ANALYSIS OF HBV ASSOCIATED ACUTE LIVER FAILURE MICROARRAY DATA

Nisan Ece Kalem

Department of Biological Sciences, Middle East Technical University
06800, Ankara, Turkey
phone: +(90)312 123 45 67, email: ece.kalem@metu.edu.tr

INTRODUCTION

The liver is an organ only found in vertebrate animals that detoxifies the body and is responsible for breaking down of certain metabolites, certain protein synthesis, and production of biochemicals that are required for digestion and detoxification. In that sense, one may claim that the total dysfunction of a liver would be deadly to the organism. [1]

Acute liver failure is a lethal condition and can be described as an immediate and massive loss of function of the liver which leads to multi-organ failure if not treated. To acknowledge the main differences in terms of the expression profiles of patients with acute liver failure and individuals with a healthy liver, the gene expression levels of acute liver failure (ALF), and health organ donors are compared. [2]

METHODS

In this study, 27 liver specimens obtained from two groups of individuals; 17 liver specimens are taken from 4 patients with ALF and 10 liver specimens are taken from 4 healthy liver donors. Gene expression profiling was performed on all of the liver specimens with Affymetrix U133 Plus 2, that is containing approximately 54 thousand probe sets which represents roughly 22 thousand human genes. The data is retrieved from the NCBI Gene Expression Omnibus database under the accession number of GSE38491. Except for enrichment analysis, all data analyses were performed on R by using the R package 'limma'. The GEO data set was automatically retrieved from the database to R workspace environment and in the database, uploaded data were already RMA normalized. Differential expression of genes between healthy donors and ALF was obtained by using eBayes method in limma. The resulting p-values are Bonferroni corrected with False Discovery Rate (FDR) of <1%. [3-5]

Next, unsupervised learning, and dimension reduction methods, are applied to the data to see if the groups of the data are appropriately clustered. Hierarchical clustering and principal component analysis were performed to visualize the clustering of the data.

In following the unsupervised learning, the genes with significant p-values are subsetted and gene set enrichment analysis was performed by using DAVID gene ontology. DAVID's enrichment analysis results were visualized by using online tool REVIGO which uses Cytoscape in the back end. [6-7]. Also, heatmap analysis was performed with this subset of genes, to see a general fashion of differential expression.

RESULTS

The general distribution of gene expression between the samples is represented as boxplots in **Fig 1.** which shows the lack of severe batch effects of noise in data, therefore ready to be further analyzed.

To observe the overall nature of differential gene expression between ALF and healthy donors, an empirical Bayesian inference t-test was applied to fold change and significance of fold change as p-values. P values are adjusted accordingly the false discovery rate < 1%, and the scatter plot was obtained which shows the differential gene expression in ALF and

healthy donors (**Fig 2.**). The scatter plot showed a similar trend of gene expression distribution between ALF and normal donors as in boxplot representation.

To visualize the subset of significantly upregulated and downregulated genes, the volcano plot was generated. The threshold was hold for LogFold change less than -1 or more than 1 and -log p-value above 10. The threshold for p-values was kept high (1e-10) because aiming to subset the appropriate number of genes for enrichment analysis(**Fig, 3.**).

To observe any clustering within the samples between the groups, unsupervised learning methods, and dimensionality reduction methods were applied to the data. For these purposes, hierarchical clustering and principal component analysis were performed. After these two analyses, it was observed that ALF and healthy donors were clearly clustered differently and ALF samples were further clustered in two different sets. These results seem to be in line with the published paper of this dataset; the authors reported that ALF patients were in two categories: massive hepatic necrosis and submassive hepatic necrosis. (**Fig 4. and Fig 5.**)[8]

To determine the enriched biological functions in ALF, on significantly differentially expressed genes subsetted in the volcano plot, gene set enrichment analysis is performed in DAVID in gene ontology biological process and results were visualized by using REVIGO. The treemap and network visualization clearly stressed out that, most enriched pathways in ALF are immune response, and MHC class II antigen recognition, which are expected processes in organ failure. [9] (**Fig 6., Fig 7**)

