



OPEN

Dysregulation of RNA splicing in early non-alcoholic fatty liver disease through hepatocellular carcinoma

Nicholas J. G. Webster^{1,2,3}✉, Deepak Kumar² & Panyisha Wu²

While changes in RNA splicing have been extensively studied in hepatocellular carcinoma (HCC), no studies have systematically investigated changes in RNA splicing during earlier liver disease. Mouse studies have shown that disruption of RNA splicing can trigger liver disease and we have shown that the splicing factor SRSF3 is decreased in the diseased human liver, so we profiled RNA splicing in liver samples from twenty-nine individuals with no-history of liver disease or varying degrees of non-alcoholic fatty liver disease (NAFLD). We compared our results with three publicly available transcriptome datasets that we re-analyzed for splicing events (SEs). We found many changes in SEs occurred during early liver disease, with fewer events occurring with the onset of inflammation and fibrosis. Many of these early SEs were enriched for SRSF3-dependent events and were associated with SRSF3 binding sites. Mapping the early and late changes to gene ontologies and pathways showed that the genes harboring these early SEs were involved in normal liver metabolism, whereas those harboring late SEs were involved in inflammation, fibrosis and proliferation. We compared the SEs with HCC data from the TCGA and observed that many of these early disease SEs are found in HCC samples and, furthermore, are correlated with disease survival. Changes in splicing factor expression are also observed, which may be associated with distinct subsets of the SEs. The maintenance of these SEs through the multi-year oncogenic process suggests that they may be causative. Understanding the role of these splice variants in metabolic liver disease progression may shed light on the triggers of liver disease progression and the pathogenesis of HCC.

Alternative RNA splicing is a crucial post-transcriptional process that plays a fundamental role in the regulation of gene expression and protein diversity. It allows a single gene to generate multiple mRNA transcripts by selectively including or excluding exons, or portions of exons, during processing of the primary RNA transcript. Alternative splicing has emerged as a key mechanism influencing the development and progression of liver diseases¹. Understanding the intricate interplay between alternative splicing and liver pathology is essential for uncovering novel therapeutic targets and advancing personalized medicine approaches for these conditions². Liver disease encompasses a wide range of conditions, such as viral hepatitis, non-alcoholic fatty liver disease (NAFLD), alcoholic liver disease (ALD), and cirrhosis³⁻⁶. Alternative splicing has been implicated in the regulation of critical genes and pathways associated with liver diseases⁷⁻¹⁰. For instance, aberrant splicing events in genes involved in hepatic lipid metabolism can contribute to the pathogenesis of NAFLD and ALD¹¹. Additionally, dysregulation of splicing factors, which control the splicing process, has been observed in liver diseases, indicating their potential as diagnostic markers or therapeutic targets¹².

Liver cancer, predominantly hepatocellular carcinoma (HCC), is a major global health concern with limited treatment options and poor prognosis. Alternative splicing also plays a significant role in the development and progression of HCC by generating transcript isoforms that contribute to tumor heterogeneity and altered cellular functions¹³⁻¹⁸. Abnormal splicing events can affect genes involved in cell cycle regulation, apoptosis, cell adhesion, and angiogenesis, all of which are crucial processes for tumor growth and metastasis¹⁹. Furthermore, alternative splicing can lead to the production of tumor-specific antigens or fusion proteins that have potential

¹Jennifer Moreno VA Medical Center, San Diego, CA 92161, USA. ²Division of Endocrinology and Metabolism, Department of Medicine, University of California, San Diego, CA 92093, USA. ³Moore's Cancer Center, University of California, San Diego, CA 92093, USA. ✉email: nwebster@health.ucsd.edu

diagnostic or therapeutic implications. Targeting splicing events associated with liver cancer holds promise for developing novel precision medicine approaches and improving patient outcomes.

Recent advancements in high-throughput sequencing technologies and bioinformatics tools have facilitated comprehensive analyses of transcriptional changes in liver disease and liver cancer. Moreover, the integration of transcriptomic, proteomic, and genomic data has provided valuable insights into the pathogenesis of liver disease^{20–22}. Most of these studies, however, have focused on gene or protein level expression and have ignored transcriptional or protein isoforms and their complexities^{23–26}. Recent studies in mice have shown that RNA splicing is crucial for liver maturation and function but studies in humans are lacking^{27–29}. These studies have revealed complex splicing patterns, unique splice variants, and splicing factor dysregulation specific to the liver^{11,30–33}. However, further investigation is necessary to decipher the regulatory mechanisms underlying alternative splicing in liver diseases in humans and identify potential therapeutic targets. A deeper understanding of the role of alternative splicing in liver disease and cancer will pave the way for the development of precision medicine strategies tailored to individual patients. To this end, we profiled alternative splicing during early liver disease in patients with non-viral, non-alcoholic fatty liver disease (NAFLD) and non-alcoholic steatohepatitis (NASH).

Results

Changes in alternative splicing are enriched for SRSF3-dependent splicing events

We have previously reported that the splicing factor Ser/Arg-rich splicing factor 3 (SRSF3) is proteolytically degraded in fatty liver disease²⁹, so we obtained 43 frozen liver samples of individuals with no known history of liver disease, or histologically verified non-alcoholic fatty liver disease (NAFLD) or non-alcoholic steatohepatitis (NASH) from the University of Minnesota Liver Tissue and Cell Distribution Service. We were able to extract high-quality RNA from 29 of these samples (13 no history, 7 NAFLD, 9 NASH) which was then subjected to high-throughput short-read sequencing. The histology of the samples was confirmed by a pathologist or assessed for those samples without pathology information (Table S1). This resulted in the identification of 8 samples with normal liver histology, 8 samples with NAFLD and 13 samples with NASH. Reads were aligned to the human hg38 genome with STAR and gene count and splice junction files analyzed with Psychomics³⁴. Principal component analysis (PCA) of the splicing data (percent-spliced-in or PSI index) indicated the presence of three clusters of samples. One cluster contained three normal samples and one NAFLD, the second was a mixture of 5 normal and 6 NAFLD samples, and the third 1 NAFLD and 13 NASH (Fig. 1A). The clusters were labeled Normal, Early and Late disease, respectively, for further analysis. The contribution of PC1 and PC2 were modest (only 12.5% and 6.2%) on a dataset that contained 78,156 alternative splicing events (SEs) derived from 753,516 splice junctions, which was likely due to the sparsity of values for many SEs. So, we used dimension reduction to filter out SEs with more than 5 missing values reducing the dataset to 6797 SEs. When these were used for PCA analysis the relative contribution of PC1 increased to 29.2% indicating that the major difference was related to the progression from normal/early disease to late disease (Fig. S1A). We also plotted the PCA data by SRSF3 expression level, which we had previously measured by western blot²⁹ (Fig. 1B). The Normal cluster showed high SRSF3 expression, whereas the Late cluster showed predominantly low SRSF3 expression, with the Early cluster showing mixed expression. The SRSF3 protein level was significantly lower in the Late cluster than the Normal cluster (Fig. S1B). We assessed which SEs contributed to the PCA distribution (Table S2) and quantification of the top two contributing SEs indicated the use of an alternative first exon in the Late cluster in the TPM1 and SLCO2B1 transcripts (Fig. S1C). To visualize the variation in splicing across the samples we used unsupervised clustering of the top 500 SEs with < 5 missing values and a PSI variance > 0.01 (Fig. 1C). The samples clustered the same way as the PCA analysis with the Late cluster showing greatest dissimilarity to the Normal and Early clusters. The SEs clustered into six groups: two large groups 1 and 2 were SEs that were different in the Late cluster representing changes that occur in later disease; two groups 3 and 4 were SEs different in the Normal cluster representing changes that occur in early disease (Fig. S1D); and two small groups 5 and 6 which contained SEs that were only different in the Early cluster representing SEs that changed early in disease but were not maintained through later disease (Fig. S1E). To identify which SEs were significantly altered we performed a statistical comparison of all three clusters using the Kruskal–Wallis rank sum test (Table S3, 4652 significant SEs), and then compared individual clusters using t-test and Wilcoxon rank sum tests (Tables S4–6). We also analyzed a dataset from the ENCODE consortium, in which SRSF3 had been knocked down in human HepG2 hepatoma cells, to provide a list of SRSF3-dependent events (Table S7)³⁵. The SEs were filtered to only analyze events that were detected in both human liver and HepG2 cells (2872 sig SEs, 37,426 total SEs). SEs that were significantly different between pairs of clusters ($p < 0.05$) were then overlapped with each other and with SRSF3-dependent events from HepG2 cells (Fig. 1D). SRSF3-dependent SEs were significantly enriched in the overall cluster data (2872 SE, odds ratio OR 1.54, $p < 0.0001$ by χ^2) and were also enriched in the individual cluster comparisons, with enrichment being greatest for the Early cluster compared to the Normal cluster (OR 1.75, $p < 0.0001$). Changes in exon skipping were the predominant form of alternative splicing in all three comparisons followed by alternative 3' splice sites and first exons (Fig. 1E).

To gain an understanding of the functional effect of these changes we mapped the genes affected by the SEs to pathways and gene ontologies using Metascape³⁶. The genes affected by these SE changes also showed considerable overlap between clusters (Fig. S1F). Nineteen of the top 20 enriched terms were shared between clusters with RNA metabolism, response to stress, protein catabolism and transport being the most significant (Fig. 1F). We also created protein–protein interaction networks out of these genes using the Molecular Complex Detection algorithm (MCODE) and the STRING and BioGRID databases^{37–39}. Three of the top 8 MCODE interaction networks were involved in mRNA splicing, 3' end processing and translation (Fig. 1G). Other networks involved organelle fusion and vesicle transport, carbohydrate metabolism, MAPK signaling and cell cycle (Fig. S1G). Inspection of the networks involved in RNA processing highlighted several splicing factors that were themselves

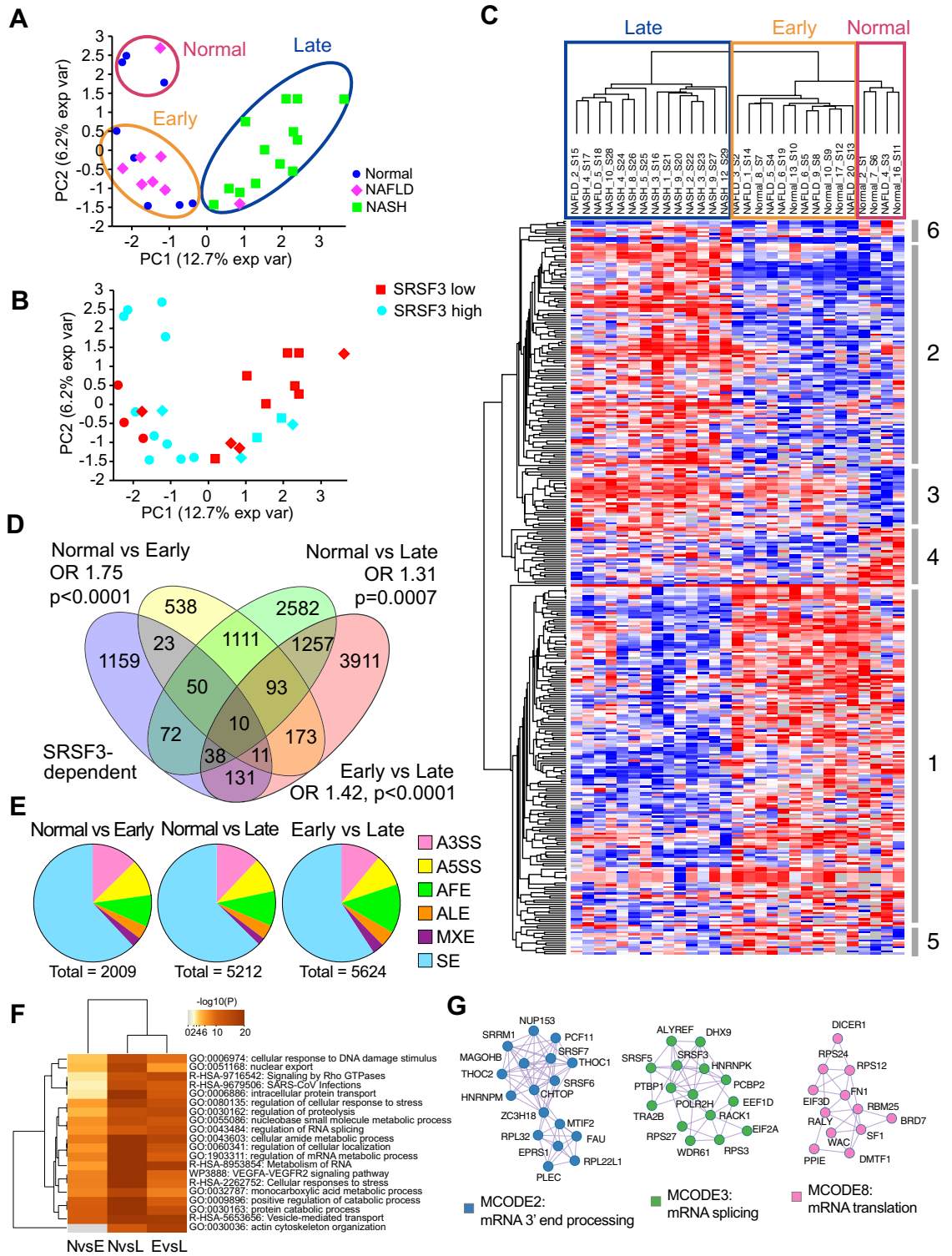


Figure 1. RNA splicing is altered in NAFLD and NASH: (A) PCA plot based on the analysis of alternative splicing events in RNA from liver samples from 8 individuals with normal liver (blue), 8 individuals with NAFLD (magenta) and 13 individuals with NASH (green). The percent-spliced-in (PSI) value for each event was calculated with Psichomics. K-means clustering highlighted three distinct clusters that are circled, which we have labelled Normal, Early or Late. (B) PCA plot color-coded for SRSF3 protein expression by western blot. SRSF3 expression below the mean is colored red, high expression above the mean is labeled cyan. (C) Heatmap of the top 500 SEs with < 5 missing values and PSI variance > 0.01. The samples clustered into the same three groups as the PCA. Red indicates high PSI and blue low PSI. (D) Venn diagram showing overlap of the statistically different SEs from the pairwise clusters comparisons and the set of SRSF3-dependent SEs from shRNA knockdown of SRSF3 in HepG2 cells. Enrichment for SRSF3-dependent SEs is indicated by the odds ratio (OR, χ^2 p value). (E) Pie charts showing distribution of splicing event types between clusters: alternative 3' splice site (A3SS), alternative 5'-splice site (A5SS), alternative first exon (AFE), alternative last exon (ALE), mutually-exclusive exons (MXE) and skipped exons (SE). (F) Heatmap showing enriched GO ontology terms from genes harboring the SEs. Color coding shows $-\log_{10}$ p-value for enrichment. NvsE, Normal vs Early; NvsL Normal vs Late; EvsL Early vs Late cluster. (G) Protein-protein interaction networks based on spliced genes. Nodes are shown as circles, interactions as lines.

subject to alternative splicing including SRSF3, SRSF6 and SRSF7, which contain poison exons causing non-sense mediated decay of the transcript, the spliceosome component SNRNP70, the ribosome subunit RPS24, and SRRM1 that bridges SR proteins to the snRNP complex (Fig. S1H).

