



Published in final edited form as:

J BioX Res. 2019 March ; 2(1): 16–24. doi:10.1097/jbr.000000000000026.

Integrative omics analysis identifies macrophage migration inhibitory factor signaling pathways underlying human hepatic fibrogenesis and fibrosis

Zhipeng Liu^a, Naga Chalasani^{b,*}, Jingmei Lin^c, Samer Gawrieh^b, Yuan He^d, Yan J. Tseng^{d,e}, Wanqing Liu^{f,g,h,*}

^aDepartment of Medicinal Chemistry and Molecular Pharmacology, College of Pharmacy, Purdue University, West Lafayette,

^bDivision of Gastroenterology and Hepatology, Indianapolis, IN

^cDepartment of Pathology, Indiana University School of Medicine, Indianapolis, IN,

^dDepartment of Biochemistry, Microbiology and Immunology, School of Medicine, Wayne State University, Detroit, MI, USA

^eCenter for Molecular Medicine and Genetics, Wayne State University, Detroit, MI, USA

^fDepartment of Pharmaceutical Sciences, Eugene Applebaum College of Pharmacy and Health Sciences, Wayne State University, Detroit, MI, USA

^gDepartment of Pharmacology, School of Medicine, Wayne State University, Detroit, MI, USA

^hCancer Biology Program, Karmanos Cancer Institute, School of Medicine, Wayne State University, Detroit, MI, USA

Abstract

The genetic basis underlying liver fibrosis remains largely unknown. We conducted a study to identify genetic alleles and underlying pathways associated with hepatic fibrogenesis and fibrosis at the genome-wide level in 121 human livers. By accepting a liberal significance level of $P < 1e-4$, we identified 73 and 71 candidate loci respectively affecting the variability in alpha-smooth muscle actin (α -SMA) levels (fibrogenesis) and total collagen content (fibrosis). The top genetic loci associated with the two markers were *BAZAI* and *NOL10* for α -SMA expression and

This is an open-access article distributed under the terms of the [Creative Commons Attribution-Non Commercial-No Derivatives License 4.0](#) (CCBY-NC-ND), where it is permissible to download and share the work provided it is properly cited. The work cannot be changed in any way or used commercially without permission from the journal.

*Corresponding authors: Wanqing Liu, iBio Building 2401, 6135 Woodward Ave, Detroit, MI 48202, USA. wliu@wayne.edu; Naga Chalasani, Rotary Building, Suite 225 702 Rotary Cir, Indianapolis, IN 46202-5175, USA. nchalasa@iu.edu.

Author contributions

ZL contributed to data analysis and manuscript drafting. NC contributed to histologic data collection and manuscript drafting. JL collected histologic data. SG, YH, and YJT participated in data analysis and manuscript drafting. WL conceived the study, data analysis, and manuscript drafting. All authors approved the final version of the paper.

Institutional review board statement

The study was reviewed and approved by the Institutional Review Board (IRB) of Wayne State University (approval No. 201842) on May 17, 2018.

Conflicts of interest

The authors have declared that no conflict of interest exists.

FAM46A for total collagen content ($P<1e-6$). We further investigated the relationship between the candidate loci and the nearby gene transcription levels (cis-expression quantitative trait loci) in the same liver samples. We found that 44 candidate loci for α -SMA expression and 44 for total collagen content were also associated with the transcription of the nearby genes ($P<0.05$). Pathway analyses of these genes indicated that macrophage migration inhibitory factor (MIF) related pathway is significantly associated with fibrogenesis and fibrosis, though different genes were enriched for each marker. The association between the single nucleotide polymorphisms, MIF and α -SMA showed that decreased MIF expression is correlated with increased α -SMA expression, suggesting that variations in MIF locus might affect the susceptibility of fibrogenesis through controlling MIF gene expression. In summary, our study identified candidate alleles and pathways underlying both fibrogenesis and fibrosis in human livers. Our bioinformatics analyses suggested MIF pathway as a strong candidate involved in liver fibrosis, thus further investigation for the role of the MIF pathway in liver fibrosis is warranted. The study was reviewed and approved by the Institutional Review Board (IRB) of Wayne State University (approval No. 201842) on May 17, 2018.

Keywords

hepatic fibrosis; genomics; macrophage migration inhibitory factor; alpha-smooth muscle actin; pathway enrichment analysis; genetic susceptibility

Introduction

Hepatic fibrosis is a highly conserved and protective response characterized by the excessive accumulation of extracellular matrix proteins following either acute or chronic liver injury.^[1] Chronic viral hepatitis, alcohol abuse, and non-alcoholic fatty liver disease are the main causes of liver fibrosis.^[2] Although normal liver composition can be restored following acute or transient insult, sustained chronic liver injury will lead to a progressive formation of fibrous scar tissue, which will destroy the liver architecture and eventually produce hepatocellular dysfunction.^[3] The pathogenesis of liver fibrosis is not fully understood, and it remains largely unclear why the severity of the disease displays substantial variability in patients with the same set of known risk factors.^[4] Therefore, there is a pressing need to understand the genetic basis underlying the liver fibrosis which may subsequently allow for identifying critical drug targets can be identified and early intervention strategies for subjects with high risk of liver disease can be developed.

The expression of alpha-smooth muscle actin (α -SMA) reflects the activation of hepatic stellate cells to myofibroblast-like cells and is closely related to human liver fibrogenesis.^[5] Sirius red staining can accurately quantify the total hepatic collagen content,^[6] which is among the predominant hepatic extracellular matrix proteins.^[7] Therefore, quantitative α -SMA expression and Sirius red staining are accurate and reliable markers indicating liver fibrogenesis and fibrosis, respectively. In this study, we explored the genetic basis underlying the variability of these two quantitative molecular phenotypes using a step-wise genome-wide analysis, using donor liver samples which reflects a general American population. We performed genome-wide association study on the quantitative level of the

two markers. We further investigated the potential function of candidate single nucleotide polymorphisms (SNPs) in regulating mRNA expression for their nearby genes. A pathway enrichment analysis was further conducted to identify critical genes and pathways that are potentially underlying the genetic susceptibility to liver fibrosis.

Materials and methods

Datasets

The tissue procurement procedure and related information of the liver samples ($n=121$) used in this cross-sectional study have been described in our previous studies.^[8–10] In brief, these liver samples were obtained from unrelated liver transplantation donors of self-reported European and African descent. Subjects with heavy alcohol consumption ($>20\text{g/day}$), hepatitis B and C virus infection, or drug-induced liver injury, were excluded from this study. The genotype and gene expression profiling of these samples have been previously analyzed and deposited to the Gene Expression Omnibus database (accession number: GSE26106, <https://www.ncbi.nlm.nih.gov/geo/>). The demographical and histological characteristics of the donor liver samples used in this study have been summarized in Table 1. Analyses in this study were performed on anonymous individuals, thus this study is not considered to involve “human subjects”. The study was reviewed and has been approved by the Institutional Review Board (IRB) of Wayne State University (approval No. 201842) on May 17, 2018.

Liver histology characterization and α -SMA and sirius red staining and quantification

Formalin-fixed, paraffin-embedded liver sections were stained with hematoxylin and eosin and Masson trichrome stains for histological evaluation. The biopsies were scored by an experienced hepatopathologist (JL) in a blinded fashion according to the non-alcoholic steatohepatitis clinical research network liver histology criteria published by Kleiner et al^[11] and it showed normal in 59% and non-alcoholic fatty liver disease in 41% [fatty liver 5%, borderline non-alcoholic steatohepatitis 23%, and definite non-alcoholic steatohepatitis in 13%]. Formalin-fixed, paraffin-embedded sections were stained for α -SMA (marker for stellate cell activation) and Sirius red (total collagen content) and were digitally quantitated and expressed as a percent of total liver biopsy area using SPSS Sigma Scan Pro 5.0 software (SPSS Inc., Chicago, IL, USA).

