Title:

Gene Expression Signatures Associated With Survival Times of Pediatric Patients With Biliary Atresia Identify Potential Therapeutic Agents

Supplementary Methods

RNA sequencing analyses

Total RNA was isolated from liver biopsies using the miRNeasy Kit (QIAGEN), verified for integrity by the Agilent 2100 Bioanalyzer, and converted to cDNA for sequencing library preparation by TruSeq Stranded mRNA Library Prep Kit from Illumina (San Diego, CA). DNA fragments were end-repaired to generate blunt ends with 5′ phosphatase and 3′ hydroxyls and adapters were ligated for single-end or paired-end sequencing on Illumina HiSeq 2500. 10 million reads with 50bp in length for single-end RNAseq and 20 million mate-paired reads with 125-bp in length for paired-end RNAseq were generated. Single-end RNAseq was used in discovery cohort and normal controls, and paired-end RNAseq was used in the validation cohort.

RNAseq reads were aligned to the human genome (GRCh37/hg19) using TopHat (version 2.1.0),[1] with transcript quantifications performed by Cufflinks v2.2.1.[2] Annotated transcripts were obtained from the UCSC genome browser (http://genome.ucsc.edu) and the Ensembl database. Transcript abundances were normalized in reads per kilobase of exon per million mapped reads (RPKM) for single-end RNAseq and fragments per kilobase of exon per million mapped reads (FPKM) for paired-end RNAseq. Genes with RPKM or FPKM <0.5 in more than half of samples were

considered not expressed and filtered out. Differentially expressed genes were identified by Cuffdiff 2.[3] The selection of genes with a fold-change cutoff at 1.5 or higher included the Benjamini–Hochberg false discovery rate adjusted $P < 0.05$.

Supervised principal components for Cox regression

Steps for Superpc Cox are summarized as follow: 1) Compute univariate Cox proportional hazard regression coefficients for each gene, 2) estimate coefficient threshold theta by 10 fold cross-validation, 3) generate a reduced gene expression matrix consisting of genes whose coefficients exceed the threshold theta, 4) compute the first principal component of the reduced matrix, 5) use the first principal component to predict the survival outcome, and 6) assign predicted scores to patients using the first principal component predictor and importance scores to genes. For the analysis, patients with higher predicted score have lower survival. The values of genes with positive importance scores means decrease in patient survival.

Gene signatures for liver diseases and animal models of hepatopathies

Gene signatures for liver diseases and their animal models are defined as differentially expressed genes between disease samples and normal control samples identified by GEO2R (https://www.ncbi.nlm.nih.gov/geo/geo2r/). We also curated signatures associated with inflammation or fibrosis in BA, and prognosis signatures in hepatocellular carcinoma. Genes with the fold-change cutoff at 2 or higher and Benjamini–Hochberg false discovery rate adjusted $P<0.05$ were considered differentially expressed. These signatures underwent pairwise overlapping comparisons between upregulated or downregulated genes in low survival group and gene signatures of human liver diseases using Fisher's exact test.

Pathway enrichment analysis and estimation of relative cell abundance

For estimation of the relative abundance of individual cell types, unique marker gene signatures were obtained for immune cell types from two different studies (Bindea et al[4] and Charoentong et al[5]); and for activated hepatic stellate cells, activated portal fibroblasts and cholangiocytes from Zhang et al[6], Iwaisako et al[7] and Dianat et al[8], respectively. Gene signature for glutathione metabolism was obtained from KEGG database (KEGG pathway: hsa00480). Levels of cell types or glutathione metabolism were quantified based on the normalized RNAseq expression values (RPKM or FPKM) of marker genes using the single sample Gene Set Enrichment Analysis (ssGSEA) implemented in GSVA-R package. ssGSEA is rank-based method that produces an enrichment score for each sample. The enrichment score was calculated by a sum of the difference between weighted Empirical Cumulative Distribution Functions of the marker genes and the other genes in an expression profile. The enrichment score represents the abundance of cells or enrichment of pathway. Differences between the low and high survival groups were tested by Wilcoxon rank sum test. To calculate the ratio of ssGSEA score for glutathione metabolism to ssGSEA score for aHSCs, aPFs or cholangiocytes, we first scaled the ssGSEA score to 0.1-1 range, then divided ssGSEA score for glutathione metabolism by ssGSEA score for aHSCs, aPFs or cholangiocytes.

Standard and modified neonatal mouse model of biliary atresia and anti-oxidant treatment

Testing the effect of anti-oxidants in this model, N-Acetyl-Cysteine (NAC; Sigma, A9165) or Manganese (III) tetrakis-(4-benzoic acid)porphyrin (MnTBAP; EMD Millipore, Burlington, MA) were reconstituted in 1xPBS, and administered in daily doses of 150

mg/Kg (for NAC) or 5 mg/Kg (for MnTBAP) beginning 12 hours after RRV inoculation; equal volumes of 1xPBS were injected into RRV infected mice serving as controls. All groups of experimental mice were phenotyped daily as described previously[9]. In brief, Mice were sacrificed at 7 and 14 days for standard model and at 7, 14 and 19 days for modified model after RRV injections, and extrahepatic bile duct, liver and plasma samples were obtained for analyses. Mice treated with 0.9% saline were used as normal controls. All BALB/c mice were maintained with normal diet in pathogen-free vivarium rooms equipped with a 12-hour dark-light cycle. Animal studies and experimental protocols were performed in strict accordance with the guidelines recommended by the Institutional Animal Care and Use Committee (IACUC; Protocol IACUC2017-0007) of Cincinnati Children's Hospital Medical Center.

Histopathology and immunofluorescence staining

Mouse livers and EHBDs were micro-dissected using a stereomicroscope, and tissues were paraffin-embedded, sectioned, and stained with hematoxylin/eosin for both tissues and with Sirius red, α-SMA staining for livers only. Human liver biopsies were obtained at the time of HPE and snap-frozen in liquid nitrogen, formalin fixed/paraffin embedded and sectioned. Liver sections were blocked from nonspecific binding by using normal goat serum and incubated with the primary antibodies rabbit anti-α-smooth muscle actin (α-SMA; ab5694 from Abcam, Cambridge, MA, USA for mouse tissues and 1A4 from Roche, Indianapolis, IN, USA for human tissues) or rabbit anti-cytokeratin 7 antibody (KRT7; ab119697 from Abcam, Cambridge, MA, USA for mouse and human livers). Secondary anti-rabbit biotinylated antibodies were from Vectastain ABC-HRP Kit (PK-

4001 from Vector Lab, Burlingame, CA, USA). Images were captured using Olympus BX51 microscope (Olympus America Inc., Center Valley, Pennsylvania, USA) and cellSens Dimension digital imaging software (Olympus corporation, Version 1.8.1); the stained areas were quantified by ImageJ software as reported previously.[10]

Human liver biopsies were obtained at the time of HPE. N=7 and N=8 human BA liver sections from low and high survival groups, respectively, and 3 normal liver sections were used for KRT7 and α-SMA staining. α-SMA or KRT7 positive areas in human liver sections were quantified by ImageJ software. Quantification of aHSCs and aPFs are represented as α-SMA positive areas in lobular and portal regions, respectively. Quantification of cholangiocytes are represented as KRT7 positive areas. α-SMA positive arteries and veins were manually excluded before quantification.

