# ‘Cytokines’: A Promising Approach for Severe Alcoholic Hepatitis

**Jesus Mario Rangel Valenzuela[1,2]; Tanubrata Dey[1,2]**

[1]Graduate School of Arts and Sciences, New York University, New York, New York, USA. [2]Department of Biology, New York University, New York, New York, United States.

## Abstract

Severe Alcoholic Hepatitis (SAH) is the final stage of Alcoholic Liver Disease (ALD). It holds a high mortality rate with a small time frame for treatment. Survivors of this stage have a high risk of developing scarring of liver tissue leading to cirrhosis. SAH-researchers are presented with an urge of developing novel treatments since current therapy approaches are ineffective. Several approaches have been proposed, including targeting of the Hippo Signaling Pathway (HSP) mechanism; as well as, factors that impede its normal functioning. This article uses raw data originating from a previously published research article. It aimed to reproduce and complement the original article with additional approaches not previously considered for treatment development. Using various bioinformatic tools, not only was this study able to replicate the findings of the original study; but also able to identify new cytokines not previously mentioned in the article. These findings place the name of those cytokines in a loop for use in future studies. All of this in hopes of developing novel effective therapies for SAH.

## Introduction

*Severe alcoholic hepatitis (SAH)* is known to be the highest manifestation of *Alcoholic Liver Disease (ALD)*. This severe stage of tissue damage, expressed through an acute symptomatology, makes SAH a life threatening disease that results in poor short-term prognosis. Serious cases of this disease hold a high mortality rate of up to 40-50% and survivors usually require monitoring because of the rapid progression of liver fibrosis leading to hepatic cirrhosis (1-3).

The results of current therapeutic methods are quite ineffective and place SAH researchers in an urge to develop novel therapies for this disease (4). Studies have demonstrated that SAH is linked to the accumulation of dedifferentiated liver cells that express, through ductular reaction, cell markers that pertain to their stem or progenitor cell lines. This process causes these dedifferentiated liver cells to lose their mature hepatocyte functions (1).

While the process of their dedifferentiation into fetal-like proliferative liver cells is well known; their excessive repopulation, which causes the liver to fail, remains unclear. For this reason, targeting the inhibition of their accumulation is not feasible at this time (1).

A proposed approach towards treatment development is through targeting of the *Hippo Signaling Pathway (HSP)*. This pathway is essential for mature hepatocytes as it maintains their differentiation and regulates their cellular proliferation (6). A disruption in this pathway’s normal activity through inactivation of its signaling, has been shown to alter the inhibition of both *Yes-associated protein (YAP)* and *Transcriptional co-activator with PDZ binding motif (TAZ)* (8).

These transcriptional regulators play an important role in organ growth and increased levels of cell proliferation during the cell’s fetal stage (1, 6). It has been found in recent studies that

*hippo-signaling* expresses low activity when the hepatocyte is in its fetal development. This enables *YAP/TAZ* to remain dephosphorylated and therefore active. In contrast, when liver cells reach maturity, *hippo-signaling* activity levels rise and remain high which renders YAP/TAZ to stay inactive and allow for healthy functioning of mature liver cells (1).

As studies relating HSP and alcohol started emerging, it became clear that excessive alcohol consumption exerted a significant impact in dysregulation of HSP (1). This started elucidating SAH researchers into investigating the root cause.

Hyun, Jeongeun et al. determined that dysregulation of HSP was a result of low expression of *epithelial splicing regulatory protein 2 (ESRP2)*. Suppression of *ESRP2* caused a reprogramming of mature hepatocytes back into fetal-like cells with replicative phenotype (1).

Furthermore, the article also concluded that ESRP2's suppression mechanism originates from the presence of two proinflammatory cytokines: *TNF-α and IL-1β*. Both of these cytokines have been found to be present in livers with an alcohol-related deterioration or damage (1).

This whole process is able to be reversed through restoring normal expression of *ESRP2*, which will then lead to reactivation of HSP, and finally to differentiate those reprogrammed hepatocytes back into their mature state for normal function (1).

It is important to assert that this article makes use of Hyun, Jeongeun's et al. raw data and aims to, not only attempt to replicate the original research findings, but to complement their analysis by presenting possible alternative cytokines, distinct differential expressed genes, or different pathways than those mentioned in the original article. All of this in hopes of providing an alternative approach for SAH future studies and treatment development.

## Results

### *Differentially Expressed Genes in Severe Alcoholic Patients (SAH) compared to healthy individuals:*

There were a total of 17676 differentially expressed genes detected in SAH patients compared to healthy individuals of which 4827 genes were found to be upregulated and 4509 genes to be downregulated (*Figure 1*).

| | SAHvsWT |
|---|---|
| Down | 4509 |
| NotSig | 8340 |
| Up | 4827 |

***Figure 1:*** *DEGs upregulated and downregulated.*

Certain genes that code for enzymes known for alcohol metabolism including *ADH4, ADH1A, ADH1B, ADH1C* were found to be highly downregulated. Many proinflammatory cytokines genes including *TNF-α, IL-1β, IL-6* and many CXC chemokine ligand genes including *CXCL1, CXCL6, CXCL8, CXCL10* were highly upregulated. Esrp2 gene was found to be highly downregulated and Esrp2 interacting genes including *NF2* and *CSNK1D* were highly dysregulated. *PTGS2, CYR61, CTGF* known for target genes by *YAP/TAZ* were also found upregulated. An MA plot was done to show how differentially expressed genes were expressed in *Figure 2*. The more plots that fall on higher than threshold in y-axis means more significant genes are upregulated whereas those fall below threshold show higher downregulation in genes.