Genome-wide association study (GWAS) analyses

Genotyping was performed using Illumina Human610-Quad v1.0 BeadChip array (Illumina, San Diego, CA, USA).^[8] The overall genotyping rate was 95.16%. After excluding rare (minor allele frequencies $<5\%$) and low quality (call rate $<90\%$ and deviation from Hardy-Weinberg equilibrium $P < 1e-3$) variants, there are 533,687 remaining SNPs for the linear regression analysis. Using an additive genetic model, each SNP was tested for association to α -SMA expression and hepatic collagen content, respectively. The phenotypes were normalized with log base 10 transformation. Age, gender, body mass index, and the first two genetic principal components were adjusted as covariates for the association. SNPs with P value less than $1e-4$ were considered as candidate loci for the following analysis. The quality control and association test was performed using the package PLINK 1.07 (<http://>

pngu.mgh.harvard.edu/purcell/plink/).[12] The regional plots were generated using the package LocusZoom (<http://csg.sph.umich.edu/locuszoom/>).[13]

Expression quantitative trait loci (eQTL) analyses

Gene expression profiling was measured using Agilent-014850 Whole Human Genome Microarray 4×44K (Agilent, Santa Clara, CA, USA) for the liver tissues of the same set of subjects.^[8] Linear regression model was used to detect transcripts significantly associated with the lead GWAS loci within ±1Mbp region. Age, gender, body mass index, and the first two genetic principal components were used as covariates. Associations with *P* value less than 0.05 were considered as significant for the following analysis. The eQTL analyses were performed using R package Matrix eQTL.^[14]

Pathway enrichment analyses

We used QIAGEN's Ingenuity® Pathway Analysis (IPA®, QIAGEN Redwood City, CA, USA; www.qiagen.com/ingenuity) to identify overrepresented signaling pathways in eQTL-controlling genes. Right-tailed Fisher's exact test was used to determine the significance level of signaling pathways, and *P*<0.05 was considered significant.

Gene interactions from curated databases and text-mining

The Gene Interactions tool in University of California, Santa Cruz Genome Browser (<https://genome.ucsc.edu>) was used to search the gene-gene interactions. Two genes were considered to be interacted if the interaction has been supported by either curated databases or text-mining. The curated databases consist of 23 pathway or protein interactions databases. A full list of databases that have been included can be found in the user guide of Gene Interactions tool (<https://genome.ucsc.edu/goldenPath/help/hgGeneGraph.html>). The text-mining supported gene interactions were generated by the Literome machine-reading program, which read and extracted the gene interactions from 20 million PubMed abstracts by the end of 2014.^[15] The gene-gene interactions among a given list of genes were visualized through igraph 1.0.0 (<http://igraph.org>).

Statistical analysis

Linear regression was used to assess the association between genetic variants and fibrosis markers or gene expression levels. Age, gender, body mass index, and the first 2 genetic principal components were used as covariates in the linear regression model. For GWAS analysis, *P* value less than 1e-4 were considered as candidate loci for the following analysis. As for the eQTL analysis, *P* value less than 0.05 were considered as significant for the following analysis. The demographical and histological features of the samples were expressed as number (percentage) for categorical variables and median (interquartile range) for continuous variables. The correlations between MIF gene expression and α -SMA expression and total collagen content were evaluated by Pearson correlation test, and *P*<0.05 was considered statistically significant. The statistical tests were performed using R 3.4 (<https://www.r-project.org>).

Results

GWAS analysis identifies multiple loci affecting α -SMA expression and total collagen content

The workflow of the study is shown in Figure 1. The two phenotypes explored by GWAS are positively correlated with each other ($r=0.31$, $P=6.6e-4$). After GWAS, no SNP was identified with a typical genome-wide significance ($5e-8$) in correlation with each phenotype. At a suggestive level of $P<1e-4$, there were 73 and 71 candidate genetic variants associated with α -SMA expression and total collagen content, respectively. Although there is a moderate correlation between the two markers, only 3 SNPs (rs1274369, rs1274351, rs1274323) were commonly associated with both markers. The Manhattan plots and Q-Q plots of the GWAS are shown in Figure 2, and the characteristics of the top 10 association loci are summarized in Table 2.

Top GWAS hits for α -SMA expression includes an intronic SNP rs8015303 in *BAZ1A* (bromodomain adjacent to zinc finger domain 1A) which encodes a protein subunit of the ATP-dependent chromatin assembly factor that involved in chromatin remodeling. An intronic SNP rs1012580 in the *NOL10* (nucleolar protein 10) was also significantly associated with α -SMA expression. A few other top hits indicated that several genes involved in inflammation and immune response especially *IL2RA* (interleukin 2 receptor subunit alpha), *HTR7* (5-hydroxytryptamine receptor 7), *ST6GALNAC3* (ST6 N-acetyl-galactosaminide alpha-2,6-sialyltransferase 3) and *AOAH* (acyloxyacyl hydrolase).

Top GWAS hits for total hepatic collagen content are mainly genes that are significantly related to cell skeleton structure, extracellular matrix and cell adhesion. These include *ZWINT* (ZW10 Interacting kinetochore protein) (spindle assembly),^[16] *BCAR3* (breast cancer anti-estrogen resistance 3) (cytoskeletal remodeling and adhesion),^[17] *EFNA5* (ephrin-A5) (cell adhesion and morphology)^[18] and *COL11A2* (collagen type XI alpha 2 chain).

Expression analysis identifies significant eQTLs for lead variants

The eQTL analysis can help establish the potential causality for the GWAS findings. To further explore the effects of GWAS loci on gene expression, we performed *cis*-eQTL analyses in the same set of liver tissues as the GWAS, by focusing on genes that are within the distance of ± 1 Mbp to GWAS identified candidate loci ($P<1e-4$). By accepting a liberal significance level of $P=0.05$ for eQTLs, we identified 102 significant eQTLs for 44 α -SMA-associated variants and 77 eQTLs for 44 candidate GWAS loci associated with collagen content. The information of the full list of eQTLs is provided in Additional Table 1, <http://links.lww.com/JR9/A4>.

Our results show that several variants are associated with the mRNA expression of their nearest genes. For example, rs1012580 in nucleolar protein 10 (*NOL10*) for α -SMA expression and rs13092046 in contactin 3 (*CNTN3*) for total collagen content are found to be correlated with the mRNA expression levels of their most nearby genes. However, other significant eQTLs may exert their effects on gene expression in a relatively broad range. For instance, fibrogenesis candidate variant rs12207 at *TUBB3* locus is associated with the

expression levels of several nearby genes including charged multivesicular body protein 1A (*CHMP1A*), Fanconi anemia complementation group A (*FANCA*), VPS9 domain containing 1 (*VPS9D1*), and paraplegin matrix AAA peptidase subunit (*SPG7*) instead of *TUBB3* itself.

Top eQTLs for α -SMA expression include several variants in 22q11.23 locus that are significantly associated with the transcription levels of macrophage migration inhibitory factor (*MIF*) and glutathione S-transferase theta 2 (*GSTT2*) (Additional Table 1, <http://links.lww.com/JR9/A4>). Top eQTL-controlling genes for total collagen content include two glycosyltransferases encoding genes, namely solute carrier family 35 member B4 (*SLC35B4*) and beta-1,3-galactosyltransferase 4 (*B3GALT4*). The transcription of prostaglandin-endoperoxide synthase 2 (*PTGS2*), which encodes a cyclooxygenase involved in the MIF-related apoptosis repression, is significantly associated with collagen-related variants in 1q31.1 locus.

