

Supplemental Figures:

Supplemental Figure 1: RNA splicing is altered in NAFLD and NASH analyzed by Psychomics.

Supplemental Figure 2: Metascape analysis of alternative splicing analyzed by MAJIQ or Whippet.

Supplemental Figure 3: Transcriptome analysis of NAFLD and NASH samples.

Supplemental Figure 4: Fatty acid treatment reduces SRSF3 protein levels in HepG2 cells.

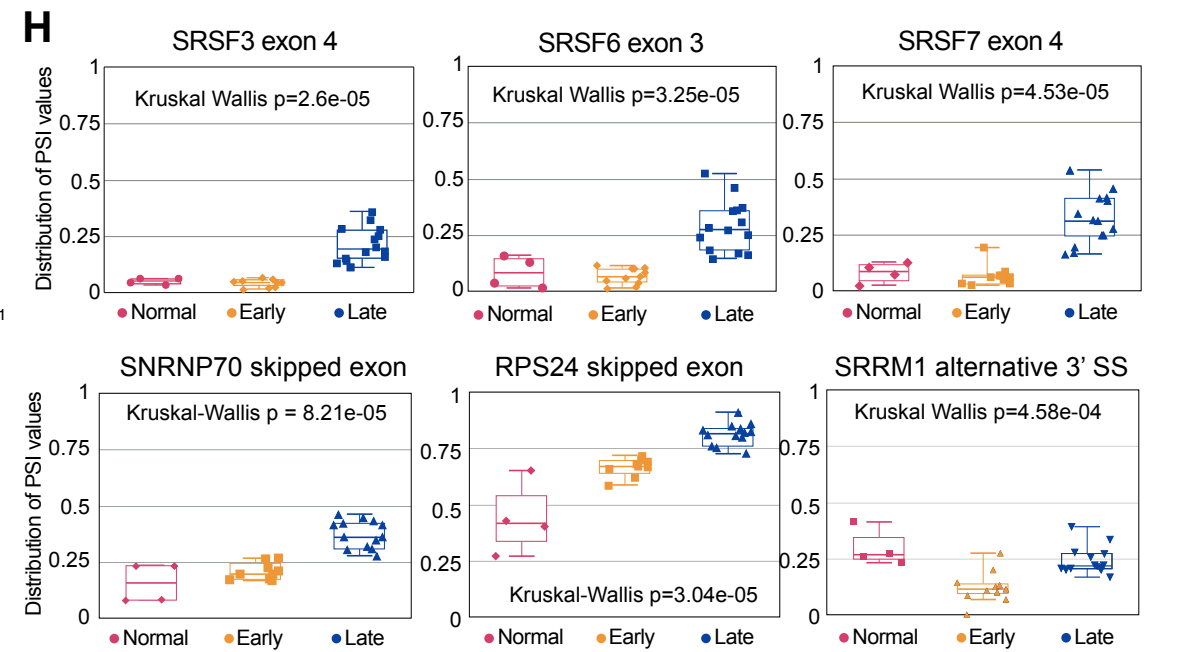
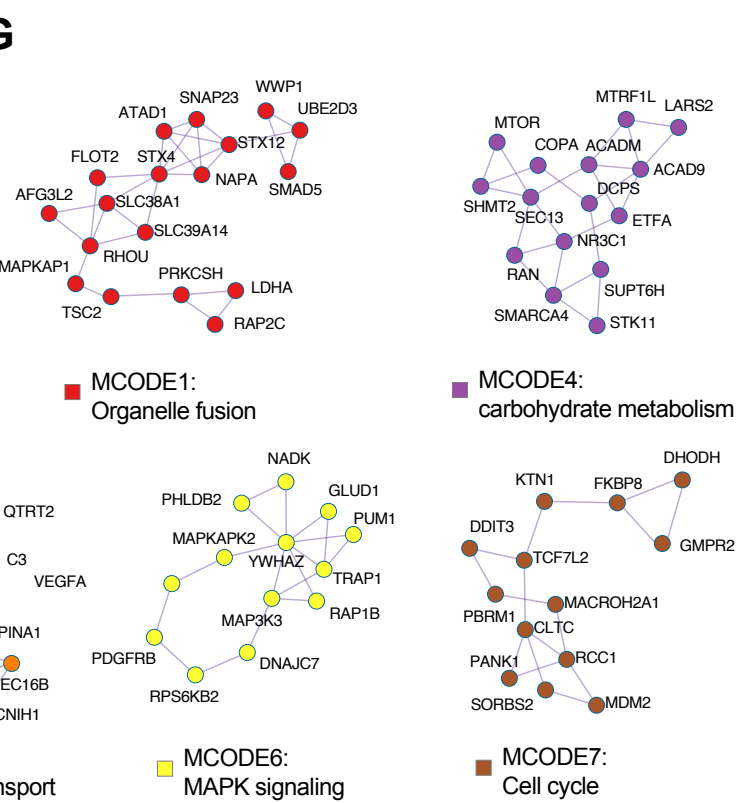
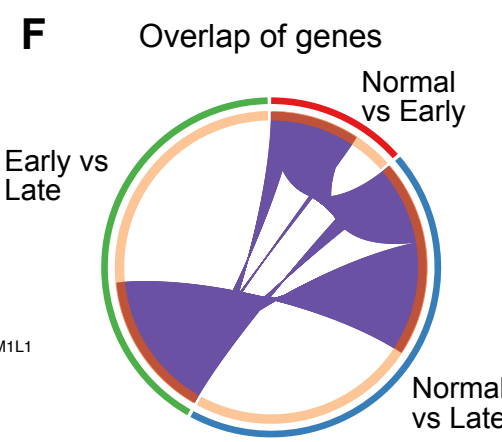
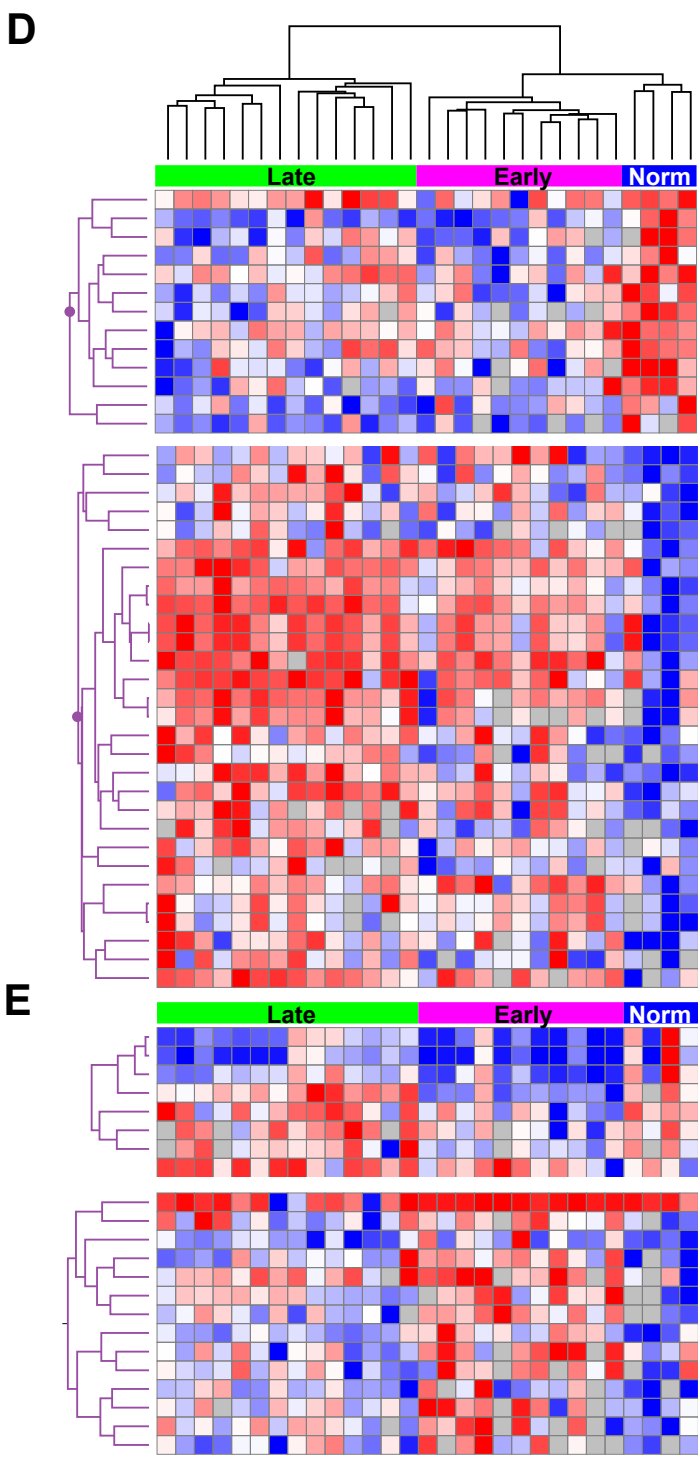
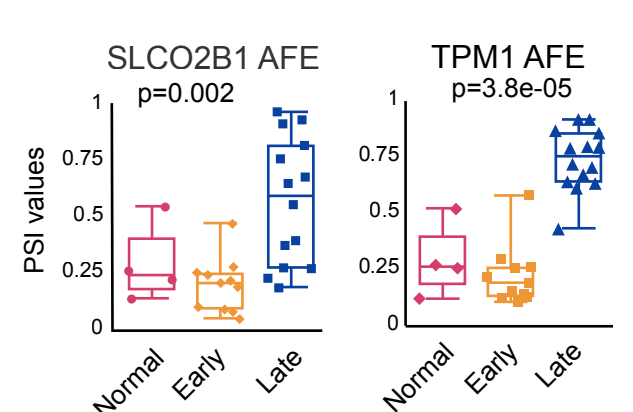
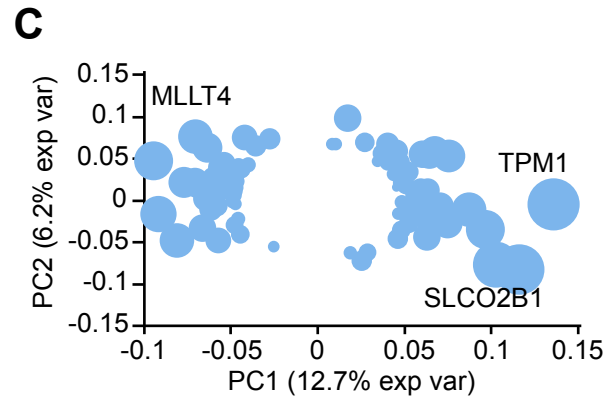
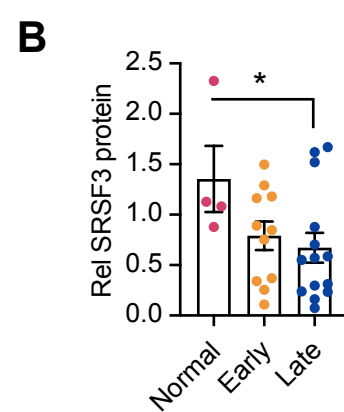
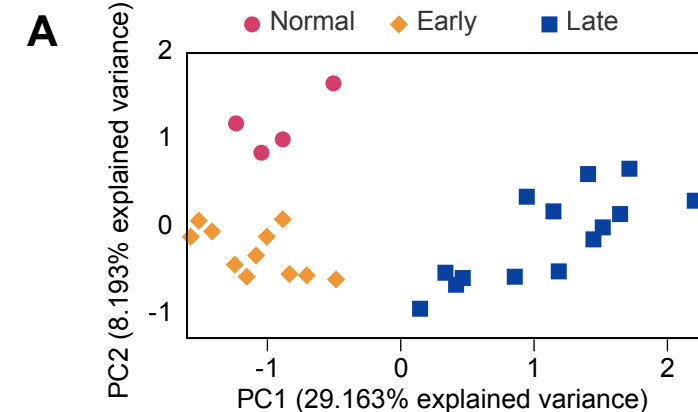
Supplemental figure 5: Comparison of selected splicing events across four datasets.