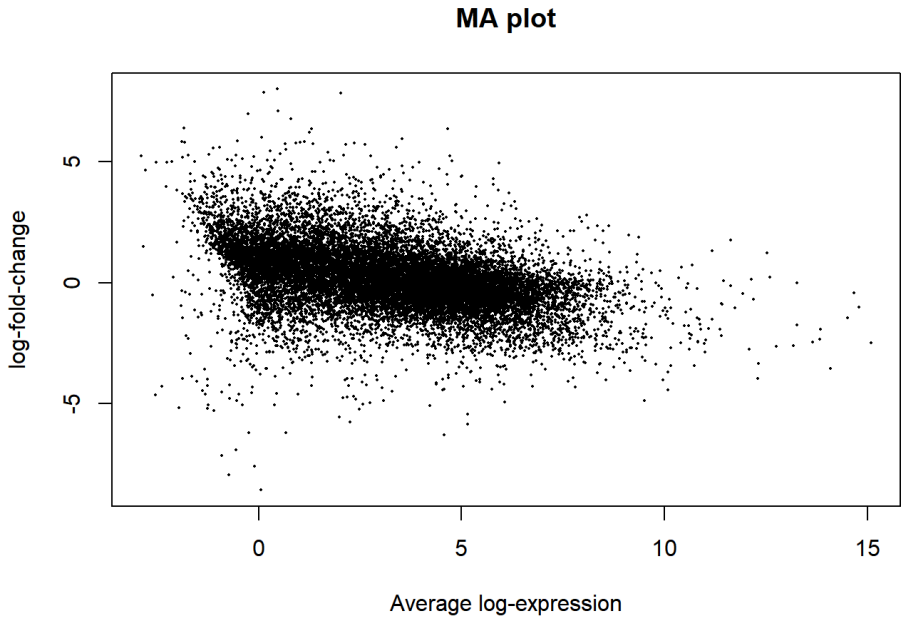


***Figure 2:*** *MA plot of DEGs in samples*

## *Gene Set Enrichment Analysis:*

Gene Set Enrichment Analysis showed gene clusters where pathways were enriched. Top 10 upregulated DEGs enriched in pathways include Hedgehog Signalling, Epithelial to Mesenchymal Transition, Inflammatory response, and Wnt-Beta Catenin Signaling pathway (*Figure 3*).

| | NGenes | Up | Down | Mixed |
|---|---|---|---|---|
| HALLMARK_HEDGEHOG_SIGNALING | 32 | 0.000999001 | 1.0000000 | 0.017982018 |
| HALLMARK_ANGIOGENESIS | 33 | 0.001998002 | 0.9990010 | 0.000999001 |
| HALLMARK_EPITHELIAL_MESENCHYMAL_TRANSITION | 187 | 0.003996004 | 0.9970030 | 0.003996004 |
| HALLMARK_MYOGENESIS | 151 | 0.027972028 | 0.9730270 | 0.001998002 |
| HALLMARK_KRAS_SIGNALING_UP | 181 | 0.040959041 | 0.9600400 | 0.002997003 |
| HALLMARK_APICAL_JUNCTION | 168 | 0.040959041 | 0.9600400 | 0.030969031 |
| HALLMARK_UV_RESPONSE_DN | 142 | 0.065934066 | 0.9350649 | 0.568431568 |
| HALLMARK_INFLAMMATORY_RESPONSE | 176 | 0.075924076 | 0.9250749 | 0.051948052 |
| HALLMARK_WNT_BETA_CATENIN_SIGNALING | 37 | 0.083916084 | 0.9170829 | 0.994005994 |
| HALLMARK_ALLOGRAFT_REJECTION | 170 | 0.097902098 | 0.9030969 | 0.070929071 |

***Figure 3:*** *Top 10 Upregulated Gene Set Enrichment (GSEA)*

Notable pathways that are enriched with downregulated DEGs include Oxidative Phosphorylation, Glycolysis, Fatty acid and Bile Acid Metabolism. Top 10 downregulated Gene Set Enrichment Pathways are shown below (*Figure 4*).

| | NGenes | Up | Down | Mixed |
|---|---|---|---|---|
| HALLMARK_OXIDATIVE_PHOSPHORYLATION | 200 | 1.000000 | 0.000999001 | 0.000999001 |
| HALLMARK_ADIPOGENESIS | 192 | 1.000000 | 0.000999001 | 0.000999001 |
| HALLMARK_XENOBIOTIC_METABOLISM | 192 | 1.000000 | 0.000999001 | 0.000999001 |
| HALLMARK_FATTY_ACID_METABOLISM | 146 | 1.000000 | 0.000999001 | 0.000999001 |
| HALLMARK_BILE_ACID_METABOLISM | 104 | 1.000000 | 0.000999001 | 0.000999001 |
| HALLMARK_PEROXISOME | 95 | 1.000000 | 0.000999001 | 0.000999001 |
| HALLMARK_MTORC1_SIGNALING | 200 | 0.980020 | 0.020979021 | 0.000999001 |
| HALLMARK_CHOLESTEROL_HOMEOSTASIS | 73 | 0.979021 | 0.021978022 | 0.000999001 |
| HALLMARK_GLYCOLYSIS | 183 | 0.966034 | 0.034965035 | 0.000999001 |
| HALLMARK_DNA_REPAIR | 145 | 0.965035 | 0.035964036 | 0.975024975 |

***Figure 4:*** *Top 10 Downregulated Gene Set Enrichment (GSEA)*

### *GO Term Enrichment Analysis:*

GO term enrichment analysis was done using the differentially expressed genes to see which biological processes were enriched with highly upregulated or downregulated genes. Biological processes that have the upregulated genes include cellular adhesion, cell motility, movement and pathways related to Extracellular Matrix. Top 10 pathways seen in GO analysis are shown below (*Figure 5*).