The Psychomics algorithm analyzes individual splicing events but does not quantify more complex splicing variation. To assess more complex splicing patterns, we took two approaches based on the generation of complex splice graphs. These approaches use graph theory to derive alternative routes to generate known and novel splicing isoforms. Analysis of the clusters using MAJIQ⁴⁰ highlighted a smaller number of local splicing variations (LSVs) that were derived from the mapping of multiple splice junctions to the splice graphs (Tables S8–10). Out of 78,179 LSVs, 887 were significantly altered between the clusters and they were significantly enriched for SRSF3-dependent events as before (OR 1.58, $p = 0.0039$) (Fig. 2A). Comparison of the Early and Late clusters showed significant enrichment ($p = 0.011$) of SRSF3-dependent splicing events, but the other comparisons did not ($p = 0.068$ and 0.065). As an example, the TPM1 splicing event that contributed most to the PCA analysis was also found in the LSVs showing a switch in splicing of upstream exons in the Late cluster using a more proximal first exon (Fig. 2B). Gratifyingly, known SRSF3-dependent events were also found. Splicing of exon 13 in the SLK1 gene was altered with greater skipping of this exon in the Late cluster (Fig. 2C). The genes affected by these LSVs also showed overlap with the greatest number of changes occurring between the Normal and Late clusters (Fig. S2A). The functional effect of the genes harboring these splicing events was again assessed using Metascape. Early changes were enriched for genes in metabolic pathways and networks including blood clotting, nitrogen and organic acid metabolism, hormone stimulation and cell shape (Fig. S2B and C) whereas later changes between the Normal or Early clusters and the Late cluster showed enrichment for genes in RNA processing including splicing, miRNA production, ribonucleoprotein complex assembly, and nucleocytoplasmic transport, and genes involved in regulating cell shape including morphogenesis, actin cytoskeletal rearrangement and cell-substrate interactions. Protein–protein interaction network analysis highlighted networks involved in mRNA splicing, cytoskeletal rearrangement, vesicle trafficking, mitotic centrosomes and xenobiotic metabolism (Fig. S2D).

A similar splice graph analysis was performed using Whippet⁴¹ and comparison with the ENCODE data again showed enrichment for SRSF3-dependent events in the overall dataset (Fig. 2D, Tables S11–13). Interestingly, this approach found more differences between the Normal and Early clusters (Figs. 2D and S2E). This analysis identifies alternative transcription start and termination sites in addition to the other SEs, and these events were more frequent than cassette exon skipping that was the major class in the Psychomics analysis (Fig. 2E). The analysis also provides a measure of splicing entropy that is related to the complexity of the splicing event. Interestingly, the splicing events that showed the greatest change in PSI had lower entropy, i.e. simpler splicing events tended to show greater changes (Fig. 2F). Functional analysis highlighted terms and pathways involving lipid metabolism, phospholipase signaling, and small molecule hydroxy acid metabolism in the early changes, with enrichment of immune response, vesicle transport, complement, chromatin modification, and mitosis in later changes (Fig. S2F and G). Protein–protein interaction network mapping with the genes harboring high entropy splicing events highlighted interaction networks incorporating proteins with known functions in the liver. The largest network included proteins involved in meta-phase progression and was centered on beta-catenin, which is the most frequently mutated gene in HCC (Fig. S2H). Other networks involved PI3K-Akt signaling, mRNA splicing and translation. These analyses showed that changes in early disease samples tended to be in metabolic pathways suggesting that changes in liver metabolism are some of the early steps, whereas changes in later disease reflected cytoskeletal and matrix changes as well as immune responses. Some of these may reflect changes in the hepatocytes but could also be indicative of the increased immune cell infiltration, stellate cell activation and fibrosis that occurs in later disease. We were not able to compare the three sets of predicted splicing events directly as each algorithm uses a different naming system and chromosomal coordinates. We could, however, compare the genes containing the alternatively spliced events; Psychomics detected the greatest number of alternatively spliced genes (2912), which included 48% of the genes detected by Majiq but only 20% of the genes detected by Whippet.

Changes in gene expression are not enriched for SRSF3-dependent events

We also analyzed changes in gene expression in these samples (Table S14). PCA analysis did not reveal the same separation as was observed with the splicing events (Fig. 3A) nor the same dependence on SRSF3 protein levels (Fig. S3A). Hierarchical clustering of the expression data, however, did show a separation of the NASH samples versus the Normal/NAFLD samples (Fig. 3B). The genes that contributed most to the PCA were distinct from the splicing events with many genes involved in inflammation and the immune response (Fig. S3C, Table S15). We also used t-distributed stochastic neighbor embedding (tSNE) (Fig. S3D) based on the expression levels and this confirmed that the samples split into two groups, the NASH group equivalent to the Late splicing cluster, and a normal/NAFLD group equivalent to the Normal and Early splicing clusters (Fig. S3B). Interestingly, six of the 20 most significant up-regulated genes in NASH encoded proteins involved in the extracellular matrix (*MMP7*, *ITGBL1*, *LUM*, *CCDC80*, *CDH6*, *MUC6*) consistent with fibrosis (Fig. 3C), whereas eight of the twenty top down-regulated genes (*ORM1*, *SAA1*, *CRP*, *SAA2*, *LRG1*, *HAMP*, *SPINK1*, *FCN2*) encode for secreted proteins suggesting impaired hepatocyte secretory function (Fig. 3D). We analyzed the changes in gene expression between the Normal/NAFLD group and the NASH group then compared those to the SRSF3-dependent gene expression changes from the ENCODE dataset in HepG2 cells (Table S16). The transcriptomes of the liver samples and the HepG2 cells only showed partial overlap with 82% of liver genes being detected in HepG2 cells but only 34% of HepG2 genes being detected in primary liver samples (Fig. 3E). Unlike the splicing events, SRSF3-dependent gene expression changes were not significantly enriched in the NASH-dependent genes (42% SRSF3-dep, $p = 0.98$ by χ^2 test). We subjected the top 1000 genes that were over-expressed or under-expressed in NASH to pathway and network mapping. The genes that were higher in the Normal/NAFLD group were enriched for genesets and pathways involved in liver function or were decreased in hepatoblastoma, whereas

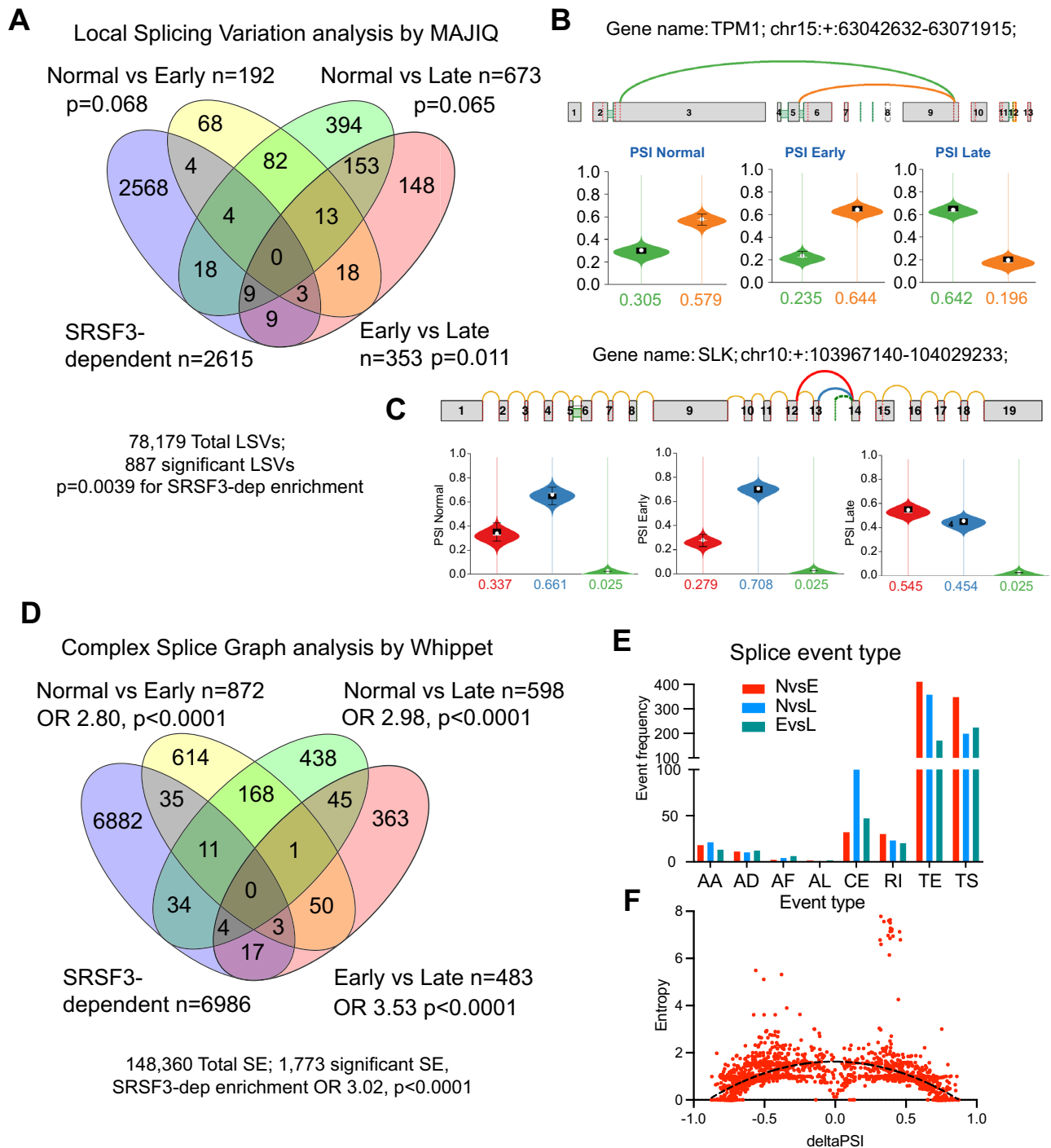


Figure 2. Changes in RNA splicing by splice graph. **(A)** Venn diagram showing overlap of local splicing variants (LSV) by MAJIQ (78,179 total LSVs, 887 significant LSVs). Overlap and enrichment χ^2 p-value with LSVs from SRSF3 shRNA knockdown in HepG2 cells is shown for individual cluster comparisons (Normal, Early or Late) and overall. **(B)** Change in upstream exon utilization in *TPM1* RNA. Use of upstream exon splice site is shown in green, downstream exon in orange. Graphs underneath show PSI values for each event in the three clusters. **(C)** Skipped exon 13 in the *SLK* RNA. Exon skipping is shown in red, inclusion in blue. Green indicates a cryptic splice exon. **(D)** Venn diagram showing overlap of alternative splicing by Whippet (148,360 total SEs, 1,773 significant SEs). Overlap and enrichment χ^2 p-value with SEs from SRSF3 shRNA knockdown in HepG2 cells is shown for individual cluster comparisons and overall. **(E)** Distribution of significant splicing events across cluster comparisons. AA alternative splice acceptor, AD alternative splice donor, AF alternative first exon, AL alternative last exon, CE cassette exon, RI retained intron, TE transcriptional termination exon, and TS transcriptional start site. NvsE, Normal vs Early; NvsL Normal vs Late; EvsL Early vs Late cluster. **(F)** Graph of splice event entropy (complexity) versus PSI for significant SEs. Dotted line is quadratic curve fit (p<0.0001).