Colorimetric assay for total bilirubin, alanine transaminase (ALT) and gamma-glutamyl transferase (GGT)

Plasma total bilirubin, ALT and GGT concentrations were measured with Total Bilirubin Reagent Set (Pointe Scientific Inc., Canton, MI), Liquid GGT Reagent Set (Pointe Scientific Inc., Canton, MI) and DiscretPak ALT Reagent Kit (Catachem, Bridgeport, CT) according to the manufacturers' instructions. Photometric absorbance for assays was read on a Synergy H1 Hybrid Reader (BioTek, Winooski, VT) at 555 nm for total bilirubin and ALT, at 405nm for GGT.

Real-time PCR

RNA was isolated from mouse liver tissues using the RNeasy Mini Kit (Qiagen, Valencia, CA) and subjected to real-time PCR with the Brilliant III SYBR Green QPCR

Master Mix Gene Expression Assay Kit and the Mx3005p system (Stratagene, La Jolla, CA), normalized with glyceraldehyde 3-phosphate dehydrogenase (*Gapdh*).

Statistical analyses

Kaplan-Meier curves were generated using Graphpad Prism 6. Receiver operating characteristic (ROC) curves were generated using survivalROC package (version 1.0.3). Gene expression heatmaps were generated using GENE-E software. Boxplots and barplots were generated using ggplot2 package (version 2.2.1), while correlation plots were generated using Corrplot package (version 0.84). Cox proportional hazards regression, log-rank test, Wilcoxon rank sum test, Fisher exact test, and Harrell's C-statistic were performed in R (version 3.3.0).

Supplementary Figure Legends

Supplementary Figure 1. Kaplan-Meier survival curves for the discovery and validation cohorts. Log-rank test was used to compare the survival distributions of the two groups.

Supplementary Figure 2. Pathway enrichment analysis of upregulated genes in the low and high survival groups in the validation cohort. Dotplots depict pathways that are enriched in upregulated genes of the low (**A**) and high (**B**) survival groups in the validation cohort. Pathway terms were ranked according to their $-\log_{10}$(FDR) values and categorized into extracellular matrix (ECM), cell-cell interaction, inflammation and other for plot **A**, and oxidation, metabolism and other for plot **B**.

Supplementary Figure 3. Correlation between ssGSEA method and section staining, and levels of aHSCs, aPFs and cholangiocytes. Scatterplots represent linear correlation between quantification by ssGSEA method and quantification by section staining for activated hepatic stellate cells (aHSC; **A**), activated portal fibroblasts (aPF; **B**), and cholangiocytes (**C**). Pearson correlation coefficient *r* is shown, with the grey area representing 95% confidence limits. Representative human liver sections with anti-α-SMA (upper panel in **D**) and anti-KRT7 staining (bottom panel in **D**). Boxplots show percentage of positive area for aHSC (α-SMA positive area in lobular region, panel **E**), aPFs (α-SMA positive area in portal region, panel **F**) and cholangiocytes (KRT7 positive area, panel **G**)**.** **P* < 0.05, ***P* < 0.01, calculated by T test. Boxplots show the expression of *ACTA2* (encoding α-SMA, marker of aHSCs and aPFs, panel **H**)

and *KRT7* and *KRT19* (markers for cholangiocytes, panel **I**). ***P* < 0.01, ****P* < 0.001, calculated by Wilcoxon rank sum test.

Supplementary Figure 4. Enrichment of CD8 T and NK cells in the low survival group and their correlation with aHSCs, aPFs, and cholangiocytes.
Boxplots in **A** and **B** show enrichment of CD8 T and NK cells, respectively, using signature from Charoentong et al. Enrichments are shown as mean ± SD of ssGSEA scores, with *P* values calculated by Wilcoxon rank sum test. Panels **C-E** show scatterplots of the correlations between NK cells and aHSCs (**C**), aPFs (**D**) and cholangiocytes (**E**) also using the method from Charoentong et al. The remaining Scatterplots **F-K** show the correlation between CD8 T cells and aHSCs, aPFs, and cholangiocytes using signatures from Bindea et al as well as Charoentong et al. Enrichments are represented as ssGSEA scores. Pearson correlation coefficient *r* is shown, with the grey area representing 95% confidence limits. **P* < 0.05, ***P* < 0.01, ****P* < 0.001.

Supplementary Figure 5. Immune cells differences between the low and the high survival group. Boxplots in panel **A** show enrichments of immune cell types using the method published in Bindea et al. Enrichments are represented as ssGSEA scores; **P* < 0.05, ***P* < 0.01, ****P* < 0.001. In panel **B,** boxplots show enrichments of immune cell types based on the method of Charoentong et al. Enrichments are represented as mean ± SD of ssGSEA score; **P* < 0.05, ***P* < 0.01, ****P* < 0.001; DC, dendritic cell; iDC, immature dendritic cell; aDC, activated dendritic cell; pDC, plasmacytoid dendritic cell;

Th1 cells, type 1 T helper cells; Th2 cells, type 2 T helper; Th17 cells, T helper 17 cells; Treg cells, regulatory T cells; Tcm cells, central memory T cells; Tem, effector memory T cells; Tfh cells, T follicular helper cells; Tgd cells, gamma delta T cells; MDSC, myeloid-derived suppressor cells; NKT, natural killer T cells; iB cells, immature B cells; aB cells, activated B cells; mB cells, memory B cells; and aT cells, activated T cells.