Pathway enrichment analysis of the eQTL-controlling genes

Given the polygenic nature of liver fibrosis, a single genetic variant or gene may only possess limited effect on the development and progression of the disease. We aim to identify pathways underlying the liver fibrosis that individual SNP-based GWAS analysis might miss. We conducted enrichment analyses on the candidate eQTL-controlling genes using the ingenuity pathway analysis package. In order to avoid missing key pathways, we assumed that those genes whose transcriptions are associated with the candidate SNPs at a nominal $P < 0.05$ level are the candidate genes for an enrichment analysis. As shown in Table 3, eumelanin biosynthesis, ataxia telangiectasia mutated signaling, glutathione redox reactions I, vascular endothelial growth factor signaling, glutathione-mediated detoxification, MIF-mediated glucocorticoid regulation, and MIF regulation of innate immunity pathways are top enriched pathways for the α -SMA activation. On the other hand, MIF-mediated glucocorticoid regulation, MIF regulation of innate immunity, branched-chain α -keto acid dehydrogenase complex, and glutathione ascorbate recycling pathways are enriched in candidate genes for total collagen content.

Interestingly, although there is only a minimal overlap between the candidate genes associated with the 2 phenotypes, the MIF related pathways, as well as glutathione-S-transferases (GSTs), are significantly enriched for both markers. A detailed inspection of the *MIF* gene SNPs identifies several variants near the *MIF* locus to be negatively associated with α -SMA levels ($\beta = -0.43$; 95% CI: $-0.64, -0.23$; $P = 8.38 \times 10^{-5}$) (Fig. 3B), but positively associated with *MIF* gene expression ($\beta = 0.37$; 95% CI: $0.30, 0.44$; $P = 6.34 \times 10^{-7}$) (Fig. 3C). We found that *MIF* gene expression is significantly correlated with decreased levels of α -SMA (Fig. 4A, $r = -0.21$, $P = 0.026$), suggesting that variations in *MIF* locus might affect the susceptibility of fibrogenesis through controlling *MIF* gene expression. There is no significant correlation between *MIF* gene expression and total collagen content (Fig. 4B, $P = 0.29$).

To further confirm this pathway enrichment, and to explore the possible mechanisms through which eQTL-controlling genes are involved in the development of fibrosis, we searched for the gene-gene interactions among candidate genes for α -SMA expression and total collagen content based on a text-mining based gene interaction database. Two genes

were considered to be interacted as long as this relationship has been supported by either curated gene interactions databases or text-mining evidence. As shown in Figure 5A, nitric oxide synthase 3 (*NOS3*) and *MIF* are the 2 genes that have the largest number of connections with other α -SMA expression-related genes. *PTGS2*, which mediates the MIF induced apoptosis suppression, is one of the most highly connected genes for the total collagen content-related gene network (Fig. 5B).

Discussion

Hepatic collagens accumulation and α -SMA activation are reliable and quantitative markers indicating the risk of developing liver fibrosis and cirrhosis,^[19,20] thus are important intermediate phenotypes for the disease. Interrogating the genetic susceptibility loci for these intermediate phenotypes will gain insight into the molecular mechanisms involved in the pathogenesis of hepatic fibrosis and provide potential targets for early diagnosis and treatment. Our study for the first time identifies candidate genetic variants, genes, and pathways contributing to human liver fibrogenesis and fibrosis in a general population.

GWAS have been widely used to investigate the genetic basis of human complex diseases, and thousands of susceptibility loci have been identified thus far.^[21] However, to date, most GWAS identified alleles only account for a modest proportion of total variance in traits. This is mainly because of the polygenic nature of complex traits and at least in part due to the multiple testing burden in test statistics.^[22] A growing body of knowledge acknowledges that there are causal variants remain undetected owe to the adoption of stringent genome-wide significance threshold level ($P < 5e-8$).^[23,24] Therefore, we set a relatively moderate threshold at $P < 1e-4$ for GWAS to systematically evaluate the overall effects of candidate genetic variants on the two phenotypes. In addition, incorporating the information of eQTLs and evaluating them in a network way is beneficial to interpret the biological mechanisms underlying the discovered loci, thus potentially identifying causal alleles underlying the genetic association. As such, although our study did not identify any SNP that reaches the typical GWAS significance level, which could be largely attributed to the small sample size, our combined analyses indeed narrowed down a few interesting genes and pathways that are broadly supported by many previous studies.

Our analyses firstly demonstrated that a few genes involved in inflammation and immune response as top GWAS hits for α -SMA activation. This is consistent with the known relationship between α -SMA expression and hepatic stellate cell activation as a key step for liver fibrogenesis. One notable loci is *IL2RA* (CD25) across which multiple SNPs were previously associated with various inflammatory disorders and traits including allergy,^[25] Epstein-Barr virus nuclear antigen-1 (EBNA-1) IgG level,^[26] diisocyanate-induced asthma,^[27] inflammatory bowel disease,^[28] autoimmune diseases,^[29,30] rheumatoid arthritis,^[31] multiple sclerosis,^[32,33] and levels of autoantibodies in type 1 diabetes.^[34] The top hit rs12722561 identified in our study is in complete linkage disequilibrium with rs12722489, a polymorphism previously identified in genome-wide meta-analyses as a susceptibility locus to multiple sclerosis^[32] and Crohn disease.^[35] A recent study has shown that variant in *IL2RA* was also significantly associated with the concentration of circulating IL2RA.^[36] Importantly, the IL-2 signaling pathway was also enriched as one of the top pathways

associated with advanced fibrosis and cirrhosis in a genome-wide pathway analysis for non-alcoholic fatty liver disease.^[37] Mechanistically, the IL-2/IL2RA signaling has been indicated to be involved in the liver fibrosis as well. The stellate cell-lymphocyte interaction in the liver plays a pivotal role in stellate cell activation and liver fibrogenesis, and the IL2RA (CD25) and CD4 positive regulatory T cells (CD4⁺/CD25⁺) have been demonstrated to be anti-fibrotic by suppression the pro-fibrotic effect of CD8 cells on stellate cells.^[38]

It is also not surprising that top GWAS hits for total collagen content are associated with multiple genes involved in cytoskeletal structure, extracellular matrix, cell migration and adhesion. In addition, two highly linked intronic SNPs at the *COL11A2* (collagen type XI alpha 2 chain) locus are strongly associated with the total collagen content as well. Interestingly, these 2 SNPs are also significantly associated with mRNA expression of multiple genes within a ~250kb region in our liver tissue set including *COL11A2* and multiple *HLA* genes. We further searched the genotype-tissue expression portal for these two SNPs, which turns out that they are also strong eQTLs for the similar set of the genes within the region among multiple tissues. Notably, this *HLA* locus has been previously linked with hepatitis B virus/hepatitis C virus-related liver cirrhosis^[39,40] as well as primary biliary cirrhosis,^[41] suggesting an essential role of these genes in increasing the susceptibility to liver fibrosis.