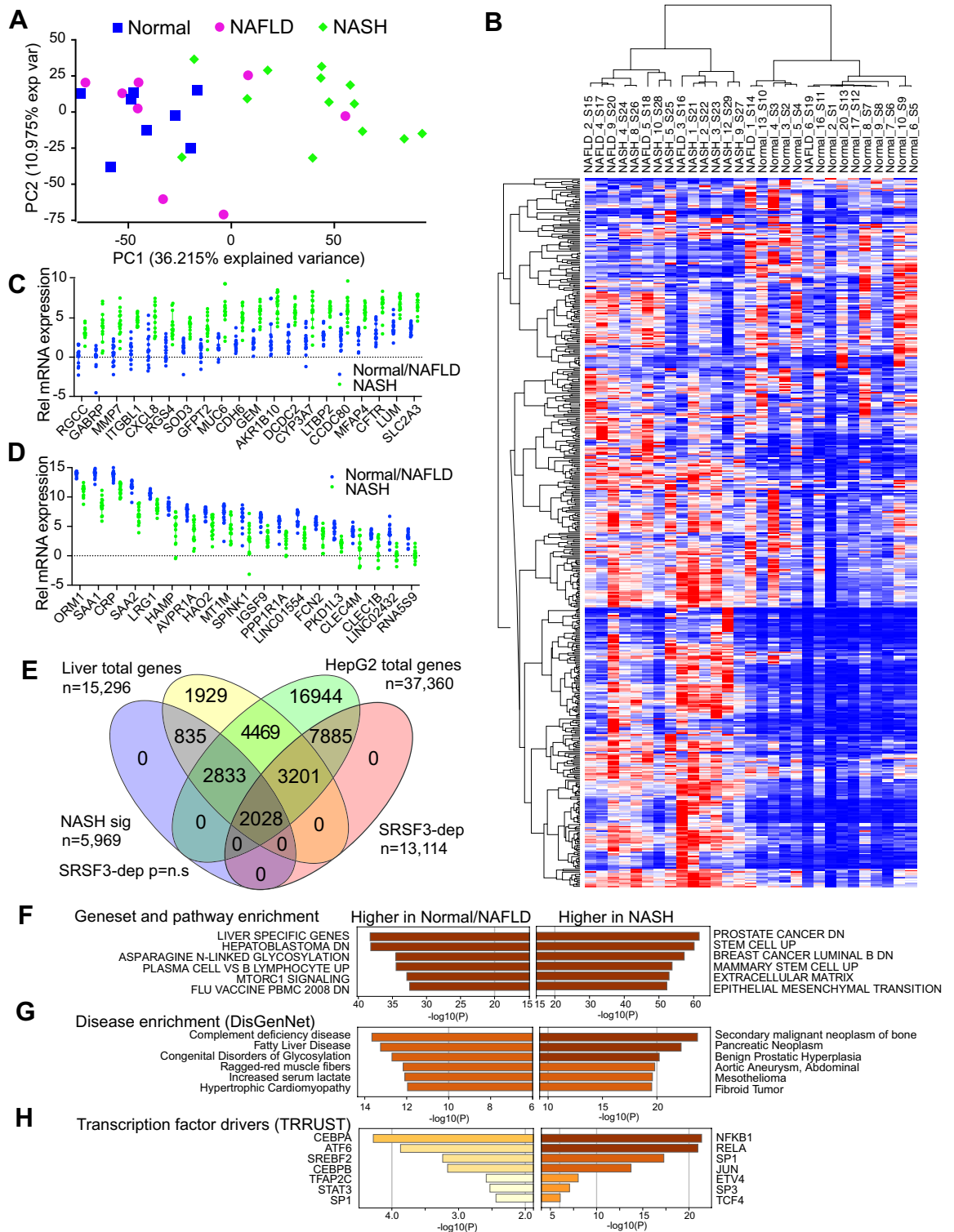


Figure 3. Transcriptional changes in liver disease. **(A)** PCA plot based on gene expression levels from gene-wise linear model fit and moderated t-test by empirical Bayes. Individuals with no history of liver disease (blue), individuals with NAFLD (magenta) and individuals with NASH (green). **(B)** Heatmap from unsupervised clustering of gene expression data. Red indicates high and blue low expression. **(C)** Relative mRNA expression of the top twenty genes showing increased expression in the NASH cluster (green) versus the Normal/NAFLD cluster (blue). **(D)** Relative mRNA expression of the top twenty genes showing decreased expression in the NASH cluster (green) versus the normal/NAFLD cluster (blue). **(E)** Venn diagram showing overlap of gene expression showing total and significant liver genes (NASH vs Normal/NAFLD) and total and significant gene from SRSF3 shRNA knockdown in HepG2 cells. $-\log_{10} P$ value is shown on x axis, color indicates level of significance. **(F)** Geneset and pathway enrichment for gene higher in Normal/NAFLD vs higher in NASH. **(G)** Disease enrichment using DisGenNet for genes higher in NAFLD/Normal vs higher in NASH. **(H)** Transcription factors driving altered gene expression in Normal/NAFLD vs NASH from TRRUST.

the genes increased in NASH were enriched for cancer, stem cell, and extracellular matrix pathways (Fig. 3F). Mapping the differential expression to the DisGenNet disease signature database⁴² showed that the Normal/NAFLD genes were enriched for signatures of complement deficiency and fatty liver disease, whereas the NASH genes were enriched for neoplastic signatures (Fig. 3G) and analysis of the altered genes for transcription factor drivers showed that genes in the Normal/NAFLD group were driven by metabolism related transcription factors such as C/EBP α / β , ATF6 and SREBP2, whereas genes in the NASH group were driven by inflammatory pathways such as NF κ B, and RelA (Fig. 3H).

At the cell level, the genes that were higher in Normal/NAFLD were enriched for hepatocyte or hepatoblast cell signatures whereas those higher in NASH were enriched for stromal, immune, and stellate cell signatures (Fig. 4A). To confirm the identity of the clusters, the expression data was deconvoluted using CIBERSORT to predict immune cell composition⁴⁴. The total predicted immune cell count was greater in the NASH cluster than the Normal/NAFLD, with an increase in CD8 and CD4 T cells and dendritic cells, and a decrease in M2 macrophages and mast cells confirming the hepatitis that is characteristic of NASH (Fig. 4B). The immune cell changes were maintained if the data were analyzed according to the splicing clusters (Fig. S3E). We also deconvoluted the expression data using a signature matrix generated from a single cell sequencing study of human liver (Fig. 4C)⁴³. All livers showed predominance of hepatocytes (Hepatocytes_6) that are characterized by expression of *ALB*, *ALDH1L1*, *ADH1B*, *XIST* and *APOC3* (Figs. 4D and S3F). The Normal/NAFLD cluster samples also showed the presence of hepatocytes that have high expression of *CYP2A6*, *CYP2A7*, *CYP2B6*, *CFHR1/3* and *GSTA1* (Hepatocytes_14) and are characteristic of hepatocytes involved in xenobiotic, steroid, fatty acid and drug metabolism (Figs. 4D and S3F). This cluster of hepatocytes decreases in the NASH samples and is replaced by a population of hepatocytes with high expression of *HMGCS1*, *SCD* and *FABP1* (Hepatocytes_5) characteristic of Zone 1 hepatocytes (Figs. 4D and S3F). Similar cell populations were identified when a cell signature matrix was derived from the human normal Liver Cell Atlas sequencing study (Fig. S3G)⁴⁵. These results indicate that the samples from NASH patients have higher immune cell infiltration, greater fibrosis, impaired secretory function, and higher neoplastic gene expression. Gene expression in the Normal/NAFLD samples is driven by liver-specific transcription factors but NASH samples are driven by inflammatory signaling transcription factors.

Alternative splicing events correlate with SRSF3 binding

As SRSF3-dependent SEs were enriched in the changes in RNA splicing during early liver disease, we sought to understand how SRSF3 could contribute to the changes in RNA splicing. We performed eCLIP with antibodies to SRSF3 in mouse primary hepatocytes⁴⁶. Most of the binding peaks mapped to the coding-region of genes (Fig. 5A) and within 100 nucleotides of the intron splice sites (Fig. 5B). We assessed SRSF3 binding to three known target genes *Fn1*, *Slk* and *Myo1b* (Fig. 5C): SRSF3 represses inclusion of exon 32 in the *Fn1* gene and SRSF3 showed binding to the flanking exons 31 and 33 but not to exon 32; SRSF3 also represses inclusion of exon 23 in the *Myo1b* gene and SRSF3 shows a strong binding peak to the upstream exon 22; in contrast SRSF3 stimulates incorporation of exon 13 in the *Slk1* gene and SRSF3 showed binding to all three exons 12, 13 and 14. Analysis of the SRSF3 eCLIP data using Pureclip⁴⁷ identified 13,119 crosslink sites encompassing 8,859 binding regions (Table S17). Sequences enriched at the crosslink sites were analyzed using DREME⁴⁸ to identify conserved motifs and FIMO⁴⁹ to find individual occurrences (Table S18). The top scoring motif CU/AUC was found in 1418 binding sites ($p = 1.3e-14$). Other motifs found include additional pyrimidine-rich motifs (UUUCUA/GC, UUU/CAUC, GGU/AUUUG) and purine-rich motifs (GAAGGA, GGUG/AAG) (Fig. 5D). The pyrimidine-rich motifs match the SRSF3 consensus motif found by SELEX (WCWWC), whereas the purine rich motifs are similar to motifs bound by other SR proteins such as SRSF1. To investigate whether SRSF3 binding correlated with SRSF3-dependent splicing we analyzed RNAseq data from mouse *Srsf3-flox* hepatocytes with Adenoviral-CRE mediated deletion of SRSF3. The RNAseq data were analyzed using the same approach used for the human samples. Splicing events were called as SRSF3-dependent if altered in the SRSF3 knockout hepatocytes with $p < 0.05$, and an SRSF3 binding peak had to map to < 1000 nucleotides from a splicing event for it to be called positive. SRSF3-dependent splicing events were significantly enriched for SRSF3 binding peaks (OR 2.2, $p < 0.0001$, Fig. 5E). We had previously published that SRSF3 protein is reduced in livers of obese mice on high-fat diet (HFD) and we recently showed that multiple lipid components of the diet can lead to proteosomal degradation of SRSF3 (Fig. S4). Consequently, we found that SRSF3-dependent splicing events were enriched in livers from the HFD mice (OR 2.65, $p < 0.0001$, Fig. 5F). SRSF3 peaks were also enriched in the altered splicing events in the HFD livers (OR 1.3, $p = 0.0007$, Fig. 5G). These results indicate that SRSF3 binding is enriched near both SRSF3- and HFD-dependent splicing events.

Changes in alternative splicing with degree of liver disease

Our data suggested that changes in SRSF3-dependent splicing events could be observed in early liver disease, so we investigated alterations in RNA splicing in three other liver transcriptome datasets to see if SRSF3-dependence correlated with early versus late disease. Suppli et al. investigated the transcriptional profile of livers from 14 healthy lean individuals, 12 obese individuals without liver disease, 15 individuals with biopsy-proven NAFLD and 16 individuals with biopsy-proven NASH²⁶. We reanalyzed their data (38,732 total SE) to identify changes in RNA splicing between the four groups (Table S19) and compared the splicing events to SRSF3-dependent events derived from HepG2 cells (Fig. 6A). SRSF3-dependent SE were almost twofold enriched in the obese group compared to the lean healthy group (OR 1.88, $p = 0.0003$), but were very significantly depleted in NAFLD and NASH groups (OR 0.19, $p < 0.0001$ and OR 0.00, $p < 0.0001$, respectively). Hoang et al. profiled 78 individuals with varying degrees of biopsy-proven NAFLD and stratified them by steatosis, fibrosis, ballooning, and inflammation scores²⁴. We analyzed this dataset (80,172 total SE) for changes in SEs with each of these scores (Table S20) and compared to the SRSF3-dependent SEs from HepG2 cells. SEs that were altered with steatosis or ballooning

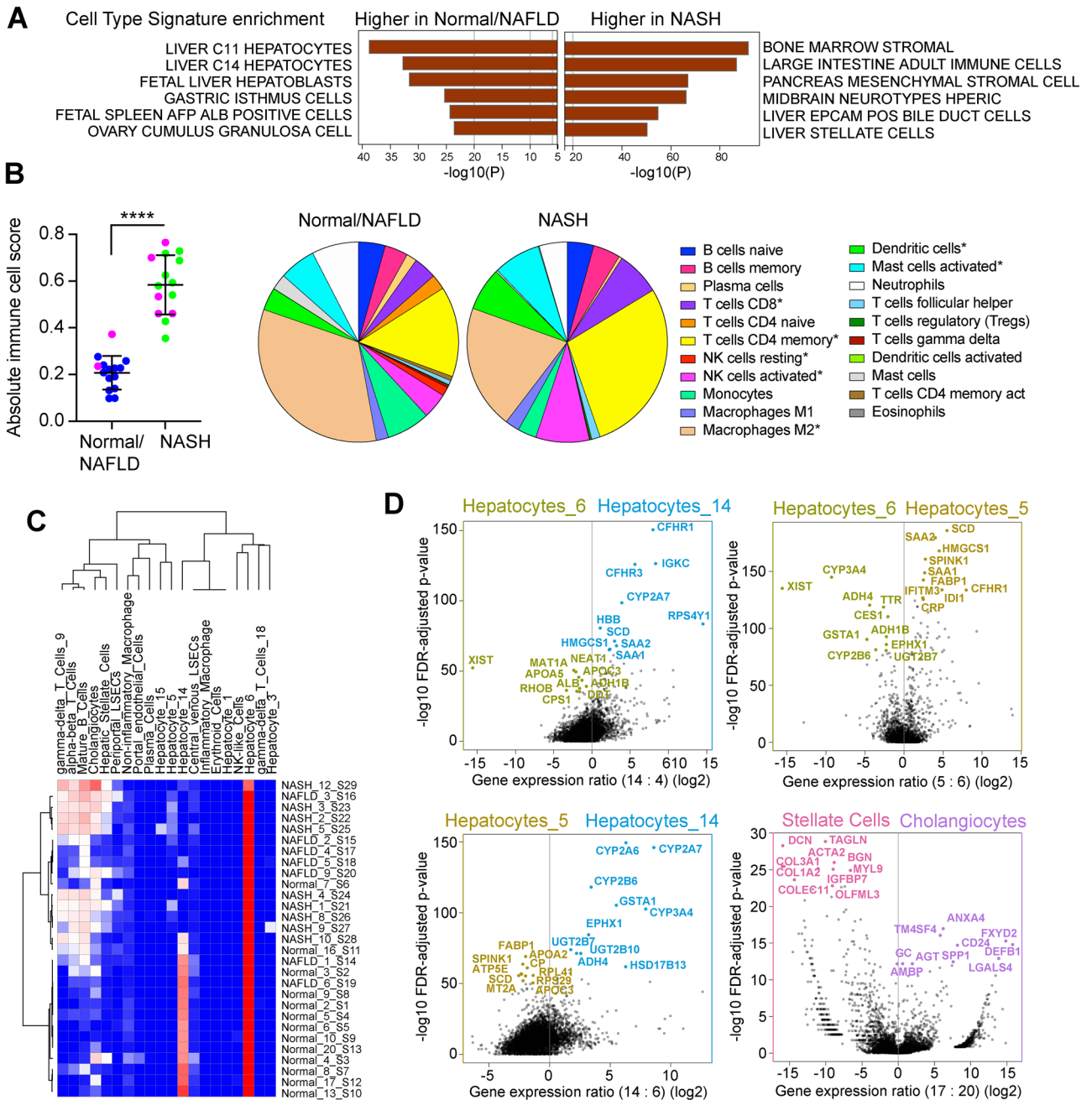


Figure 4. Liver and immune cells signatures. (A) Enrichment of cell-type signatures in gene higher in Normal/NAFLD vs higher in NASH. $-\log_{10}$ P value is shown on x axis. (B) Immune cell score and immune cell composition in Normal/NAFLD and NASH groups using CIBERSORT. Pie charts show relative immune cell composition. Asterisks indicate statistical significance difference in NASH vs Normal/NAFLD by t-test, **** $p < 0.0001$, * $p < 0.05$. (C) Heatmap showing unsupervised clustering of liver cell composition of each sample using CIBERSORT and the cell signatures from the Bader Lab Liver Atlas⁴³. Samples cluster into Normal/NAFLD and NASH groups as before. Blue indicates low relative cell content, red high. (D) Pairwise comparisons of gene expression in hepatocyte clusters 6, 5 and 14 or a comparison of stellate cells vs cholangiocytes. Graphs plot $-\log_{10}$ FDR-adjusted P value against \log_2 gene expression ratio. Top differentially expressed genes are shown.

were significantly enriched for SRSF3-dependent SEs (OR 1.43, $p = 0.0018$; OR 1.31, $p = 0.021$; respectively) but were not enriched when analyzed by fibrosis or inflammation scores (Fig. 6B). In the largest study, Govaere et al. analyzed 206 individuals with liver disease (51 NAFLD, 155 NASH) and 10 healthy controls²⁵. SEs that were significantly altered (out of 85,026 total SE) in the comparison between groups (Table S21) or between NAS scores were enriched for SRSF3-dependent SEs (OR 1.55, $p < 0.0001$ and OR 1.38, $p = 0.0007$, respectively) but no enrichment was seen based on fibrosis scores (Fig. 6C). In all three studies, enrichment of SRSF3-dependent