Lastly, to observe overall significantly expressed gene expression among the samples, a heatmap analysis was applied according to the subset given in the volcano plot. The heatmap showed that the two biological groups are separated in terms of differential gene expression. (**Fig. 8**)

DISCUSSION

The analysis has revealed organ failure related cellular processes such as immune response, inflammatory response, antigen recognition and MHC class II antigen recognition were highly accented in acute liver failure. With the help of unsupervised learning methods, it made it possible to visually acknowledge the subgroups of the samples. Overall, the dataset made possible to understand the molecular processes during the organ failure are not stochastic, rather highly active and complex by nature.

REFERENCES

[1] Hoffman, Matthew. “Liver (Anatomy): Picture, Function, Conditions, Tests, Treatments.” WebMD, WebMD, 18 May 2019, https://www.webmd.com/digestive-disorders/picture-of-the-liver.

[2] “Acute Liver Failure.” Mayo Clinic, Mayo Foundation for Medical Education and Research, 29 Aug.2017,https://www.mayoclinic.org/diseases-conditions/acute-liver-failure/symptoms-causes/syc-20352863.

[3] Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, Marshall KA, Phillippy KH, Sherman PM, Holko M, Yefanov A, Lee H, Zhang N, Robertson CL, Serova N, Davis S, Soboleva A.
NCBI GEO: archive for functional genomics data sets--update.
Nucleic Acids Res. 2013 Jan;41(Database issue):D991-5.

[4] R Core Team (2019). R: A language and environment for statistical computing. R Foundation
for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

[5] Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015).
limma powers differential expression analyses for RNA-sequencing and microarray studies.
Nucleic Acids Research 43(7), e47.

[6] Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID Bioinformatics Resources. Nature Protoc. 2009;4(1):44-57.

[7] Supek F, Bošnjak M, Škunca N, Šmuc T.
"REVIGO summarizes and visualizes long lists of Gene Ontology terms"
PLoS ONE 2011. doi:10.1371/journal.pone.0021800

[8] [1] O. Nissim et al., “Liver Regeneration Signature in Hepatitis B Virus (HBV) - Associated Acute Liver Failure Identified by Gene Expression Profiling,” vol. 7, no. 11, 2012.

[9]K. Murphy and C. Weaver, JANEWAY ’ S 9 TH EDITION. .