It is known that the typical GWAS approach may miss important genetic loci as only the very top significant SNPs are selected. We therefore further performed a pathway enrichment analysis by focusing on the genes whose mRNA expression is likely to be affected by the candidate SNPs that are associated with the two markers using a liberal cut-off of $P < 1e-4$. Interestingly, the glutathione-related genes and MIF-related pathways are the two common major pathways enriched for both markers. It should be noted that there is only a minimal overlap between the candidate genes associated with each of the 2 markers, suggesting that despite different genes, the underlying pathways still stand out. While it is known that GSTs are protective for oxidative stress-mediated liver damage,^[42] MIF related signaling pathway has been strongly linked to liver fibrosis as well. A recent study demonstrated that MIF-deleted mice (*Mif*^{-/-}) tend to show exaggerated fibrogenic phenotypes in 2 chronic liver injury models.^[43] This study is consistent with our finding that MIF may exert anti-fibrotic effects in human livers. Moreover, other eQTL-controlling genes in MIF pathway, including *PLA2* and *PTGS2*, are key mediators of MIF induced apoptosis suppression signaling, indicating apoptosis repression might be one of the mechanisms for the antifibrotic effects of MIF.^[44]

In addition, angiogenesis has been significantly involved in liver fibrosis.^[45,46] Indeed, our pathway enrichment analysis identified that NOS3 and vascular endothelial growth factor signaling pathways are significantly associated with α -SMA expression, while the endothelin-1 signaling pathway is significantly pathway is associated with total collagen level. A considerable body of evidence demonstrates that vascular endothelial growth factor signaling promotes liver fibrogenesis by stimulating activated stellate cells growth, migration, and collagen production.^[47,48] It has been also shown that *NOS3* expression and activation plays a critical role in the development of FLD and liver fibrosis.^[49-51]

We further explore the potential interaction network between the key genes and pathways. Again, the *NOS3* and *MIF* are the most highly connected genes for the α -SMA expression gene network. Our results suggested that the 2 hub genes, *NOS3* and *MIF*, and their related genes were connected through MutS homolog 4 (*MSH4*), suggesting that the crosstalk between *NOS3* and *MIF* signaling pathways might be critical for the liver fibrogenesis. However, this hypothesis needs to be validated through further investigations. As for the gene interaction network of the total collagen content related genes, *PTGS2* has the largest number of connections. Again, it is notable that *PTGS2* is also involved in both the *MIF* and endothelin-1 signaling pathway, which may further indicate that the *MIF* and angiogenesis signaling are both involved in liver fibrosis. It should be noted again that the interaction networks for both phenotypes highlighted the *MIF* signaling pathway although there is limited overlap between eQTL-controlling genes for the two phenotypes. Therefore, the role of *MIF* signaling, especially its potential interaction with the angiogenesis pathway needs to be further investigated.

There are also several limitations of our study. Due to the moderate sample size, our analyses are limited in its power for identifying genetic risk loci. In addition, we adopted a generous threshold to incorporate more potential causal variants into analysis. This will inevitably increase the false positive rate of our test. Therefore, the susceptibility loci identified in our study should be considered as suggestive and need to be validated independently and within more diverse cohorts.

In conclusion, our study identified candidate genetic variants and pathways significantly associated with intermediate markers for liver fibrogenesis and fibrosis. Our findings would be helpful to elucidate the genetic basis underlying the inter-individual differences in the development of liver fibrosis and provide candidate targets for developing therapeutic strategies.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

None.

Financial support

This study was supported in part by a NIH grant, No. R01 DK106540 (to WL).

References

- [1]. Batailler R, Brenner DA. Liver fibrosis. *J Clin Invest* 2005;115:209–218. [PubMed: 15690074]
- [2]. Hernandez-Gea V, Friedman SL. Pathogenesis of liver fibrosis. *Annu Rev Pathol* 2011;6:425–456. [PubMed: 21073339]
- [3]. Benyon RC, Iredale JP. Is liver fibrosis reversible? *Gut* 2000;46:443–446. [PubMed: 10716665]
- [4]. Pellicoro A, Ramachandran P, Iredale JP, et al. Liver fibrosis and repair: immune regulation of wound healing in a solid organ. *Nat Rev Immunol* 2014;14:181–194. [PubMed: 24566915]

- [5]. Mann DA, Smart DE. Transcriptional regulation of hepatic stellate cell activation. *Gut* 2002;50:891–896. [PubMed: 12010897]
- [6]. Lattouf R, Younes R, Lutomski D, et al. Picosirius red staining: a useful tool to appraise collagen networks in normal and pathological tissues. *J Histochem Cytochem* 2014;62:751–758. [PubMed: 25023614]
- [7]. Schuppan D Structure of the extracellular matrix in normal and fibrotic liver: collagens and glycoproteins. *Semin Liver Dis* 1990;10:1–10. [PubMed: 2186485]
- [8]. Innocenti F, Cooper GM, Stanaway IB, et al. Identification, replication, and functional fine-mapping of expression quantitative trait loci in primary human liver tissue. *PLoS Genet* 2011;7:e1002078. [PubMed: 21637794]
- [9]. Gamazon ER, Innocenti F, Wei R, et al. A genome-wide integrative study of microRNAs in human liver. *BMC Genomics* 2013;14:395. [PubMed: 23758991]
- [10]. Wang L, Athinarayanan S, Jiang G, et al. Fatty acid desaturase 1 gene polymorphisms control human hepatic lipid composition. *Hepatology* 2015;61:119–128. [PubMed: 25123259]
- [11]. Kleiner DE, Brunt EM, Van Natta M, et al. Design and validation of a histological scoring system for nonalcoholic fatty liver disease. *Hepatology* 2005;41:1313–1321. [PubMed: 15915461]
- [12]. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007;81:559–575. [PubMed: 17701901]
- [13]. Pruim RJ, Welch RP, Sanna S, et al. LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics* 2010;26:2336–2337. [PubMed: 20634204]
- [14]. Shabalin AA. Matrix eQTL: ultra fast eQTL analysis via large matrix operations. *Bioinformatics* 2012;28:1353–1358. [PubMed: 22492648]
- [15]. Poon H, Quirk C, DeZiel C, et al. Literome: PubMed-scale genomic knowledge base in the cloud. *Bioinformatics* 2014;30:2840–2842. [PubMed: 24939151]
- [16]. Woo Seo D, Yeop You S, Chung WJ, et al. Zwint-1 is required for spindle assembly checkpoint function and kinetochore-microtubule attachment during oocyte meiosis. *Sci Rep* 2015;5:15431. [PubMed: 26486467]
- [17]. Wilson AL, Schrecengost RS, Guerrero MS, et al. Breast cancer antiestrogen resistance 3 (BCAR3) promotes cell motility by regulating actin cytoskeletal and adhesion remodeling in invasive breast cancer cells. *PLoS One* 2013;8:e65678. [PubMed: 23762409]
- [18]. Buensuceso AV, Deroo BJ. The ephrin signaling pathway regulates morphology and adhesion of mouse granulosa cells in vitro. *Biol Reprod* 2013;88:25. [PubMed: 23242526]
- [19]. Yamaoka K, Nouchi T, Marumo F, et al. Alpha-smooth-muscle actin expression in normal and fibrotic human livers. *Dig Dis Sci* 1993;38:1473–1479. [PubMed: 8344103]
- [20]. Huang Y, de Boer WB, Adams LA, et al. Image analysis of liver collagen using sirius red is more accurate and correlates better with serum fibrosis markers than trichrome. *Liver Int* 2013;33:1249–1256. [PubMed: 23617278]
- [21]. Welter D, MacArthur J, Morales J, et al. The NHGRI GWAS catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res* 2014;42:D1001–D1006. [PubMed: 24316577]
- [22]. Visscher PM, Brown MA, McCarthy MI, et al. Five years of GWAS discovery. *Am J Hum Genet* 2012;90:7–24. [PubMed: 22243964]
- [23]. Amin N, van Duijn CM, Janssens AC. Genetic scoring analysis: a way forward in genome wide association studies? *Eur J Epidemiol* 2009;24:585–587. [PubMed: 19728114]
- [24]. Sveinbjornsson G, Albrechtsen A, Zink F, et al. Weighting sequence variants based on their annotation increases power of whole-genome association studies. *Nat Genet* 2016;48:314–317. [PubMed: 26854916]
- [25]. Hinds DA, McMahon G, Kiefer AK, et al. A genome-wide association meta-analysis of self-reported allergy identifies shared and allergy-specific susceptibility loci. *Nat Genet* 2013;45:907–911. [PubMed: 23817569]
- [26]. Zhou Y, Zhu G, Charlesworth JC, et al. Genetic loci for Epstein-Barr virus nuclear antigen-1 are associated with risk of multiple sclerosis. *Mult Scler* 2016;22:1655–1664. [PubMed: 26819262]