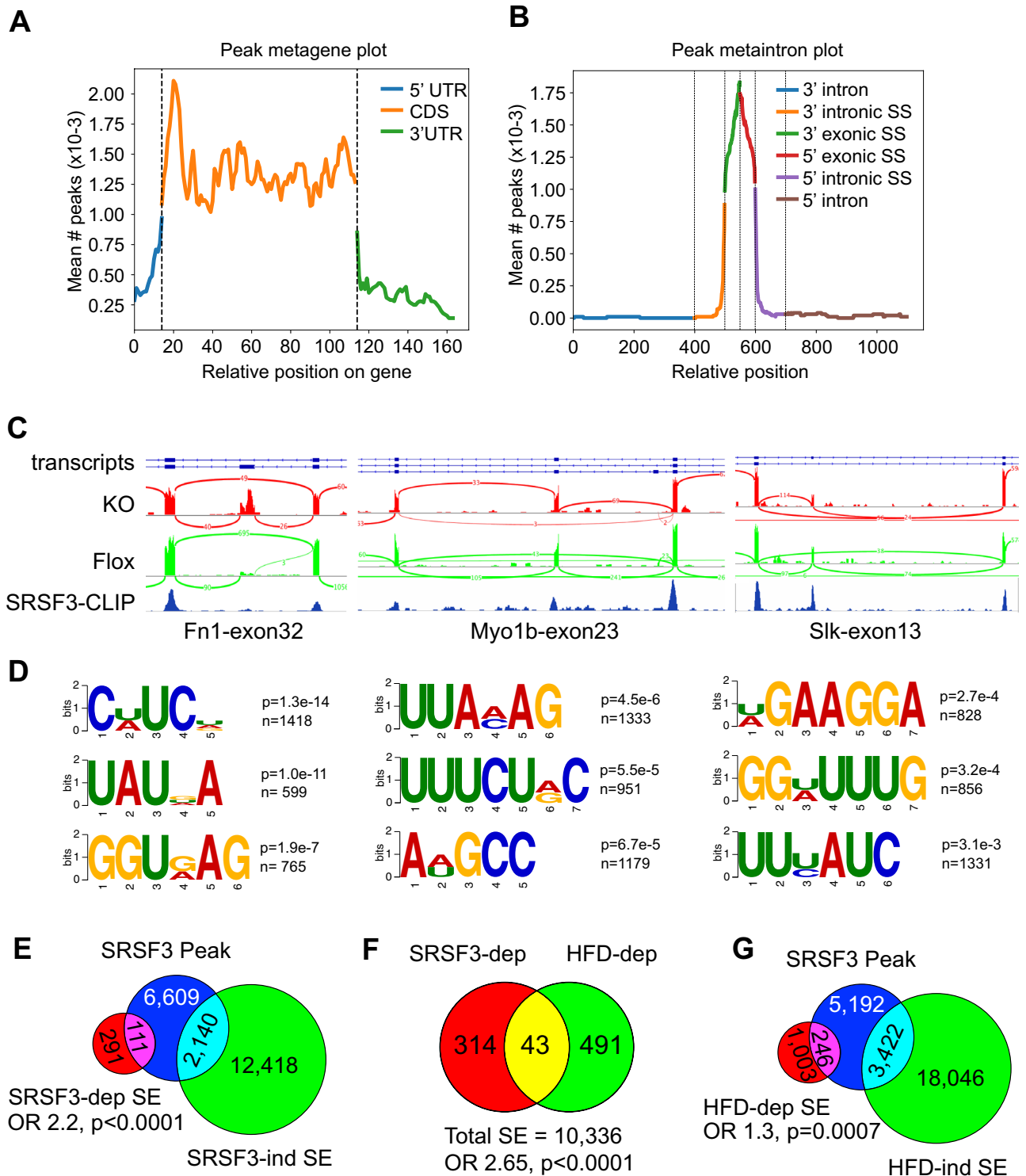


Figure 5. eCLIP analysis of SRSF3 binding sites in mouse hepatocytes. **(A)** Metagene plot showing distribution of SRSF3 binding peaks by eCLIP relative to gene 5'UTR, coding region or 3'UTR. **(B)** Metaintron plot showing position of SRSF3 binding peaks relative to exon/intron boundaries. **(C)** SRSF3 binding peaks in the Fn1, Myo1b and Slk compared to expression in flox and SRSF3 knockout hepatocytes. Transcript structures are shown above, sashimi plot of RNAseq reads are shown in red for SRSF3 KO and green for Flox mice, SRSF3-CLIP peaks are shown underneath in blue. **(D)** SRSF3 binding motifs from Pureclip analysis of the cross-linking sites. Enrichment p value and number of sites is given. **(E)** Overlap of SRSF3 CLIP peaks with SRSF3-dependent (Kruskal $p < 0.05$) and independent SEs. **(F)** Overlap of SRSF3-dependent and HFD-dependent liver splicing events. **(G)** Overlap of SRSF3 CLIP peaks with HFD-dependent (Kruskal $p < 0.05$) and HFD-independent SEs in liver.

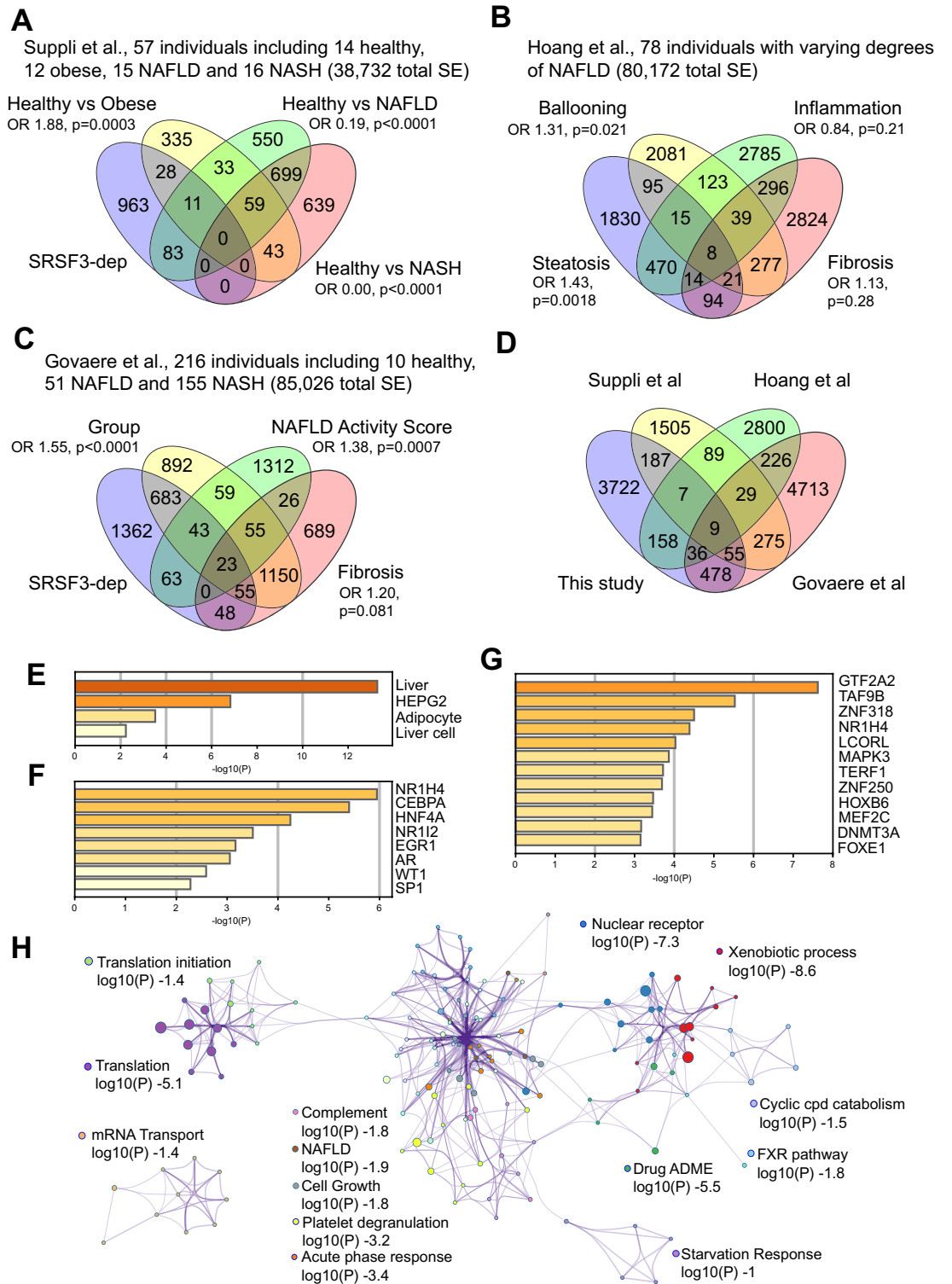


Figure 6. Confirmation of altered RNA splicing and SRSF3-dependence in three other NASH datasets. (A) Analysis of RNA splicing in Suppli et al. in 57 individuals using the PSI approach in psychomics. Venn diagram shows overlap of SEs altered in Obese, NAFLD or NASH groups relative to Healthy lean controls with SRSF3-dependent SEs from HepG2 cells. Odds ratio (OR) for enrichment of SRSF3-dep SEs in each group is shown with p value. (B) Analysis of RNA splicing in Hoang et al. in 78 individuals with varying degrees of NAFLD using the PSI approach. Venn diagram shows overlap of SEs altered based on steatosis score, hepatocyte ballooning, inflammation score, or fibrosis score with SRSF3-dependent SEs from HepG2 cells. (C) Analysis of RNA splicing in Govaere et al. of 216 individuals with NAFLD or various degrees of NASH using the PSI approach. Venn diagram shows overlap of SEs altered based on group, fibrosis score or NAFLD Activity Score with SRSF3-dependent SEs from HepG2 cells. (D) Overlap of altered SEs in the four studies based on a Kruskal $p < 0.05$ threshold. (E) Analysis of 136 SEs that are common to 3 or more studies by cell type. (F) Analysis of transcription factor driving expression of the genes harboring the 136 SEs. (G) Analysis of transcription factor targets of the genes with the 136 SEs. (H) Network showing clusters of GO terms that are significantly enriched in the 136 common SEs.

SEs was seen with NAFLD (NAS, steatosis, ballooning) but not in NASH (fibrosis, inflammation) suggesting that changes in SRSF3-dependent splicing occur early in liver disease. We then compared the significant SEs between the four studies (Fig. 6D). Of the 30,436 SEs that were detected in the four datasets, only 9 SEs were common to all four studies, but 136 were significantly altered in at least three studies. For many of these, the SE did not quite reach significance for one out of the four studies. For example, alternative splicing of an upstream exon in the TPM1 gene was significantly increased in three datasets but did not quite reach significance in the Suppli dataset ($p=0.053$), whereas that in the PANK1 did not reach significance in the Govaere dataset ($p=0.055$) (Fig. S5A and B). In contrast, alternative splicing of a mutually exclusive exon in the IRF3 gene was significant in three datasets but was not detected in our dataset (Fig. S5C).

The differences in SEs between the groups, especially those with more advanced liver disease, could be explained by the presence of infiltrating immune cells or activation of Kupffer or hepatic stellate cells. So, we analyzed the tissue or cell of origin of the genes harboring the 136 common SE changes. The greatest enrichment was for SEs in genes that were specific to liver tissue, HepG2 cells or liver cells (Fig. 6E). We also observed enrichment of genes found in adipocytes, which would be consistent with the steatosis. The cell type signatures were primarily derived from hepatocytes (Fig. S5D). We confirmed this conclusion by analyzing the transcription factors that regulate expression of the genes harboring the 136 SEs. The top four enriched transcription factors based on TTRUST⁵⁰ are liver specific (NR1H4/FXR, HNF4A, CEBPA, NR1I2/PXR) (Fig. 6F) and the top transcription factor targets based on GTRD ChIP-seq data are GTF2A2, TAF9B, ZNF318, and NR1H4/FXR (Fig. 6G). To assess which pathways or processes might be affected by the changes in SE, the genes containing the 136 common SEs were mapped to gene ontologies then clusters of significant GO terms and their interactions were plotted as a network (Fig. 6H). This analysis highlighted a large network of GO clusters representing metabolic functions of the liver, including complement, platelet degranulation, acute phase response, xenobiotic and drug metabolism, nuclear receptor and FXR pathways, and NAFLD (Fig. 6H). There were also subclusters centered on mRNA transport and translation. Mapping the SEs to DisGeNet highlighted chronic liver disease, liver injury, hepatitis, and hypoglycemia among the major disease association (Fig. S5E). From this we conclude that the SEs common to multiple studies are likely derived from hepatocytes rather than infiltrating immune cells.

Early versus late disease changes in alternative splicing

As the previous analysis indicated that SRSF3-dependent SEs were more enriched with steatosis rather than with fibrosis or inflammation, we created heatmaps for the three additional datasets based on the top 500 SEs. For each dataset, the SEs were further filtered to remove SEs found in less than half the healthy control samples, or with PSI variances less than 0.001 (0.1% of max PSI). Visual inspection of the heatmaps shows that there is a major change in splicing in the transition from control to NAFLD in the Govaere dataset (Fig. 7A), from obese to NAFLD in the Suppli dataset (Fig. 7B), or from NAS 0 to NAS 1 in the Hoang dataset (Fig. S6A). We then focused on these early changes in splicing by comparing the SEs altered in these transitions (Fig. 7C). Only 145 early SEs were significantly altered in three or more datasets, and these showed enrichment for genes involved in carbohydrate metabolism, gluconeogenesis and glycogen synthesis, protein localization to organelles and membranes, and autophagy and ferroptosis (Fig. 7D) and were also associated with acute drug- or chemical-induced liver injury and hepatitis, as well as fibrosis and malignant neoplasms (Fig. 7E). To compare the common early SE changes to the total altered early SEs in the four individual datasets, we mapped all the early SEs for each dataset (FDR-adjusted) to GO biological processes. The Circos plot shows the overlap of the genes harboring the SEs based on shared functions or pathways (Fig. S6B). Mapping to GO terms showed enrichment of vesicle and Golgi transport, cellular responses to stress, and lipid metabolism in all four datasets and three of the datasets showed enrichment of small molecule metabolism (Fig. S6C). These results suggest that these early splicing alterations are involved in normal liver function and response to injury.