Supplemental Figure 6: Early changes in splicing.

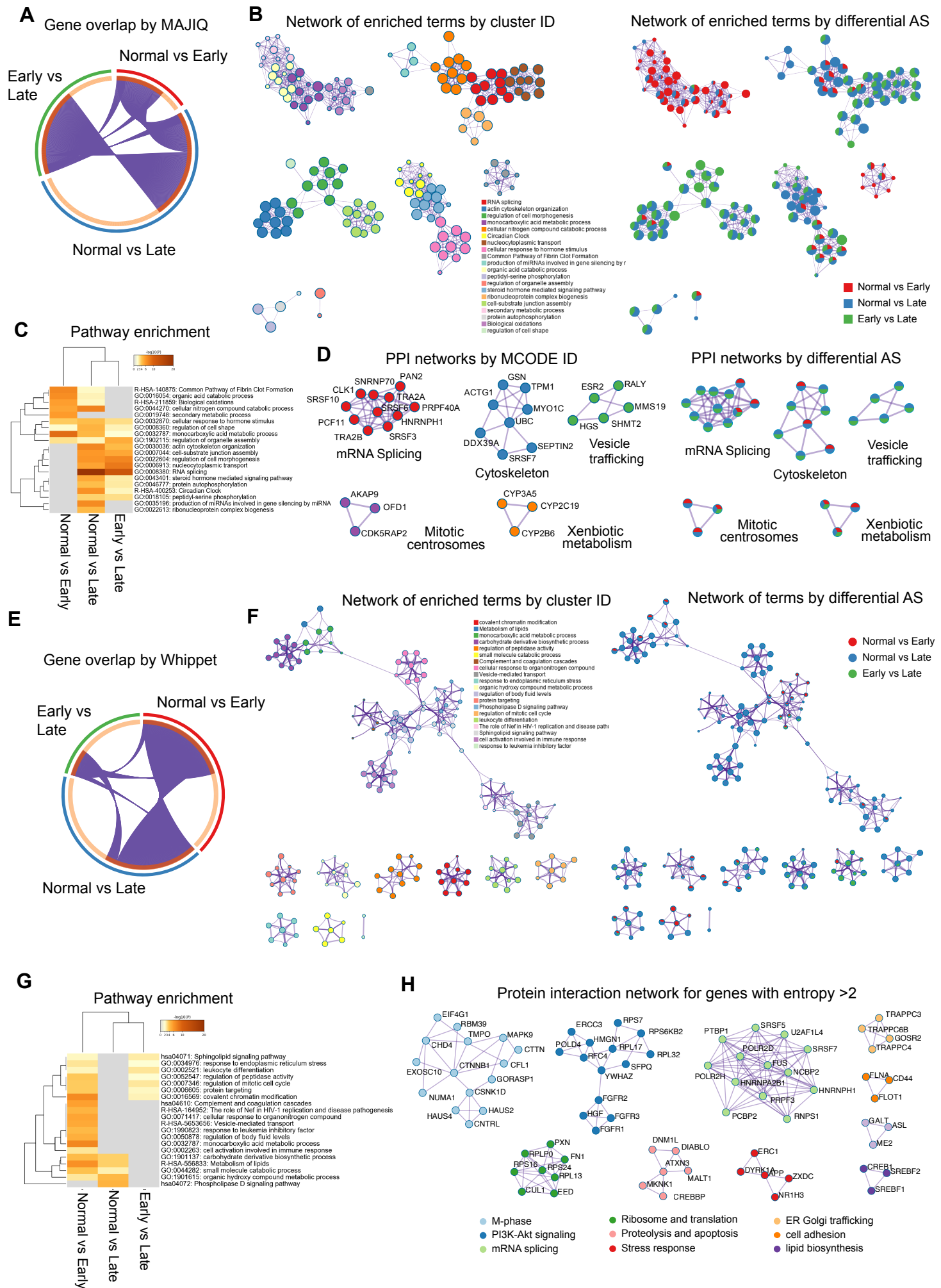
Supplemental Figure 7: Late changes in splicing.

Supplemental Figure 8: Survival curves for common splicing factors.

