

Supplemental Materials

Dysregulation of murine long non-coding single cell transcriptome in non-alcoholic steatohepatitis and liver fibrosis

Kritika Karri and David J. Waxman

Department of Biology and Bioinformatics Program
Boston University, Boston, MA 02215 USA

Supplemental Tables S1-S8 (Excel workbooks)

GTF file: gtf76k_mm10_Liver_scRNAseq.gtf

Loupe Browser files:

- 1_Control_Liver_scRNAseq.cloupe (300 MB)
- 2_Chow_NASH_NAFLD.cloupe (525 MB)
- 3_Healthy_CCl4_Mesenchyme.cloupe (250 MB)

Supplementary Figure legends

Fig. S1 – Identification of liver cell clusters from healthy mouse liver. A. Dot plot showing average expression values for marker genes (shown along X-axis) across the 13 hepatic cell clusters (Y-axis) shown in the UMAP in Fig. 2A. Those cell clusters were based on hepatocytes and NPCs from four datasets from control mouse livers. **B.** Dot plot showing average expression values for marker genes (shown along X-axis) across the 12 hepatic cell clusters (Y-axis) shown in the UMAP in Fig. 4A. Those cell clusters were based on liver cells isolated from healthy mice (chow-fed diet), and from mice fed HFHFD for 15 wk to induce simple steatosis (NAFLD), or for 30-34 wk to induce NASH.

Fig. S2 – Kupffer cell subpopulations. Feature plots presenting marker gene expression for Trem2, and for 4 lncRNAs that are markers for different Kupffer subpopulations. UMAP is as shown in Fig. 4G.

Fig. S3 – Shared network-essential regulators in healthy and diseased (NAFLD and NASH) mouse livers. Venn diagram showing overlap between network-essential PCGs and lncRNAs from control, NAFLD and NASH gene regulatory networks from Fig. 6 and Table S5.

Fig. S4 – Ctflos regulation of Pck1. Subnetwork showing Ctflos (lnc2065), an essential network regulatory gene in the NASH network, which makes connections with Pck1 via lnc10621.

Fig. S5 – Differentiation of quiescent HSCs into myofibroblasts. A. UMAP showing HSCs zonation clusters (top left) and corresponding DotPlot of known HSC marker genes (top right). Shown are feature plots for genes that are markers for quiescent HSCs (Ecm1, Rgs5) and for activated HSCs (myofibroblast) (Col1a1, Col1a2). **B.** Heatmap showing relative expression patterns for PCGs that are differentially expressed between quiescent pericentral HSCs and myofibroblasts from CCl₄-exposed mouse liver mesenchymal cells, along with their top functional enrichment terms, listed on the right. See Fig. 7E for corresponding heat map for lncRNA genes.

Fig. S6 – Feature plots showing examples of lncRNAs that are differentially expressed across multiple mesenchymal cell subpopulations.

Fig. S7 – Shared gene regulatory network regulators in healthy and CCl₄-induced fibrotic liver. Venn diagram showing overlap between network-essential PCGs and lncRNAs from the healthy and fibrotic liver gene regulatory networks shown in Fig. 8. See details in Table S5.

Fig. S8 – Clustering of 39 regulatory lncRNAs from chow-fed (control) liver gene regulatory network. **Top:** PCG targets that make a direct connection with regulatory lncRNAs in the chow fed liver network of Fig. 6A were provided as input to Metascape to obtain top functional enrichment terms based on $-\log_{10}$ (P-value). The color scale represents $-\log_{10}$ (P) values, with darker purple coloration indicating more significant p-values. Two-way hierarchical clustering of the functional enrichment heatmap split the 39 lncRNAs into 3 clusters (column colors) based on their patterns of shared functions. Target genes of lncRNAs from two of the major functional clusters (green, blue) were variously enriched for diverse metabolic processes, and a third (yellow) was functionally enriched for angiogenesis and vasculature development, extracellular matrix, and cell-cell adhesion, as shown on the right. **Bottom:** subnetwork of the network shown in Fig. 6A. This subnetwork is comprised of the network-essential 39 regulatory lncRNAs and their target genes, which clustered to give gene modules that mirror the clustering results obtained for the same lncRNAs based on their functional group enrichments, shown on top. The subnetwork was derived from the network in Fig. 6A by removing all PCG regulators and their gene targets, and by retaining regulatory lncRNAs and their direct protein coding genes connections, which were used as input for the functional enrichment analysis shown at the top. The regulatory lncRNAs (nodes) are color coded to match the lncRNA cluster colors shown in the heat map. Gene targets for lnc6251 (bottom, gray node) did not show enrichment for any functional terms and was therefore excluded from the enrichment heatmap (top).

Fig. S9 – Clustering of 28 regulatory lncRNAs from NAFLD gene regulatory network. **Top:** Analysis of network-essential regulatory lncRNAs from NAFLD network of Fig. 6B, as described in Fig. S8. Two-way hierarchical clustering identified 3 clusters of regulatory lncRNAs, with target genes of lncRNAs from cluster blue enriched for carboxylic acid and other metabolic processes, cluster green lncRNA targets enriched for vasculature development, angiogenesis, hippo signaling, extracellular matrix and cell motility, and the target genes of lncRNAs from cluster yellow, which is comprised of only 2 lncRNAs (lnc6840, lnc12750), specifically enriched for import across plasma membrane and ion transport. **Bottom:** subnetwork of the network shown in Fig. 6B. It is comprised of the 28 regulatory lncRNAs and their target genes (excluding regulatory PCGs) and was obtained as described in Fig. S8. The 28 regulatory lncRNAs clustered to give gene modules that mirror the clustering results obtained for the same lncRNAs based on their functional group enrichments, shown on top.

Fig. S10 – Clustering of 25 regulatory lncRNAs from NASH gene regulatory network. **Top:** Analysis of network-essential regulatory lncRNAs from NASH network of Fig. 6C, as described in Fig. S8. Two-way hierarchical clustering identified 4 clusters of regulatory lncRNAs, with the target genes of lncRNAs from cluster green enriched for vasculature development, angiogenesis, and cell motility, and with a subset of lncRNAs in this cluster enriched for extracellular matrix, wound healing, and regulation of Wnt signaling. The target genes of lncRNAs from cluster yellow (n=3 lncRNAs) were enriched for inflammatory response gene targets, while the target genes of lncRNAs from cluster orange (n=2 lncRNAs) were enriched for cation transmembrane transport. The targets of cluster blue lncRNAs (n=5) were enriched for carboxylic acid and other metabolic processes. **Bottom:** subnetwork of the network shown in Fig. 6C. It is comprised of the 25 regulatory lncRNAs and their target genes (excluding regulatory PCGs) and was obtained as described in Fig. S8. The 25 regulatory lncRNAs clustered to give gene modules that mirror the clustering results obtained for the same lncRNAs based on their functional group enrichments, shown on top.