- [27]. Yucesoy B, Kaufman KM, Lummus ZL, et al. Genome-wide association study identifies novel loci associated with diisocyanate-induced occupational asthma. *Toxicol Sci* 2015;146:192–201. [PubMed: 25918132]
- [28]. Liu JZ, van Sommeren S, Huang H, et al. Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. *Nat Genet* 2015;47:979–986. [PubMed: 26192919]
- [29]. Jin Y, Birlea SA, Fain PR, et al. Variant of TYR and autoimmunity susceptibility loci in generalized vitiligo. *N Engl J Med* 2010;362:1686–1697.
- [30]. Li YR, Li J, Zhao SD, et al. Meta-analysis of shared genetic architecture across ten pediatric autoimmune diseases. *Nat Med* 2015;21:1018–1027. [PubMed: 26301688]
- [31]. Orozco G, Viatte S, Bowes J, et al. Novel rheumatoid arthritis susceptibility locus at 22q12 identified in an extended UK genome-wide association study. *Arthritis Rheumatol* 2014;66:24–30. [PubMed: 24449572]
- [32]. Patsopoulos NA; Bayer Pharma MS Genetics Working Group; Steering Committees of Studies Evaluating IFN β -1 β and a CCR1-Antagonist, et al. Genome-wide meta-analysis identifies novel multiple sclerosis susceptibility loci. *Ann Neurol* 2011;70:897–912. [PubMed: 22190364]
- [33]. International Multiple Sclerosis Genetics Consortium; Wellcome Trust Case Control Consortium 2, Sawcer S, et al. Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. *Nature* 2011;476:214–219. [PubMed: 21833088]
- [34]. Plagnol V, Howson JM, Smyth DJ, et al. Genome-wide association analysis of autoantibody positivity in type 1 diabetes cases. *PLoS Genet* 2011;7:e1002216. [PubMed: 21829393]
- [35]. Franke A, McGovern DP, Barrett JC, et al. Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. *Nat Genet* 2010;42:1118–1125. [PubMed: 21102463]
- [36]. Ahola-Olli AV, Wurtz P, Havulinna AS, et al. Genome-wide association study identifies 27 loci influencing concentrations of circulating cytokines and growth factors. *Am J Hum Genet* 2017;100:40–50. [PubMed: 27989323]
- [37]. Chen QR, Braun R, Hu Y, et al. Multi-SNP analysis of GWAS data identifies pathways associated with nonalcoholic fatty liver disease. *PLoS One* 2013;8:e65982. [PubMed: 23894275]
- [38]. Horani A, Muhanna N, Pappo O, et al. Beneficial effect of glatiramer acetate (Copaxone) on immune modulation of experimental hepatic fibrosis. *Am J Physiol Gastrointest Liver Physiol* 2007;292:G628–G638. [PubMed: 17038628]
- [39]. Mangia A, Gentile R, Cascavilla I, et al. HLA class II favors clearance of HCV infection and progression of the chronic liver damage. *J Hepatol* 1999;30:984–989. [PubMed: 10406174]
- [40]. Zhang X, Wan Y, Zhang S, et al. Nonalcoholic fatty liver disease prevalence in urban school-aged children and adolescents from the Yangtze River delta region: a cross-sectional study. *Asia Pac J Clin Nutr* 2015;24:281–288. [PubMed: 26078245]
- [41]. Almasio PL, Licata A, Maida M, et al. Clinical course and genetic susceptibility of primary biliary cirrhosis: analysis of a prospective cohort. *Hepat Mon* 2016;16:e31681. [PubMed: 28070198]
- [42]. Li S, Tan HY, Wang N, et al. The role of oxidative stress and antioxidants in liver diseases. *Int J Mol Sci* 2015;16:26087–26124. [PubMed: 26540040]
- [43]. Heinrichs D, Knaul M, Offermanns C, et al. Macrophage migration inhibitory factor (MIF) exerts antifibrotic effects in experimental liver fibrosis via CD74. *Proc Natl Acad Sci U S A* 2011;108:17444–17449. [PubMed: 21969590]
- [44]. Elsharkawy AM, Oakley F, Mann DA. The role and regulation of hepatic stellate cell apoptosis in reversal of liver fibrosis. *Apoptosis* 2005;10:927–939. [PubMed: 16151628]
- [45]. Srivastava A, Shukla V, Tiwari D, et al. Targeted therapy of chronic liver diseases with the inhibitors of angiogenesis. *Biomed Pharmacother* 2018;105:256–266. [PubMed: 29859468]
- [46]. Thomas H Liver: delineating the role of angiogenesis in liver fibrosis. *Nat Rev Gastroenterol Hepatol* 2018;15:6.
- [47]. Yoshiji H, Kuriyama S, Yoshii J, et al. Vascular endothelial growth factor and receptor interaction is a prerequisite for murine hepatic fibrogenesis. *Gut* 2003;52:1347–1354. [PubMed: 12912869]

- [48]. Novo E, Cannito S, Zamara E, et al. Proangiogenic cytokines as hypoxia-dependent factors stimulating migration of human hepatic stellate cells. *Am J Pathol* 2007;170:1942–1953. [PubMed: 17525262]
- [49]. Leung TM, Tipoe GL, Liang EC, et al. Endothelial nitric oxide synthase is a critical factor in experimental liver fibrosis. *Int J Exp Pathol* 2008;89:241–250. [PubMed: 18429990]
- [50]. Sheldon RD, Laughlin MH, Rector RS. Reduced hepatic eNOS phosphorylation is associated with NAFLD and type 2 diabetes progression and is prevented by daily exercise in hyperphagic OLETF rats. *J Appl Physiol* 19852014;116:1156–1164.
- [51]. Persico M, Masarone M, Damato A, et al. Non alcoholic fatty liver disease and eNOS dysfunction in humans. *BMC Gastroenterol* 2017;17:35. [PubMed: 28264657]