Supplemental Figure 9: Original blots for Supplemental Figure 4

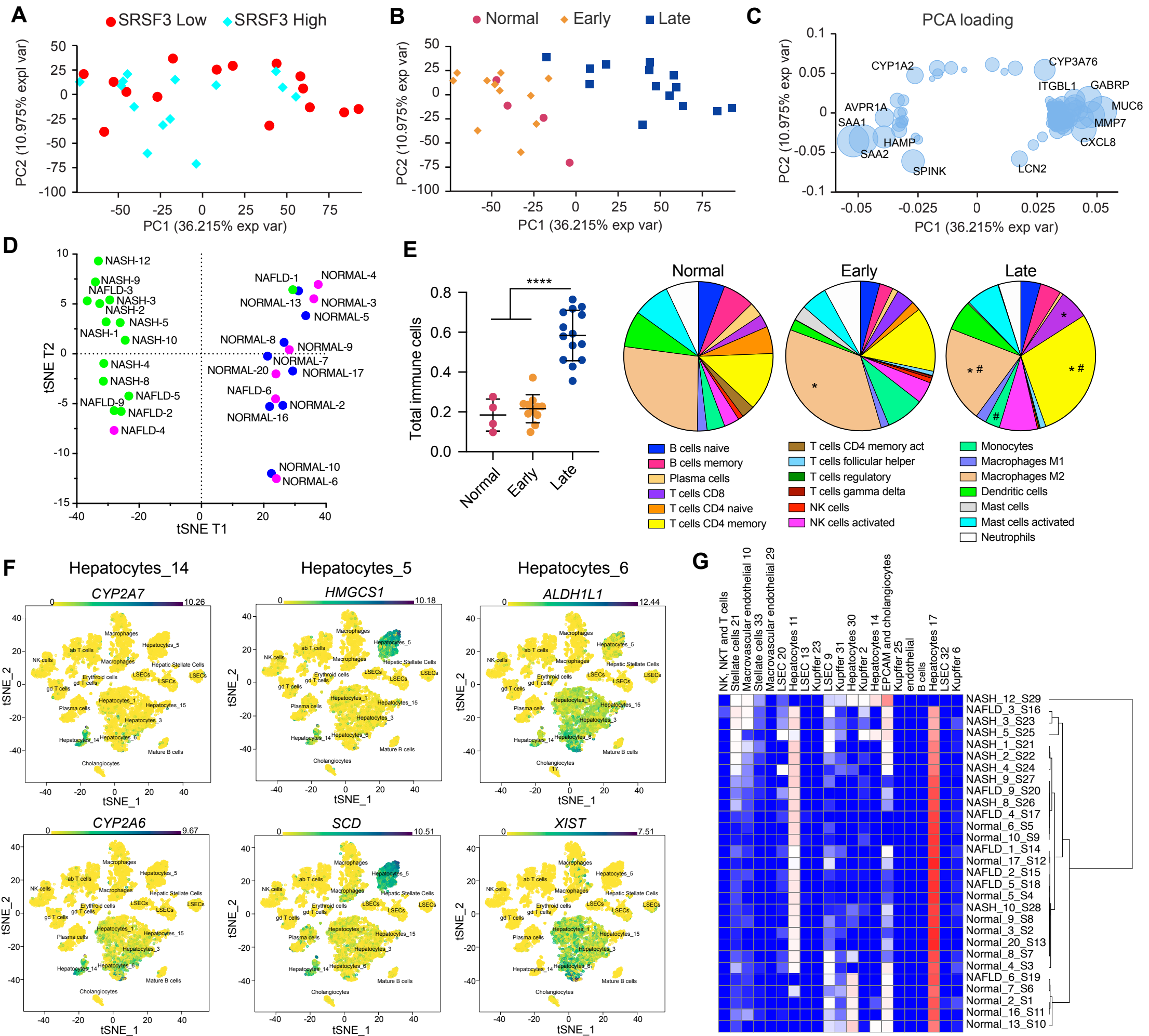


Supplemental Figure 1: RNA splicing is altered in NAFLD and NASH analyzed by psichomics: A: PCA plot based on the analysis of alternative splicing events in RNA from 29 liver samples in three clusters after dimension reduction by filtering out splicing events with more than 5 missing values which reduced the dataset to 6,797 SEs. B: SRSF3 protein expression in each cluster quantified by western blot. C: Plot showing top 100 SEs that contribute the most to the PCA. Each SE is shown by a blue circle, the size of the circle denoting the magnitude of the contribution. PSI values for the two top SEs, SLCO2B1 AFE and TPM1 AFE are shown in column graphs for the three clusters. D: Exploded view of heatmap of the clusters of SEs that are altered in the Normal cluster rather than the Early or Late clusters. Red indicates high PSI and blue low PSI. E: Exploded view of heatmap of the clusters of SEs that are altered in the Early cluster rather than the Normal or Late clusters. F: Circos plot showing overlap of genes harboring the altered SEs in pairwise comparisons of the three clusters. Purple lines connect the same genes. G: MCODE protein interaction networks for genes harboring the altered SEs. Networks are colored individually and functionally annotated. Nodes are shown as circles, interactions as lines. H: PSI values for alternative splicing of poison exons in the SRSF3, SRSF6 and SRSF7 genes, and in the splicing-related proteins SNRNP70, RPS24, and SRRM1 according to cluster. Kruskal-Wallis pvalues are given for the three group comparison.



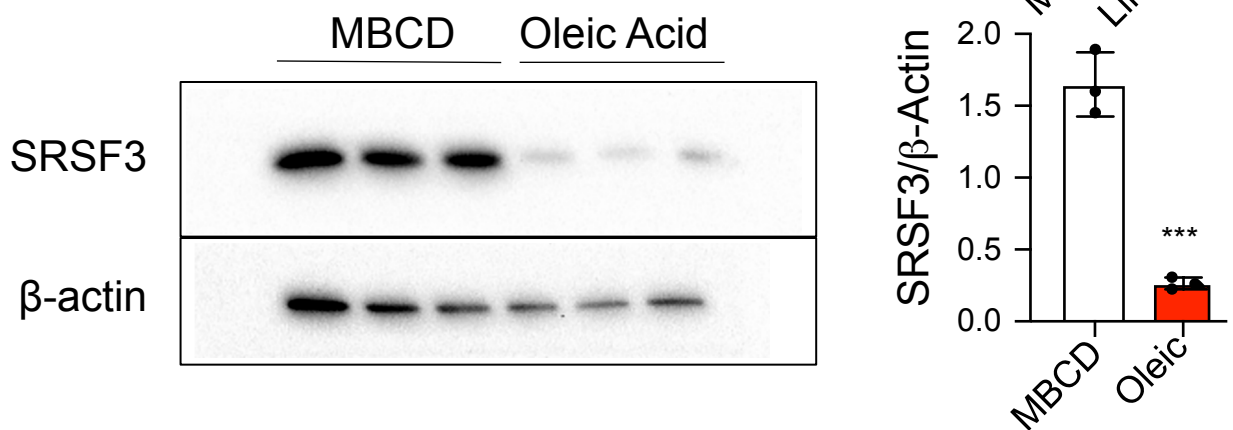
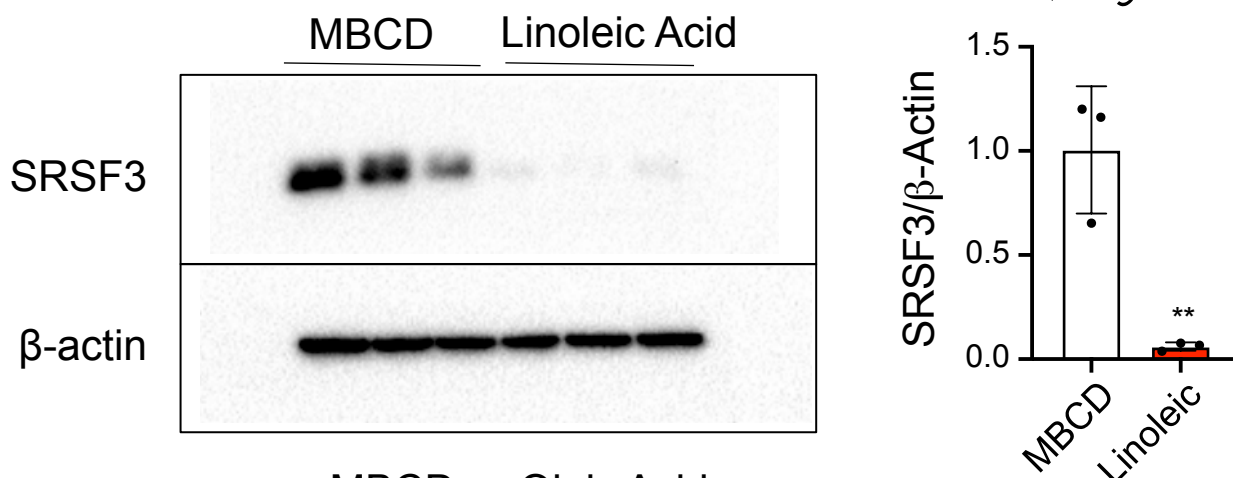
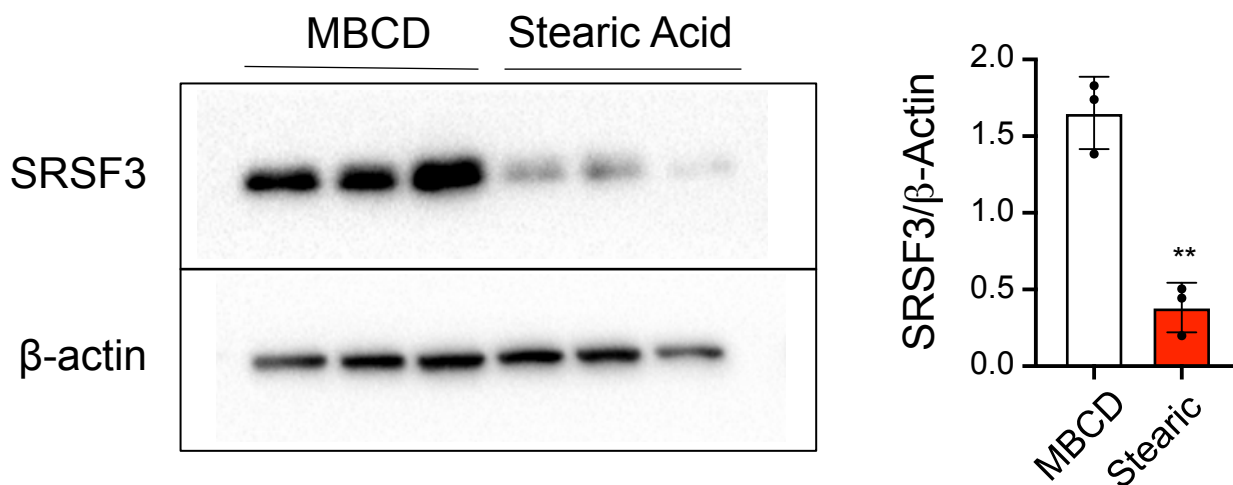
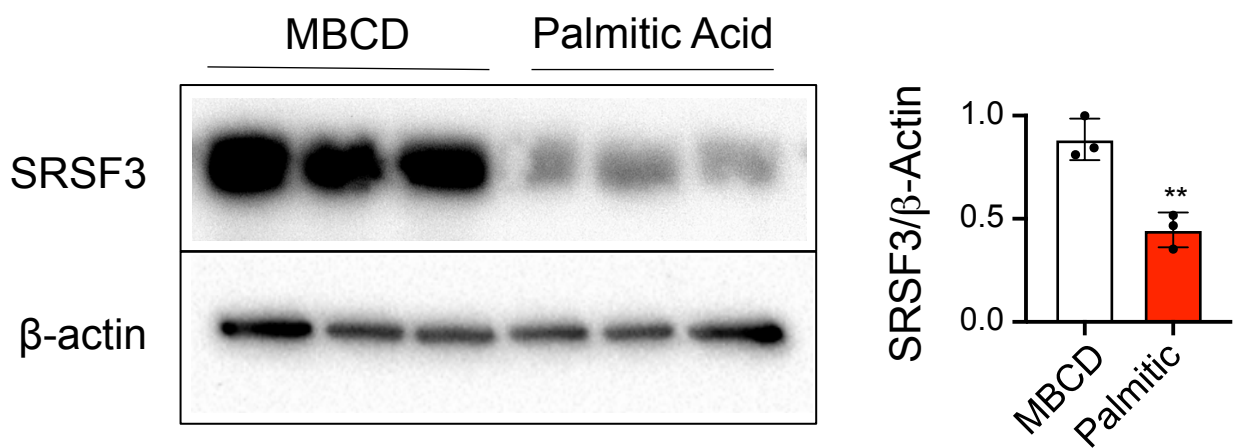
Supplemental Figure 2: Metascape analysis of alternative splicing analyzed by MAJIQ or