Fig. S11 – Combined clustering of 92 regulatory lncRNAs from Chow, NAFLD and NASH liver gene regulatory networks. Shown is a merged functional enrichment heatmap showing the top enriched terms (as rows) for genes targets of all network-essential regulatory lncRNAs (as columns) from the chow, NAFLD, and NASH networks combined and analyzed as described in Fig. S8. The NAFLD and NASH network-derived regulatory lncRNAs showed more extensive and more significant enrichment for vasculature development and angiogenic processes than those from control liver, whereas the NAFLD and control network regulatory lncRNAs both

showed enrichment for diverse metabolic processes. Many fewer regulatory lncRNAs showed target gene enrichment for immune-related processes. Overall, this figure compares regulatory lncRNAs and their target functional terms across biological conditions.

Fig. S12 – Clustering of 19 regulatory lncRNAs from healthy liver mesenchymal cell gene regulatory network.

Top: Analysis of network-essential regulatory lncRNAs from the healthy liver mesenchymal network of Fig. 8A, as described in Fig. S8. Two-way hierarchical clustering identified 2 clusters of regulatory lncRNAs, with the target genes of lncRNAs from cluster blue (n=5 lncRNAs) significantly enriched for terms related to muscle contraction, notch signaling, and ion homeostasis, and those of cluster yellow (n=14 lncRNAs) strongly enriched for diverse biological processes, including cell morphogenesis, regulation of defense response, vascular development, and interferon responses. **Bottom:** subnetwork of the network shown in Fig. 8A. It is comprised of the 19 regulatory lncRNAs and their target genes (excluding regulatory PCGs) and was obtained as described in Fig. S8. The 19 regulatory lncRNAs clustered to give gene modules that mirror the clustering results obtained for the same lncRNAs based on their functional group enrichments, shown on top.

Fig. S13 – Clustering of 26 regulatory lncRNAs from CCl₄-induced liver fibrosis gene regulatory network. Top:

Analysis of network-essential regulatory lncRNAs from the CCl₄-treated mesenchymal cell network of Fig. 8B, as described in Fig. S8. Two-way hierarchical clustering identified 3 clusters of regulatory lncRNAs (blue, green, yellow), with the target genes of lncRNAs from these clusters respectively showing strongest enrichments and greatest specificities for: muscle contraction and notch signaling and related processes (blue); extracellular matrix organization and other terms (green); and cell junction assembly, response to interferon-beta, and other terms (yellow). **Bottom:** subnetwork of the network shown in Fig. 8B. It is comprised of the 26 regulatory lncRNAs and their target genes (excluding regulatory PCGs) and was obtained as described in Fig. S8. The 26 regulatory lncRNAs clustered to give gene modules that mirror the clustering results obtained for the same lncRNAs based on their functional group enrichments, shown on top.

Fig. S14 – Combined clustering of 45 network-essential regulatory lncRNAs from Healthy and CCl₄-induced liver fibrosis gene regulatory networks. Shown is a merged functional enrichment heatmap showing the top enriched terms (as rows) for genes targets of all network-essential regulatory lncRNAs (as columns) from the healthy and fibrotic liver mesenchymal gene regulatory networks combined and analyzed as described in Fig. S8. Shown are pathways that were either common or specific to control versus CCl₄-treated mesenchymal lncRNA gene targets. Overall, this figure compares regulatory lncRNAs and their target functional terms across two biological conditions.

Fig. S15 – lncRNA-PCG regulatory network based on genes predicted to be involved in lncRNA-genomic DNA triplex binding.

Shown are regulatory subnetworks for those lncRNAs that are enriched for forming triplex interactions (based on TDF analysis) with their protein coding gene (PCG) targets identified using bigScale2 gene regulatory networks from Chow diet (control) liver (**A**), NAFLD liver (**B**), and NASH liver (**C**) (see Fig. 6). The networks shown here are subsets of those shown at the bottom of Fig. S12 and Fig. S13, respectively, insofar as only the regulatory lncRNAs and their protein-coding gene targets that form triplexes based on TDF are shown.

Fig. S16 – lncRNA-PCG regulatory network based on genes predicted to be involved in lncRNA-genomic DNA triplex binding.

Shown are regulatory subnetworks for those lncRNAs that are enriched for forming triplex interactions (based on TDF analysis) with their PCG targets identified using bigScale2 gene regulatory networks for mesenchymal cells from control liver (**A**) and from CCl₄-induced fibrotic liver (**B**) (see Fig. 8). The networks shown here are subsets of those shown at the bottom of Figs. S8, S9 and S10, respectively, insofar as only the regulatory lncRNAs and their protein-coding gene targets that form triplexes based on TDF are shown.

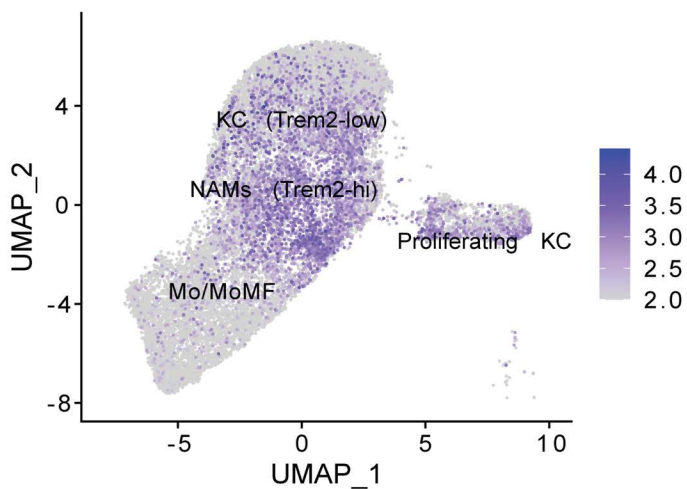
Example: In Fig. 8A, the lncRNA Meg3 is connected to many PCGs and lncRNAs, some of which have their own connections. In Fig. S12, where only regulatory lncRNAs and their direct connections with other regulatory lncRNAs or protein coding genes are shown, Meg3 is in an isolated subnetwork comprised only of its direct target genes. Finally, in Fig. S16A, the subnetwork with Meg3 is also in an isolated subnetwork, one that has

even fewer genes than in Fig. S12, because only the targets that are predicted to form a triplex with Meg3 are shown.