| | Term <chr> |
|---|---|
| GO:0040011 | locomotion |
| GO:0006928 | movement of cell or subcellular component |
| GO:0007155 | cell adhesion |
| GO:0022610 | biological adhesion |
| GO:0048870 | cell motility |
| GO:0051674 | localization of cell |
| GO:0016477 | cell migration |
| GO:0006935 | chemotaxis |
| GO:0042330 | taxis |
| GO:0040012 | regulation of locomotion |

***Figure 5:*** *Top 10 Biological pathways with upregulated DEGs*

Downregulated differentially expressed genes that are enriched in biological processes include various metabolic processes specific to materials (*Figure 6*).

| | Term <chr> | Ont <chr> | N <dbl> | Up <dbl> | Down <dbl> |
|---|---|---|---|---|---|
| GO:0044281 | small molecule metabolic process | BP | 1743 | 283 | 935 |
| GO:0019752 | carboxylic acid metabolic process | BP | 951 | 129 | 589 |
| GO:0043436 | oxoacid metabolic process | BP | 1028 | 152 | 617 |
| GO:0006082 | organic acid metabolic process | BP | 1044 | 156 | 622 |
| GO:0055114 | oxidation-reduction process | BP | 899 | 123 | 536 |
| GO:0044282 | small molecule catabolic process | BP | 405 | 46 | 273 |
| GO:0008152 | metabolic process | BP | 9406 | 2086 | 3038 |
| GO:0006520 | cellular amino acid metabolic process | BP | 319 | 22 | 230 |
| GO:0046395 | carboxylic acid catabolic process | BP | 266 | 20 | 203 |
| GO:0016054 | organic acid catabolic process | BP | 266 | 20 | 203 |

***Figure 6:*** *Top10 Biological pathways with downregulated DEGs*

### *KEGG Pathway Analysis:*

Using the differentially expressed genes as an input and filtering for *Homo sapiens*, it was possible to sort the top ten upregulated and downregulated pathways. These pathways help us understand the molecular interactions in between all our differentially expressed genes and serve as a starting point in determining the best approach for a potential development of a treatment for SAH.

| | Pathway <chr> | N <dbl> | Up <dbl> | Down <dbl> |
|---|---|---|---|---|
| path:hsa04060 | Cytokine-cytokine receptor interaction | 171 | 84 | 30 |
| path:hsa04512 | ECM-receptor interaction | 71 | 44 | 8 |
| path:hsa05168 | Herpes simplex virus 1 infection | 435 | 162 | 46 |
| path:hsa05414 | Dilated cardiomyopathy | 69 | 40 | 8 |
| path:hsa04080 | Neuroactive ligand-receptor interaction | 130 | 63 | 25 |
| path:hsa05200 | Pathways in cancer | 437 | 165 | 92 |
| path:hsa05410 | Hypertrophic cardiomyopathy | 64 | 37 | 9 |
| path:hsa04514 | Cell adhesion molecules | 114 | 56 | 10 |
| path:hsa04061 | Viral protein interaction with cytokine and cytokine receptor | 65 | 36 | 8 |
| path:hsa04510 | Focal adhesion | 175 | 76 | 29 |

***Figure 7:*** *Top 10 Upregulated KEGG pathways*

Notable pathways that were seen upregulated include Cytokine-Cytokine Receptor Interaction, ECM-receptor interaction, adhesions and Pathways in Cancer (*Figure 7*).

| | Pathway <chr> | N <dbl> | Up <dbl> | Down <dbl> |
|---|---|---|---|---|
| path:hsa01100 | Metabolic pathways | 1309 | 206 | 712 |
| path:hsa05012 | Parkinson disease | 223 | 21 | 143 |
| path:hsa04146 | Peroxisome | 80 | 4 | 68 |
| path:hsa00190 | Oxidative phosphorylation | 120 | 3 | 89 |
| path:hsa01200 | Carbon metabolism | 105 | 10 | 81 |
| path:hsa01240 | Biosynthesis of cofactors | 141 | 7 | 98 |
| path:hsa05016 | Huntington disease | 269 | 38 | 149 |
| path:hsa04714 | Thermogenesis | 215 | 32 | 126 |
| path:hsa05010 | Alzheimer disease | 312 | 54 | 165 |
| path:hsa05020 | Prion disease | 240 | 44 | 134 |

***Figure 8:*** *Top 10 Downregulated KEGG pathways*

Important pathways that were seen downregulated include certain metabolic pathways (*Figure 8*).

This analysis also focussed on Hippo Pathway which was severely downregulated. *CSNK1D* and *NF12* were found to be highly dysregulated in SAH and *YAP/TAZ* was highly upregulated leading to upregulation of many pro-proliferative and anti-apoptotic genes.