To understand which splicing factors might be driving these changes, we then analyzed expression of splicing factors and RNA-binding proteins during the early obese to NAFLD transition (Fig. 7F). Only four splicing factors, *A1CF*, *HNRNPAB*, *PABPN1* and *PCBP2* were altered in early disease in our data (Fig. 7G) and the other three datasets (Fig. S6D–F). An additional 24 splicing factors (*PCBP1*, *SFPQ*, *SRSF8*, *KHSRP*, *ESRP2*, *HNRNP2*, *KHDRBS1* (Sam68), *CELF2*, *ZCRB1*, *HNRNPUL1*, *RBFOX2*, *RBMS1*, *MATR3*, *RBM8A*, *RBM7*, *RBM38*, *HNRNPD*, *CPEB3*, *SART3*, *ZC3H14*, *CELF1*, *HNRNPM*, *CNOT4* and *SRRM2*) were altered in three out of four datasets.

We also analyzed the late changes in SEs that occur during the NASH progression with increased inflammation and fibrosis; NASH_F0/1 vs NASH_F3/4 in the Govaere dataset, NAFLD vs NASH in the Suppli dataset, or NAS 1–3 vs NAS 4–6 in the Hoang dataset. Thirty-six SEs were altered in three out of four datasets (Fig. 8A). The 36 genes harboring the common late SEs show enrichment in immune response genes and RHO GTPase signaling (Fig. 8B) that is not observed in the early SE changes and are associated with chronic liver disease, hepatitis, fatty liver disease, NASH, obesity, and many inflammatory and infectious diseases (Fig. 8C). Comparing all the late SEs in the four individual datasets, the Circos plot again shows extensive overlap (Fig. S7A) and the enriched terms now include chromatin and cytoskeletal organization and DNA damage in addition to the previous metabolic terms (Fig. S7B). These results suggest that the late splicing alterations are also involved in cytoskeletal and genomic changes, and DNA damage, indicative of more advanced disease. There were no splicing factors that were altered in all four late datasets but *HNRNPAB*, *RBM6*, *RBM5*, *SRSF5* and *RBFOX2* were altered in three out of four (Fig. 8D). *HNRNPAB* and *RBFOX2* were the only genes whose expression was altered in both early and late disease, and *RBM6*, *RBM5* and *SRSF5* were only altered in late disease (our data in Figs. 7F and 8E, other datasets in Fig. S7C–E).

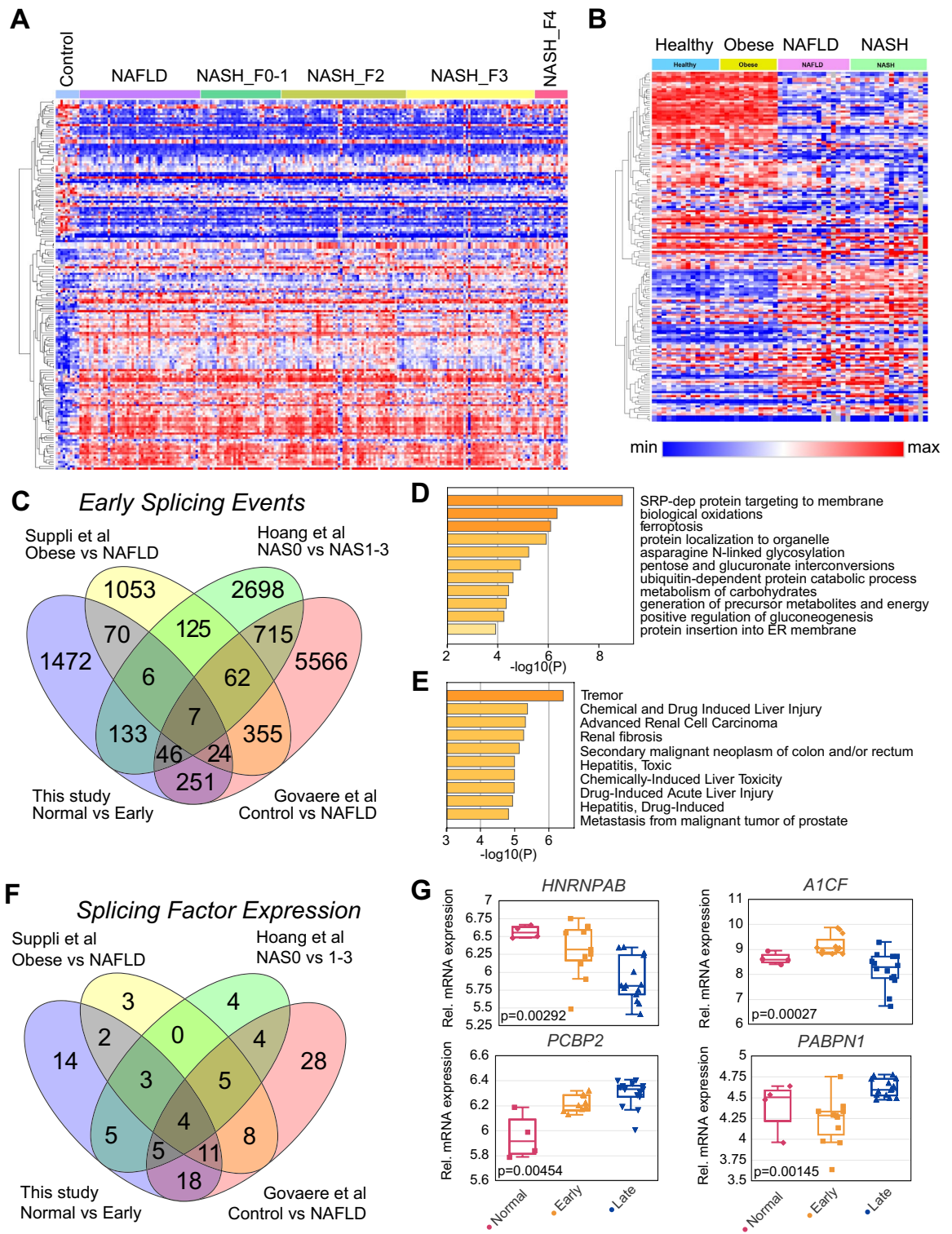


Figure 7. Changes in RNA splicing in early disease. **(A)** Heatmap showing top SEs in the Govaere et al. dataset clustered according to group. SEs cluster into two groups which either have increased or decreased PSI going from healthy controls to NAFLD. Red indicates high PSI, blue low PSI. **(B)** Heatmap showing top SEs in the Suppli et al. dataset clustered according to group. SEs cluster into two groups which either have increased or decreased PSI going from Obese to NAFLD. **(C)** Overlap of early SEs from the four studies based on a Kruskal $p < 0.05$ threshold. The groups compared are shown for each study. **(D)** and **(E)** Pathway and disease enrichment of genes harboring 145 early SEs common to 3 or more studies. **(F)** Overlap of changes in splicing factor expression during early disease in the four studies. Splicing factors analyzed were taken from Sebestyén et al.⁸⁶. **(G)** Relative expression of the four splicing factors (*HNRNPAB*, *A1CF*, *PCBP2* and *PABPN1*) that are altered in all four datasets in normal early or late clusters in our study. Statistical significance is for Cluster A vs Cluster B.

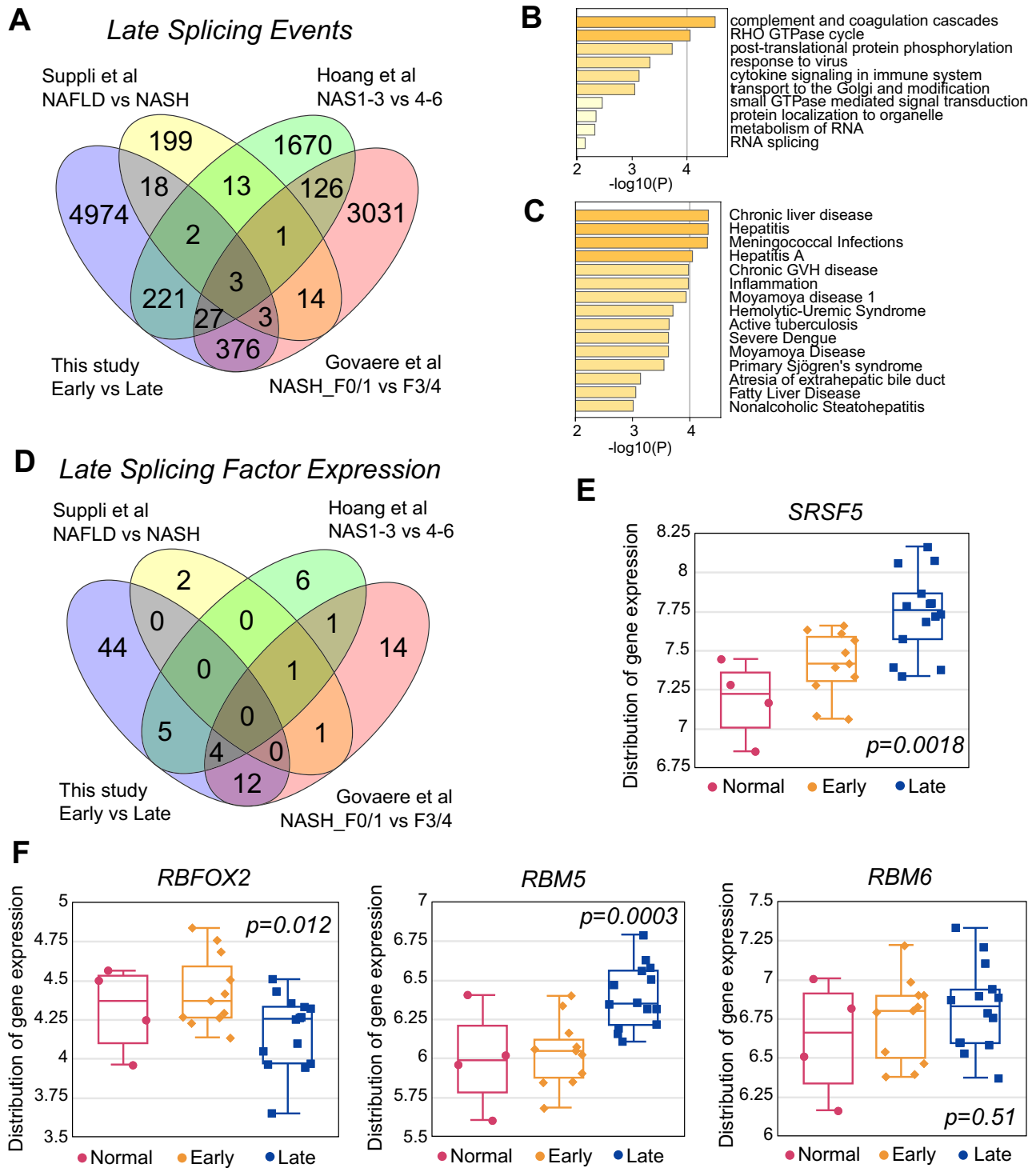


Figure 8. Changes in RNA splicing in late disease. (A) Overlap of late SEs from the four studies. The groups compared are shown for each study. (B) and (C) Pathway and disease enrichment of genes harboring 36 late SEs common to 3 or more studies. (D) Overlap of changes in splicing factor expression during late disease in the four studies. Splicing factors analyzed were taken from Sebestyén et al.⁸⁶. (E) and (F) Relative expression of the four splicing factors (*SRSF5*, *RBFOX2*, *RBM5* and *RBM6*) that are altered in all four datasets in the normal, early or late clusters in our study. Statistical significance is for early vs late clusters.

Changes in alternative splicing in hepatocellular carcinoma

The changes in RNA splicing events that we identified occurred mainly during early liver disease (obesity to simple steatosis), so we then investigated whether any of the early changes would persist in hepatocellular

carcinoma. Analysis of the TCGA LIHC dataset⁵¹ using the same approach identified 159,714 splicing events of which 18,118 were altered (Kruskal $p < 0.05$) in tumors compared to normal adjacent tissue (Table S22). These SEs were compared to the significant SEs identified for each of the four NAFLD datasets (Fig. 9A). Thirty percent of the significant SEs from this study and Govaere et al. were also significantly altered in the LIHC data compared to 55% of the SEs from Suppli et al. and Hoang et al., indicating that many of these NASH-associated SEs are still present in HCC. Overall, 22% of the tumor-associated SEs were found in the NASH-associated SEs compared to 12% of total SEs, hence tumor-associated SEs were significantly enriched for NASH-associated SEs (OR 2.13, $p < 0.0001$). Looking more closely at early changes, we mapped the common 145 early SEs that were found in three or more NAFLD datasets to the HCC data; 75 out of the 145 SEs (52%) were significantly altered in the tumors in the TCGA LIHC dataset (Fig. 9B) compared to only 22% for other SEs (OR = 3.9, $\chi^2 p < 0.0001$). We calculated the survival log-rank p-value for each of these early SEs based on the optimal PSI cut-off. Twenty-eight SEs (37%) out of the 75 SEs were significantly associated with survival compared to only 26% of other tumor-associated SEs. Of the 28 SEs that are associated with survival, 23 SEs cause alterations in the protein sequence (Table S23). Six of the top early NAFLD-associated SEs that are conserved in HCC and associated with survival are (i) an alternative first exon that deletes a ubiquitin-like domain in the *FBXO7* gene, an F-box containing protein that interacts with the ubiquitin ligase complex SCF, (ii) an alternative first exon in the *EPHX1* gene that encodes epoxide hydrolase 1, (iii) an alternative first exon on the myosin-1b (*MYO1B*) mRNA that leads to the production of a protein containing only the myosin head domain, (iv) an alternative 3' splice site in the first exon of the *FUOM* fucose mutarotase protein, (v) an alternative first exon in the *ATP6V0E1* gene that deletes a transmembrane helix, and (vi) a profibrogenic cassette exon in the fibronectin 1 (*FN1*) gene (Fig. 9C).

We also looked at expression of the splicing factors that were altered in the NAFLD datasets in the HCC. The four splicing factors *HNRNPAB*, *PCBP2*, *AICF* and *PABPN1*, which were altered in all four NAFLD datasets, also have altered expression in HCC tumors but only *HNRNPAB* and *PCBP2* are associated with survival (Fig. 9D). We extended this analysis to splicing factors conserved in three out of four datasets (Fig. S8). For ten splicing factors, *SFPQ*, *KHDRBS1*, *HNRNPM*, *HNRNPUL1*, *RBMS1*, *MATR3*, *RBM8A*, *SRSF8*, *SART3* and *ZCRB1*, higher expression was associated with poorer survival (Fig. S8A) whereas for six splicing factors, *CEL2F2*, *CPEB3*, *SRSF5*, *RBFOX2*, *ESRP2*, and *SRRM2*, higher expression was associated with better survival (Fig. S8B). Expression of *PCBP1*, *KHSRP*, *HNRNP2*, *RBM7*, *RBM38*, *HNRNPD*, *ZC3H14*, *CEL1*, *CNOT4*, *RBM5*, or *RBM6* was not associated with survival, independent of any change in expression in HCC. These results suggest that many early alterations in RNA splicing are conserved through disease progression and can influence survival.