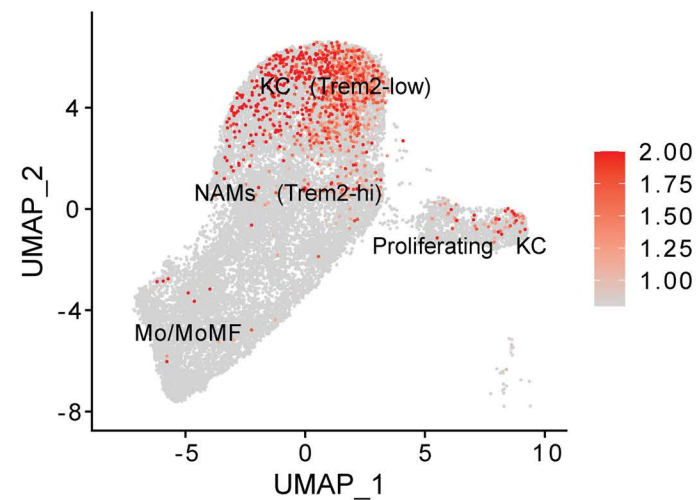
Fig. S17 – Macrophage proliferation. Violin plots showing expression of genes related to cell division (Top2a and Stmn1) in proliferating Kupffer cells (from Fig. 4G).