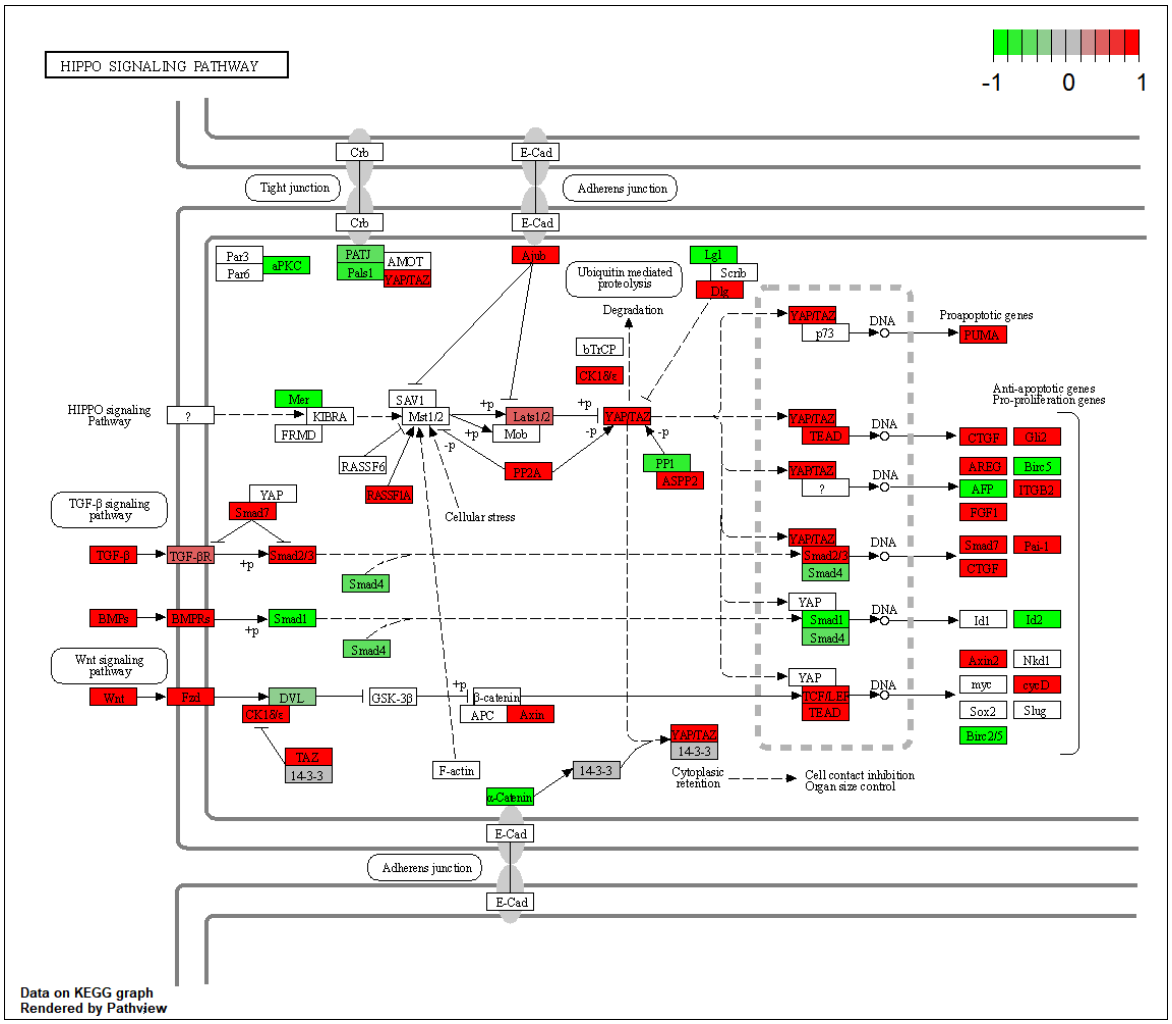


***Figure 9:*** *Hippo Signaling Pathway in KEGG*

### *Alternative Splicing Analysis:*

Alternative splicing analysis was done on specific genes including *ESRP2, YAP1* and *TAZ* to see how exons are expressed in these genes and to complement with the author's study.

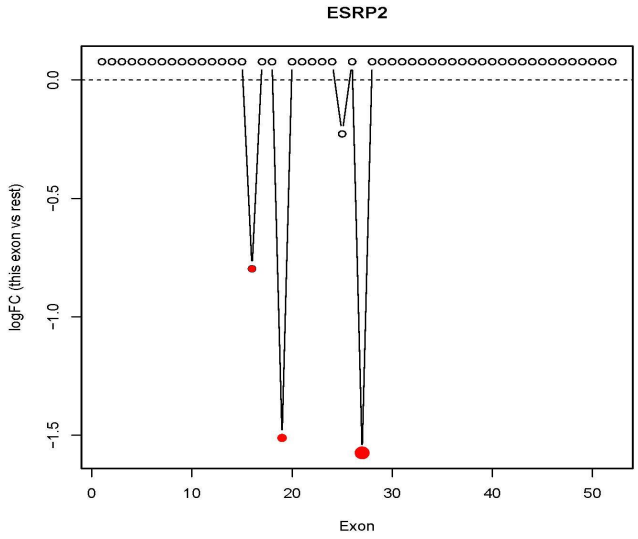


***Figure 10:*** *ESRP2 Alternative Splicing Plot*

Since *ESRP2* is a negative stranded gene, transcription should be from right to left. 3-4 exons are seen to be downregulated in severe alcoholic hepatitis patients.

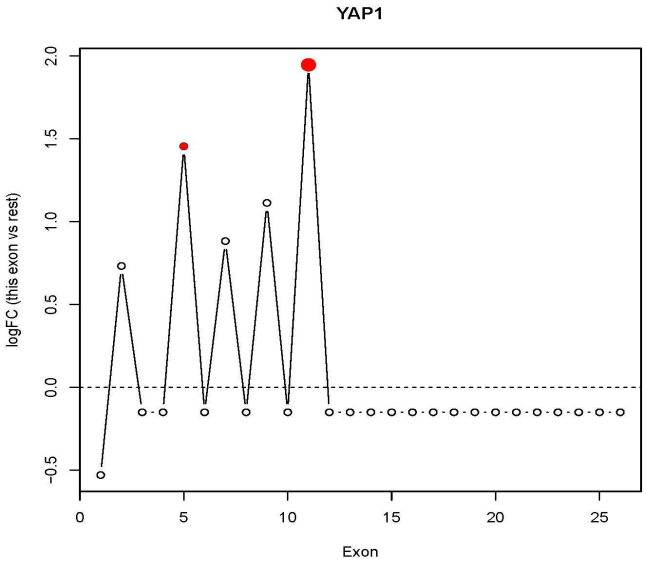


***Figure 11:*** *YAP1 Alternative Splicing Plot*

Since *YAP1* is a positive stranded gene, transcription should be from left to right. The very first exon is seen down regulated but certain earlier 5 exons are seen upregulated (*Figure 10*).