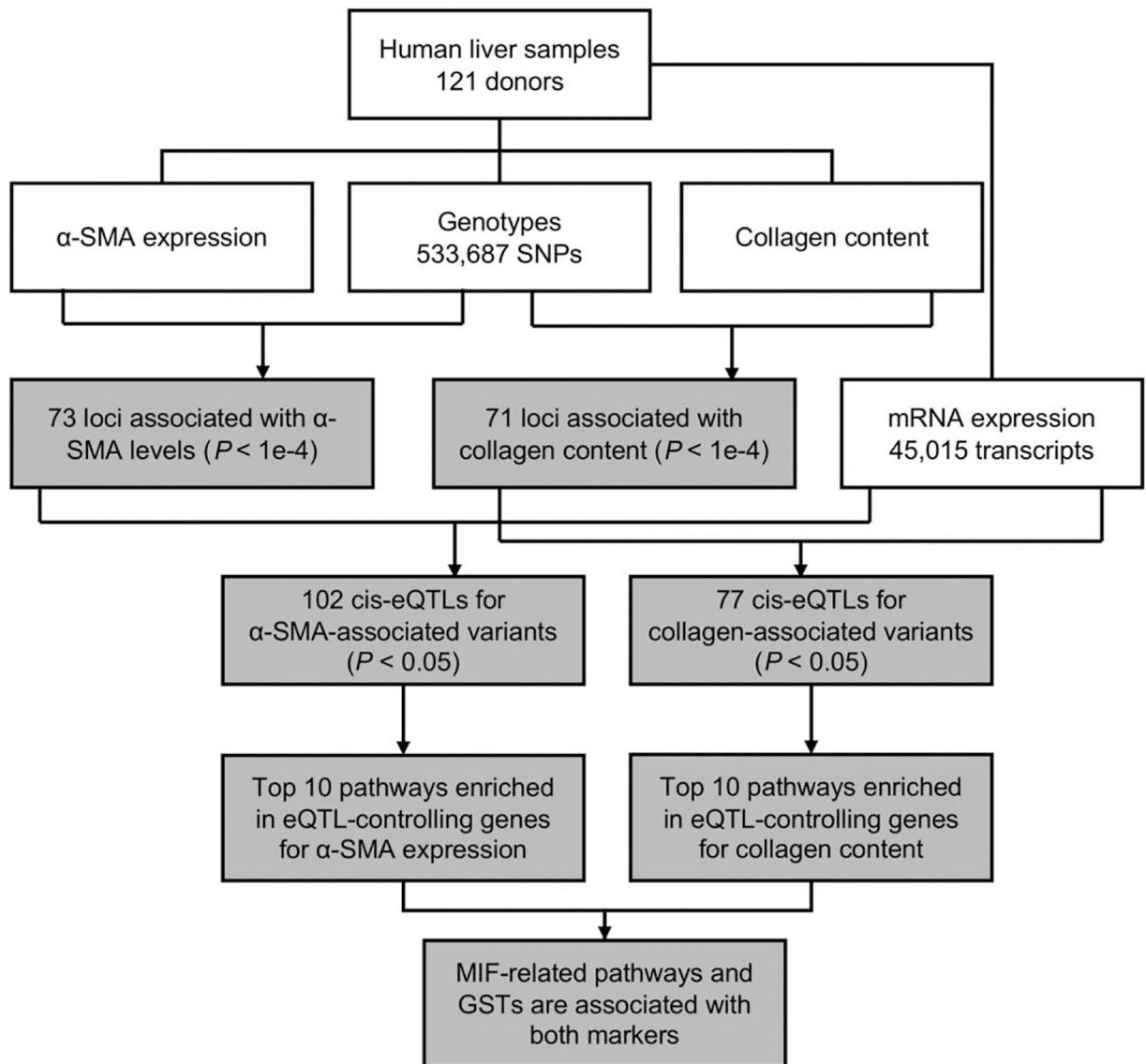


Figure 1.

Flowchart showing the workflow of the analysis. White boxes represent the input data of human liver samples. Grey boxes indicate the significant output of analysis. α -SMA=alpha-smooth muscle actin, eQTL=expression quantitative trait loci, GST=glutathione S-transferase, MIF=macrophage migration inhibitory factor, SNP=single nucleotide polymorphism.

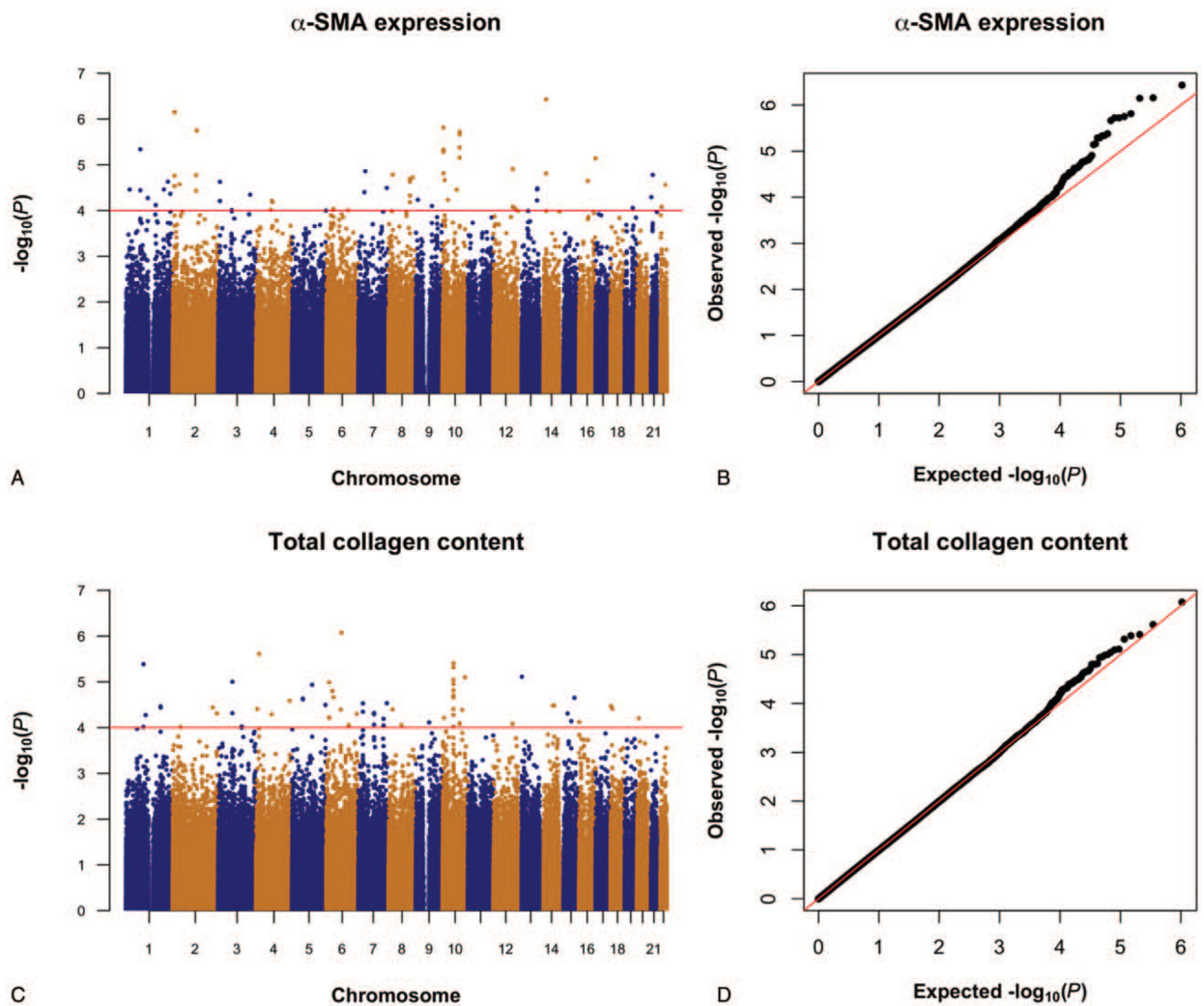


Figure 2.

Genome-wide association studies of α -SMA expression and total collagen content. (A, B) Manhattan plot and quantile-quantile plot for GWAS of α -SMA expression. (C, D) Manhattan plot and quantile-quantile plot for GWAS of total collagen content. P values were calculated by using multiple linear regression model adjusted for age, gender, body mass index, and first two principle components. The horizontal red lines indicate the suggestive significance threshold ($P < 1e-4$) used for further analysis. α -SMA=alpha-smooth muscle actin, GWAS=genome-wide association study.