Fig. S2

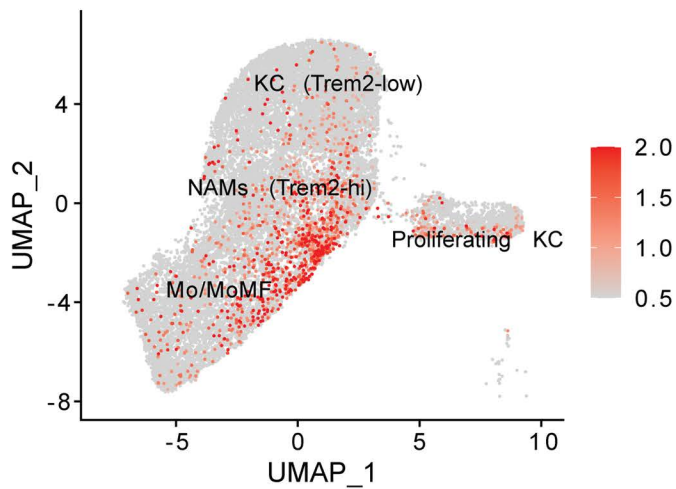
Trem2



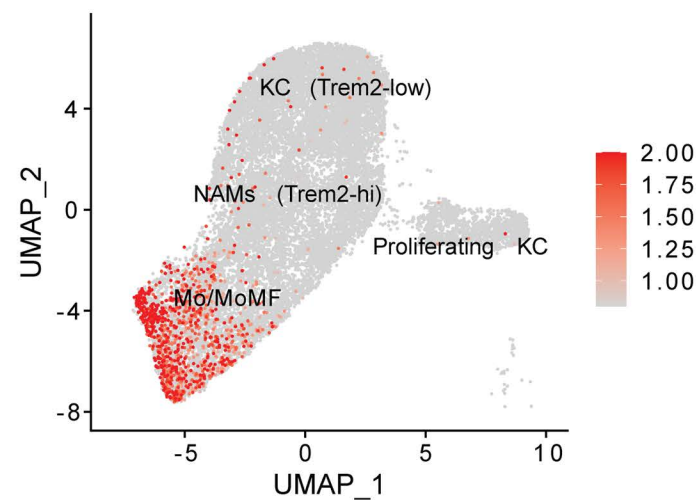
Inc10020



Inc7979



Inc5646



Inc38082

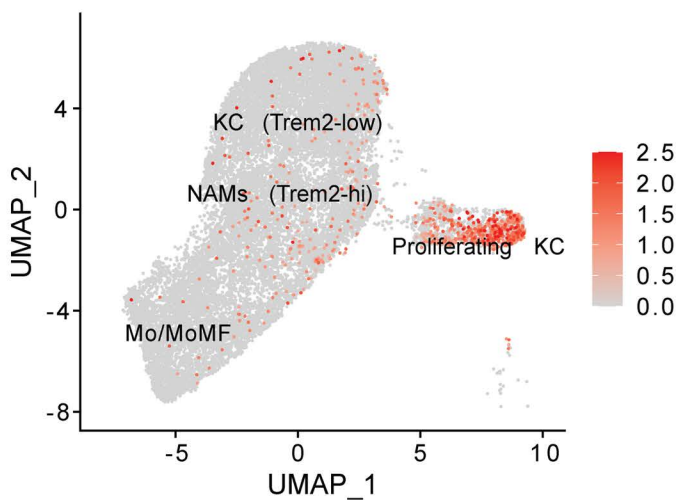


Fig. S3

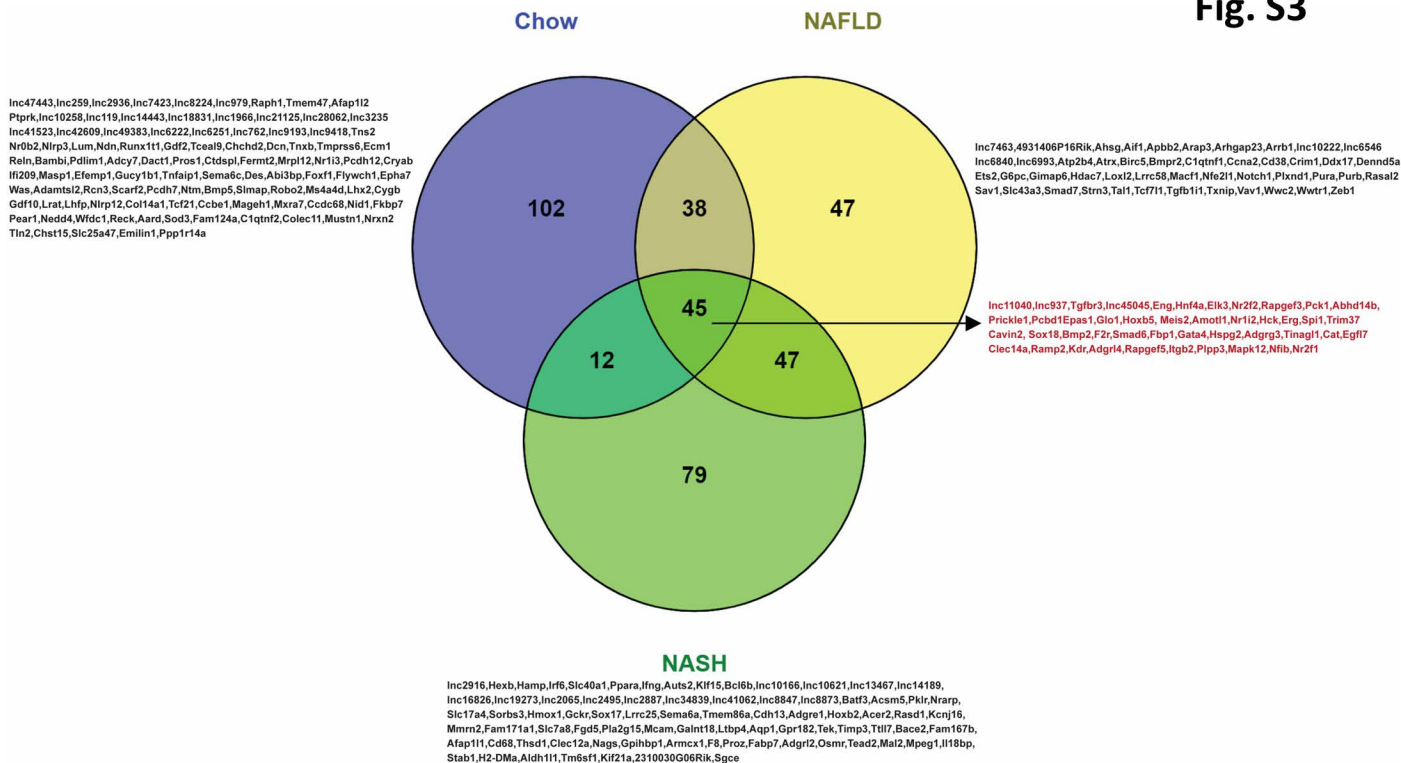


Fig. S4

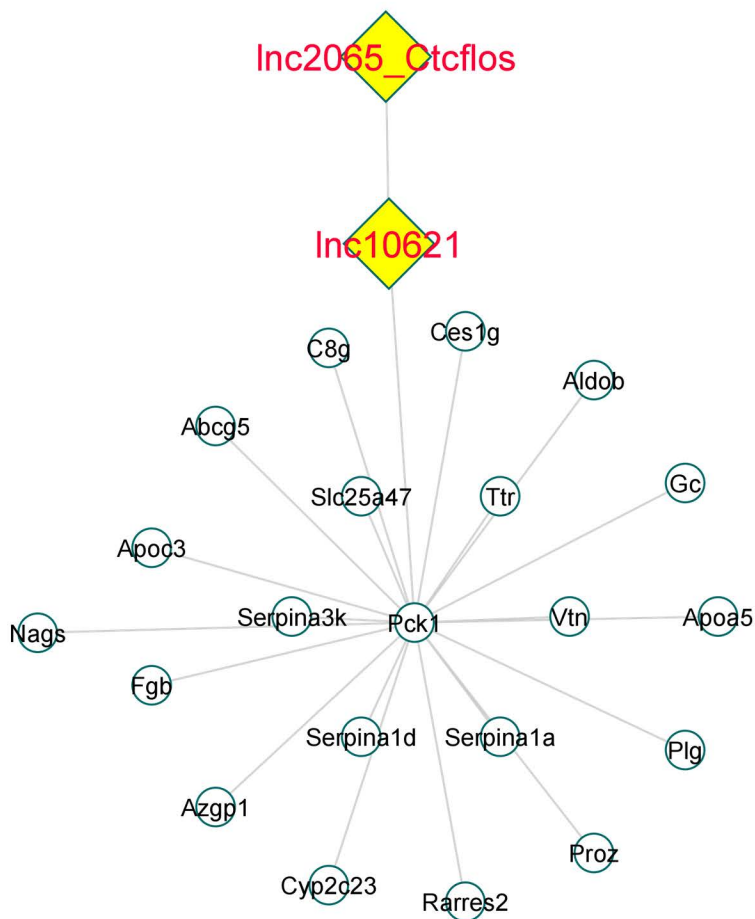


