, macrophages, and neutrophils—and the production of inflammatory cytokines through pathogen-recognition receptors like Toll-like receptors (TLRs) and nucleotide-binding oligomerization domain-like receptors (NLRs) [6]. Other pivotal mechanisms in alcoholic liver disease (ALD) include hepatocellular injury and death [7], metabolic reprogramming [8], inflammation [9], oxidative stress [10], and a loss of liver regeneration [11]. Given the complex mechanisms of liver disease, current preclinical animal models struggle to replicate severe AH entirely [12].

To address the complex mechanisms of ALD and vulnerability to its development, genome-wide association studies (GWAS) have identified critical single nucleotides (SNP) associated with alcohol-related liver cirrhosis, such as rs738409 (PNPLA3), rs10401969 (SUGP1), rs58542926 (TM6SF2), and rs626283 (MBOAT7) [13]. However, GWAS faces limitations due to the missing heritability issue. SNPs provide only partial information about complex phenotypes and the topological challenge of many SNPs residing in non-coding regions [14]. To unbiasedly identify signatures associated with AH, liver gene expression profiling identified approximately 200 differentially expressed genes (DEGs), which were enriched in several biological pathways, including "cytokine-cytokine receptor interaction" [15]. A recent integrative study of the hepatic transcriptome and metabolome pinpointed extensive dysregulation of glucose metabolism, explicitly proposing the hexokinase domain containing 1 (HKDC1) as an essential gene related to AH [16].

Research involving genetic data from human tissues or animal experiments demands substantial effort and cost for sample collection. Individual studies were rarely able to attain the necessary sample size for generalized results, especially when faced with potential incongruent outcomes [12]. Meta-analyses are commonly employed to mitigate the lack of replication in small, heterogeneous datasets and to derive generalized findings for targeted phenotypes, particularly in GWAS studies [13]. While meta-analyses are prevalent in the GWAS domain, they are less frequently applied in transcriptomic analyses, especially in gene expression datasets profiling patients with ALD.

This study aims to identify AH-related genes and cell therapy targets based on a four-stage consecutive meta-analysis. First, a multidisciplinary team, including hepatologists, data scientists, and database administrators, conducted literature and database reviews and collected liver and blood AH-related transcriptomic datasets (Fig. 1). Using a meta-analysis approach, we identified genes with convergent patterns of change for the three lists summarizing AH-associated transcriptome changes in liver tissues as AH liver meta genes. In the second and third phases of meta-analysis, AH blood, and prognosis-related datasets were pooled to discern liver-blood AH and mortality-severe meta genes, respectively (Fig. 1). In a subsequent fourth meta-analysis step, multi-tissue gene expression datasets, comprising mesenchymal stem cell (MSC) therapy (MSC-Tx) for infected humans and mice, were retrieved and
processed for pooled analysis to narrow down the MSC-Tx target meta-genes. In the final step, various lines of evidence, including the transcription factor (TF) database [14], disease-gene network, AH-related SNPs, liver cis-expression quantitative trait loci (eQTL), and protein-protein interaction (PPI) network, were utilized to provide potential upstream AH-related genes (Fig. 1).

**Methods**

**Collection of AH-related transcriptome datasets**

To obtain the meta-analyzed AH-related signatures, we collected three liver AH transcriptome datasets (GSE142530, GSE167308, and GSE28619), along with a blood AH dataset (GSE135285). Additionally, we compiled two liver gene expression datasets (GSE94397 and GSE94399) to identify pooled signatures associated with the poor prognosis of AH. To identify robust biomarkers related to MSC intervention, we curated three cohorts, each comprising subjects treated with stem cells (GSE40180, GSE108685, and GSE121970). GSE108685 profiled the genome-wide transcriptome of human blood, GSE40180 measured gene expression in various murine tissues (liver, spleen, lung, kidney, and heart), and GSE121970 included mouse lung mRNA profiling.