Supplementary Figure 6. Neonatal mouse model of liver fibrosis. Panel **A** shows representative sections stained with hematoxylin and eosin at different time points after administration of $1.875x10^6$ ffu of RRV in 25 µl volume at day 3 of life; an equal volume of 0.9% normal saline was injected to control mice. EHBDs and livers were harvested from both groups at days 7, 14 and 19 after injections. H&E staining of EHBDs show normal bile duct epithelium in control mice (top panel) while RRV challenge (bottom panel) induce inflammatory obstruction of the duct lumen at 7 days followed by restoration of duct epithelium and lumen at days 14 and 19. Arrows denote areas of periductal inflammation; asterisk denotes duct lumen. In panel **B**, liver sections show normal parenchyma and portal triads. Infection with RRV (bottom panel) induces significant portal inflammation, cholangitis and expansion of portal spaces. Arrows indicate inflammatory cell infiltrations. Panel **C** shows Sirius red stainings of normal livers (top panel) and after RRV infection (lower panel) resulting in progressive accumulation of the red staining localized to portal areas that is mild at day 7 and extensive by day 19 after RRV challenge. Arrrowheads denote Sirius red staining; N=8-28 mice per group/time-point from 3-6 independent experiments.

Supplementary Figure 7. N-acetyl-cysteine (NAC) suppresses expression of fibrotic genes in modified mouse model of BA.

Graphs show the expression levels of fibrosis-related genes *Acta2, Col1a1*, *Timp1,* and *Tgfb1* in Standard RRV model (**A**) and modified model (**B**). Data are shown as mean ± SD. N=6-12 mice per group from three independent experiments are shown. **P* < 0.05, ***P* < 0.01, ****P* < 0.001.

References

1. Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. Bioinformatics 2009;25:1105-11.
2. Trapnell C, Roberts A, Goff L, et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat Protoc 2012;7:562-78.
3. Trapnell C, Hendrickson DG, Sauvageau M, et al. Differential analysis of gene regulation at transcript resolution with RNA-seq. Nat Biotechnol 2013;31:46-53.
4. **Bindea G, Mlecnik B**, Tosolini M, et al. Spatiotemporal dynamics of intratumoral immune cells reveal the immune landscape in human cancer. Immunity 2013;39:782-95.
5. **Charoentong P, Finotello F, Angelova M**, et al. Pan-cancer Immunogenomic Analyses Reveal Genotype-Immunophenotype Relationships and Predictors of Response to Checkpoint Blockade. Cell Rep 2017;18:248-262.
6. Zhang DY, Goossens N, Guo J, et al. A hepatic stellate cell gene expression signature associated with outcomes in hepatitis C cirrhosis and hepatocellular carcinoma after curative resection. Gut 2016;65:1754-64.
7. **Iwaisako K, Jiang C, Zhang M**, et al. Origin of myofibroblasts in the fibrotic liver in mice. Proc Natl Acad Sci U S A 2014;111:E3297-305.
8. Dianat N, Dubois-Pot-Schneider H, Steichen C, et al. Generation of functional cholangiocyte-like cells from human pluripotent stem cells and HepaRG cells. Hepatology 2014;60:700-14.

9. **Yang L, Mizuochi T, Shivakumar P**, et al. Regulation of epithelial injury and bile duct obstruction by NLRP3, IL-1R1 in experimental biliary atresia. J Hepatol 2018.

10. Koyama Y, Wang P, Liang S, et al. Mesothelin/mucin 16 signaling in activated portal fibroblasts regulates cholestatic liver fibrosis. J Clin Invest 2017;127:1254-1270.

Author names in bold designate shared co-first authorship

Supplementary Table 1 Information of studies for liver diseases and their animal models

Name in figure 3	Disease or model	Control	Species	GEO accession	Reference PMID
Up in cirrhosis	Cirrhotic liver	Normal liver	Human	GSE6764	17393520
Down in cirrhosis	Cirrhotic liver	Normal liver	Human	GSE6764	17393520
Poor prognosis in HCC	Hepatocellular carcinoma	NA	Human	GSE10140	18923165
Good prognosis in HCC	Hepatocellular carcinoma	NA	Human	GSE10140	18923165
Up in NASH	Nonalcoholic steatohepatitis	Normal liver	Human	NA	25581263
Down in NASH	Nonalcoholic steatohepatitis	Normal liver	Human	NA	25581263
Inflammation signature in BA	Biliary atresia	NA	Human	GSE15235	20465800
Fibrosis signature in BA	Biliary atresia	NA	Human	GSE15235	20465800
Up in PBC	Primary biliary cholangitis	Normal liver	Human	NA	11559656
Down in PBC	Primary biliary cholangitis	Normal liver	Human	NA	11559656
Up in PSC	Primary sclerosing cholangitis	Normal liver	Human	NA	11559656
Down in PSC	Primary sclerosing cholangitis	Normal liver	Human	NA	11559656
Up in BDL rat	Bile duct ligation for two weeks	Normal liver	Rat	GSE13747	20077562
Down in BDL rat	Bile duct ligation for two weeks	Normal liver	Rat	GSE13747	20077562
Up in DEN rat	100 mg/kg diethylnitrosamine for 18 weeks	Normal liver	Rat	GSE27641	24677197
Down in DEN rat	100 mg/kg diethylnitrosamine for 18 weeks	Normal liver	Rat	GSE27641	24677197
Up in CCl4 rat	Carbon tetrachloride	Normal liver	Rat	GSE73499	27659347
Down in CCl4 rat	Carbon tetrachloride	Normal liver	Rat	GSE73499	27659347
Up in CCl4 mouse	0.1cc of a 40% CCL4 for 18 weeks	Normal liver	Mouse	GSE27641	24677197
Down in CCl4 mouse	0.1cc of a 40% CCL4 for 18 weeks	Normal liver	Mouse	GSE27641	24677197
Up in EHBD: BA RRV D3	EHBD and gallbladder with RRV infection at day 3	Normal EHBD	Mouse	GSE46995	24493287
Down in EHBD: BA RRV D3	EHBD and gallbladder with RRV infection at day 3	Normal EHBD	Mouse	GSE46995	24493287
Up in EHBD: BA RRV D7	EHBD and gallbladder with RRV infection at day 7	Normal EHBD	Mouse	GSE46995	24493287
Down in EHBD: BA RRV D7	EHBD and gallbladder with RRV infection at day 7	Normal EHBD	Mouse	GSE46995	24493287
Up in EHBD: BA RRV D14	EHBD and gallbladder with RRV infection at day 14	Normal EHBD	Mouse	GSE46995	24493287
Down in EHBD: BA RRV D14	EHBD and gallbladder with RRV infection at day 14	Normal EHBD	Mouse	GSE46995	24493287
Hepatic stellate cell signature	NA	NA	Mouse	NA	26045137
Portal fibroblast signature	NA	NA	Mouse	NA	25074909

Supplementary Table 2. Forward and reverse oligonucleotide primer sequences to quantify the expression of fibrotic markers.