Discussion

In this paper, we analyzed changes in RNA splicing in 29 samples of liver tissue from individuals with non-alcoholic fatty liver disease and found extensive changes in alternative splicing. We compared the splicing results to a parallel transcriptome analysis and found that while the transcriptome data showed good separation of individuals with NASH from those with NAFLD or normal individuals, there was no separation of normal versus NAFLD. The splicing analysis on the other hand allowed separation of the normal and NAFLD samples in addition to separating from the NASH samples. We demonstrated alterations in RNA splicing using three independent bioinformatic approaches and to test the generality of our findings, we confirmed changes in RNA splicing in three large liver transcriptome datasets. Most of the splicing changes in these larger datasets occurred with the transition to steatosis, with fewer changes occurring with the onset of fibrosis and inflammation. As might be expected, common splicing events that are altered early in liver disease are associated with metabolic pathways such as lipid and carbohydrate metabolism and acute liver injury, whereas those in later disease also show enrichment of chronic liver injury and inflammation, DNA damage and cytoskeletal reorganization. The changes in RNA splicing during early liver disease were significantly enriched for targets of the splicing factor SRSF3, which is consistent with our previously published finding that the SRSF3 protein is degraded during lipid overload or in NAFLD in both human and mouse samples²⁹.

In contrast, the late splicing changes that occurred with inflammation or fibrosis were not enriched for SRSF3 targets. The SRSF3 targets, however, only represented a small fraction of the total changes in splicing so other splicing factors might be involved and, indeed, we found that many RNA-binding proteins showed altered splicing in the NAFLD and NASH samples suggesting multifactorial dysregulation of RNA splicing. This conclusion is supported by other studies showing dysregulation of components of the splicing apparatus in NAFLD/NASH^{1,11,33}. The majority of these studies analyzed RNA expression by microarray or RNAseq but interpretation of the results is difficult as many RNA binding proteins autoregulate their own expression through inclusion of poison exons^{52–54}. Studies that evaluated splicing factor protein expression are far fewer, but Wang et al. observed the upregulation of NONO, SRSF6, HNRNPA2/B1, HNRNPH and confirmed the downregulation of SRSF3 protein in livers of mice with NAFLD³³. Indeed, DRK2 prevents nuclear localization of SRSF6, altering RNA splicing and aggravating NAFLD⁵⁵.

Many of the observed changes in splicing in early NAFLD/NASH are also observed in HCC and associated with survival. That these changes are maintained throughout the oncogenic process suggests that they may contribute to cancer development or support tumor growth. A limitation of this study is that it is difficult to predict the effect of altered splicing events on cellular or tissue phenotypes as the splicing events in most cases do not uniquely define the transcript isoform, and most alternative transcript isoforms have no functional annotation. Transcript isoform ratios can be estimated by deconvoluting the data but that only generates known isoforms and cannot predict novel transcript isoforms. Furthermore, without knowledge of the function of the protein isoforms it is not possible to predict functional changes. Given these limitations, interpretation is limited to assessing the functional roles of the genes that harbor the splicing events without knowing whether an event increases or decrease gene function. Nevertheless, the top NAFLD-associated splicing events that were conserved

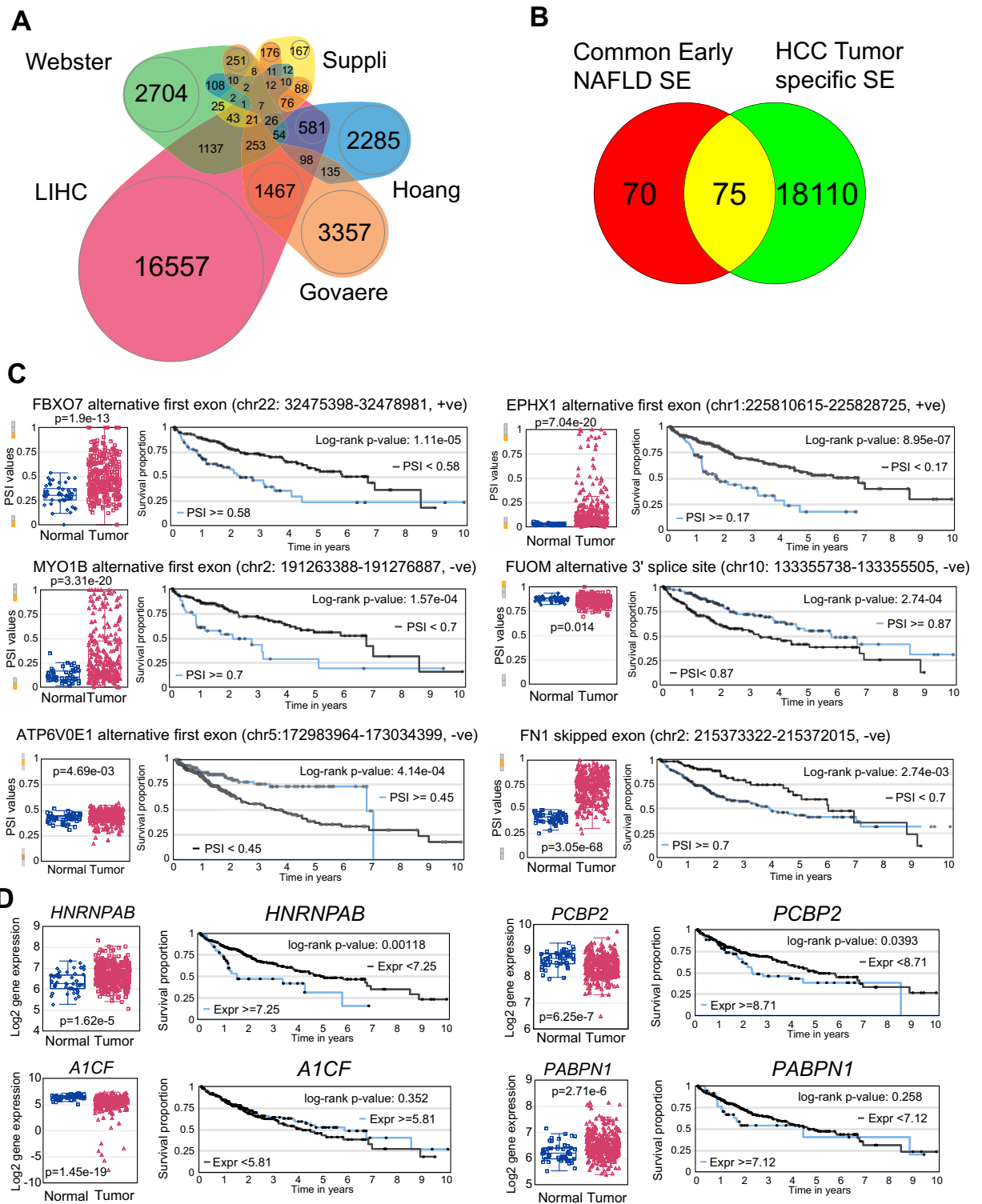


Figure 9. Altered RNA splicing events in early metabolic liver disease are conserved in HCC. **(A)** Overlap of total SEs in the four liver datasets with changes in splicing in the TCGA HCC data (LIHC). **(B)** Overlap of the 145 common early SEs with changes in splicing in HCC data. **(C)** Six of the top common early SEs that correlate with survival. Box plot shows individual PSI values for normal liver and tumor in the HCC data. Survival curves are based on optimal PSI values for each SE. Log-rank p-value is shown. **(D)** Expression of the four common early splicing factors in the HCC data. Box plot shows individual log2 expression values for normal liver and tumor. Survival curves are based on optimal expression cut-off for each gene. Log-rank p-value is shown.

in HCC and correlated with survival have been implicated in either cellular proliferation or cancer. FBXO7 gene regulates ubiquitin-dependent CyclinD/CDK6 degradation and maintains mesenchymal gene expression^{56,57} and mutations have been associated with Parkinson's disease but its role in the liver has not been determined. EPHX1 is downregulated in HCC and reduces tumorigenicity in mice but may contribute to chemoresistance⁵⁸. FUOM isomerizes α - and β -fucose for N- and O-linked fucosylation that is increased in HCC⁵⁹. Myosin-1b (MYO1B) promotes cancer by targeting HIF1a and SNAI2/cyclinD1^{60,61}. ATP6V0E1 encodes the E subunit of the vacuolar ATPase that is important for cancer cell survival in pancreatic cancer⁶². Lastly, Fibronectin 1 (FN1) can have positive or negative effects on cancer growth depending on the tissue⁶³.

The observed changes in splicing factor expression are also maintained in HCC and are associated with survival. Some of these have documented roles in HCC development. For example, A1CF promotes NASH and HCC⁶⁴, ESRP2 loss promotes HCC and lower expression predicts worse survival³², HNRNPAB induces EMT and promotes HCC metastasis⁶⁵, the poly(C)-binding protein PCBP2 is upregulated in HCC and promotes proliferation⁶⁶, SFPQ and RBM8A mediate resistance to platinum drugs in HCC^{67,68}, RBMS1 blocks ferroptosis in HCC⁶⁹, MATR3 promotes HCC progression⁷⁰, SART3 is an immunotherapy target in HCC⁷¹, ZCRB1 is a signature gene for HCC⁷², and down regulation of CELF2 or CPEB3 promotes HCC⁷³, but the targets for many of these splicing factors remain to be determined. The role of these splicing factors in early liver disease is less clear but mouse studies have highlighted a potential causative role for some splicing factors in the pathogenesis of NAFLD and NASH. KHDRBS1 (Sam68) and SRSF3 regulate gluconeogenesis⁹ and loss of SRSF3, A1CF, SRSF2, SRSF1, and SRSF10 causes steatosis and changes in lipid metabolism⁷⁴ but loss of NONO, the binding partner for SFPQ, increases fat catabolism and protects against steatosis⁷⁵. Splicing factors are also involved in liver maturation as loss of SRSF3, SRSF7, ESRP2 or RBM15 reduces hepatocyte differentiation^{9,32,76,77}. Indeed, many of these factors can have opposing effects. PCBP2 overcomes palmitate-induced insulin resistance in HepG2 cells via inhibition of HIF1a and STAT3⁷⁸ but, in more advanced disease, PCBP2 promotes collagen production and liver fibrosis⁷⁹. The contribution of each of these splicing factors to the dysregulation of RNA splicing in human NAFLD and NASH, however, remains to be determined. An alternative approach is to assess the splicing of known target genes but that requires identification of direct targets through binding and knockout studies. We have shown that known SRSF3 target genes are altered in NAFLD, NASH and HCC suggesting that SRSF3 splicing function is reduced but that remains to be demonstrated for other factors.

In summary, dysregulated RNA splicing is observed early in metabolic liver disease and many of these changes are maintained during disease progression and are even found in HCC. The fact that the altered splicing is maintained through a long disease progression over many years argues for their importance in the disease pathogenesis. While the function of the individual transcriptional isoforms is not known for most splicing events, the genes themselves have documented roles in liver metabolism, cell proliferation and tumorigenesis supporting the importance of these events. Further studies investigating individual mRNA isoforms and the proteins produced will be required for a full understanding of the disease process.

Materials and methods

Human tissue samples

Human liver samples were obtained from the Liver Tissue and Cell Distribution Service at the University Minnesota, the Department of Pathology at the VASDHS, and the Moores Cancer Center Biorepository. Subject characteristics are provided in Supplemental Table 1. Liver and other tissues were harvested at sacrifice for both histology and RNA extraction. Livers were fixed in formalin for 24 h followed by washing in 70% ethanol. Paraffin embedded sections (5 μ m) were cut, dewaxed, and stained with hematoxylin and eosin. All human studies were approved by the UCSD Institutional Review Board.

Isolation of murine primary hepatocytes

Primary hepatocytes were obtained by two-step perfusion with liver perfusion medium, SC-1 (137 mM NaCl, 5.4 mM KCl, 0.56 mM NaH₂PO₄·H₂O, 0.85 mM Na₂HPO₄, 10 mM HEPES, 4.2 mM NaHCO₃, 0.5 mM EGTA, 5 mM Glucose) followed by digestion medium, SC-2 (137 mM NaCl, 5.4 mM KCl, 0.56 mM NaH₂PO₄·H₂O, 0.85 mM Na₂HPO₄, 10 mM HEPES, 4.2 mM NaHCO₃, 12 mM CaCl₂·H₂O) containing 0.5 mg/mL collagenase D (Roche, Indianapolis, IN). Liver cells were disaggregated by passing through a 100 μ m pore nylon mesh Cell Strainer (BD Biosciences, San Jose, CA) and centrifuged at 100 \times RCF for 10 min at 4 °C. Cells pellets were suspended in 36% percoll and centrifuged at 60 \times RCF for 6 min. The subsequent cell pellets were washed with SC-2 buffer without collagenase-D, the numbers of total viable cells were determined by Trypan blue staining, and cells were plated on collagen-coated plates (Invitrogen, Carlsbad, CA). Primary hepatocyte cells were cultured in William E culture medium supplemented with 10% fetal bovine serum, and 100 U mL⁻¹ of penicillin and 100 μ g mL⁻¹ of streptomycin at 37 °C. All animal studies were approved by the UCSD Institutional Animal Care and Use Committee.