The RNA expression levels in GSE28619, GSE94397, GSE94399, GSE40180, GSE108685, and GSE121970 were profiled using microarrays. GSE94397 and GSE94399 underwent scaling and normalization using log-transformation and quantile methods, respectively. Other microarray-based transcriptomic datasets had already undergone normalization and were directly subjected to differential expression (DE) analysis. The gene expression levels of GSE142530, GSE167308, and GSE135285 were profiled using RNA-sequencing platforms, and the gene-specific counts in each RNA-seq dataset were transferred for DE analysis.

To systematically compare the compiled gene expression datasets, we unified the transcript IDs in each dataset using Entrez Gene [15]. Chip-specific transcripts, probes, and probe-set IDs were contained in the microarray datasets, while Ensembl IDs were implemented for all genes in the RNA-seq datasets. We systematically annotated them into Entrez ID using the org.Hs.eg.db, org.Mm.eg.db, and Orthology.eg.db packages to perform the meta-analysis.

**Differential expression analysis**

Differential expression (DE) analyses of microarrays and RNA-seq were performed to compare specific phenotypes (e.g., AH and death) with their matched controls. The limma and DESeq2 methods were employed for microarray and RNA-seq data, respectively. Both DE tools provided gene-specific fold changes (FC) and matched p-values for comparing disease and control samples. Genes with a false discovery rate (FDR)-adjusted p-value of less than 0.5 were identified as differentially expressed genes (DEGs) between the two conditions.
A hypergeometric test was applied to identify enriched biological pathways in the candidate gene sets. Enrichment analyses were carried out using the Gene Ontology (GO) [16] and Kyoto Encyclopedia of Genes and Genomes (KEGG) [17] databases. Biological pathways were considered enriched when they met the criteria of a 0.5 FDR-adjusted p-value determined by the hypergeometric test.

Meta-analysis

We conducted a meta-analysis to identify biomarkers commonly associated with diseases in the collected candidate gene expression datasets. The meta-analysis utilized a fixed-effect inverse-variance-weighted (IVW) method adopted in METAL software [18]. The IVW method requires an effect similar to the FC between the two conditions and a matched standard error. We selected gene-specific FC between the two conditions for each transcriptomic dataset and matched the standard errors obtained from summary statistics calculated using limma or DESeq2 in the previous step. The generated METAL method execution commands based on DOS prompts are presented as example combinations for specific gene expression datasets (Figure S1). The significance level for associations calculated through meta-analysis was determined by FDR adjustment for multiple comparison test.

Implementation of Prior Knowledge for Identifying Upstream Biomarkers

Hägg et al. [19] curated gene sets involved in transcription activities such as transcription activator activity, transcription coactivator activity, and TF binding from the GO database [16] to identify upstream genes related to the coronary artery. Lee and Lee [20] used a TF-related gene set obtained from TRANSFACT [21] to identify disease-related genes. Recently, Lambert et al. [14] proposed a TF catalog integrating prominent TF databases, and Lee and Lee [22] used it to narrow down the potential gene sets correlated to comorbidity. Motivated by these studies, we implemented the TF catalog manually updated by Lambert et al. [14] to identify the upstream genes related to AH (Fig. 1).

We implemented the DigSee database by curating the relationships of approximately 4500 disease types with about 13000 genes by integrating text mining and machine learning methods [23, 24]. Using a “Liver Diseases, Alcoholic” query, approximately 500 ALD-related genes were compiled from the DigSee.

A GWAS was conducted on alcohol-related cirrhosis in separate German and UK cohorts (712 cases and 1,426 controls) [25]. Full summary statistics of the alcohol-related cirrhosis GWAS were downloaded from the supplementary website (http://gengastro.med.tu-dresden.de/suppl/alc_cirrhosis/) [25]. Among the 6,502,449 SNPs, approximately 5500 variants with uncorrected p-values < 0.001 for the association between genetic variants and alcohol-related cirrhosis were selected. The 5500 SNPs were assigned to their corresponding genes using the Gsnpense function in the gprofiler2 package [26]. Among the candidate AH genes, those with evidence from GWAS were selected as potential biomarkers of AH (Fig. 1).