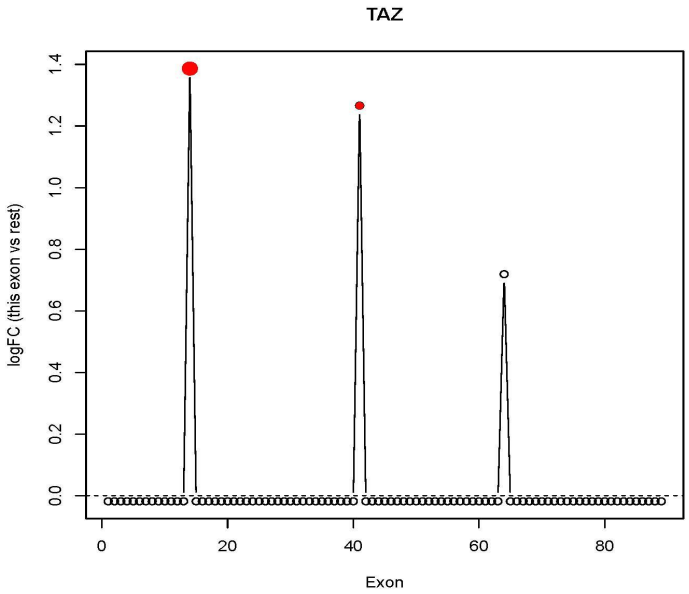


***Figure 12:*** *TAZ Alternative Splicing Plot*

Since *TAZ* is a positive stranded gene, transcription starts from left to right. 3 of the exons were seen upregulated as shown in Figure 11 in SAH patients compared to healthy ones.

### *Trimmomatic in RNA Seq Data:*

Trimmomatic was done using two different parameters in the reads twice and analysis was done with each of the sets individually and compared to the untrimmed set which was also analyzed the same way. On the first approach around 2.51% on average of bases were dropped and on the second approach around 4.3% on average of bases were dropped. Alignment percentage for both pseudo alignment using Kallisto and regular alignment with STAR had the same percentage of aligned reads on all the 3 processes for each dataset. At the gene-level counts for DEGs, there were only 2-3 less genes recovered using both trimmomatic datasets compared to the untrimmed. No significant difference in exon counts was also observed in alternative splicing analysis.

## Discussion

Testing for trimmomatic using various parameters at the gene datasets there was significantly no difference in output observed. Though one study found that doing trimmomatic actually helps in removing PCR duplicates and adapter contents as well as remove low quality reads, but it was shown that their downstream analysis took higher time compared to untrimmed making the data and many genes were not counted thus making the dataset redundant (18). Our downstream analysis did not become slow but we were missing 2-3 genes. Based on the results it is evident that RNA seq data trimming is not necessary or highly recommended for RNA sequencing.

High alcohol consumption is known to be the driving cause for most of the liver related diseases in the world (5). Since metabolism of alcohol is mainly done in the liver, excessive amounts of ethanol leads to stress in the liver. Ethanol is mainly broken down by alcohol dehydrogenase (*ADH*) and aldehyde dehydrogenase (*ALDH*) into acetaldehyde (7). Differentially expressed genes showed *ADH4*, *ADH1A, ADH1B, ADH1C* genes which code for these metabolic enzymes are highly down regulated meaning alcohol metabolism is totally diminished in SAH livers. GO term enrichment biological processes also shows many metabolic pathways are

downregulated which provides support that alcohol metabolism is highly dysregulated in SAH.

Bangru et al. showed that the main cause of liver failure in SAH is the excessive proliferation of hepatocytes as fetal-like cells that accumulates the liver (1). Their outcome proposed that Esrp2 is a splicing factor that regulates the Hippo Signaling Pathway, which is known earlier to maintain homeostasis of hepatocytes in the liver. Esrp2 was seen to interact with *NF2* and *CSNK1D* of Hippo Pathway which downregulated the activity of *YAP/TAZ* complex by phosphorylating it which is responsible for proliferation of fetal-like hepatocytes mainly seen during liver development or regeneration (1).

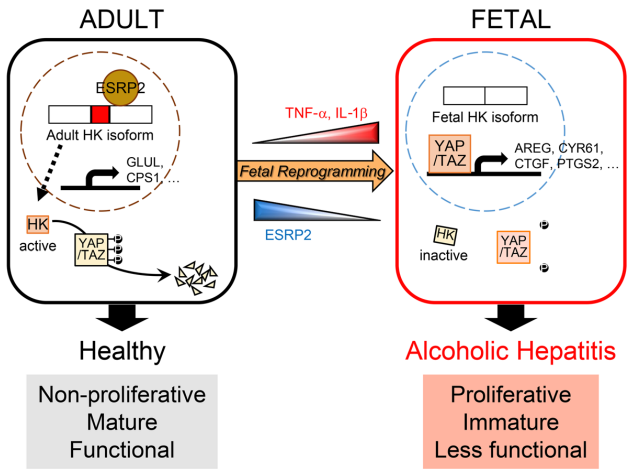


***Figure 13:*** *Esrp2 & Yap/Taz role in liver failure for SAH (Picture courtesy: Bangru et al.)*

They proposed that toxicity by ethanol in liver was leading to an inflammatory response leading to production of cytokines *TNF-A* and *IL-1B* which are damaging *ESRP2* leading to downregulation of Hippo pathway which in turn failing to phosphorylate *YAP/TAZ* leading to activation of *YAP/TAZ* which is activating the fetal-like immature hepatocyte proliferation leading to liver damage (Figure 12).