Name	Sequence 5' to 3'
Col1a1	Forward: TGGTGCTAAGGGTGAAGCTG
	Reverse: TCCATCAGCACCAGGGTTTC
Acta2	Forward: GGCATCCACGAAACCACCTA
	Reverse: AATGCCTGGGTACATGGTGG
Timp1	Forward: TTCTTGGTTCCCTGGCGTAC
	Reverse: ACTCTCCAGTTTGCAAGGGA
Tgfb1	Forward: CCGCAACAACGCCATCTATG
	Reverse: TGCCGTACAACTCCAGTGAC
Gapdh	Forward: TGGTTTGACAATGAATACGGCTAC
	Reverse: GGTGGGTGGTCCAAGGTTTC

Supplementary Table 3 Characteristics of patients at the time of diagnosis

Patient characteristic	Discovery cohort (n=121)	Validation cohort (n=50)	P value*
Sex, N (%)			
Female	65 (54)	27 (54)	1
Male	56 (46)	23 (46)	
Age (days), median (25%-75%)	63 (46-76)	63 (46-77)	0.7859
TB (Mean±SD, mg/dL)	7.4±2.8	7.9±3.2	0.8894
AST (Mean±SD, IU)	172±137.1	206.8±160.3	0.4621
ALT (Mean±SD, IU)	128±124.4	133.0±103.3	0.3169
GGT (Mean±SD, IU)	605±530.6	718.6±514.8	0.8264
Platelets	490.0±215.4	481.8±175.3	0.4275
BASM, N (%)	12 (10)	7 (14)	0.4336
Histological inflammation, N (%)			
0	18 (15)	2 (4)	0.0544
1	31 (26)	18 (36)	
2	64 (53)	23 (46)	
3	8 (7)	7 (14)	
Histological fibrosis, N (%)			
0	3 (2)	0 (0)	0.1471
1	33 (27)	6 (12)	
2	62 (51)	31 (62)	
3	18 (14)	11 (22)	
4	5 (4)	2 (4)	

* P value for Sex, BASM (biliary atresia splenic malformation syndrome), histological inflammation and fibrosis were calculated by Fisher exact test, P values for others were calculated by Wilcoxon rank sum test.

Supplementary Table 4 Information of 14 prognostic genes.

Gene Symbol	Gene Name	Entrez ID	Importance score
LOXL1	Lysyl oxidase like 1	4016	68.95
C1QTNF5	C1q and TNF related 5	114902	67.246
3H4P	Prolyl 3-hydroxylase family member 4 (non-enzymatic)	10609	66.785
HSDL1	Hydroxysteroid dehydrogenase like 1	83693	58.281
SFRP4	Secreted frizzled related protein 4	6424	57.239
SLC44A3	Solute carrier family 44 member 3	126969	55.351
GLMN	Glomulin, FKBP associated protein	11146	53.919
DPY30	Dpy-30, histone methyltransferase complex regulatory subunit	84661	47.122
ABHD4	Abhydrolase domain containing 4	63874	46.62
KIF7	Kinesin family member 7	374654	45.807
EDIL3	EGF like repeats and discoidin domains 3	10085	44.443
PLOD1	Procollagen-lysine,2-oxoglutarate 5-dioxygenase 1	5351	41.664
LGALS4	Galectin 4	3960	35.479
HSF2	Heat shock transcription factor 2	3298	30.262

Supplementary Table 5 Characteristics of patients at the time of diagnosis

Patient characteristic	Discovery cohort (n=121)			Validation cohort (n=50)		
	Low survival (n=60)	High survival (n=61)	P value*	Low survival (n=25)	High survival (n=25)	P value*
Sex, N (%)						
Female	34 (57)	31 (51)	0.5859	15 (60)	12 (48)	0.571
Male	26 (43)	30 (49)		10 (40)	13 (52)	
Age (days), median (25%-75%)	68 (55-78)	52 (40-73)	0.0001	59 (55-82)	50 (37-71)	0.034
TB (Mean±SD, mg/dL)	8.2±2.9	7.2±2.7	0.0849	8.1±2.7	7.7±3.6	0.187
AST (Mean±SD, IU)	210.7±123.8	208.8±149.9	0.5053	252.9±201.9	163.0±90.8	0.274
ALT (Mean±SD, IU)	156.3±124.3	150.6±125.4	0.6822	167.4±123.6	101.4±68.8	0.068
GGT (Mean±SD,IU)	825.3±593.1	629.4±448.3	0.0535	881.2±592.7	548.9±357.7	0.077
Platelets	477.3±203.0	467.5±198.6	0.6793	445.8±127.8	515.0±206.9	0.187
BASM, N (%)	4 (7)	8 (13)	0.3626	3 (25)	4 (25)	1.000
Histological inflammation, N (%)						
0	10 (17)	8 (13)	0.4857	0 (0)	2 (8)	0.237
1	12 (20)	19 (31)		11 (44)	7 (28)	
2	34 (56)	30 (49)		10 (40)	13 (52)	
3	3 (5)	5 (8)		2 (8)	5 (20)	
Histological fibrosis, N (%)						
0	1 (2)	2 (3)	0.0039	0 (0)	0 (0)	0.016
1	13 (22)	20 (33)		1 (4)	5 (20)	
2	26 (43)	36 (59)		13 (52)	18 (72)	
3	14 (23)	4 (7)		9 (36)	2 (8)	
4	5 (8)	0 (0)		2 (8)	0 (0)	

* P values for Sex, BASM (biliary atresia splenic malformation syndrome), histological inflammation and fibrosis were calculated by Fisher exact test, P values for others were calculated by Wilcoxon rank sum test.