RNAseq analysis

RNAseq was performed on RNA from frozen human liver samples and on RNA from primary mouse hepatocytes with adenoviral-mediated acute loss of SRSF3 using an Illumina Nextgen short-read sequencer (paired-end 100). Raw reads were trimmed and QC'd. Sequencing reads from HepG2 cells with shRNA knockdown of SRSF3 or control shRNA were downloaded from ENCODE (ENCSR376FGR). Fastq reads were aligned to the human (hg38) or mouse (mm10) genomes using the STAR-2pass method⁸⁰ then alternative splicing (AS) events identified using the percent spliced in (PSI) metric with the Psychomics software in R³⁴. STAR gene counts were normalized using trimmed mean of M-values (TMM), log-transformed, and IDs converted to gene symbols. STAR splice-junction counts were converted to the percent-spliced in metric (PSI) and annotated using human

hg38. Principal component analysis (PCA) was performed on PSI and gene expression data. Many splicing events were either not detected or had fewer than 10 reads, so dimension reduction was performed by filtering out events with more than 5 missing values. Non-parametric tests (Wilcoxon or Kruskal–Wallis rank sum) were used to assess differential splicing as the percent spliced in (PSI) data are not normally distributed. A p-value of 0.05 was used to identify potential differential AS events which were further refined using an FDR of 0.05. Data were also analyzed by MAJIQ⁴⁰ and Whippet⁴¹ using default parameters. Mouse genes showing HFD-induced AS events were compared to genes with SRSF3-dependent AS events to identify SRSF3-dependent changes in mouse NAFLD, and genes with SRSF3-dependent splicing events were mapped to human homologs to allow comparison with the human NAFLD data. Heatmaps were generated using Morpheus (<https://software.broadinstitute.org/morpheus>). Immune and liver cell composition was analyzed using CIBERSORTx⁴⁴ using the Human Liver Cell Atlas⁴⁵ or the Bader Human Liver Atlas⁴³. Enrichment analysis and protein–interaction network generation was performed using Metascape³⁶.

SRSF3-seCLIP analysis

Primary mouse hepatocytes were isolated and subjected to seCLIP following published procedures by Eclipse Biosciences^{46,81} (La Jolla, CA). SRSF3 was precipitated using mouse monoclonal 7B4 anti-SRSF3 antibody (1:100 dilution, ATCC CRL-2384, Manassas, VA). CLIP reads were aligned to the mouse mm10 genome then peaks of SRSF3 binding were identified using the HOMER suite⁸² or using Pureclip⁴⁷. Sequences enriched at the crosslink sites identified by Pureclip were analyzed using DREME⁴⁸ to identify conserved motifs and FIMO⁴⁹ to find individual occurrences. Alternative splicing events and SRSF3 binding were co-visualized on individual genes using IGV⁸³. Correlation of SRSF3 binding peaks and alternative splicing events was performed using custom scripts in R. Splicing events were called as SRSF3-dependent if they were altered in the SRSF3 knockout hepatocytes with $p < 0.05$ and contained a SRSF3 binding peak that mapped to < 1000 nt from the splicing event.

TCGA liver cancer data analysis (LIHC)

HCC gene counts, exon–junction counts, and metadata files were downloaded from the TCGA website. The STAR hg38-aligned TCGA gene count, splice junction and clinical data were downloaded and formatted for the Psychomics program. Gene counts and splice–junction counts were imported and annotated using human hg38 as above. Individual gene expression or alternative splicing events were visualized on violin plots. Survival analysis was performed using selected gene expression or PSI cutoffs, generating Kaplan–Meier curves and deriving Cox proportional hazard models. Correlation analysis between gene expression and an alternative splicing event was performed using Spearman rank correlation.

Immunoblot analysis

Equal amounts of cellular protein (10 μ g) were separated by SDS–PAGE using 4–15% or 20% Criterion precast polyacrylamide gels (Bio–Rad), transferred to PVDF membranes (MilliporeSigma, Burlington, MA), blocked with 5% BSA for 1 h at RT and immunoblotted with primary antibodies overnight at 4 °C followed by HRP-conjugated secondary antibodies at room temperature for 1 h, washed 3 \times in TBS–Tween20 then developed using a chemiluminescent substrate kit (Pierce, Rockford, IL). Antibodies used for immunoblotting were mouse monoclonal 7B4 anti-SRSF3 antibody (1:1000 dilution, ATCC CRL-2384, Manassas, VA), HRP labeled anti-mouse (sc516102) secondary antibody (1:5000 dilution, Santa Cruz Biotechnology, Santa Cruz, CA). Blots were quantified using a Gel–Doc imaging system (Bio–Rad).

Statistical analysis

Data was analyzed by 1-way or 2-way ANOVA followed by Tukey multiple comparison post-test, or Students' t-test as appropriate using Prism (Graph Pad, La Jolla, CA) or in R statistical software (v3.4.4). Normality was assessed by D'Agostino–Pearson omnibus normality test. Results were expressed as mean \pm standard error and considered significant with $p < 0.05$. Linear regression and Chi-squared contingency table analysis was performed using Prism. Venn diagrams were generated using Venny⁸⁴ or nVenn⁸⁵.

Data availability

Supporting Information is available from the author. All the data supporting the findings of this study are available within the article and its supplementary information files and from the corresponding author upon reasonable request. Sequence reads are available from the Sequence Read Archive (SRA BioProject ID PRJNA1040226). A reporting summary for this article is available as a Supplementary Information file.

Received: 6 September 2023; Accepted: 16 January 2024

Published online: 30 January 2024

References

1. Wu, P., Zhang, M. & Webster, N. J. G. Alternative RNA splicing in fatty liver disease. *Front. Endocrinol.* **12**, 613213. <https://doi.org/10.3389/fendo.2021.613213> (2021) (Epub 2021/03/16).
2. Oltean, S. & Bates, D. O. Hallmarks of alternative splicing in cancer. *Oncogene.* **33**(46), 5311–5318. <https://doi.org/10.1038/ncr.2013.533> (2014).
3. Qian, Y. & Fan, J. G. Obesity, fatty liver and liver cancer. *Hepatobiliary Pancreat. Dis. Int.* **4**(2), 173–177 (2005).
4. Rinella, M. E. Nonalcoholic fatty liver disease: A systematic review. *JAMA.* **313**(22), 2263–2273. <https://doi.org/10.1001/jama.2015.5370> (2015).
5. Saitta, C., Pollicino, T. & Raimondo, G. Obesity and liver cancer. *Ann. Hepatol.* **18**(6), 810–815. <https://doi.org/10.1016/j.aohep.2019.07.004> (2019) (Epub 2019/09/24).

6. Streba, L. A., Vere, C. C., Rogoveanu, I. & Streba, C. T. Nonalcoholic fatty liver disease, metabolic risk factors, and hepatocellular carcinoma: An open question. *World J. Gastroenterol.* **21**(14), 4103–4110. <https://doi.org/10.3748/wjg.v21.i14.4103> (2015).
7. Berasain, C. *et al.* Impairment of pre-mRNA splicing in liver disease: Mechanisms and consequences. *World J. Gastroenterol.* **16**(25), 3091–3102 (2010) (Epub 2010/07/02).
8. Jimenez, M., Arechederra, M., Avila, M. A. & Berasain, C. Splicing alterations contributing to cancer hallmarks in the liver: Central role of dedifferentiation and genome instability. *Transl. Gastroenterol. Hepatol.* **3**, 84. <https://doi.org/10.21037/tgh.2018.10.11> (2018) (Epub 2018/12/07).
9. Sen, S., Jumaa, H. & Webster, N. J. Splicing factor SRSF3 is crucial for hepatocyte differentiation and metabolic function. *Nat. Commun.* **4**, 1336. <https://doi.org/10.1038/ncomms2342> (2013) (Epub 2013/01/10).
10. Sen, S., Langiewicz, M., Jumaa, H. & Webster, N. J. Deletion of serine/arginine-rich splicing factor 3 in hepatocytes predisposes to hepatocellular carcinoma in mice. *Hepatology.* **61**(1), 171–183. <https://doi.org/10.1002/hep.27380> (2015).
11. Paterson, H. A. B. *et al.* Liver RBFOX2 regulates cholesterol homeostasis via Scarb1 alternative splicing in mice. *Nat. Metab.* **4**(12), 1812–1829. <https://doi.org/10.1038/s42255-022-00681-y> (2022) (Epub 2022/12/20).
12. Wang, Y., Yang, F., Shang, J., He, H. & Yang, Q. Integrative analysis reveals the prognostic value and functions of splicing factors implicated in hepatocellular carcinoma. *Sci. Rep.* **11**(1), 15175. <https://doi.org/10.1038/s41598-021-94701-8> (2021) (Epub 2021/07/28).
13. Tremblay, M. P. *et al.* Global profiling of alternative RNA splicing events provides insights into molecular differences between various types of hepatocellular carcinoma. *BMC Genomics.* **17**, 683. <https://doi.org/10.1186/s12864-016-3029-z> (2016).
14. Liu, L., Xie, S., Zhang, C. & Zhu, F. Aberrant regulation of alternative pre-mRNA splicing in hepatocellular carcinoma. *Crit. Rev. Eukaryot. Gene Expr.* **24**(2), 133–149. <https://doi.org/10.1615/critrevukaryotgeneexpr.2014007702> (2014) (Epub 2014/06/19).
15. Wu, F. *et al.* Profiles of prognostic alternative splicing signature in hepatocellular carcinoma. *Cancer Med.* **9**(6), 2171–2180. <https://doi.org/10.1002/cam4.2875> (2020) (Epub 2020/01/25).
16. Zhang, D., Duan, Y., Wang, Z. & Lin, J. Systematic profiling of a novel prognostic alternative splicing signature in hepatocellular carcinoma. *Oncol. Rep.* **42**(6), 2450–2472. <https://doi.org/10.3892/or.2019.7342> (2019) (Epub 2019/10/04).
17. Xiong, Y. *et al.* Genome-wide transcriptional analysis reveals alternative splicing event profiles in hepatocellular carcinoma and their prognostic significance. *Front. Genet.* **11**, 879. <https://doi.org/10.3389/fgene.2020.00879> (2020) (Epub 2020/08/28).
18. Luo, D. *et al.* Alternative splicing-based differences between hepatocellular carcinoma and intrahepatic cholangiocarcinoma: Genes, immune microenvironment, and survival prognosis. *Front. Oncol.* **11**, 731993. <https://doi.org/10.3389/fonc.2021.731993> (2021) (Epub 2021/11/12).
19. Shkreta, L. *et al.* Cancer-associated perturbations in alternative pre-messenger RNA splicing. *Cancer Treat. Res.* **158**, 41–94. https://doi.org/10.1007/978-3-642-31659-3_3 (2013).
20. Shiraiishi, Y. *et al.* Integrated analysis of whole genome and transcriptome sequencing reveals diverse transcriptomic aberrations driven by somatic genomic changes in liver cancers. *PLoS ONE.* **9**(12), e114263. <https://doi.org/10.1371/journal.pone.0114263> (2014).
21. Wruck, W. *et al.* Multi-omic profiles of human non-alcoholic fatty liver disease tissue highlight heterogenic phenotypes. *Sci. Data.* **2**, 150068. <https://doi.org/10.1038/sdata.2015.68> (2015).
22. Yu, J., Shen, J., Sun, T. T., Zhang, X. & Wong, N. Obesity, insulin resistance, NASH and hepatocellular carcinoma. *Semin. Cancer Biol.* **23**(6 Pt B), 483–491. <https://doi.org/10.1016/j.semcancer.2013.07.003> (2013).
23. Shackel, N. A., Seth, D., Haber, P. S., Gorrell, M. D. & McCaughan, G. W. The hepatic transcriptome in human liver disease. *Comp. Hepatol.* **5**, 6. <https://doi.org/10.1186/1476-5926-5-6> (2006).
24. Hoang, S. A. *et al.* Gene expression predicts histological severity and reveals distinct molecular profiles of nonalcoholic fatty liver disease. *Sci. Rep.* **9**(1), 12541. <https://doi.org/10.1038/s41598-019-48746-5> (2019) (Epub 2019/08/31).
25. Govaere, O. *et al.* Transcriptomic profiling across the nonalcoholic fatty liver disease spectrum reveals gene signatures for steatohepatitis and fibrosis. *Sci. Transl. Med.* <https://doi.org/10.1126/scitranslmed.aba4448> (2020) (Epub 2020/12/04).
26. Suppli, M. P. *et al.* Hepatic transcriptome signatures in patients with varying degrees of nonalcoholic fatty liver disease compared with healthy normal-weight individuals. *Am. J. Physiol. Gastrointest. Liver Physiol.* **316**(4), G462–G472. <https://doi.org/10.1152/ajpgi.00358.2018> (2019) (Epub 2019/01/18).
27. Cheng, Y. *et al.* Liver-specific deletion of SRSF2 caused acute liver failure and early death in mice. *Mol. Cell Biol.* **36**(11), 1628–1638. <https://doi.org/10.1128/MCB.01071-15> (2016).
28. Elizalde, M. *et al.* Splicing regulator SLU7 is essential for maintaining liver homeostasis. *J. Clin. Investig.* **124**(7), 2909–2920. <https://doi.org/10.1172/JCI74382> (2014) (Epub 2014/05/29).
29. Kumar, D. *et al.* Degradation of splicing factor SRSF3 contributes to progressive liver disease. *J. Clin. Investig.* <https://doi.org/10.1172/JCI127374> (2019).
30. Wu, X. *et al.* CUG-binding protein 1 regulates HSC activation and liver fibrogenesis. *Nat. Commun.* **7**, 13498. <https://doi.org/10.1038/ncomms13498> (2016).
31. Bhat, A. *et al.* ESRP2 controls an adult splicing programme in hepatocytes to support postnatal liver maturation. *Nat. Commun.* **6**, 8768. <https://doi.org/10.1038/ncomms9768> (2015).
32. Hyun, J. *et al.* Dysregulation of the ESRP2-NF2-YAP/TAZ axis promotes hepatobiliary carcinogenesis in non-alcoholic fatty liver disease. *J. Hepatol.* **75**(3), 623–633. <https://doi.org/10.1016/j.jhep.2021.04.033> (2021) (Epub 2021/05/09).
33. Wang, H. *et al.* Alteration of splicing factors' expression during liver disease progression: Impact on hepatocellular carcinoma outcome. *Hepatol. Int.* <https://doi.org/10.1007/s12072-019-09950-7> (2019).
34. Saraiva-Agostinho, N. & Barbosa-Morais, N. L. psychomics: Graphical application for alternative splicing quantification and analysis. *Nucleic Acids Res.* **47**(2), e7. <https://doi.org/10.1093/nar/gky888> (2019) (Epub 2018/10/03).
35. Consortium EP. An integrated encyclopedia of DNA elements in the human genome. *Nature.* **489**(7414), 57–74. <https://doi.org/10.1038/nature1247> (2012) (Epub 2012/09/08).
36. Zhou, Y. *et al.* Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat. Commun.* **10**(1), 1523. <https://doi.org/10.1038/s41467-019-09234-6> (2019) (Epub 2019/04/05).
37. Oughtred, R. *et al.* The BioGRID interaction database: 2019 update. *Nucleic Acids Res.* **47**(D1), D529–D541. <https://doi.org/10.1093/nar/gky1079> (2019) (Epub 2018/11/27).
38. Szklarczyk, D. *et al.* STRING v11: Protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* **47**(D1), D607–D613. <https://doi.org/10.1093/nar/gky1131> (2019) (Epub 2018/11/27).
39. Bader, G. D. & Hogue, C. W. An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinform.* **4**, 2. <https://doi.org/10.1186/1471-2105-4-2> (2003) (Epub 2003/01/15).
40. Norton, S. S., Vaquero-Garcia, J., Lahens, N. F., Grant, G. R. & Barash, Y. Outlier detection for improved differential splicing quantification from RNA-Seq experiments with replicates. *Bioinformatics.* **34**(9), 1488–1497. <https://doi.org/10.1093/bioinformatics/btx790> (2018) (Epub 2017/12/14).
41. Sterne-Weiler, T., Weatheritt, R. J., Best, A. J., Ha, K. C. H. & Blencowe, B. J. Efficient and accurate quantitative profiling of alternative splicing patterns of any complexity on a laptop. *Mol. Cell.* **72**(1), 187–200.e6. <https://doi.org/10.1016/j.molcel.2018.08.018> (2018) (Epub 2018/09/18).