DEGs analysis showed downregulation of *ESRP2* and upregulation of genes like *CTGF, CYR61, PTGS2* which are known to be activated by *YAP/TAZ* were upregulated meaning this analysis output complements with that of the authors Esrp2 was downregulated whereas *YAP/TAZ* was upregulated. GSEA analysis also showed upregulation of Epithelial to Mesenchymal transition pathway supporting the idea that mature parenchymal hepatocytes are being converted as proliferative less functional mesenchymes. KEGG pathway analysis of the Hippo Signaling pathway also showed that *YAP/TAZ* are highly overexpressed whereas *NF2* and *CSNK1D* were dysregulated meaning ESRP2 is not interacting with them as it is damaged. To add up, alternative splicing analysis of *ESRP2, YAP1* and *TAZ* showed some exons are downregulated in *ESRP2* gene whereas few exons in *YAP1* and *TAZ* are upregulated supporting the results to the authors' claim about hepatocytes.

Since this study was not mainly focussed on how *ESRP2* acts in maintaining homeostasis of hepatocytes, it was also focussed on any particular reason for such early prognosis and high mortality rate in SAH. Apart from TNF-A and IL-1B, there were many differentially expressed cytokines and chemokines genes seen highly expressed including many genes from the CXC and CC chemokine family. These genes include IL-6, CXCL1, CXCL6, CXCL8, CXCL10 which are known to be highly proinflammatory. Earlier studies have shown that certain chemokines upregulation including CCL2, CXCL1 and CXCL6 are the main cause of worst prognosis and high mortality in SAH (9). GSEA enrichment analysis also showed some upregulated DEGs enriched in Inflammatory response pathway. KEGG pathway analysis also showed Cytokine-Cytokine receptor interaction and ECM-receptor interaction pathway to be highly regulated which is due to inflammatory response elicited by immune response due to ethanol toxicity. Visualizing the Cytokine-Cytokine Receptor interaction pathway showed certain CXC and CC ligands to be highly upregulated (Figure 14).

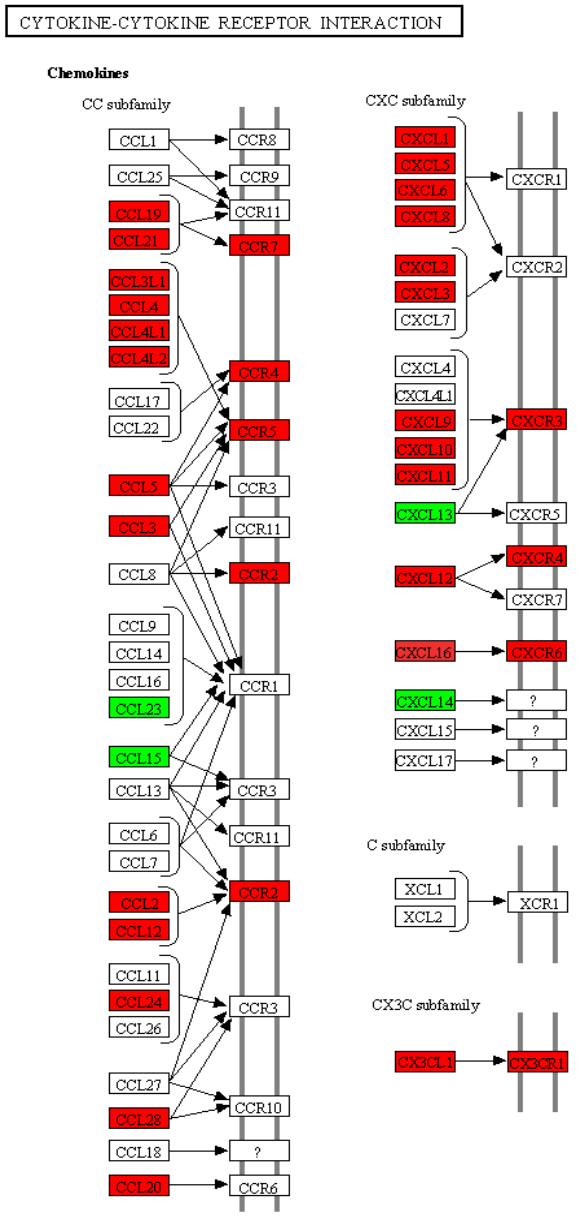


***Figure 14 :*** *Upregulated chemokines in cytokine-cytokine receptor interaction pathway*

There were many other cytokines being upregulated but these are some of the important ones known to be related with SAH. Results indicate that there might be some kind of cytokine storm happening due to inflammation from ethanol which is destroying the liver homeostasis as Bangru et al. showed how TNF-A and IL-1B cytokines are destroying ESRP2 and making the hepatocyte homeostasis completely dysregulated.

Certain limitations that were faced include doing some wet-lab to see at what time point these cytokines start to express in higher amounts so that controlling these cytokines and chemokines at a timely manner could give some time for treatment. With only the datasets from other research it is only possible to make analysis more focused on what the original authors' main focus and point of view was rather than having more diversity. Furthermore research on how these excessive cytokines are activated at what timeframes and how they can be

controlled should be studied. The cytokine-cytokine receptor interaction pathway can be concentrated more for future studies to see for any therapeutic targets possible to downregulate the cytokine and chemokine formation pathways so that the short mortality time can be increased to give a timeline for treatment.

## Methods

### 1. TOOLS: A Brief Description

#### *1.1 Fast-QC*

Fast-QC is a popular tool used in bioinformatics to perform basic quality control in raw data output by next generation sequence technologies (13). Following analysis of the sequenced reads in a FASTQ file format, it outputs an html which is viewable in your browser.

The Fast-QC html displays various modular analysis. This aids the researcher in determining if the data is of good quality to continue further analysis or if it has any problems that need to be addressed before continuing with the analysis pipeline.

#### *1.2 Kallisto*

Kallisto is known for being a fast and accurate quantification tool. It accomplishes its speed by applying the novel concept of *pseudo-aligning* the reads to find indels and base errors, and by its capacity of performing many rounds of an iterative method known as *expectation-maximization*. In fact, because of this approach, it makes it more accurate than many existing tools (11).