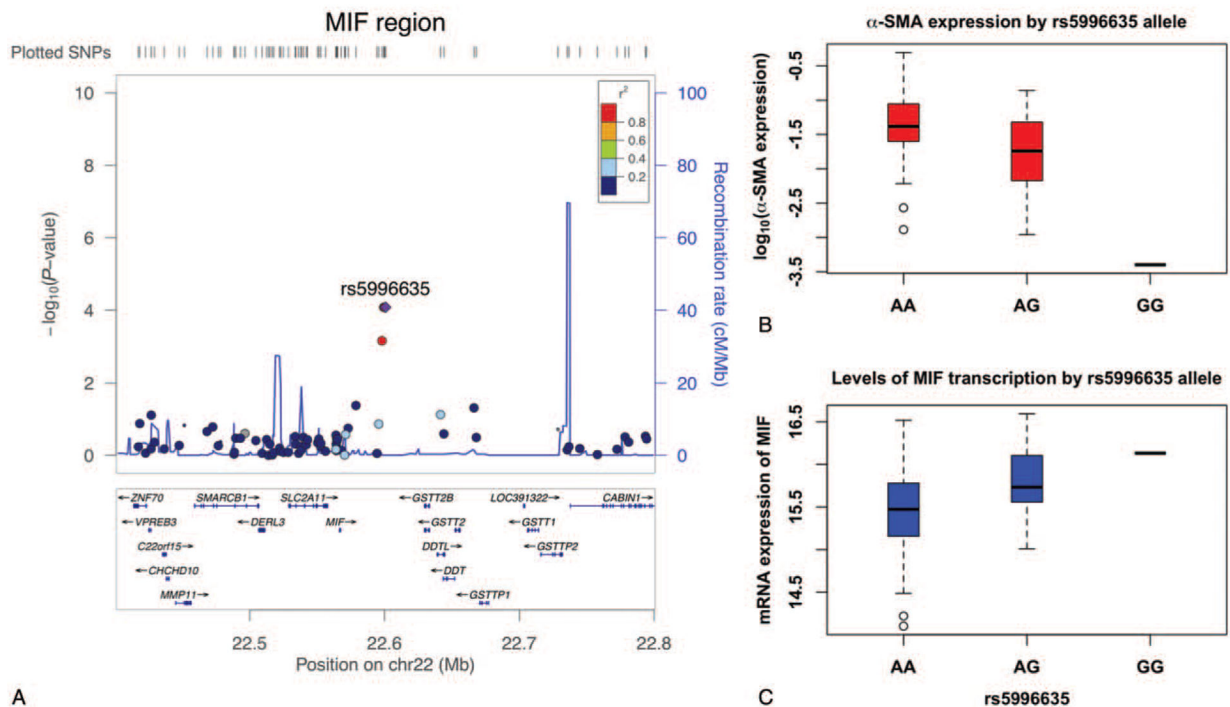


Figure 3.

Association between variants near MIF locus and MIF gene expression and α -SMA levels. (A) Regional plot for MIF locus. Each SNP is plotted with its P value (shown as $-\log_{10}(P$ value)) as a function of its genomic coordinate (Hg 18). The local LD structure was estimated based on 1000 Genomes 2010 CEU; B: Negative association between *MIF* rs5996635 and α -SMA expression levels ($\beta = -0.43$; 95%CI: $-0.64, -0.23$; $P = 8.38e-5$). P values were calculated by using multiple linear regression model adjusted for age, gender, BMI, and first 2 principle components. (C) Positive association between *MIF* rs5996635 and MIF transcription ($\beta = 0.37$; 95%CI: $0.30, 0.44$; $P = 6.34e-7$). A multiple linear regression model was used to test the association between MIF expression levels and nearby GWAS loci within ± 1 Mbp region. Age, gender, body mass index, and the first two genetic principal components were used as covariates. Numbers of individuals with AA, AG, and GG genotypes are 88, 32, and 1, respectively. α -SMA=alpha-smooth muscle actin, GWAS=genome-wide association study, MIF=macrophage migration inhibitory factor, SNP=single nucleotide polymorphism.

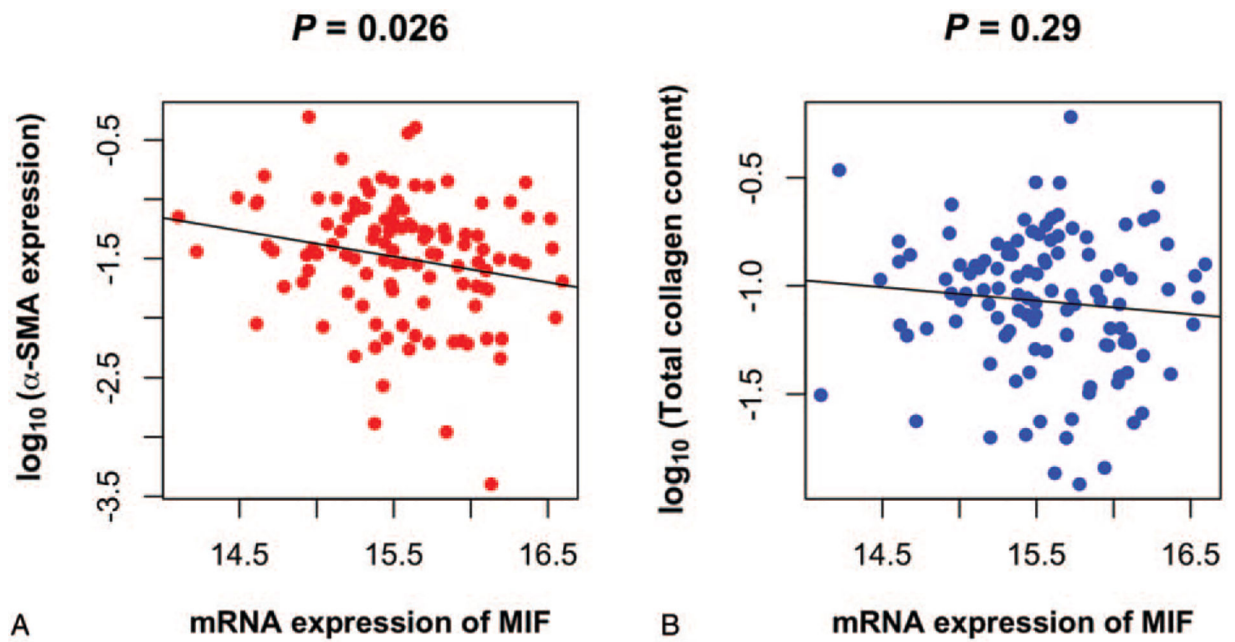


Figure 4.

Correlation between *MIF* gene expression and α -SMA expression and total collagen content. (A) Negative correlation ($r=-0.21$, $P=0.026$) between *MIF* gene expression and α -SMA expression. (B) Correlation ($r=-0.10$, $P=0.29$) between *MIF* gene expression and total collagen content. P values were calculated by Pearson correlation test. α -SMA=alpha-smooth muscle actin, MIF=macrophage migration inhibitory factor.