42. Pinero, J., Sauch, J., Sanz, F. & Furlong, L. I. The DisGeNET cytoscape app: Exploring and visualizing disease genomics data. *Comput. Struct. Biotechnol. J.* **19**, 2960–2967. <https://doi.org/10.1016/j.csbj.2021.05.015> (2021) (**Epub 2021/06/18**).
43. MacParland, S. A. *et al.* Single cell RNA sequencing of human liver reveals distinct intrahepatic macrophage populations. *Nat. Commun.* **9**(1), 4383. <https://doi.org/10.1038/s41467-018-06318-7> (2018) (**Epub 2018/10/24**).
44. Newman, A. M. *et al.* Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nat. Biotechnol.* **37**(7), 773–782. <https://doi.org/10.1038/s41587-019-0114-2> (2019) (**Epub 2019/05/08**).
45. Aizarani, N. *et al.* A human liver cell atlas reveals heterogeneity and epithelial progenitors. *Nature*. **572**(7768), 199–204. <https://doi.org/10.1038/s41586-019-1373-2> (2019) (**Epub 2019/07/12**).
46. Van Nostrand, E. L. *et al.* Robust, cost-effective profiling of RNA binding protein targets with single-end enhanced crosslinking and immunoprecipitation (seCLIP). *Methods Mol. Biol.* **1648**, 177–200. https://doi.org/10.1007/978-1-4939-7204-3_14 (2017).
47. Krakau, S., Richard, H. & Marsico, A. PureCLIP: Capturing target-specific protein-RNA interaction footprints from single-nucleotide CLIP-seq data. *Genome Biol.* **18**(1), 240. <https://doi.org/10.1186/s13059-017-1364-2> (2017) (**Epub 2017/12/30**).
48. Bailey, T. L. DREME: Motif discovery in transcription factor ChIP-seq data. *Bioinformatics*. **27**(12), 1653–1659. <https://doi.org/10.1093/bioinformatics/btr261> (2011) (**Epub 2011/05/06**).
49. Grant, C. E., Bailey, T. L. & Noble, W. S. FIMO: Scanning for occurrences of a given motif. *Bioinformatics*. **27**(7), 1017–1018. <https://doi.org/10.1093/bioinformatics/btr064> (2011) (**Epub 2011/02/19**).
50. Han, H. *et al.* TRRUST v2: An expanded reference database of human and mouse transcriptional regulatory interactions. *Nucleic Acids Res.* **46**(D1), D380–D386. <https://doi.org/10.1093/nar/gkx1013> (2018) (**Epub 2017/11/01**).
51. Cancer Genome Atlas Research Network. Electronic address wbe, Cancer Genome Atlas Research N. Comprehensive and integrative genomic characterization of hepatocellular carcinoma. *Cell*. **169**(7), 1327–41.e23. <https://doi.org/10.1016/j.cell.2017.05.046> (2017) (**Epub 2017/06/18**).
52. Vaquero-Garcia, J. *et al.* A new view of transcriptome complexity and regulation through the lens of local splicing variations. *Elife*. **5**, e11752. <https://doi.org/10.7554/eLife.11752> (2016).
53. Leclair, N. K. *et al.* Poison exon splicing regulates a coordinated network of SR protein expression during differentiation and tumorigenesis. *Mol. Cell*. **80**(4), 648–65.e9. <https://doi.org/10.1016/j.molcel.2020.10.019> (2020) (**Epub 2020/11/12**).
54. de Oliveira Freitas Machado, C. *et al.* Poison cassette exon splicing of SRSF6 regulates nuclear speckle dispersal and the response to hypoxia. *Nucleic Acids Res.* **51**(2), 870–890. <https://doi.org/10.1093/nar/gkac1225> (2023) (**Epub 2023/01/10**).
55. Li, Y. *et al.* DRAK2 aggravates nonalcoholic fatty liver disease progression through SRSF6-associated RNA alternative splicing. *Cell Metab.* **33**(10), 2004–20.e9. <https://doi.org/10.1016/j.cmet.2021.09.008> (2021) (**Epub 2021/10/07**).
56. Laman, H. *et al.* Transforming activity of Fbxo7 is mediated specifically through regulation of cyclin D/cdk6. *Embo J.* **24**(17), 3104–3116. <https://doi.org/10.1038/sj.emboj.7600775> (2005) (**Epub 2005/08/13**).
57. Shen, J. Z. *et al.* A FBXO7/EYA2-SCF(FBXW7) axis promotes AXL-mediated maintenance of mesenchymal and immune evasion phenotypes of cancer cells. *Mol. Cell*. **82**(6), 1123–39.e8. <https://doi.org/10.1016/j.molcel.2022.01.022> (2022) (**Epub 2022/02/20**).
58. Sun, R. *et al.* Proteomic analysis reveals that EPHX1 contributes to 5-fluorouracil resistance in a human hepatocellular carcinoma cell line. *Proteomics Clin. Appl.* **14**(4), e1900080. <https://doi.org/10.1002/prca.201900080> (2020) (**Epub 2020/02/19**).
59. Norton, P. A. & Mehta, A. S. Expression of genes that control core fucosylation in hepatocellular carcinoma: Systematic review. *World J. Gastroenterol.* **25**(23), 2947–2960. <https://doi.org/10.3748/wjg.v25.i23.2947> (2019) (**Epub 2019/06/30**).
60. Li, S. H. *et al.* Targeting MYO1B impairs tumorigenesis via inhibiting the SNAI2/cyclin D1 signaling in esophageal squamous cell carcinoma. *J. Cell Physiol.* **237**(9), 3671–3686. <https://doi.org/10.1002/jcp.30831> (2022) (**Epub 2022/07/22**).
61. Chen, Y. H. *et al.* Myo1b promotes tumor progression and angiogenesis by inhibiting autophagic degradation of HIF-1alpha in colorectal cancer. *Cell Death Dis.* **13**(11), 939. <https://doi.org/10.1038/s41419-022-05397-1> (2022) (**Epub 2022/11/09**).
62. Haugen, O. P., Khuu, C., Weidemann, H. M., Utheim, T. P. & Bergersen, L. H. Transcriptomic and functional studies reveal miR-431-5p as a tumour suppressor in pancreatic ductal adenocarcinoma cells. *Gene*. **822**, 146346. <https://doi.org/10.1016/j.gene.2022.146346> (2022) (**Epub 2022/02/20**).
63. Lin, T. C. *et al.* Fibronectin in cancer: Friend or foe. *Cells*. <https://doi.org/10.3390/cells9010027> (2019) (**Epub 2019/12/22**).
64. Blanc, V. *et al.* Apobec1 complementation factor overexpression promotes hepatic steatosis, fibrosis, and hepatocellular cancer. *J. Clin. Investig.* <https://doi.org/10.1172/JCI138699> (2021) (**Epub 2021/01/15**).
65. Zhou, Z. J. *et al.* HNRNPAB induces epithelial-mesenchymal transition and promotes metastasis of hepatocellular carcinoma by transcriptionally activating SNAIL. *Cancer Res.* **74**(10), 2750–2762. <https://doi.org/10.1158/0008-5472.CAN-13-2509> (2014) (**Epub 2014/03/19**).
66. Zhang, X. *et al.* Overexpression of PCBP2 contributes to poor prognosis and enhanced cell growth in human hepatocellular carcinoma. *Oncol. Rep.* **36**(6), 3456–3464. <https://doi.org/10.3892/or.2016.5167> (2016) (**Epub 2016/10/18**).
67. Ru, Y. *et al.* NEAT1_2-SFPQ axis mediates cisplatin resistance in liver cancer cells in vitro. *Onco Targets Ther.* **11**, 5695–5702. <https://doi.org/10.2147/OTT.S163774> (2018) (**Epub 2018/09/27**).
68. Liang, R. *et al.* Mechanism and molecular network of RBM8A-mediated regulation of oxaliplatin resistance in hepatocellular carcinoma. *Front. Oncol.* **10**, 585452. <https://doi.org/10.3389/fonc.2020.585452> (2020) (**Epub 2021/02/09**).
69. Zhai, H. *et al.* Suppressing circIDE/miR-19b-3p/RBMS1 axis exhibits promoting-tumour activity through upregulating GPX4 to diminish ferroptosis in hepatocellular carcinoma. *Epigenetics*. **18**(1), 2192438. <https://doi.org/10.1080/15592294.2023.2192438> (2023) (**Epub 2023/03/30**).
70. He, H. *et al.* MatrIn-3 promotes hepatocellular carcinoma progression by regulating cell cycle. Research Square (2022).
71. Kaji, K. *et al.* Cellular immune responses for squamous cell carcinoma antigen recognized by T cells 3 in patients with hepatocellular carcinoma. *PLoS ONE*. **12**(1), e0170291. <https://doi.org/10.1371/journal.pone.0170291> (2017) (**Epub 2017/01/24**).
72. Li, G. *et al.* Large-scale transcriptome analysis identified a novel cancer driver genes signature for predicting the prognostic of patients with hepatocellular carcinoma. *Front. Pharmacol.* **12**, 638622. <https://doi.org/10.3389/fphar.2021.638622> (2021) (**Epub 2021/08/03**).
73. Li, Z. *et al.* MiR-20b-5p promotes hepatocellular carcinoma cell proliferation, migration and invasion by down-regulating CPEB3. *Hepatol.* **23**, 100345. <https://doi.org/10.1016/j.aohep.2021.100345> (2021) (**Epub 2021/04/04**).
74. Pihlajamaki, J. *et al.* Expression of the splicing factor gene SFRS10 is reduced in human obesity and contributes to enhanced lipogenesis. *Cell Metab.* **14**(2), 208–218. <https://doi.org/10.1016/j.cmet.2011.06.007> (2011).
75. Benegiamo, G. *et al.* The RNA-binding protein NONO coordinates hepatic adaptation to feeding. *Cell Metab.* **27**(2), 404–18.e7. <https://doi.org/10.1016/j.cmet.2017.12.010> (2018) (**Epub 2018/01/24**).
76. Kadota, Y. *et al.* Srsf7 establishes the juvenile transcriptome through age-dependent alternative splicing in mice. *iScience*. **23**(3), 100929. <https://doi.org/10.1016/j.isci.2020.100929> (2020) (**Epub 2020/03/09**).
77. Hu, L., Li, H., Chi, Z. & He, J. Loss of the RNA-binding protein Rbm15 disrupts liver maturation in zebrafish. *J. Biol. Chem.* **295**(33), 11466–11472. <https://doi.org/10.1074/jbc.RA120.014080> (2020) (**Epub 2020/06/11**).
78. Xia, N. *et al.* PCBP2 regulates hepatic insulin sensitivity via HIF-1alpha and STAT3 pathway in HepG2 cells. *Biochem. Biophys. Res. Commun.* **463**(1–2), 116–122. <https://doi.org/10.1016/j.bbrc.2015.04.150> (2015) (**Epub 2015/05/24**).
79. Jain, A. *et al.* Targeted delivery of an siRNA/PNA hybrid nanocomplex reverses carbon tetrachloride-induced liver fibrosis. *Adv. Ther.* <https://doi.org/10.1002/adtp.201900046> (2019) (**Epub 2019/08/01**).
80. Dobin, A. *et al.* STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics*. **29**(1), 15–21. <https://doi.org/10.1093/bioinformatics/bts635> (2013).

81. Van Nostrand, E. L. *et al.* Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). *Nat. Methods*. **13**(6), 508. <https://doi.org/10.1038/Nmeth.3810> (2016).
82. Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell*. **38**(4), 576–589. <https://doi.org/10.1016/j.molcel.2010.05.004> (2010).
83. Thorvaldsdottir, H., Robinson, J. T. & Mesirov, J. P. Integrative Genomics Viewer (IGV): High-performance genomics data visualization and exploration. *Brief. Bioinform.* **14**(2), 178–192. <https://doi.org/10.1093/bib/bbs017> (2013).
84. Oliveros, J. C. Venny. An Interactive Tool for Comparing Lists with Venn's Diagrams. (2007). <https://bioinfogp.cnb.csic.es/tools/venny/index.html>.
85. Perez-Silva, J. G., Araujo-Voces, M. & Quesada, V. nVenn: Generalized, quasi-proportional Venn and Euler diagrams. *Bioinformatics*. **34**(13), 2322–2324. <https://doi.org/10.1093/bioinformatics/bty109> (2018) (Epub 2018/06/29).
86. Sebestyén, E. *et al.* Large-scale analysis of genome and transcriptome alterations in multiple tumors unveils novel cancer-relevant splicing networks. *Genome Res.* **26**(6), 732–744. <https://doi.org/10.1101/gr.199935.115> (2016). Erratum in: *Genome Res.* **28**(9), 1426 (2018). PMID: 27197215; PMCID: PMC4889968.

Acknowledgements

This work was funded primarily by a VA Merit Review Award (I01BX004848) and Senior Research Career Scientist Award (IBX005224) and in part by NIH grants (R01CA196853 and R01HL141999) to NJGW, and by Center NIH grants (P30CA023100, P30DK063491, and P30DK120515). We would like to acknowledge the assistance of Dr. Valeria Estrada and the Biorepository and Tissue Technology Shared Resource, and the Genomics and Computational Biology Shared Resource at the Moores' Cancer Center. Normal and pathologic human liver were obtained through the Liver Tissue Cell Distribution System, Minneapolis, MN, and Pittsburgh, PA, which was funded by NIH Contract #HHSN276201200017C.

Author contributions

N.J.G.W. conceived of the experiments, analyzed the data and wrote the manuscript; D.K. processed tissues and cells, and prepared the RNA; P.W. performed western blots.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-52237-7>.

Correspondence and requests for materials should be addressed to N.J.G.W.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

This is a U.S. Government work and not under copyright protection in the US; foreign copyright protection may apply 2024