Whippet. A: Circos plot showing overlap of genes harboring the altered SEs in pairwise comparisons of the three clusters using MAJIQ. Purple lines connect the same genes. B: Network of enriched GO terms derived from genes harboring altered SEs according to MAJIQ. Left panel shows individual network clusters colored by functional annotation. Right panel shows networks colored by differential alternative splicing in pair-wise comparisons. C: Pathway enrichment for each pairwise comparison. D: Protein interaction networks colored by MCODE functional annotation (left) or differential alternative splicing (right). E: Circos plot showing overlap of genes harboring the altered SEs in pairwise comparisons of the three clusters using Whippet. F: Network of enriched GO terms derived from genes harboring altered SEs according to Whippet. Left panel shows individual network clusters colored by functional annotation. Right panel shows networks colored by differential alternative splicing in pair-wise comparisons. G: Pathway enrichment for each pairwise comparison by Whippet. H: Protein interaction networks for Genes with altered SEs with entropy >2 colored by MCODE functional annotation.



Supplemental Figure 3: Transcriptome analysis of NAFLD and NASH samples. A: PCA plot based on gene expression across all 29 samples colored by SRSF3 expression. B: PCA plot colored by splicing clusters. C: Plot showing top 100 genes that contribute the most to the PCA. Each gene is shown by a blue circle, the size of the circle denoting the magnitude of the contribution. Top thirteen genes are labeled. D: Analysis of gene expression by tSNE. Samples are colored according to original designation: Normal = no history of liver disease. E: Graph shows total immune cell content after deconvolution of the data using CIBERSORT. Pie charts show immune cells fractions for individual splicing clusters. Asterisks indicated statistical significance vs Cluster A, # indicates significance vs Cluster B by t-test with *,# $p < 0.05$. F: Single cell tSNE maps derived using the HumanLiver package in R developed by the Bader lab [44]. Maps are colored according to expression of CYP2A7 and CYP2A6 genes enriched in Hepatocyte cluster 14, HMGCS1 and SCD genes enriched in Hepatocyte cluster 5, and ALDH1L1 and XIST genes enriched in Hepatocyte cluster 6. Yellow signifies low expression, dark blue high expression. G: Heatmap showing liver cell composition based on cell signatures derived from the Human Liver Cell Atlas (<http://human-liver-cell-atlas.ie-freiburg.mpg.de/>). Blue indicates low relative cell content, red high.

Treatment Fatty Acid: 0.5mM, 12h.

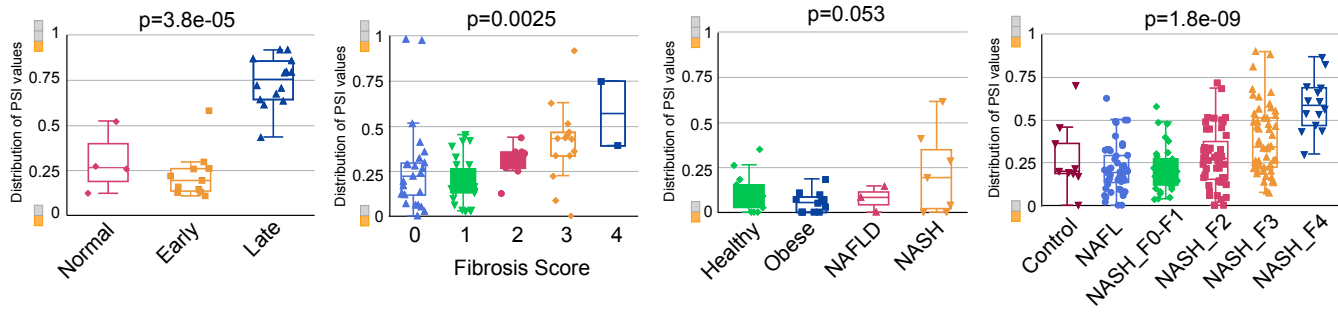


Supplemental Figure 4: Fatty acid treatment reduces SRSF3 protein levels in HepG2

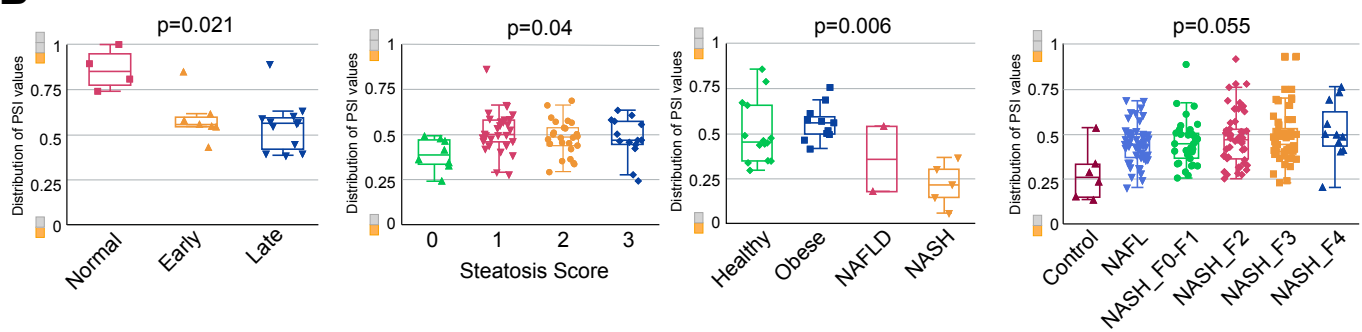
cells. Hep G2 cells were treated with 500 μ M of the four major fatty acids in the high fat diet, palmitic acid, stearic acid, linoleic acid or oleic acid, for 12 h. Fatty acid were solubilized using methyl- β -cyclodextrin (MBCD) which was used as a vehicle control. SRSF3 levels were measured by western blot and normalized to β -actin.

A This study Hoang et al. Suppli et al. Govaere et al.

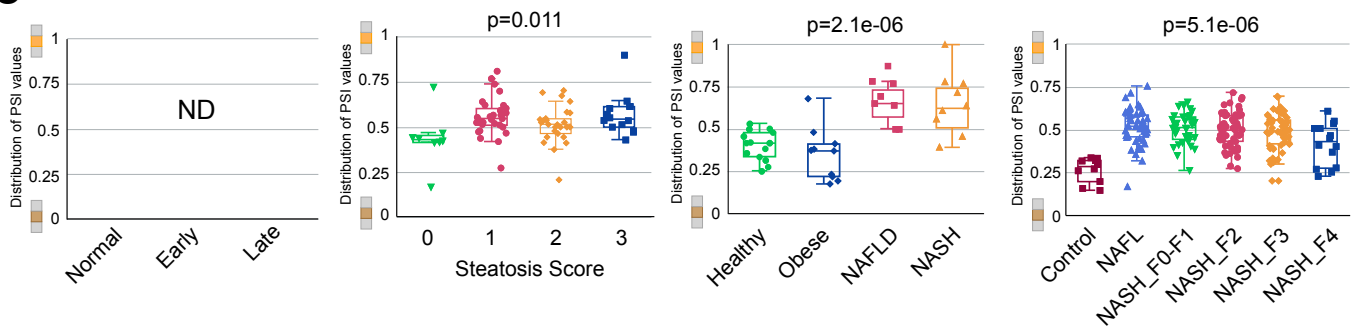
TPM1 alternative first exon (AFE) (chr15:63044152-63056985, + strand)