Liver cis-eQTL data were obtained from the Genotype-Tissue Expression (GTEx) project [27]. In detail, full summary statistics of the GTEx liver eQTL were obtained with an access ID (study ID: QTS000015;
dataset ID: QTD000266) from the eQTL catalog database [28]. Cis-associations between gene expression and variant types with an uncorrected p-value < $10^{-5}$ were selected, yielding 317,324 gene-SNP pairs that accounted for 2829 genes (annotated by symbols assigned by HGNC). Among the candidates the meta-analysis chose, genes showing evidence of liver cis-eQTL were designated candidate upstream genes (Fig. 1).

A PPI network was obtained from the STRING database, which collected the interactome from various sources, including automated text mining of scientific and medical literature, computational predictions of co-expression and co-occurrence across genomes, PPI experiment databases, and known biological pathways [29]. STRING consists of approximately 12 million edges of 40,000 proteins (based on the Ensembl Protein (ESPN)). The PPI network introduced by the STRING database included protein-protein pairs and their matched scores. The distribution of PPI scores did not follow a Gaussian distribution; therefore, PPIs with the top 90 percent of the interactome scores were selected. Among significant genes in meta-analysis, genes with $\geq 200$ edges were chosen as the candidate upstream genes (Fig. 1).

**Results**

**Whole transcriptomic signatures of alcoholic hepatitis liver tissues.**

Hierarchical clustering and dimensionality reduction (principal component analysis, PCA) were applied to genes exhibiting the top 20 percent variance values across samples in GSE142530 (liver transcriptome). Clustering and PCA led to a distinct separation between ALD and control samples, as depicted in Figure S2. In the additional liver datasets, GSE167308 and GSE28619, hierarchical clustering and PCA revealed a noticeable classification between ALD and control groups (Figures S3 and 4).

Based on an FDR-adjusted p-value threshold of 0.05, approximately 6000, 6500, and 6500 DEGs were selected between the AH and control groups in the three liver transcriptomic datasets (Figs. 2A and S5). The dysregulated genes in GSE142530 maintained a balance between upregulated and downregulated genes. Besides, the DEGs in GSE167308 and GSE28619 were biased towards upregulation. We applied a hypergeometric test-based enrichment analysis for each DEG list from the three datasets. Three DEG lists in the three liver AH gene expression datasets were commonly estimated to be involved in mitochondria-related pathways (Fig. 2B). Among the three transcriptomic datasets, the DEGs in the two liver sets participated in various biological pathways, including mitochondrial protein-containing complex (GO:0098798), Fatty acid degradation (hsa00071), Oxidative phosphorylation (hsa00190), Citrate cycle (TCA cycle) (hsa00020), Lipid oxidation (GO:0034440), Inner mitochondrial membrane protein complex (GO:0098800), and PPAR signaling pathway (hsa03320). As is widely known, factors related to mitochondrial, glucose, and fatty acid metabolism are dysregulated in ALD samples (Fig. 2B).
Three liver gene expression datasets were meta-analyzed to identify generalized liver genes associated with AH status. Specifically, the summary statistics from the three gene expression datasets, calculated using the limma and DESeq2 algorithms [30, 31], were integrated using the Inverse Variance Weighting (IVW) method in METAL [18]. This analysis identified approximately 8600 liver genes related to AH, based on an FDR-adjusted p-value < 0.05, and these genes were annotated as liver AH meta-genes (Fig. 3). Additionally, seven genes that were not identified as DEGs in any of the three liver datasets were discovered through the meta-analysis (Fig. 3).

The liver AH meta-genes were found to be associated with cancer pathways ((hsa05200) Pathways in cancer, (hsa05210) Colorectal cancer, (hsa05211) Renal cell carcinoma, and (hsa05212) Pancreatic cancer), aging-related pathways ((GO:0090672) telomerase RNA localization and (hsa05016) Huntington disease), metabolism ((hsa00020) Citrate cycle (TCA cycle), (hsa03320) PPAR signaling pathway, and (GO:0034440) lipid oxidation), as well as mitochondrial-related pathways ((GO:0009055) electron transfer activity, (GO:0004033) aldo-keto reductase (NADP) activity, and (GO:0001836) release of cytochrome c from mitochondria) (Fig. 3).