Supplementary Table 6. Association of 14-gene signature and clinical variables at diagnosis and at 1 and 3 months after HPE with survival in the discovery cohort

Variable	Univariable analysis		Multivariable analysis	
	Hazard Ratio (95% CI)	P value	Hazard Ratio (95% CI)	P value
14-gene signature	2.463 (1.754-3.459)	<0.0001	2.200 (1.3500-3.587)	0.002
Age at diagnosis (days)	1.009 (0.998-1.019)	0.109		
Sex	1.608 (0.922-2.804)	0.094		
Platelets at diagnosis	1.000 (0.999-1.002)	0.602		
APRI at diagnosis	1.272 (0.8432-1.92)	0.251		
Histological inflammation	0.959 (0.699-1.316)	0.797		
Histological fibrosis stage	0.909 (0.5943-1.391)	0.660		
TB at diagnosis	1.000 (0.909-1.100)	0.999		
AST at diagnosis	1.000 (0.998-1.002)	0.930		
ALT at diagnosis	1.001 (0.999-1.003)	0.522		
GGT at diagnosis	1.000 (0.999-1.001)	0.435		
TB at 1 month	1.271 (1.175-1.376)	<0.0001	0.953 (0.8143-1.114)	0.543
AST at 1 month	1.007 (1.003-1.010)	0.0002	1.000 (0.993-1.008)	0.940
ALT at 1 month	1.001 (0.998-1.003)	0.717		
GGT at 1 month	1.000 (0.999-1.000)	0.473		
TB at 3 months	1.229 (1.173-1.289)	<0.0001	1.205 (1.130-1.284)	<0.0001
AST at 3 months	1.004 (1.003-1.005)	<0.0001	1.003 (0.997-1.009)	0.309
ALT at 3 months	1.003 (1.001-1.005)	0.0009	0.999 (0.993-1.005)	0.633
GGT at 3 months	1.000 (0.999-1.000)	0.717		

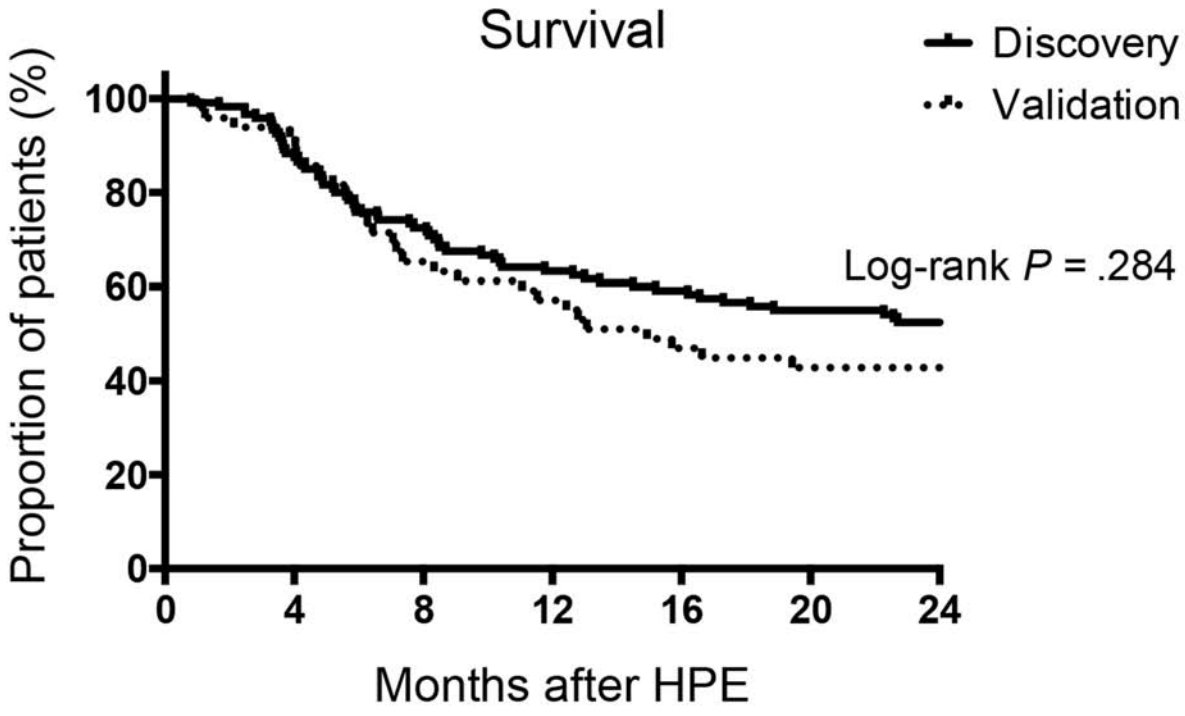
Supplementary Table 7 Association of 14-gene signature and clinical variables at diagnosis and at 1 and 3 months after HPE with survival in the validation cohort

Variable	Univariable analysis		Multivariable analysis	
	Hazard Ratio (95% CI)	P value	Hazard Ratio (95% CI)	P value
14-gene signature	1.873 (1.079- 3.25)	0.003	1.795 (1.329-1.921)	0.006
Age at diagnosis (days)	1.006 (0.992-1.020)	0.431		
Sex	1.397 (0.617-3.165)	0.423		
Platelets at diagnosis	1.000 (0.998-1.003)	0.798		
APRI at diagnosis	0.912 (0.573-1.45)	0.697		
Histological inflammation	1.153 (0.688-1.934)	0.589		
Histological fibrosis	1.078 (1.053-1.781)	0.390		
TB at diagnosis	1.068 (0.954-1.195)	0.253		
AST at diagnosis	1.002 (0.999-1.005)	0.205		
ALT at diagnosis	1.002 (0.998-1.006)	0.407		
GGT at diagnosis	1.000 (0.999-1.001)	0.582		
TB at 1 month	1.255 (1.093-1.441)	0.001	0.953 (0.801- 1.375)	0.727
AST at 1 month	1.005 (1.002-1.009)	0.002	1.005 (0.999-1.011)	0.135
ALT at 1 month	1.005 (0.998-1.011)	0.162		
GGT at 1 month	1.000 (0.999-1.000)	0.883		
TB at 3 months	1.148 (1.067-1.234)	0.0002	1.301 (1.039-1.631)	0.002
AST at 3 months	1.002 (1.000-1.004)	0.010	0.993 (0.984-1.003)	0.164
ALT at 3 months	1.002 (0.999-1.005)	0.033	1.006 (0.995-1.018)	0.292
GGT at 3 months	1.000 (0.999-1.000)	0.896		