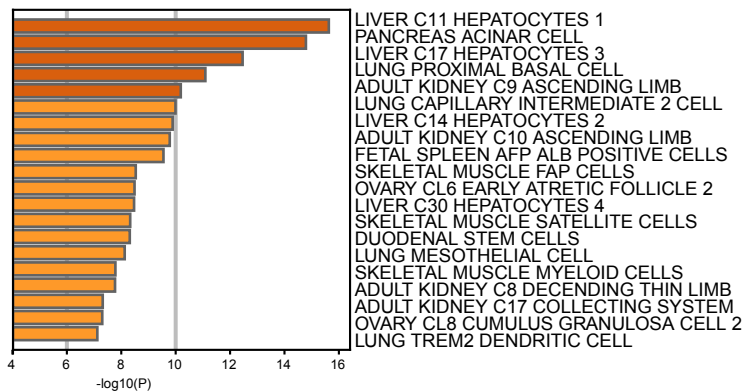
B PANK1 alternative first exon (AFE) (chr10:89612048-89644600, - strand)



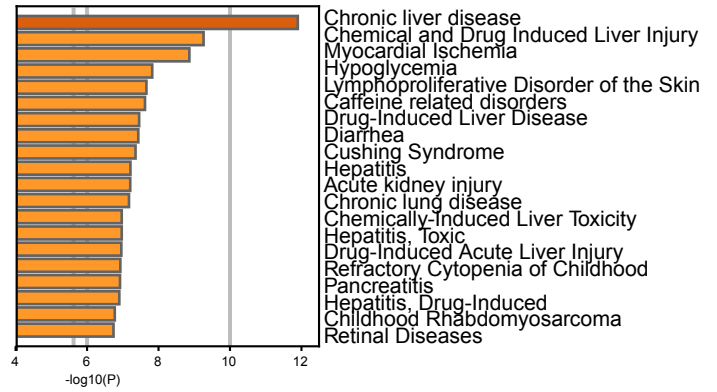
C IRF3 mutually exclusive exon (MXE) (chr19:49659833-49662425, - strand)



D

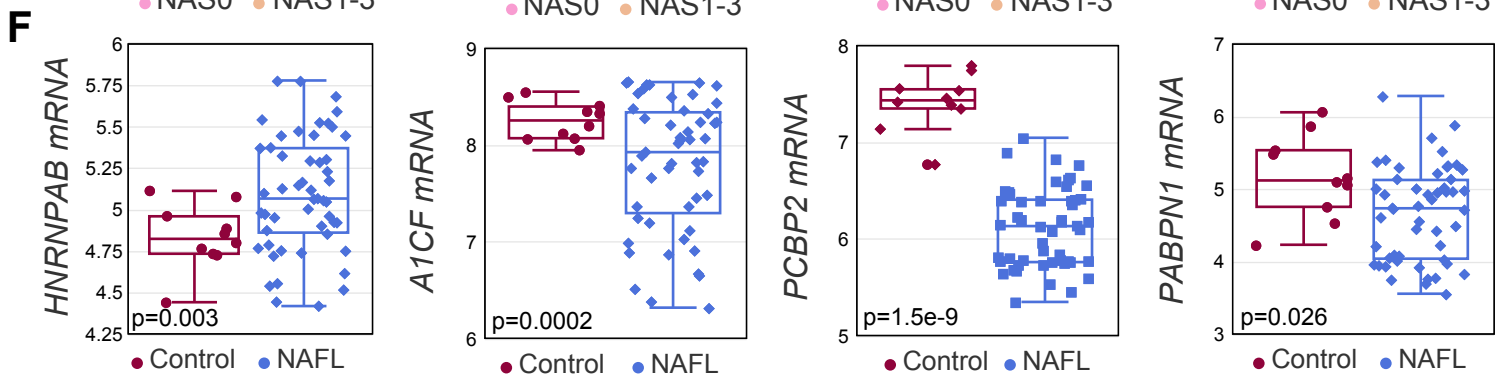
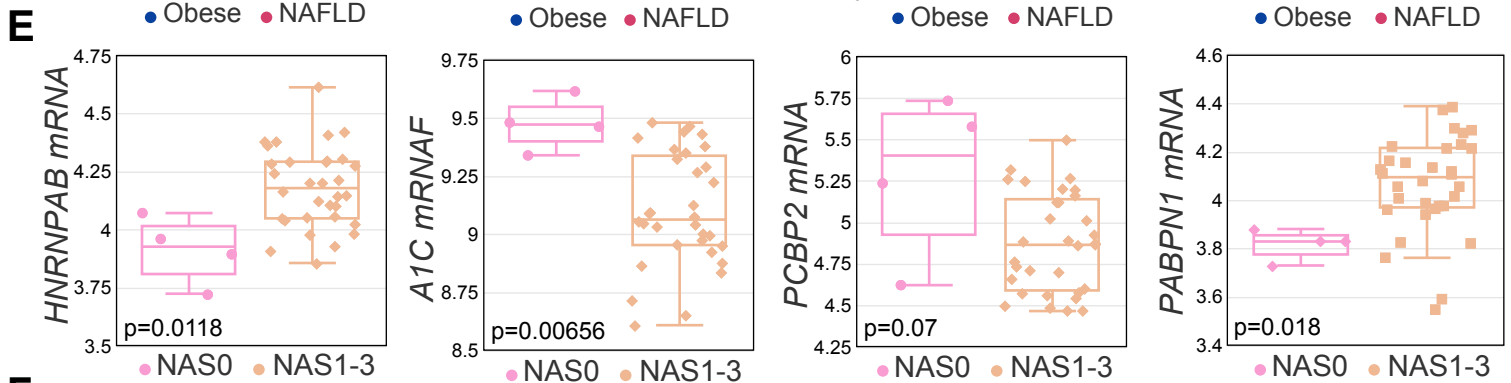
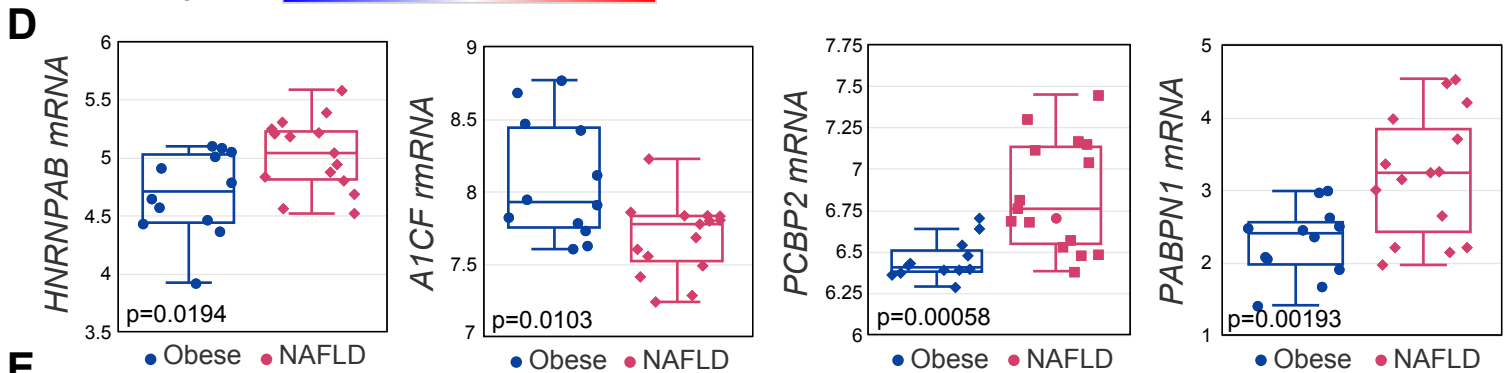
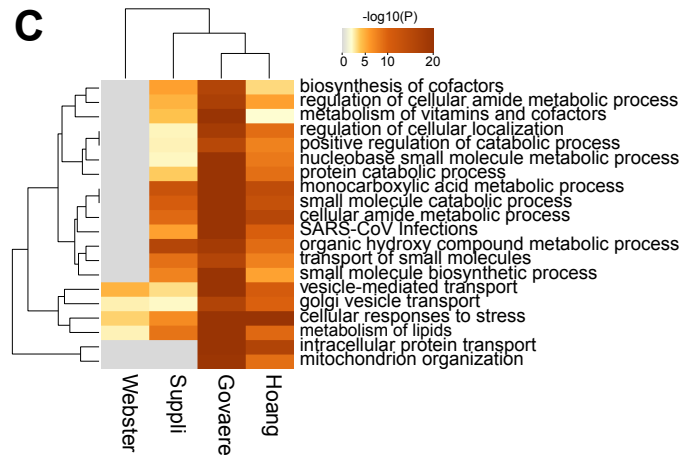
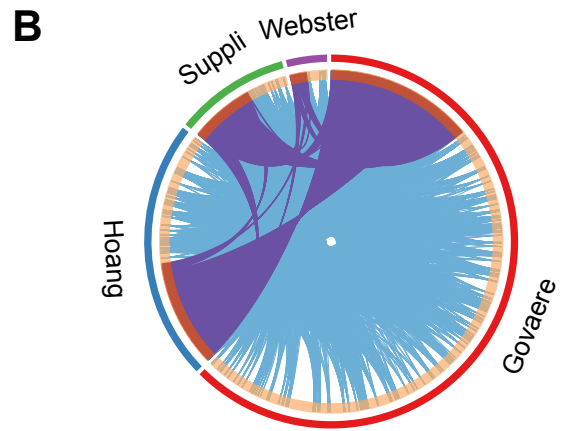
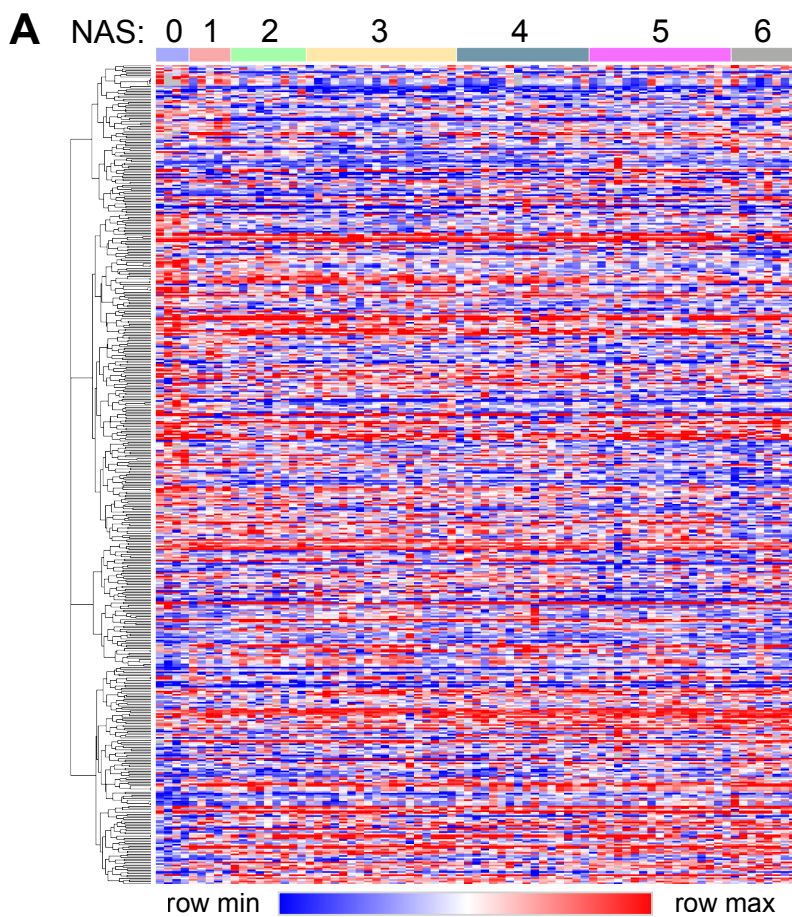