We integrated meta-z-scores by combining three liver AH summary statistics with blood AH summary statistics obtained from GSE135285. We identified approximately 4300 DEGs with an FDR-adjusted p-value < 0.05, categorizing them as liver-blood AH meta-genes (Fig. 4). While most liver-blood AH meta-genes originated from liver AH meta-genes, 111 genes were from the blood AH dataset (Fig. 4). Furthermore, 370 genes were not observed in individual liver and blood datasets within the liver-blood meta-genes but were newly identified in the meta-analysis (Fig. 4).

For the two liver AH prognosis-related gene expression datasets (GSE94397 and GSE94399), we applied the Limma method, curating two lists of FC values and matched standard errors for all genes in the form of summary statistics. Using IVW-based meta-analysis, we integrated three lists of summary statistics for all genes, including liver-blood AH meta-genes, and two liver summary statistics encompassing AH-related death signatures. The IVW method identified approximately 1800 genes related to AH and death due to AH after steroid treatment (Fig. 5). We defined these about 1800 genes as mortality-severe AH meta-genes, primarily involved in macro- and micro-nutrient metabolism, diabetes-related, and mitochondria-related pathways (Fig. 5).

**Meta-analysis of AH and MSC therapy-related transcriptomic signatures.**

We collected three genome-wide transcriptomic datasets to investigate the effects of MSC-Tx: GSE108685 (human blood); GSE40180 (murine spleen, liver, lung, kidney, and heart tissues); and GSE121970 (mouse lung). Experimental subjects, both humans and mice in GSE108685 [36], GSE40180 [37], and GSE121970 [38], were induced into an infected state through intravenous injection of lipopolysaccharide, cecal ligation and puncture, and inoculation of Klebsiella pneumoniae via the airway, respectively. Using the limma method, we curated three lists of FC and paired standard error (SE) values for all genes between the MSC-Tx and matched control groups.
We integrated the mortality-severe meta-genes (liver + blood + death datasets) and the three MSC-related summary statistics (FC between MSC treated group and no intervention group and its standard error) using the IVW method in METAL. 719 genes were identified as MSC-Tx target meta-genes, exhibiting opposite altered patterns (up- and downregulation) between AH and MSC therapy (Fig. 6). For instance, NFE2L2 was the downregulated severe AH meta-gene and the upregulated gene in the MSC-Tx group of GSE40180, GSE108685, and GSE121970 (Figs. 6 and 7). The 719 MSC-Tx target meta-genes were implicated in numerous biological pathways (Fig. 6), with 18 involved in the telomere maintenance pathway.

**External Biological Evidence for Identifying Upstream AH-related Genes**

To identify regulators or upstream genes, computational studies have implemented TF, disease-gene networks, GWAS, eQTL, and PPI databases [19, 20, 22, 32–34]. Taking advantage of these frameworks, prior knowledge, including the TF catalog [14], disease-gene relationship database [23, 24], full summary statistics of GWAS [25], liver eQTL [27], and protein interactome [29] was implemented to uncover the upstream AH-related biomarkers. As a result, 47 genes (PSMC5, SPHK1, CEBPA, AKT1, KLF11, NFIA, NR3C2, NFE2L2, DDIT3, LGALS9, MED24, PGS1, PLXNA3, ZRSR2, PFDN1, STAT5B, SIRT1, SUMO1, MAPK1, LYAR, NPM1, BLM, PCNA, DDB1, PSMD11, FLNA, RBBP7, NOP2, RPL7A, PA2G4, DNMT1, MYOM2, ADH1C, RABEP1, SOD1, GRB2, SHC1, CASP3, CASP8, CYCS, PSMC2, FTSJ3, POLR1E, ALDH18A1, GNAI3, TYMS, DHX15) had one or more of external evidence (Fig. 7), therefore were selected as the candidate upstream genes.