Supplementary Table 8 Predictive accuracy of prognostic models

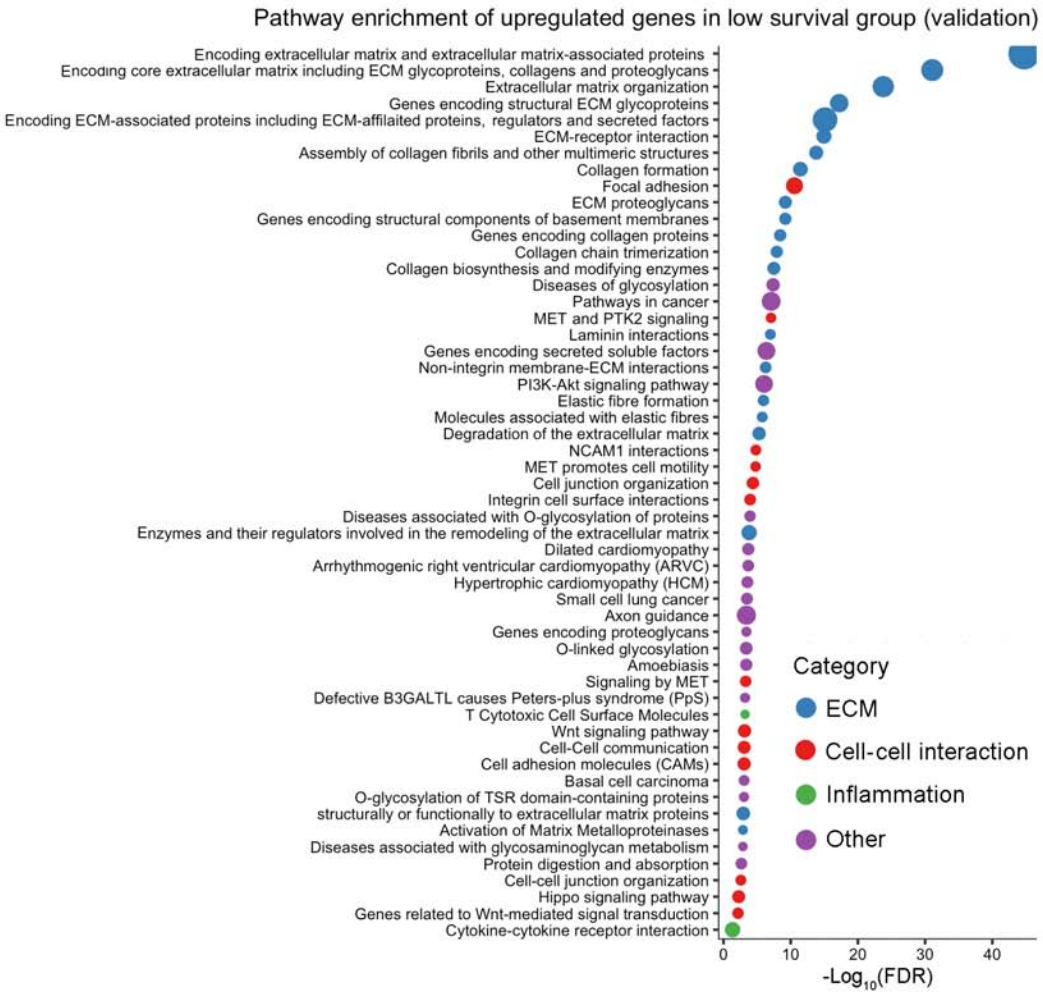
Parameter	Discovery						Validation					
	AUC	C-statistic	Sensi-tivity	Speci-ficity	PPV	NPV	AUC	C-statistic	Sensi-tivity	Speci-ficity	PPV	NPV
14-gene signature	0.734	0.711	0.748	0.792	0.791	0.758	0.732	0.704	0.733	0.755	0.741	0.821
Total bilirubin at 3 months (3M TB)	0.920	0.844	0.771	0.844	0.771	0.814	0.780	0.746	0.801	0.762	0.746	0.813
Index (14-gene signature + 3M TB)	0.948	0.862	0.886	0.867	0.822	0.907	0.813	0.763	0.825	0.821	0.801	0.811