E

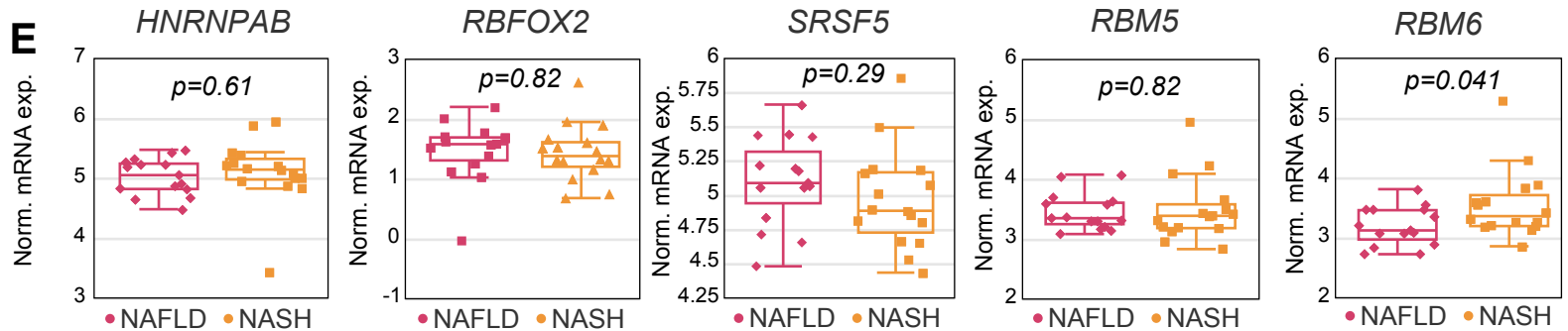
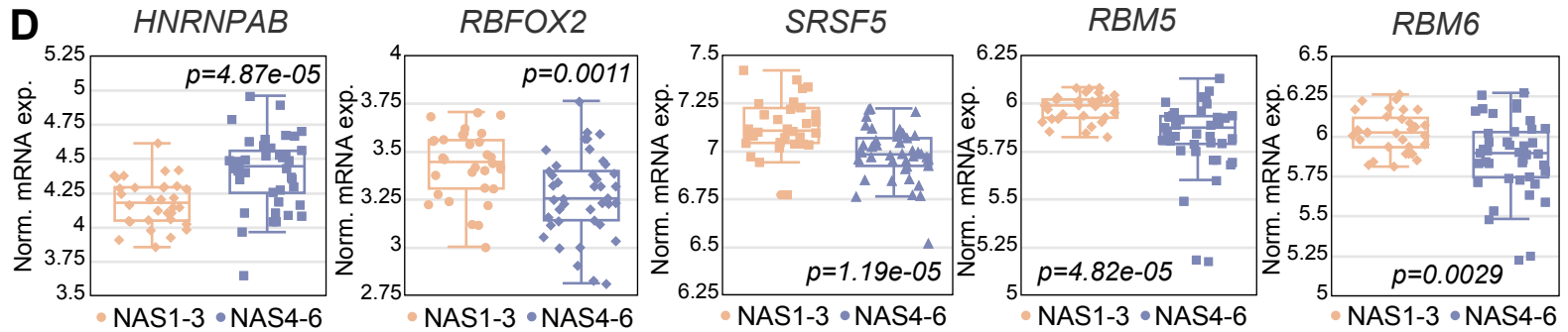
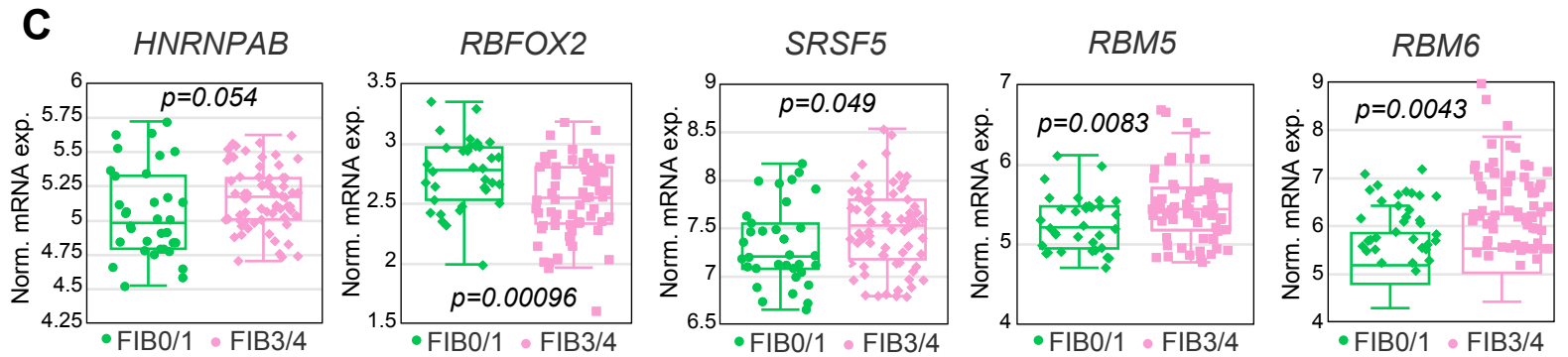
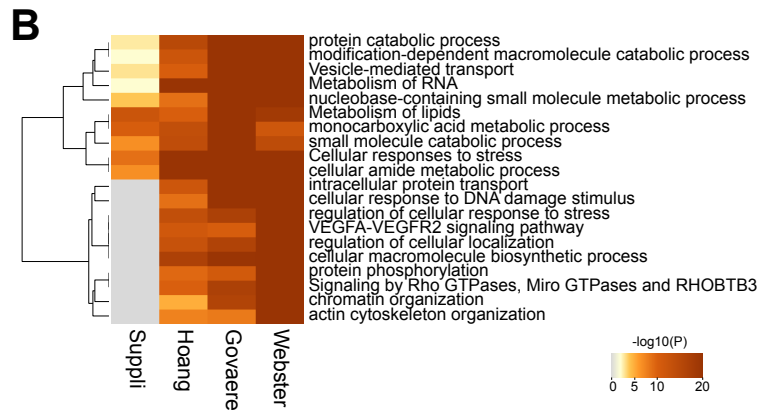
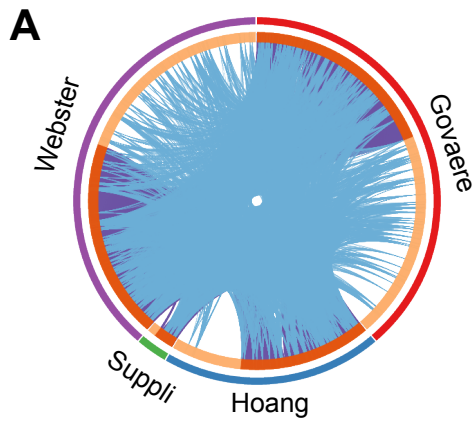


Supplemental figure 5: Comparison of selected splicing events across four datasets.