**Discussion**

We identified three lists of meta-genes related to AH, designated as liver AH (Fig. 3), liver blood AH (Fig. 4), and mortality-severe meta-genes (Fig. 5), using a pooled approach. Subsequently, gene expression datasets associated with MSC therapy were utilized to generate candidate lists of genes targeted by MSC therapy, annotated as MSC-Tx target meta-genes (Fig. 6). Employing validated and globally recognized databases, such as the TF catalog [14], disease-gene network [23, 24], and protein interactome [29], we selected candidate upstream genes crucial in AH that could potentially serve as therapeutic targets for MSCs (Fig. 7).

The analysis included some genes not previously characterized in the context of ALD. For instance, SPHK1 (sphingosine kinase 1) plays a role in liver fibrosis. In patients with cirrhosis, SPHK1 levels significantly increase in fibrotic compared to normal human livers [42]. Additionally, it mediates CCL2 secretion and activates Kupffer cells and hepatic stellate cells in mouse models [42]. In our study, SPHK1 expression was downregulated in AH and upregulated after stem cell therapy. These changes suggest that SPHK1 may be a target mechanism in stem cell therapy for AH.

The gene AKT1 is on chromosome 14 and exhibits ubiquitous expression across various tissues, including the arteries, esophagus, lung, colon, adrenal gland, prostate, and cervix [31]. Recent
investigations highlight the significant involvement of the phosphatidylinositol 3 kinase (PI3K)/Akt pathway in the pathophysiology of acute ethanol-induced fatty liver in mice [43]. In a murine model, the inhibition of both AKT1 and AKT2 demonstrated a restraining effect on alcohol-induced fibrosis progression [44]. The impact of AKT1 inhibitors on ALD was corroborated in an animal model. Although the specific influence of AKT1 inhibitors on ALD remains unclear, our observations indicating the upregulation of AKT1 in the AH group and its downregulation in the MSC-Tx group could provide preliminary support for future investigations aimed at developing AKT1 inhibitors as a potential therapeutic strategy for AH.

LGALS9, also known as Galectin-9, belongs to the galectin protein family, comprising fifteen members identified in mammals, with 11 proteins present in humans. Involved extensively in immune modulation and tumor pathogenesis, any mutation or dysregulation of LGALS9 is linked to the prognosis of various cancers. Elevated serum levels of galectin-9 positively correlate with the risk of autoimmune liver disease and hepatocellular carcinoma, underscoring the connection between fibrosis and LGALS9 [45]. LGALS9 has recently emerged as a promising immunotherapy target [46, 47]. However, the association between LGALS9 and ALD has only been substantiated through SNP-based studies [48].

MAPK1, a mitogen-activated protein (MAP) kinase family member, is typically expressed at lower levels in liver tissue [27]. Stress responses, such as obesity, diabetes, and metabolic dysfunction-associated steatotic liver disease, activate hepatic MAPK, subsequently impairing the insulin signaling pathway and lipid metabolism [35]. MAPK inhibitors, or drugs that regulate the MAPK signaling pathway, have been developed to treat various cancers, including lung cancer, breast cancer, renal cell carcinoma, gastric tumors, meningiomas, and pancreatic tumors [36]. However, specific inhibitors or drugs targeting MAPK1 have not yet been developed. This study presents MAPK1 results obtained through meta-analysis and external evidence, suggesting its potential as a target gene for MSC therapy.

This study employed a multidisciplinary approach to meticulously screen candidate genetic data, followed by their integration and analysis. We aimed to minimize potential biases within each cohort by utilizing four datasets based on statistical methods. A notable strength of our research lies in the combination of blood-derived gene data, a strategy employed to overcome the limitations associated with liver tissue analysis. Previous studies have suggested the utility of genomic indicators from blood as biomarkers for predicting liver disease [37]. Consistent with a prior study, blood-based diagnostic gene biomarkers derived from peripheral blood mononuclear cells (PBMC) demonstrated a 90% overall accuracy in differentiating AH from alcoholic cirrhosis [38]. Consequently, we incorporated this hypothesis into our study, leveraging blood-derived samples to reflect liver disease and its state. Furthermore, genes identified through this analysis underwent cross-analysis with datasets related to prognosis, enabling the selection of genes capable of predicting AH prognosis or monitoring treatment response in clinical settings. This allows for establishing a model predicting the response to AH treatment or providing a targeted mechanism for its management. Finally, we conducted a meta-analysis of genes associated with AH and genetic changes during MSC therapy, confirming the genetic responses
related to MSC therapy in AH. These candidate genes present potential targets for enhancing stem cell treatment functions.