Fig. S6

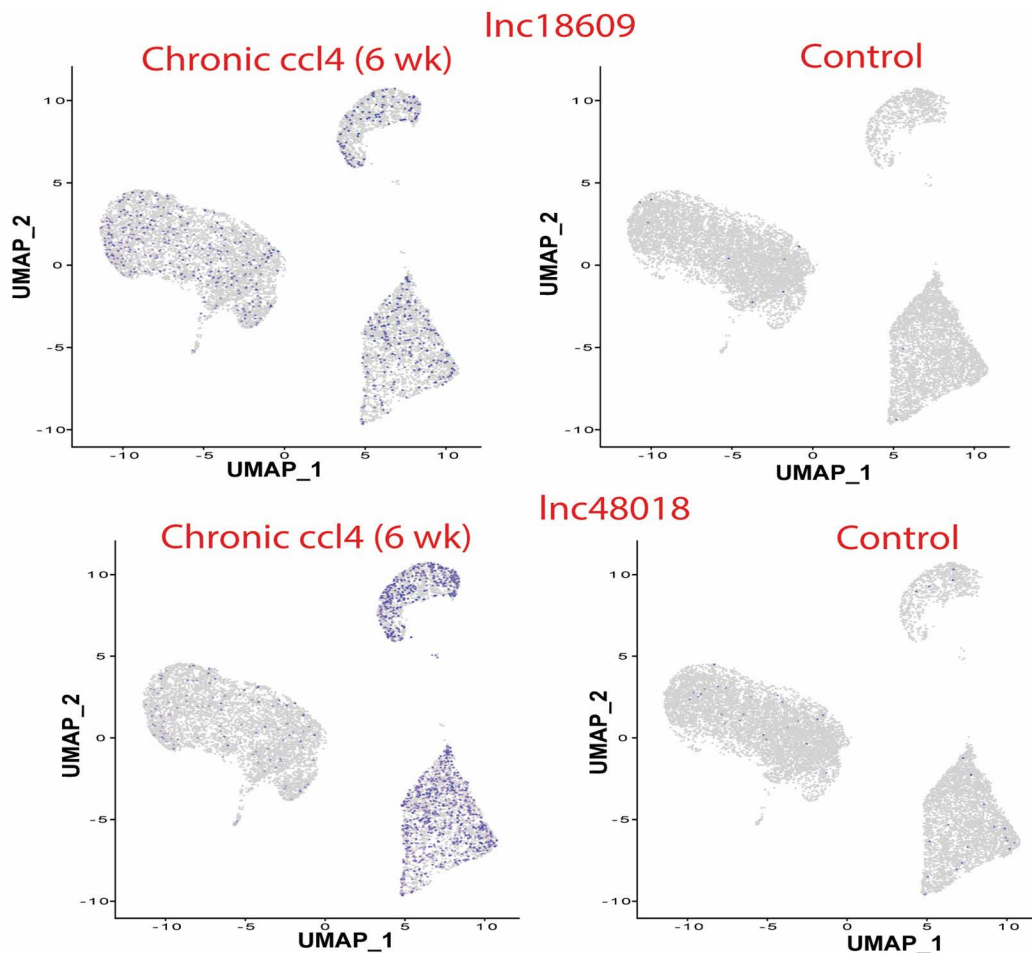


Fig. S7

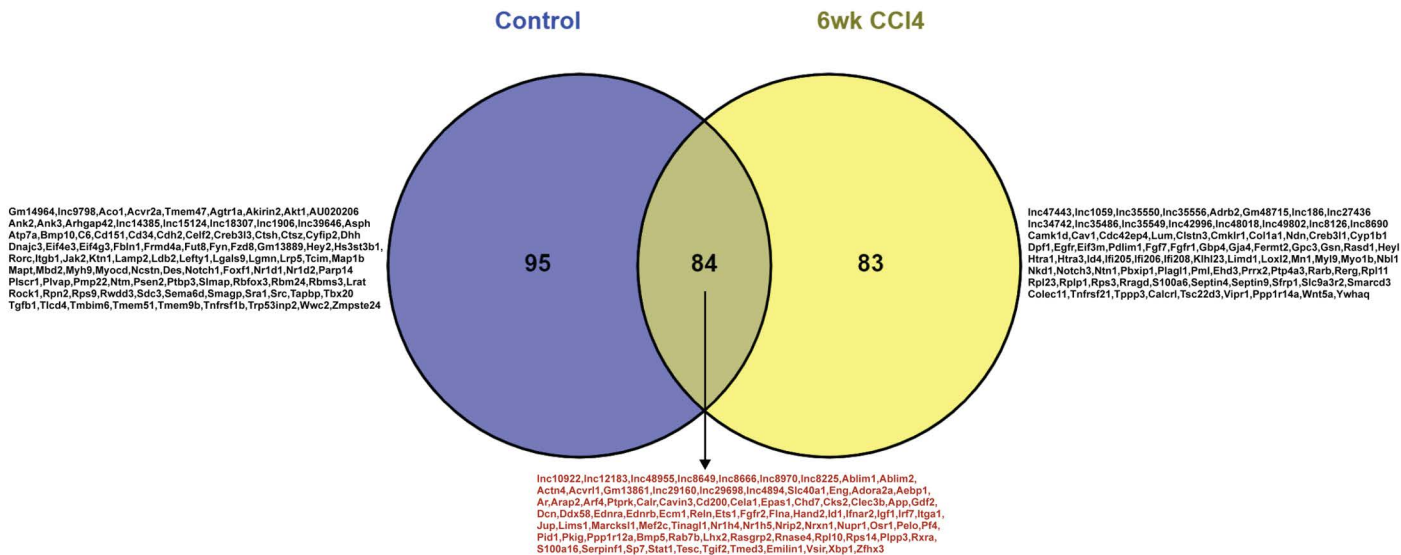


Fig. S8 (Chow diet)



Fig. S9 (NAFLD)



Fig. S10 (NASH)



Fig. S11 (Chow-NAFLD-NASH)

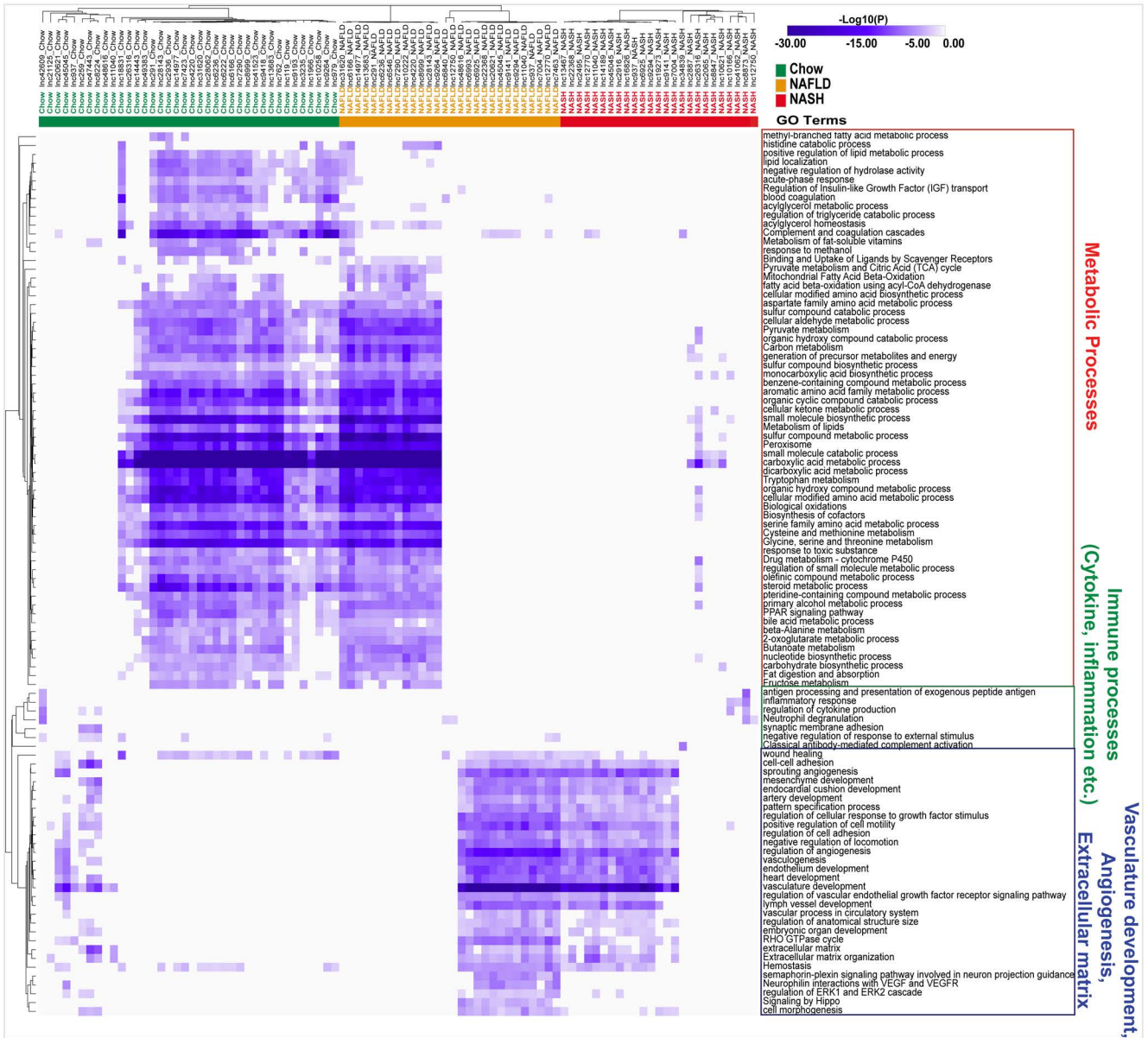


Fig. S12
(Healthy mesenchyme)