Supplementary Figure 1

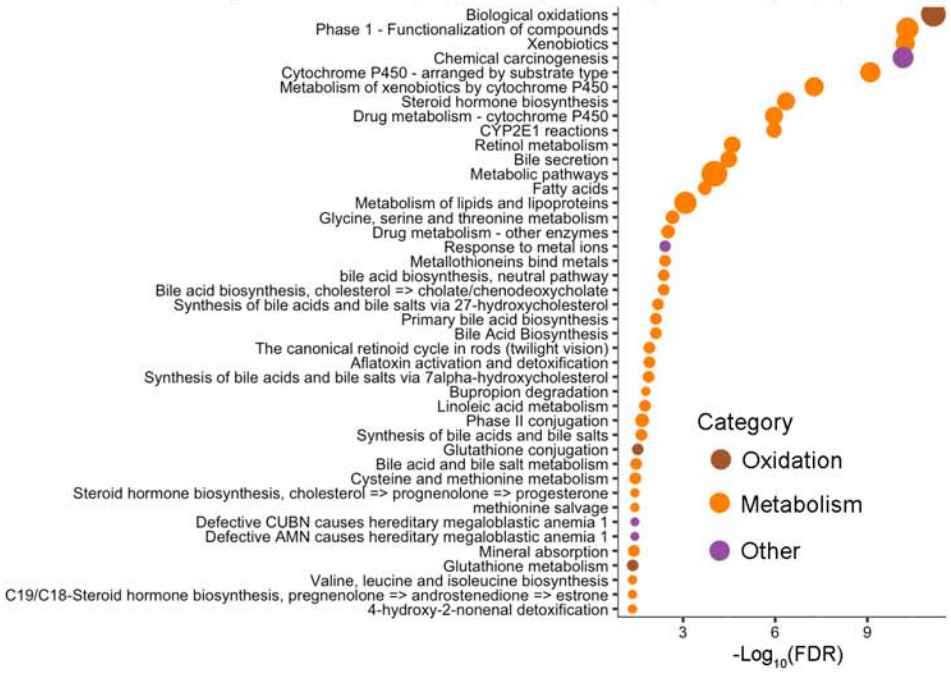


Supplementary Figure 2

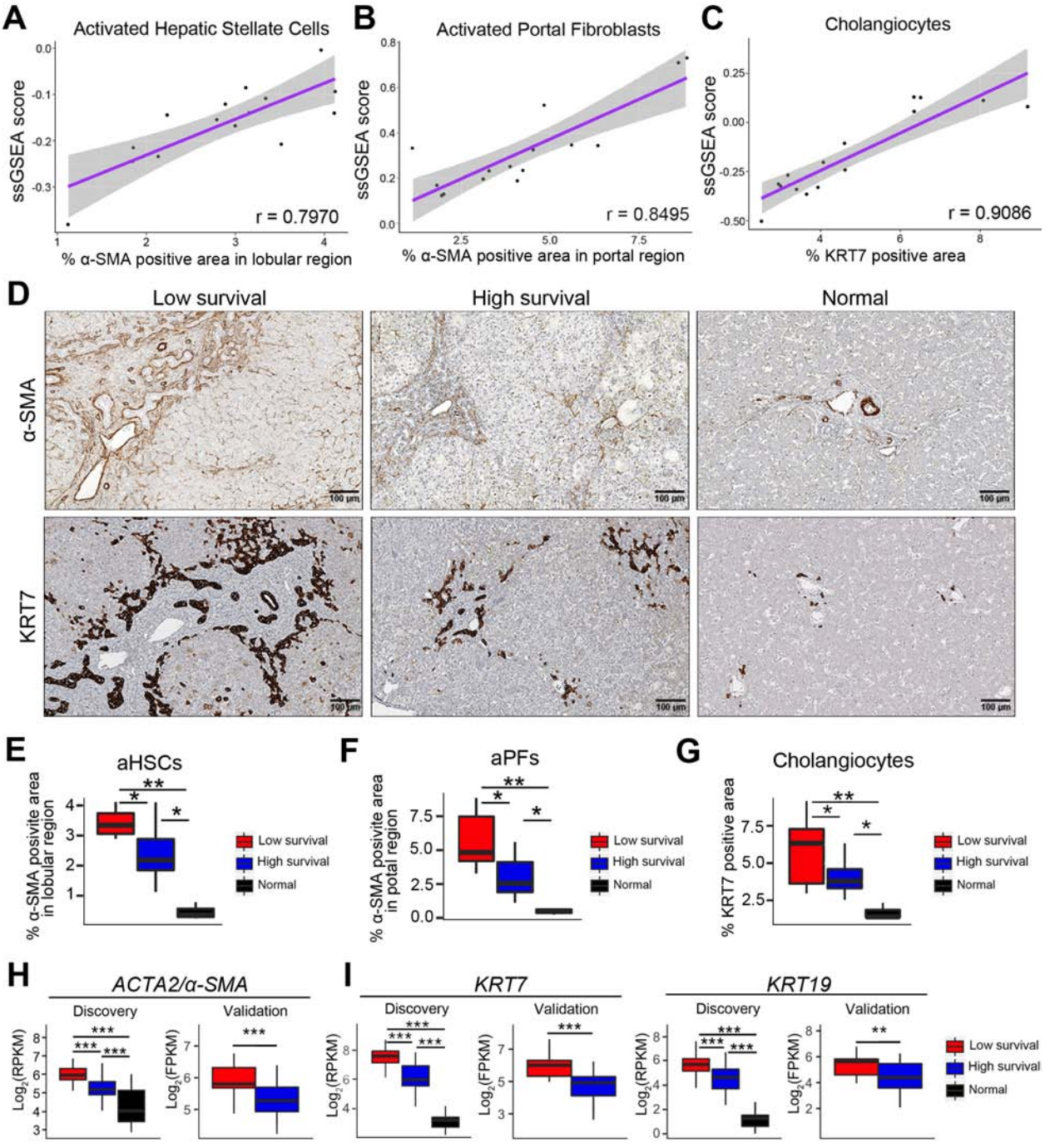
A



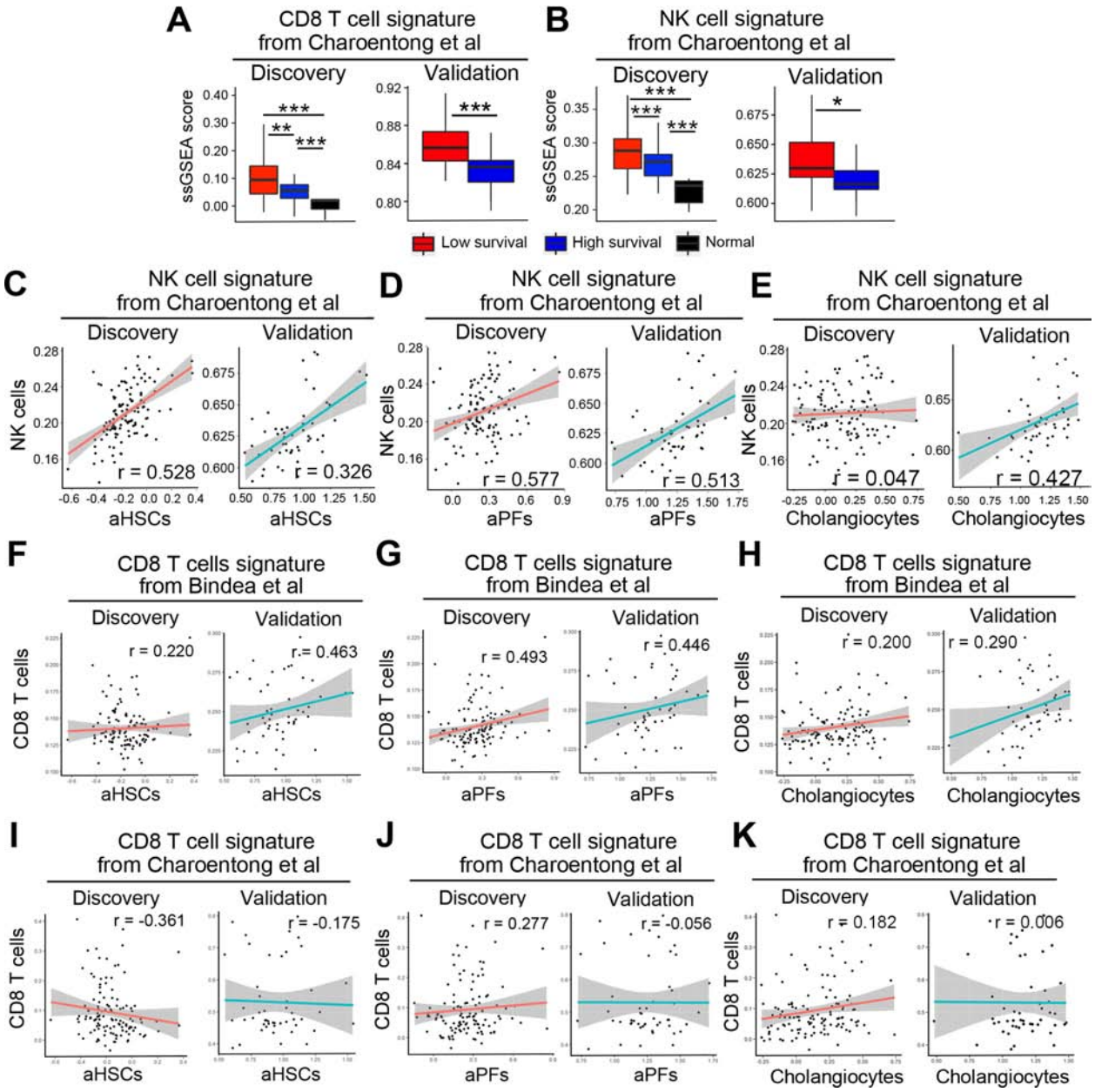
B