In conclusion, our meta-analysis aimed to identify therapeutic target genes for patients with AH undergoing MSC therapy. Additionally, by incorporating multiple lines of external evidence as prior knowledge, we estimated upstream markers of AH and its therapeutic targets.

Declarations

Conflict of Interest Statement

All authors have no conflicts of interest to declare.

Author Contribution

All authors reviewed the manuscript. S.H. and T.L. wrote the main manuscript text. S.H. and T.L. conducted all bioinformatics analyses. J.L. and H.K. reviewed and collected candidate gene expression datasets. M.K. and Y.Uh. reviewed and supervised all works of the current study.

References


**Figures**
Figure 1

Schematic representation outlining the present study to identify the AH-related genes.

Step 1: Three rounds of meta-analysis were conducted to screen for AH meta genes. In the first round, three liver gene expression datasets related to AH were meta-analyzed using the IVW in METAL [18]. In the subsequent rounds, the meta-analyzed statistics and blood AH-related signatures were combined using the IVW. In the final round in the first step, two liver gene expression datasets, including information on death after receiving steroids for AH, were pooled to identify mortality-severe AH meta-
genes. Step 2: Multi-tissue transcriptomic datasets, including human and mouse sepsis models treated with MSC, were combined by meta-analysis to select MSC-Tx target meta-genes. Step 3: We utilized multiple lines of evidence, such as TF catalogue, disease-gene association, and SNP summary statistics to narrow down the candidate upregulators for AH-MSC Tx genes.

Abbreviations: AH, alcoholic hepatitis; GSE, gene expression data series; MSC, mesenchymal stem cell; Tx, therapy; GWAS, genome-wide association study; eQTL, expression quantitative trait loci; PPI protein-protein interaction; AH, alcoholic hepatitis; Tx, therapy; SNP, single nucleotide polymorphism; IVW, inverse variance-weighted average method.

Figure 2

Hepatic transcriptomic alterations in AH patients.
(A) DEGs between AH patients and matched controls were selected based on FDR-adjusted p-value < 0.05. The DE analyses of RNA-seq (GSE142530 and GSE167308) and microarray (GSE28619) were conducted using DEseq2 [31] and Limma [30] methods, respectively. (B) Pathway enrichment analysis was conducted using AH-DEGs as candidate genes, GO and KEGG pathways as a library, and hypergeometric test as an assessment method. Only gene sets with an FDR < 5% are represented. Numbers in each cell indicate common genes between the AH-DEGs and genes annotated in each pathway.

Abbreviations: AH, alcoholic hepatitis; GSE, gene expression data series; DEG, differentially expressed gene; GO, gene ontology database; has, homo sapiens.
Figure 3

Meta-analysis to identify the hepatic AH-related genes.

(A) The first row indicates the z-score calculated by IVW in METAL [18]. The z-scores are meta-analyzed FC values among those in GSE142530, GSE167308, and GSE28619. The second through fourth rows show the binomial distribution (increase or decrease) of AH-DEGs. Only genes with FDR-adjusted p-value < 0.05 are represented in the second through fourth rows. (B) Venn diagram displays the intersections of a list of meta-analyzed AH-related genes and three AH-DEGs lists from each gene expression dataset.
(C) Only gene sets with an FDR < 5% in pathway enrichment analysis are represented. Numbers in each cell indicate common genes between the AH-DEGs and genes annotated in each pathway.

Abbreviations: AH, alcoholic hepatitis; GSE, gene expression data series; GO, gene ontology database; has, homo sapiens; IVW, inverse variance-weighted average method; FC, fold change; DEG, differentially expressed gene; FDR, false discovery rate.