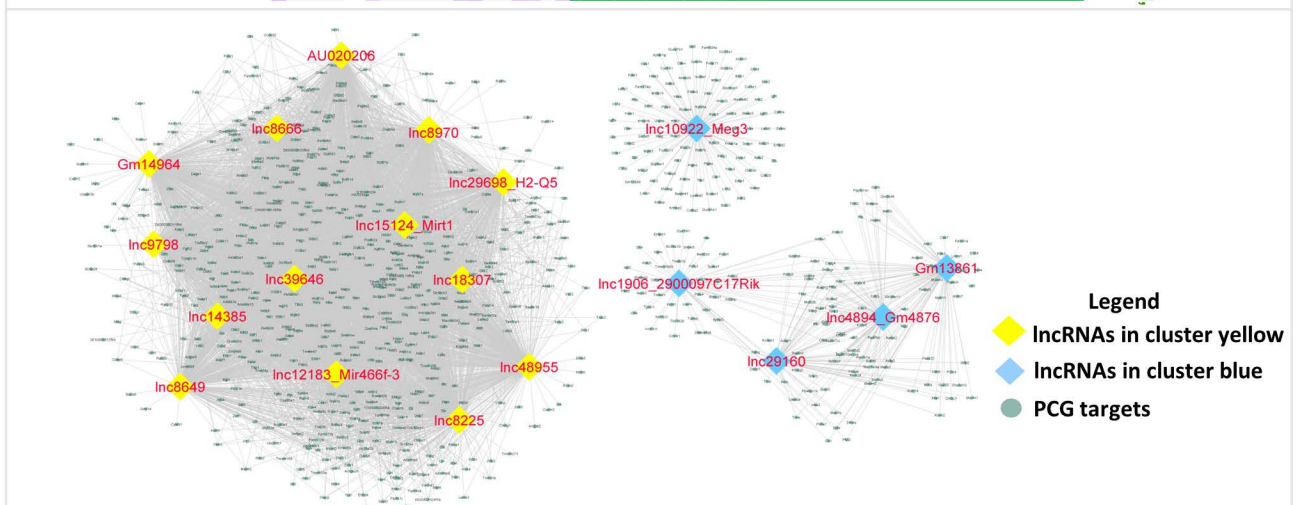
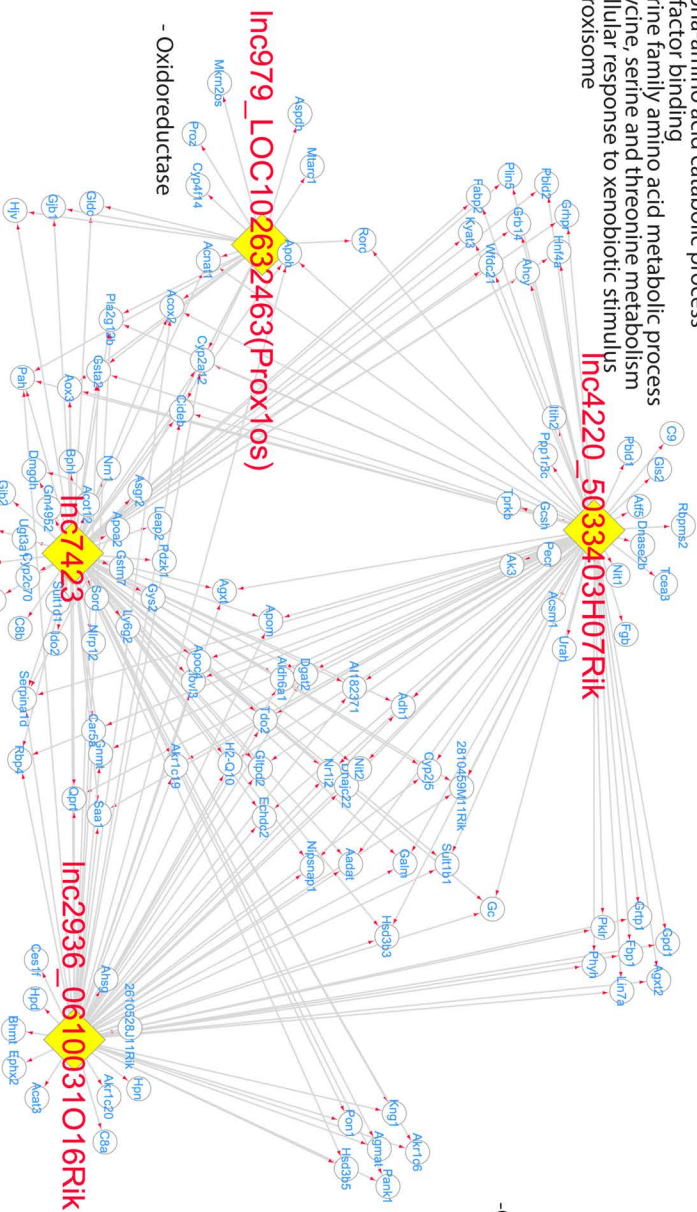


Fig. S13
(CCl₄ mesenchyme)



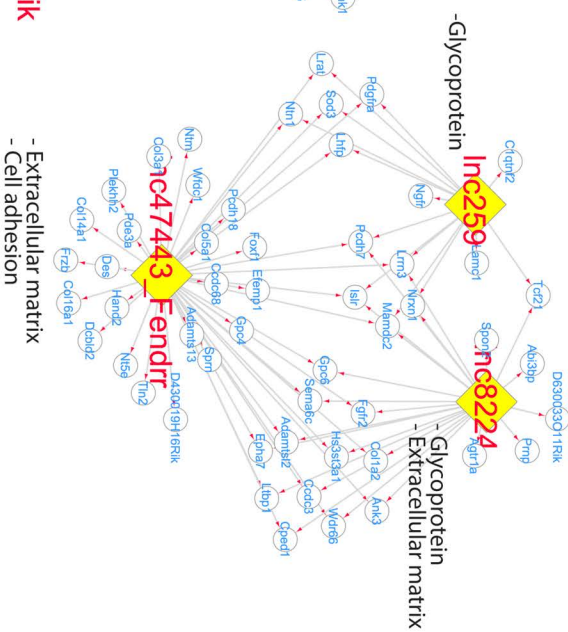


- Organic acid metabolic process
- Alpha-amino acid catabolic process
- Cofactor binding
- Serine family amino acid metabolic process
- Glycine, serine and threonine metabolism
- Cellular response to xenobiotic stimulus
- Peroxisome