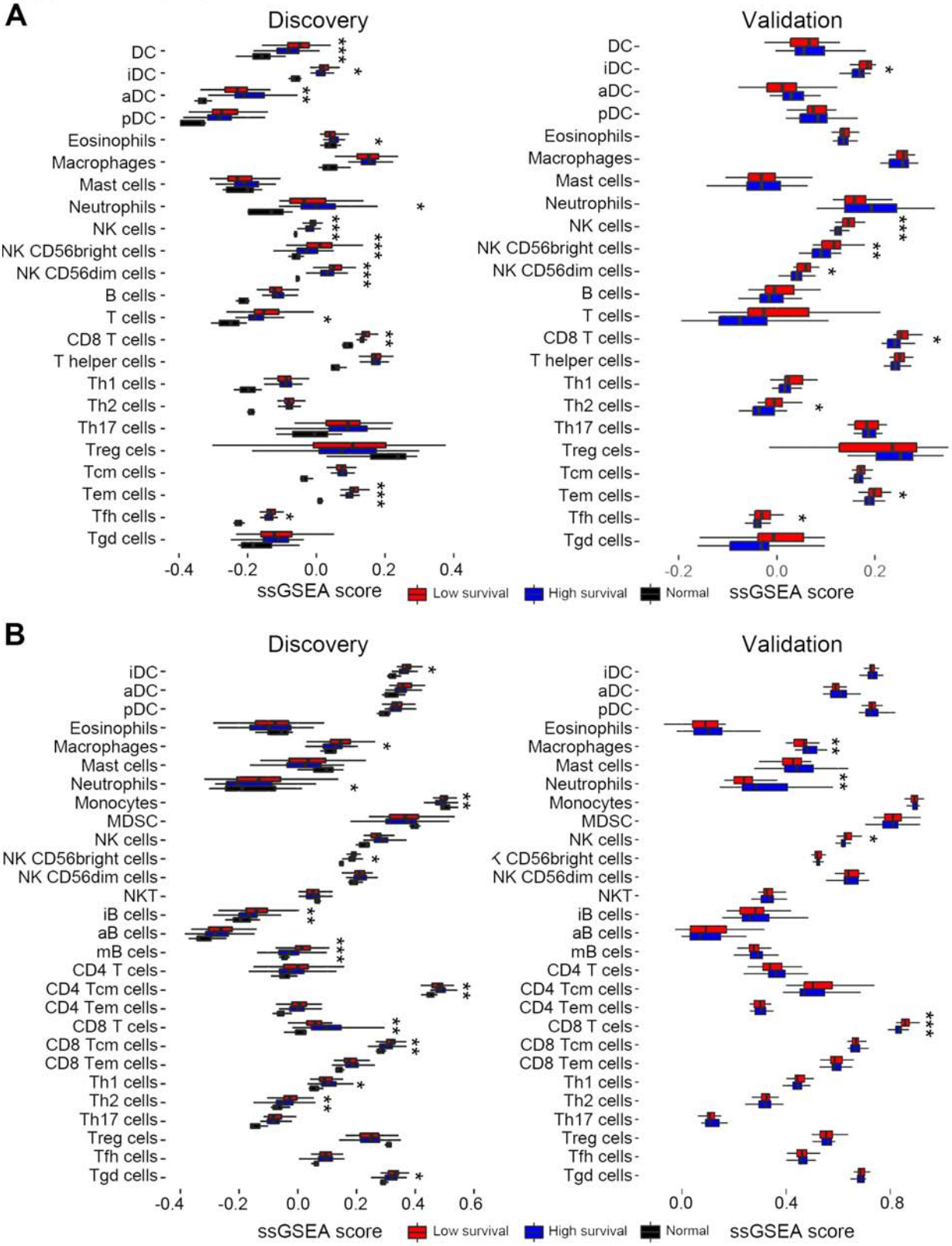
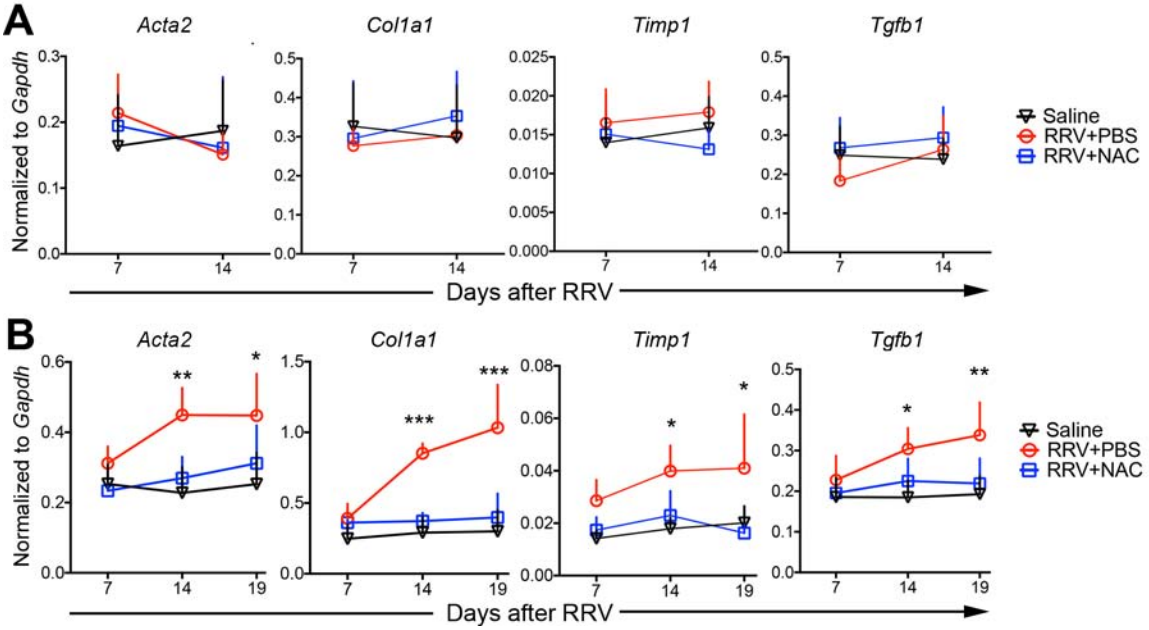
Supplementary Figure 3



Supplementary Figure 4



Supplementary Figure 5



Supplementary Figure 7