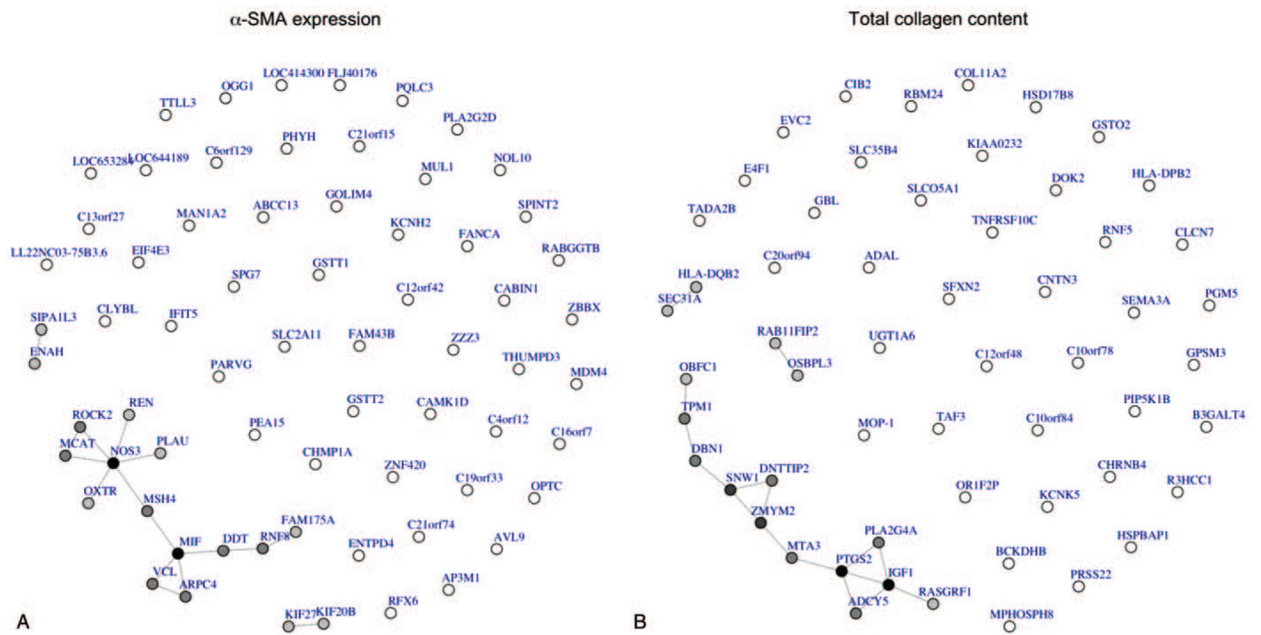


Figure 5.

Gene interaction networks for eQTL-controlling genes. (A) Gene interactions for α -SMA expression-related genes. (B) Gene interactions for total collagen content-related genes. The grayscale intensity of the node represents the number of its edges. The darker the node is, the more connections it has. α -SMA=alpha-smooth muscle actin, eQTL=expression quantitative trait loci.

Table 1

Demographical and histological characteristics of the donor liver tissues.

Item	Data
Male	82 (67.8)
Age (year)	40 (17–57)
body mass index (kg/m ²)	25.8 (21.8–29.8)
Race	
White	102 (84.3)
Black	19 (15.7)
Percentage of alpha-smooth muscle actin expression	3.6 (1.8–7.1)
Percentage of total collagen content	9.4 (5.5–14.1)
Fibrosis stage*	
Focal perisinusoidal	1(0.9)
Perisinusoidal	6 (5.2)
No fibrosis	109 (93.9)

N = 121.

* missing information for 5 participants. Data are expressed as number (percent) in male, race, and fibrosis stage, and median (interquartile range) in others.

Table 2

Top 10 GWAS Loci associated with α -SMA expression and total collagen content.

SNP	Gene	Locus	Position	A1	A2	MAF	Beta	SE	P
α-SMA expression									
rs8015303	<i>BAZ1A</i>	14q13.2	34316223	G	A	0.06967	-0.7229	0.1313	3.70e-7
rs1012580	<i>NOL10</i>	2p25.1	10725199	A	G	0.2869	-0.4247	0.07935	6.96e-7
rs12722561	<i>IL2RA</i>	10p15.1	6109899	A	G	0.1557	-0.5144	0.09968	1.54e-6
rs12479413	<i>AK311291</i>	2q14.3	129343258	G	A	0.05328	-0.9579	0.1869	1.78e-6
rs1274351	<i>BC037970</i>	10q23.31	92237989	A	G	0.123	-0.5068	0.09921	1.90e-6
rs1274323	<i>HTR7</i>	10q23.31	92264462	G	A	0.127	-0.5038	0.09925	2.17e-6
rs724999	<i>ST6GALNAC3</i>	1p31.1	76961614	A	G	0.07438	-0.8518	0.1741	4.58e-6
rs12207	<i>TUBB3</i>	16q24.3	88529861	A	G	0.05785	-0.9667	0.2023	7.23e-6
rs1345546	<i>NUAK1</i>	12q23.3	104826063	A	G	0.09426	-0.6404	0.1381	1.23e-5
rs10281171	<i>AOAH</i>	7p14.2	36773906	G	A	0.3852	-0.3295	0.07149	1.37e-5
Total collagen content									
rs6908031	<i>FAM46A</i>	6q14.1	81900082	G	A	0.2992	-0.2554	0.04814	8.46e-7
rs2108935	<i>LDB2</i>	4p15.32	16453008	G	A	0.2541	-0.2545	0.05043	2.44e-6
rs7089692	<i>ZWINT</i>	10q21.1	58603555	G	A	0.2459	-0.2431	0.0493	3.89e-6
rs3950119	<i>BCAR3</i>	1p22.1	93821825	G	A	0.2336	0.2674	0.05438	4.10e-6
rs9512950	<i>TPTE2</i>	13q12.11	18870233	C	A	0.2992	-0.2572	0.05406	7.74e-6
rs181662	<i>EMX2</i>	10q26.11	119372445	G	A	0.3279	0.2253	0.04743	7.97e-6
rs13092046	<i>CNTN3</i>	3p12.3	74617642	A	G	0.4262	0.2266	0.04828	9.97e-6
rs2237172	<i>ATXN1</i>	6p22.3	16709566	G	A	0.3512	-0.2444	0.05214	1.03e-5
rs6894788	<i>EFNA5</i>	5q21.3	106875173	G	A	0.2645	0.2476	0.05317	1.16e-5
rs9368758	<i>COL11A2</i>	6p21.32	33245999	A	G	0.06557	-0.4443	0.09714	1.58e-5

The single nucleotide polymorphism is mapped to its nearest gene; A1 is the minor allele, and A2 is the major allele; All positions refer to hg18. α -SMA=alpha-smooth muscle actin, MAF=minor allele frequencies, SE=standard error.

Table 3

Results from ingenuity pathway analysis analyses.

Item	Ingenuity canonical pathways	P	Molecules
α-SMA expression	Eumelanin biosynthesis	7.079e-5	MIF, DDT
	ATM signaling	1.318e-3	MDM4, RNF8, ZNF420
	Glutathione redox reactions I	1.585e-3	GSTT2/GSTT2B, GSTT1
	Vascular endothelial growth factor signaling	2.692e-3	ROCK2, VCL, NOS3
	Glutathione-mediated detoxification	2.754e-3	GSTT2/GSTT2B, GSTT1
	MIF-mediated glucocorticoid regulation	3.548e-3	MIF, PLA2G2D
	MIF regulation of innate immunity	5.495e-3	MIF, PLA2G2D
	Actin nucleation by ARP-WASP complex	1.000e-2	ROCK2, ARPC4
	Citruelline-nitric oxide cycle	1.349e-2	NOS3
	Remodeling of epithelial adherens junctions	1.445e-2	VCL, ARPC4
Total collagen accumulation	MIF-mediated glucocorticoid regulation	2.884e-3	PLA2G4A, PTGS2
	MIF regulation of innate immunity	4.365e-3	PLA2G4A, PTGS2
	Branched-chain α-keto acid dehydrogenase complex	9.550e-3	BCKDHB
	Ascorbate recycling (cytosolic)	9.550e-3	GSTO2
	Endothelin-1 signaling	1.047e-2	PLA2G4A, ADCY5, PTGS2
	Eicosanoid signaling	1.148e-2	PLA2G4A, PTGS2
	Role of MAPK signaling in the pathogenesis of influenza	1.318e-2	PLA2G4A, PTGS2
	Estrogen-dependent breast cancer signaling	1.479e-2	IGF1, HSD17B8
	Adenine and adenosine salvage III	1.660e-2	ADAL
	Purine ribonucleosides degradation to ribose-1-phosphate	1.905e-2	ADAL

α-SMA=alpha-smooth muscle actin, ATM=ataxia telangiectasia mutated, MAPK =mitogen-activated protein kinase, MIF=macrophage migration inhibitory factor.