- Oxoaacid metabolic process
- Lipid metabolic process
- Carboxylic acid binding
- Small molecule catabolic process
- Sulfur compound metabolic process
- Serine family amino acid metabolic process
- Cofactor binding
- Mitochondrion
- Oxidoreductase

- Oxoaacid metabolic process
- Monocarboxylic acid metabolic process
- Carboxylic acid binding
- Idenical protein binding
- Alpha-amino acid catabolic process
- Vitamin binding
- Sulfur compound metabolic process
- Lipid homeostasis
- Steroid metabolic process



- Extracellular matrix
- Cell adhesion

- Glycoprotein
- Glycoprotein
- Extracellular matrix

Fig. S15A
(Chow)

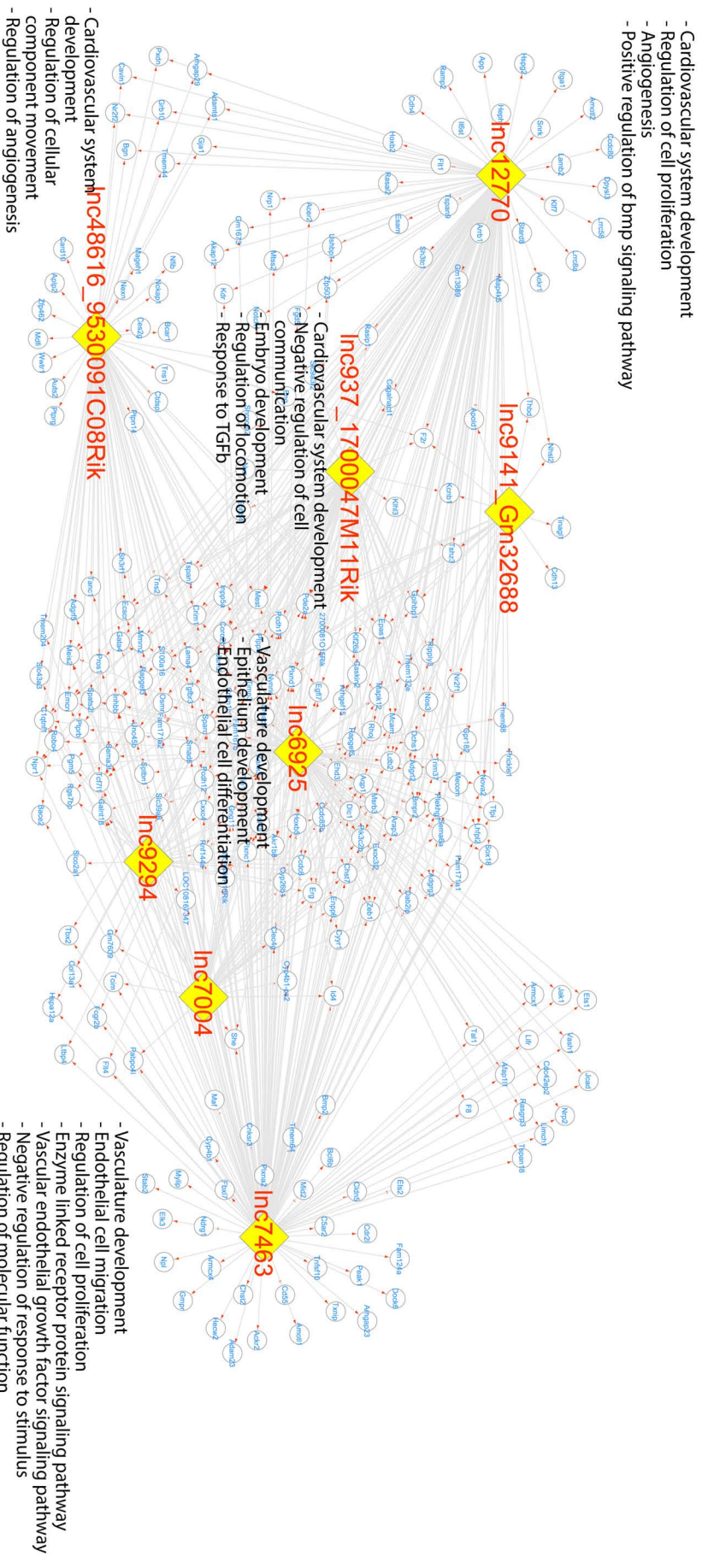


Fig. S15B
(NAFLD)

- Angiogenesis
- Vasculature development
- Negative regulation of developmental process
- Endothelial cell differentiation