Figure 4

Meta-analysis to identify the AH-related genes in liver and blood tissues.
(A) The first and second rows indicate the z-score calculated by IVW in METAL [18].

The z-scores in the first and second rows are the meta-analyzed combined liver-blood AH-related signatures and blood AH-related patterns (Figure 3), respectively. The second to fourth rows show the binomial distribution (increase or decrease) of significant AH-DEGs. (B) Venn diagram illustrating the comparisons of two lists of meta-analyzed genes, including liver-blood and liver meta-genes, and a list of blood AH-DEGs. (C) Only gene sets with FDR < 5% in the pathway enrichment analysis are presented. The numbers in each cell indicate common genes between AH-DEGs and the genes annotated in each pathway.

Abbreviations: AH, alcoholic hepatitis; GSE, gene expression data series; GO, gene ontology database; has, homo sapiens; IVW, inverse variance-weighted average method; FC, fold change; DEG, differentially expressed gene; FDR, false discovery rate.
Figure 5

Meta-analysis to identify the AH-related genes in liver and blood tissues.

(A) The first and second rows denote the z-score calculated by IVW in METAL [18]. The z-scores in the first and second rows are the pooled liver-blood-death AH-related signatures (mortality-severe AH meta genes) and liver-blood AH meta genes (Figure 4), respectively. The second through fourth rows show the binomial distribution (increase or decrease) of significant AH-DEGs. (B) Venn diagram illustrates the comparisons of two lists of meta-analyzed genes, including mortality-severe and liver-blood meta genes,
and a list of AH-related mortality genes. (C) Only gene sets with an FDR < 5% in pathway enrichment analysis are represented. Numbers in each cell indicate common genes between the AH-DEGs and genes annotated in each pathway.

Abbreviations: AH, alcoholic hepatitis; GSE, gene expression data series; GO, gene ontology database; has, homo sapiens; IVW, inverse variance-weighted average method; FC, fold change; DEG, differentially expressed gene; FDR, false discovery rate.
Meta-analysis to identify the MSC therapeutic target genes for AH.

(A) The first and second rows indicate the z-score calculated by IVW in METAL [18]. The z-scores in the first and second rows are the meta-analyzed MSC-Tx and AH-related signatures (MSC-Tx target meta genes) and mortality-severe AH meta genes (Figure 5), respectively. The second through fourth rows show the binomial distribution (increase or decrease) of significant AH-DEGs. (B) Venn diagram illustrates the comparisons of two lists of meta-analyzed genes, including MSC-Tx target and mortality-severe meta genes, and three lists of MSC-related genes. (C) Only gene sets with an FDR < 5% in pathway enrichment analysis are represented. Numbers in each cell indicate common genes between the MSC-Tx targets and genes annotated in each pathway.

Abbreviations: MSC, mesenchymal stem cell; AH, alcoholic hepatitis; GSE, gene expression data series; GO, gene ontology database; has, homo sapiens; IVW, inverse variance-weighted average method; FC, fold change; DEG, differentially expressed gene; FDR, false discovery rate.
Figure 7

Identification of upstream MSC therapeutic target genes using multiple lines of external evidence.

The MSC-Tx targets identified by the four steps of meta-analysis (Figure 6) were evaluated by prior knowledge, including the human TF catalog [14], disease-gene association database (referred to as Digsee) [23, 24], disease- [25], expression-related SNP summary statistics [27, 28], and protein interactome [29]. Genes with two or more external evidence were selected as upstream targets of MSC-Tx. Cells colored in dark grape and emerald represent genes with at least three and two of the five
prominent lines of evidence, respectively. Cells colored in red and blue are differentially expressed genes in the meta-analysis, with genes showing increased and decreased expression in the AH group or MSC-Tx group, respectively.

Abbreviations: MSC, mesenchymal stem cell; FC, fold change; TF, transcription factor; GWAS, genome-wide association study; eQTL, expression quantitative trait loci; PPI protein-protein interaction; AH, alcoholic hepatitis; Tx, therapy; SNP, single nucleotide polymorphism.

**Supplementary Files**

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

- [Supplementaryfile230930.pdf](#)