This tool may be used in both single-ended and pair-ended reads. Although it is unable to perform *Differential Expressed Genes* analysis by itself, its bootstrapped output may be used as input for tools that are able to do DEG analysis.

#### *1.3 Limma*

*Limma* is a package that uses linear models to analyze differentially expressed genes using data from high throughput RNA-seq technologies (15).

#### *1.4 STAR*

The *Spliced Transcripts Alignment to a Reference (STAR) Aligner* is a very important tool that aids the researcher in determining the origin of his reads in the human genome (13). It is an ideal tool since its algorithm is able to maintain a high accuracy when mapping while still considering spliced alignments.

#### *1.4 HT-Seq*

HT-Seq library is a python based aligner used to count different genomic features which include exons, introns, peaks from ChiP-Seq etc. It takes file in SAM or BAM file as input as well as GTF file format for gene annotations and gives output as number of genes that gets mapped each time in reads (16).

#### *1.5 Picard*

Picard tools is a java based command line high-throughput sequencing tool used for modifying files at different levels (17). Picard tools accept input as SAM or BAM or VCF. Picard tools were used to remove duplicates after STAR alignment in this study.

#### *1.6 Trimmomatic*

Trimmomatic is a tool that is used to filter out low quality reads and adapters and Polymerase Chain Reaction Duplicates in a sequence (14).

### *2. PROCEDURES:*

#### *2.1 Data Origin:*

The expression dataset of genes was collected from National Center for Biotechnology Information (NCBI) at Gene Expression Omnibus (GEO) database with accession number GSE143318. The

dataset was prepared using GPL18573 Illumina NextSeq 500 (Homo sapiens) platform. The dataset includes 10 single-ended reads: 5 reads were prepared using cells from explanted livers of 5 Severe Alcoholic Hepatitis (SAH) patients and 5 reads were prepared using cells from livers of healthy individuals who had no previous history or have liver disease. All the samples were used in accordance with NIH institutional guidelines and the study was approved by the Human Ethics Committee of John Hopkins University. A written informed consent was taken from each of the patients from whom the samples were collected for this study.

### *2.2 Data Pre-processing and Quality Check :*

All of the 10 reads used in this study were run for quality assurance with FastQC to check adapter content and Trimmomatic was used for soft-trimming of the reads. Trimmomatic was done twice: one with the following parameters- LEADING:3, TRAILING:3, SLIDINGWINDOW:4:16, MINLEN:36 ; and for the other one the parameters used were- HEADCROP:15, TRAILING:3, SLIDINGWINDOW: 4:15, MINLEN:36. And then compared with the untrimmed data by running the whole analysis thrice using each of the trimmomatic and untrimmed reads set. Since the original reads had no adapter content, we focussed our study on the untrimmed RNA seq data and used that data only for downstream analysis.

### *2.3 Data Alignment:*

To obtain transcript counts for downstream DEGs and Enrichment Analysis, alignment of all the reads was done using *Kallisto* aligner against cDNA and ncRNA reference fasta files of Human *GRCh38.p13 (GCA_000001405.28)* genome assembly obtained from *Ensembl Database*. To obtain exon-level counts for alternative splicing analysis, STAR aligner was used to align the reads against the primary assembly genome fasta file of Human *GRCh38.p13 (GCA_000001405.28)* genome assembly obtained from *Ensembl Database* and then used *Picard tools* to remove duplicates followed by *HTSeq* which was used using Picard output to obtain exon-level counts.

### *2.4 Annotation:*

Gene annotation for Differentially Expressed Genes (DEGs) and Alternative Splicing Analysis were performed using annotation files obtained from the HGNC database. Hallmark genesets file for gene set enrichment analysis (GSEA) was obtained in R format from Walter+Eliza Hall Bioinformatics website.

### *2.5 Differential Gene Expression Analysis:*

Differential Gene Expression Analysis was done in R using the Limma package available on Bioconductor webpage. A design matrix was made based on the experimental analysis in R. Gene counts obtained from Kallisto were used and converted into a dge list using the DGEList function from Limma and then normalized the data using TMM. Filtering was done by cpm>1 in at least 3 samples and low quality genes were dropped from the list. Voom from Limma was used to convert dge list of counts into a voom object to aid in fitting a linear model. lmFit function from Limma was used to fit the design matrix into the voom processed data and next ebayes from Limma was used on lmFit data to add statistics. Then the topTable function from Limma was used to get Differentially Expressed Genes with adjusted p-value < 0.05 in severe alcoholic hepatitis patients compared to healthy patients.

### *2.6 Enrichment Analysis:*

Gene Set Enrichment Analysis was done using Romer function from Limma to see where the differentially expressed genes are clustered as genesets. GO term Enrichment Analysis was also performed using Goana function from Limma to see the enriched biological pathways of differentially expressed genes. Pathway enrichment analysis was performed using Kyoto Encyclopedia of Genes and Genomes (KEGG) to visualize in which particular pathways the differentially expressed genes were mostly enriched.

### *2.7 Alternative Splicing Analysis:*

Alternative Splicing Analysis of specific genes was performed with the Limma pipeline in R and followed the same procedure as getting the differentially expressed genes but this time the data was the counts at the exon-level from HTSeq. Diffsplice function of Limma was used to get the genes having differentially spliced exons and plotSplice function of Limma was used to plot specific genes with alternatively spliced exons.

## References

1. Hyun J, Sun Z, Ahmadi AR, Bangru S, Chembazhi UV, Du K, Chen T, Tsukamoto H, Rusyn I, Kalsotra A, Diehl AM. Epithelial splicing regulatory protein 2-mediated alternative splicing reprograms hepatocytes in severe alcoholic hepatitis. J Clin Invest. 2020 Apr 1;130(4):2129-2145. doi: 10.1172/JCI132691. PMID: 31945016; PMCID: PMC7108908.