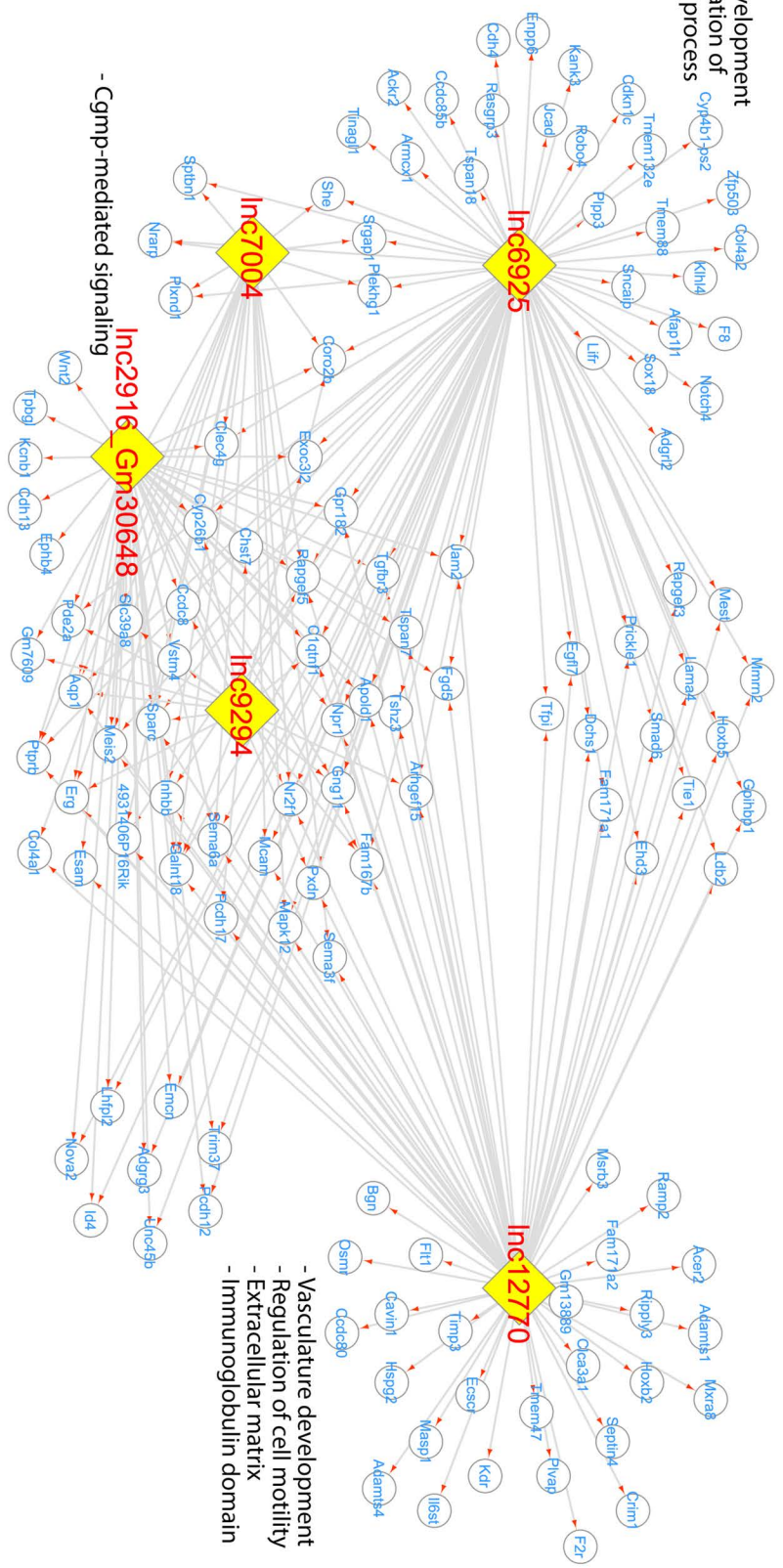


Fig. S15C (NASH)

Fig. S16A

(Healthy Mesenchyme)

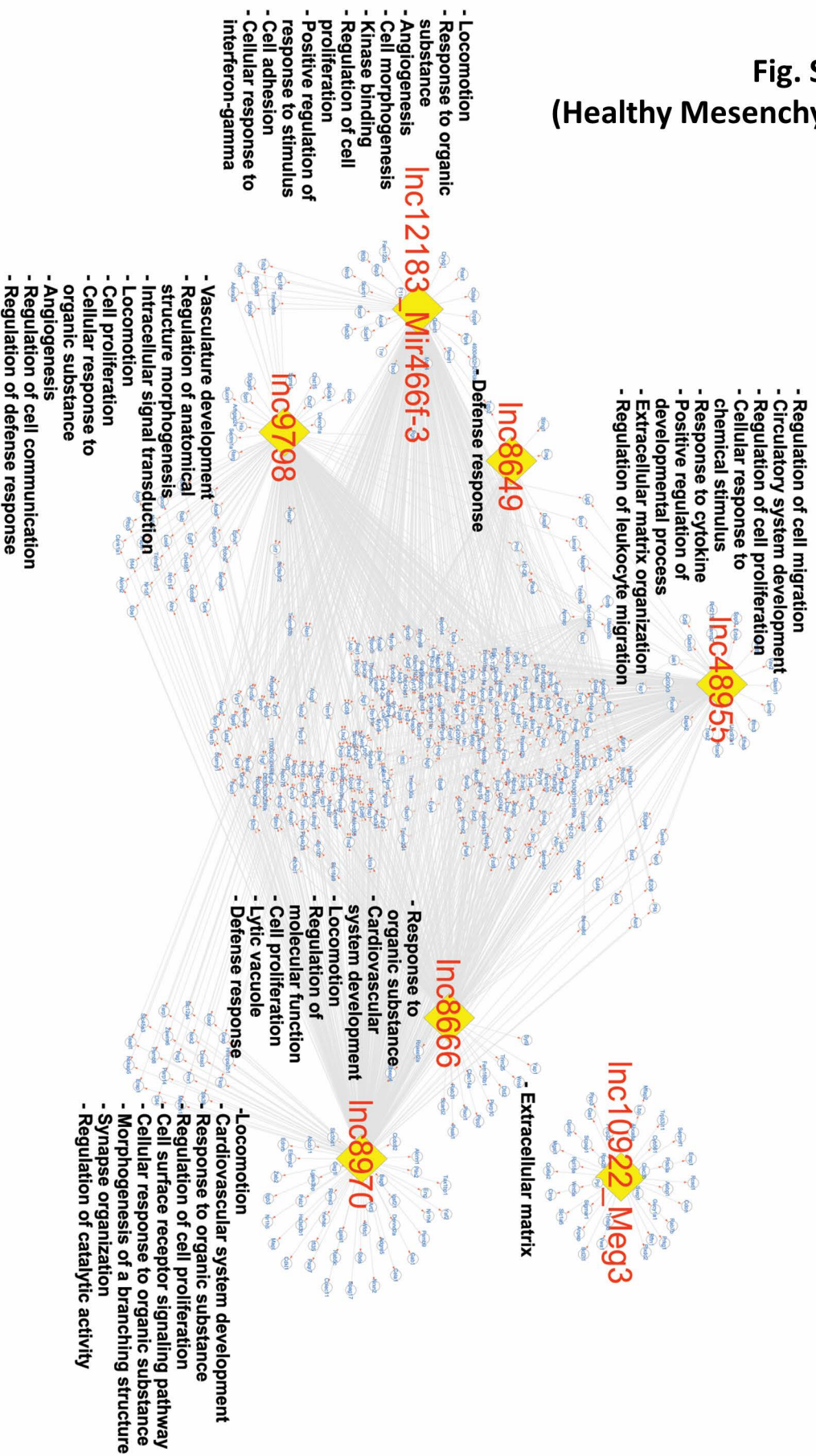


Fig. S16B
(CCl₄ Mesenchyme)

Fibrosis (6wk CCl₄)