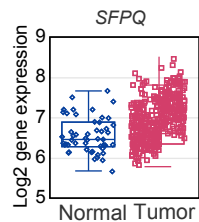
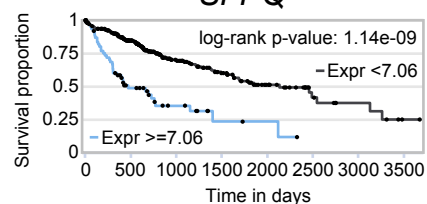
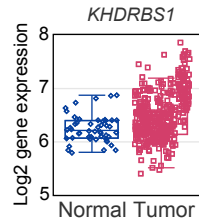
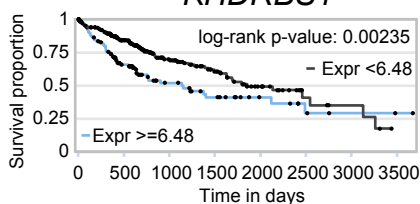
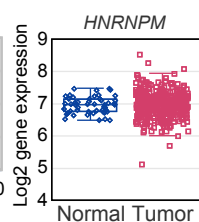
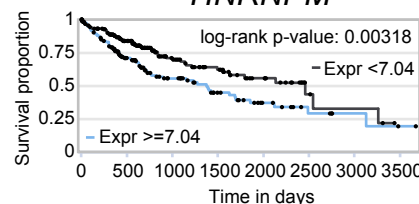
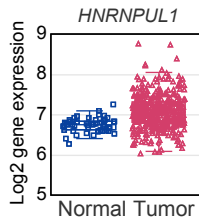
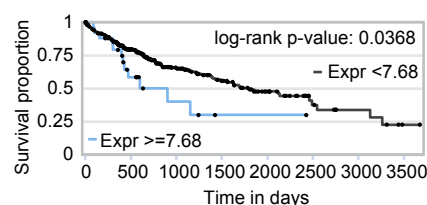
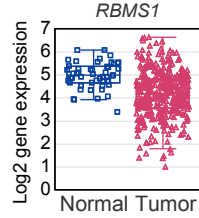
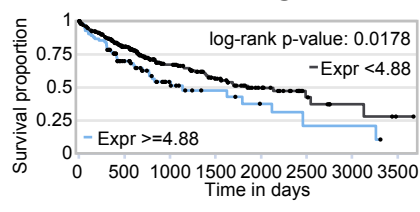
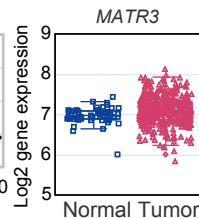
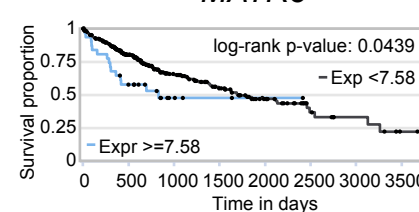
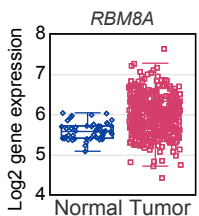
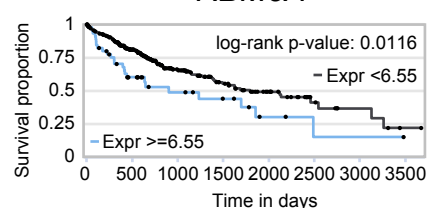
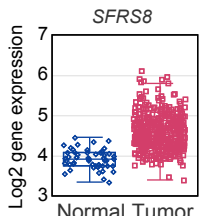
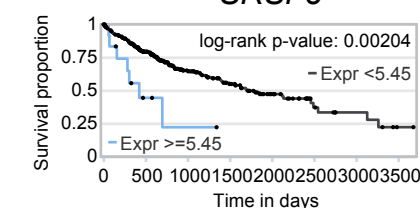
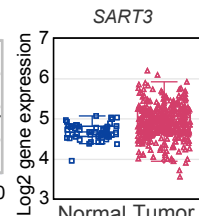
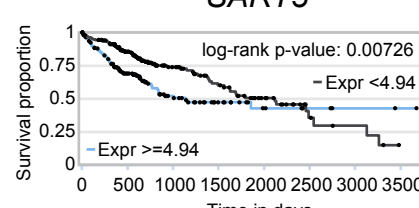
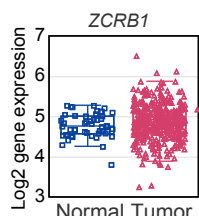
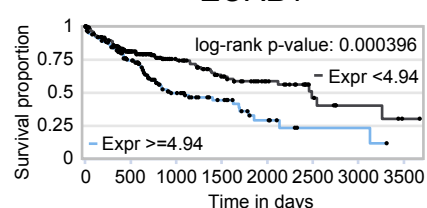
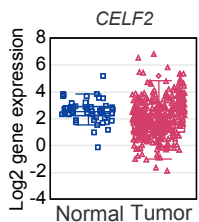
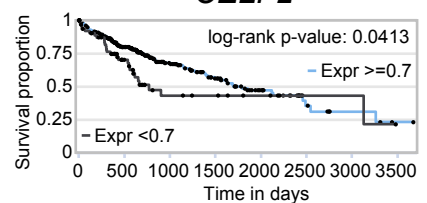
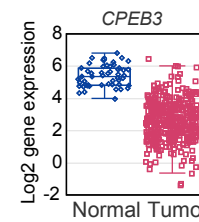
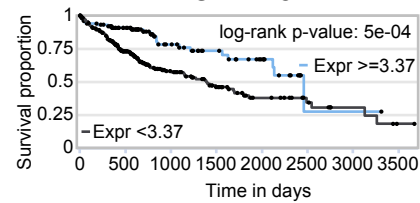
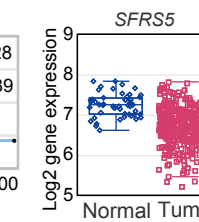
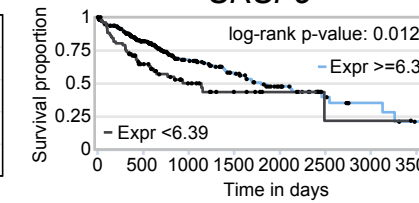
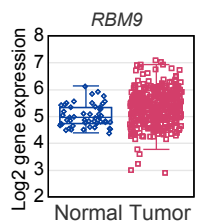
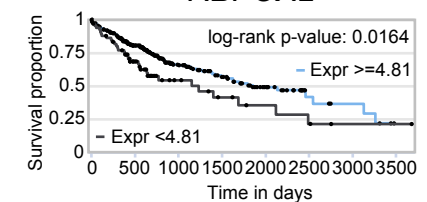
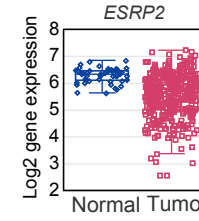
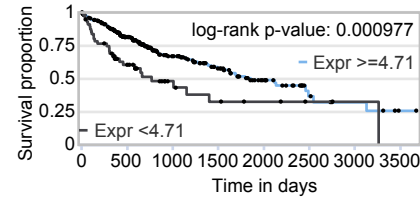
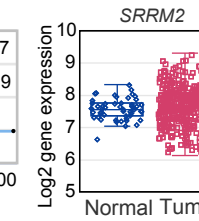
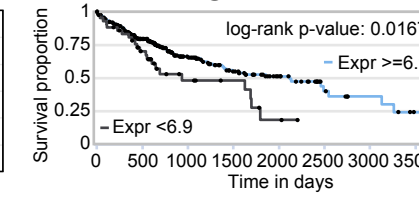
A: Splicing of TPM1 alternative first exon across the four datasets (this study, Hoang et al, Suppli et al. or Govaere et al.). Kruskal-Wallis p-value is given for each dataset. Hoang et al. data is shown based on fibrosis score. B: Splicing of PANK1 alternative first exon across the four datasets Hoang et al. data is shown based on steatosis score. C: Splicing of IRF3 mutually exclusive exon across the four datasets Hoang et al. data is shown based on steatosis score. D: Analysis of the cell of origin for the genes harboring the 136 SEs that are common to the four datasets. E: Analysis of the disease associations for the genes harboring the 136 SEs that are common to the four datasets.