2. Liu M, Dou Y, Sun R, Zhang Y, Liu Y. Molecular mechanisms for alcoholic hepatitis based on analysis of gene expression profile. Hepat Mon. 2015 May 23;15(5):e27336..doi:10.5812/hepatmon.15(5)2015.27336. PMID: 26045708; PMCID: PMC4451276.

3. Nagata K, Suzuki H, Sakaguchi S. Common pathogenic mechanism in development progression of liver injury caused by non-alcoholic or alcoholic steatohepatitis. J Toxicol Sci. 2007 Dec;32(5):453-68. doi: 10.2131/jts.32.453. PMID: 18198478.

4. Affò S, Dominguez M, Lozano JJ, Sancho-Bru P, Rodrigo-Torres D, Morales-Ibanez O, Moreno M, Millán C, Loaeza-del-Castillo A, Altamirano J, García-Pagán JC, Arroyo V, Ginès P, Caballería J, Schwabe RF, Bataller R. Transcriptome analysis identifies TNF superfamily receptors as potential therapeutic targets in alcoholic hepatitis. Gut. 2013 Mar;62(3):452-60. doi: 10.1136/gutjnl-2011-301146. Epub 2012 May 25. PMID: 22637703; PMCID: PMC4064940.

5. Martin J, Dufour JF. Tumor suppressor and hepatocellular carcinoma. World J Gastroenterol. 2008 Mar 21;14(11):1720-33. doi: 10.3748/wjg.14.1720. PMID: 18350603; PMCID: PMC2695912.

6. Yimlamai D, Christodoulou C, Galli GG, Yanger K, Pepe-Mooney B, Gurung B, Shrestha K, Cahan P, Stanger BZ, Camargo FD. Hippo pathway activity influences liver cell fate. Cell. 2014 Jun 5;157(6):1324-1338. doi: 10.1016/j.cell.2014.03.060. PMID: 24906150; PMCID: PMC4136468.

7. Hurley T.D. Genes encoding enzymes involved in ethanol metabolism. *Alcohol Res. Health.* 2012;34:339–344.

8. Yu FX, Zhao B, Guan KL. Hippo Pathway in Organ Size Control, Tissue Homeostasis, and Cancer. Cell. 2015 Nov 5;163(4):811-28. doi: 10.1016/j.cell.2015.10.044. PMID: 26544935; PMCID: PMC4638384.

9. Dominguez M, Miquel R, Colmenero J, Moreno M, García-Pagán JC, Bosch J, Arroyo V, Ginès P, Caballería J, Bataller R. Hepatic expression of CXC chemokines predicts portal hypertension and survival in patients with alcoholic hepatitis. Gastroenterology. 2009 May; 136(5):1639-50. doi: 10.1053/j.gastro.2009.01.056. Epub 2009 Jan 31. PMID: 19208360.

10. Zhang JM, An J. Cytokines, inflammation, and pain. Int Anesthesiol Clin. 2007 Spring;45(2):27-37. doi: 10.1097/AIA.0b013e318034194e. PMID: 17426506; PMCID: PMC2785020.

11. Nicolas L Bray, Harold Pimentel, Páll Melsted and Lior Pachter, Near-optimal probabilistic RNA-seq quantification, Nature Biotechnology **34**, 525–527 (2016), doi:10.1038/nbt.3519

12. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013 Jan 1;29(1):15-21. doi: 10.1093/bioinformatics/bts635. Epub 2012 Oct 25. PMID: 23104886; PMCID: PMC3530905.

13. Andrews, S. (2010). FastQC: A Quality Control Tool for High Throughput Sequence Data [Online]. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

14. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014 Aug 1;30(15):2114-20. doi: 10.1093/bioinformatics/btu170. Epub 2014 Apr 1. PMID: 24695404; PMCID: PMC4103590.

15. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK (2015). "limma powers differential expression analyses for RNA-sequencing and

microarray studies." *Nucleic Acids Research*, **43**(7), e47. doi: 10.1093/nar/gkv007.

16. Anders S, Pyl PT, Huber W. HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics. 2015 Jan 15;31(2):166-9. doi: 10.1093/bioinformatics/btu638. Epub 2014 Sep 25. PMID: 25260700; PMCID: PMC4287950.

17. "Picard Toolkit." 2019. Broad Institute, GitHub Repository. http://broadinstitute.github.io/picard/; Broad Institute

18. Yang Liao, Wei Shi-Read trimming is not required for mapping and quantification of RNA-seq reads bioRxiv 833962; doi: https://doi.org/10.1101/833962

## TERMINOLOGY

**DEGs:** Differentially Expressed Genes

**SAH:** Severe Alcoholic Hepatitis

**GO:** Gene Ontology

**KEGG:** Kyoto Encyclopedia of Genes and Genomes

**GSEA:** Gene Set Enrichment Analysis

**LIMMA:** Linear Models for Microarray Data

**STAR:** Spliced Transcripts Alignment to a Reference

**ESRP2:** Epithelial Splicing Regulatory Protein 2

**YAP1:** Yes-associated protein 1

**TAZ:** Tafazzin

**CXCL1:** CXC motif Ligand 1

**CXCL6:** CXC motif Ligand 6

**CXCL10:** CXC motif Ligand 10

**IL-6:** Interleukin 6

**IL-1B:** Interleukin 1 Beta

**NF2:** Neurofibromin 2

**CSNK1D:** Casein Kinase 1 Delta

**ADH4:** Alcohol dehydrogenase 4

**ADH1A:** Alcohol Dehydrogenase 1A

**ADH1B:** Alcohol Dehydrogenase 1B

**ADH1C:** Alcohol Dehydrogenase 1C

**PTGS2:** Prostaglandin-endoperoxide synthase 2

**CYR61:** Cysteine-rich angiogenic inducer 61

**CTGF:** Connective Tissue Growth Factor