Supplemental Figure 6: Early changes in splicing. A: Heatmap for the top 500 SEs in Hoang et al., after filtering 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). Samples are clustered according to the NAFLD activity score (NAS). Blue indicates low PSI, red high. B: Circos plot showing the overlap of genes harboring the altered SEs in the four studies. Purple lines indicate shared genes, blue lines indicate shared pathways or functions. C: Enrichment of GO terms for genes harboring the altered SEs from the four datasets. Color indicates significance ($-\log_{10}P$). D: Expression of the common early splicing factors *HNRNPAB*, *A1CF*, *PCBP2* and *PABPN1* in obese and NAFLD groups in Suppli et al. E: Expression of *HNRNPAB*, *A1CF*, *PCBP2* and *PABPN1* in NAS0 vs NAS1-3 groups in Hoang et al.. F: Expression of *HNRNPAB*, *A1CF*, *PCBP2* and *PABPN1* in control and NAFLD groups in Govaere et al.

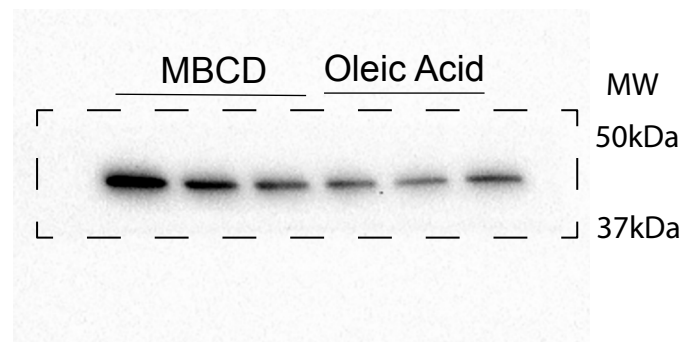
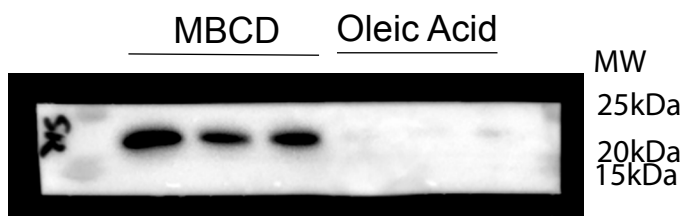
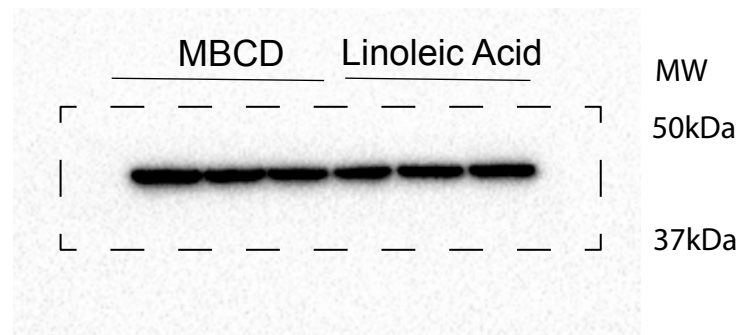
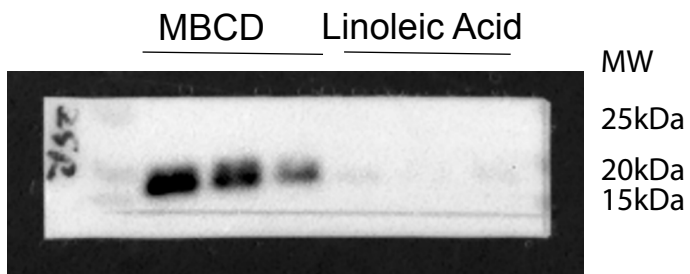
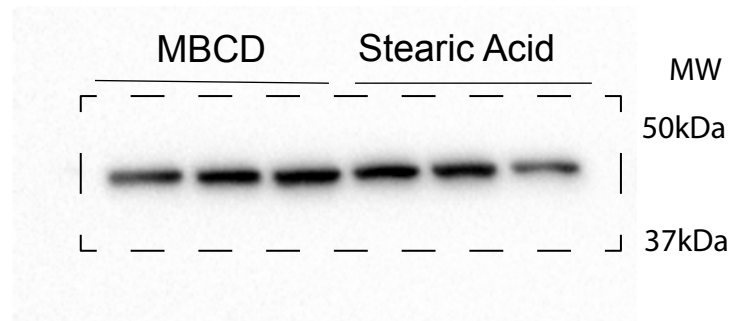
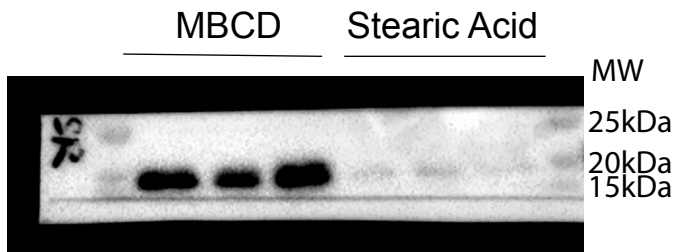
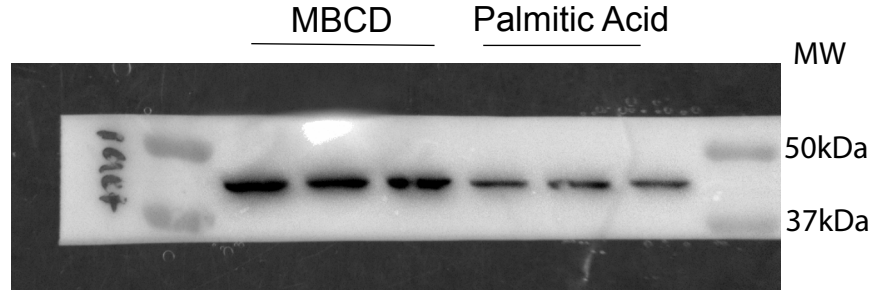
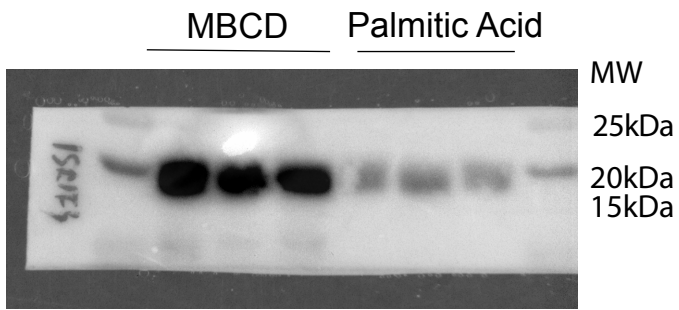


Supplemental Figure 7: Late changes in splicing. A: Circos plot showing the overlap of genes harboring the altered late SEs in the four studies. Purple lines indicate shared genes, blue lines indicate shared pathways or functions. C: Enrichment of GO terms for genes harboring the altered late SEs from the four datasets. Color indicates significance ($-\log_{10}P$). D: Expression of *HNRNPAB*, *RBFOX2*, *SRSF5*, *RBM5* and *RBM6* in FIB0/1 vs FIB3/4 groups in Hoang et al.. E: Expression of *HNRNPAB*, *RBFOX2*, *SRSF5*, *RBM5* and *RBM6* in NAS1-3 and NAS4-6 groups in Govaere et al. F: Expression of the common late splicing factors *HNRNPAB*, *RBFOX2*, *SRSF5*, *RBM5* and *RBM6* in NAFLD and NASH groups in Suppli et al.

A*SFPQ**KHDRBS1**HNRNPM**HNRNPUL1**RBMS1**MATR3**RBM8A**SRSF8**SART3**ZCRB1***B***CELF2**CPEB3**SRSF5**RBFOX2**ESRP2**SRRM2*

Supplemental Figure 8: Survival curves for common splicing factors. A: Kaplan-Meier survival curves for HCC patients and gene expression of splicing factors *SFPQ*, *KHDRBS1*, *HNRNPM*, *HNRNPUL1*, *RBMS1*, *MATR3*, *RBM8A*, *SRSF8*, *SART3* and *ZCRB1* where lower expression is associated with better survival. B: Kaplan-Meier survival curves for HCC patients and gene expression of splicing factors *CELF2*, *CPEB3*, *SRSF5*, *RBFOX2*, *ESRP2*, and *SRRM2* where higher expression is associated with better survival. Log-rank p-values are given.

SRSF3 blots



Supplemental Figure 9: Original blots for Figure S4. Equal amounts of protein were separated by SDS-PAGE and transferred to PVDF membranes. Membranes are cut below the 37kDa marker and above the 50 kDa marker (pre-stained markers) so that SRSF3 and β -actin could be quantified from the same gel. The lower portion was used for immunoblotting for SRSF2 (21 kDa) and the upper portion used for immunoblotting for β -actin (42 kDa) as a loading control. Chemiluminescence and brightfield images were taken of the SRSF3 blots (left panels) and merged to allow positioning of the molecular weight markers relative to the SRSF3 protein. β -actin always runs between the 37 and 50 kDa MW markers so merged images are not routinely taken, only chemiluminescence. Top panel for palmitic acid shows a representative merged images for β -actin (right panels). Dotted outlines below show the dimensions of the membranes in the chemiluminescence images.