FIGURES

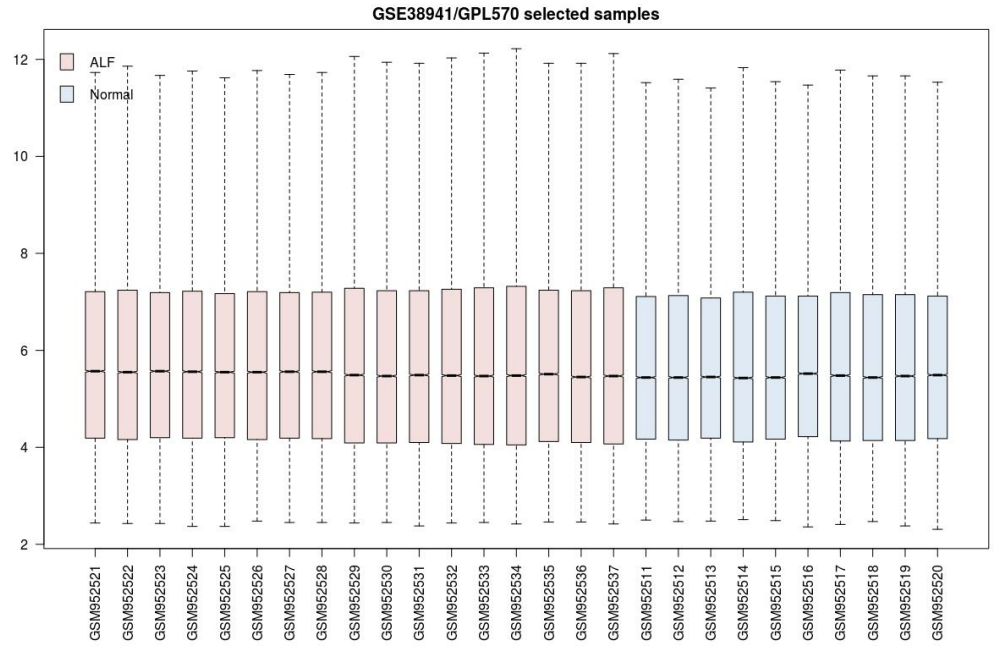

Fig. 1: Boxplot representation for visualization of overall distribution of gene expression across the samples

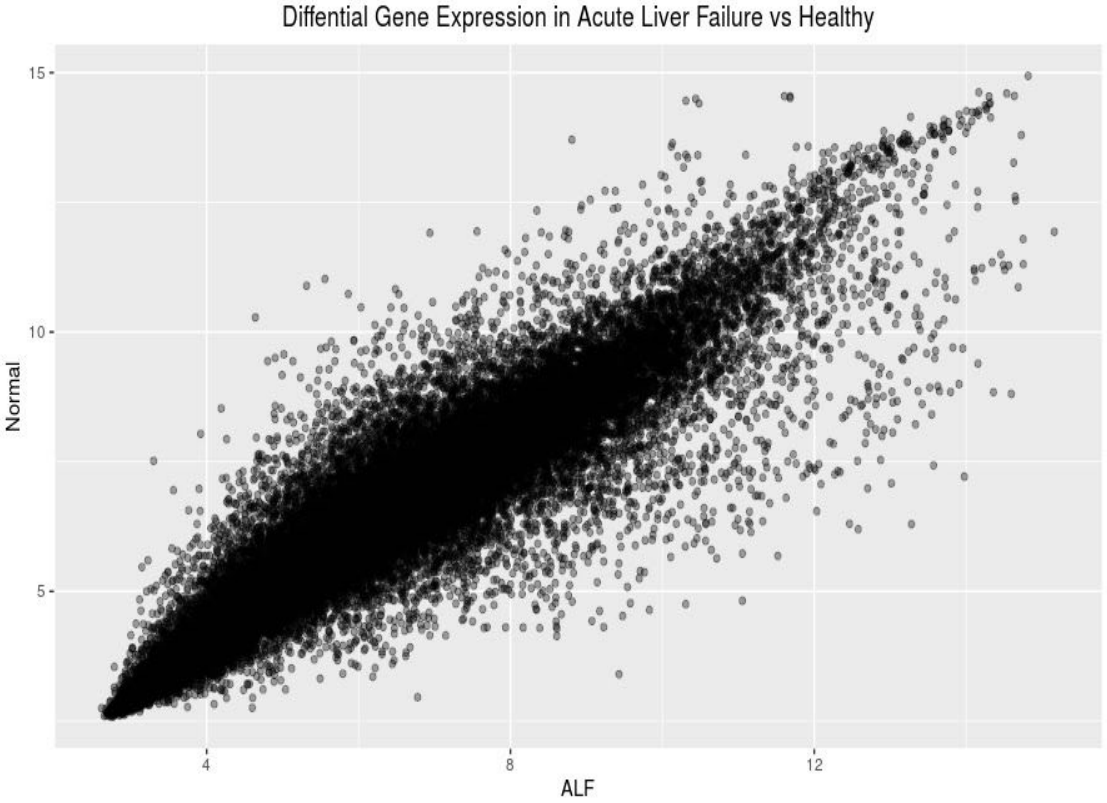

Fig 2: Scatter plot of differential gene expression in Acute liver failure and healthy donors.

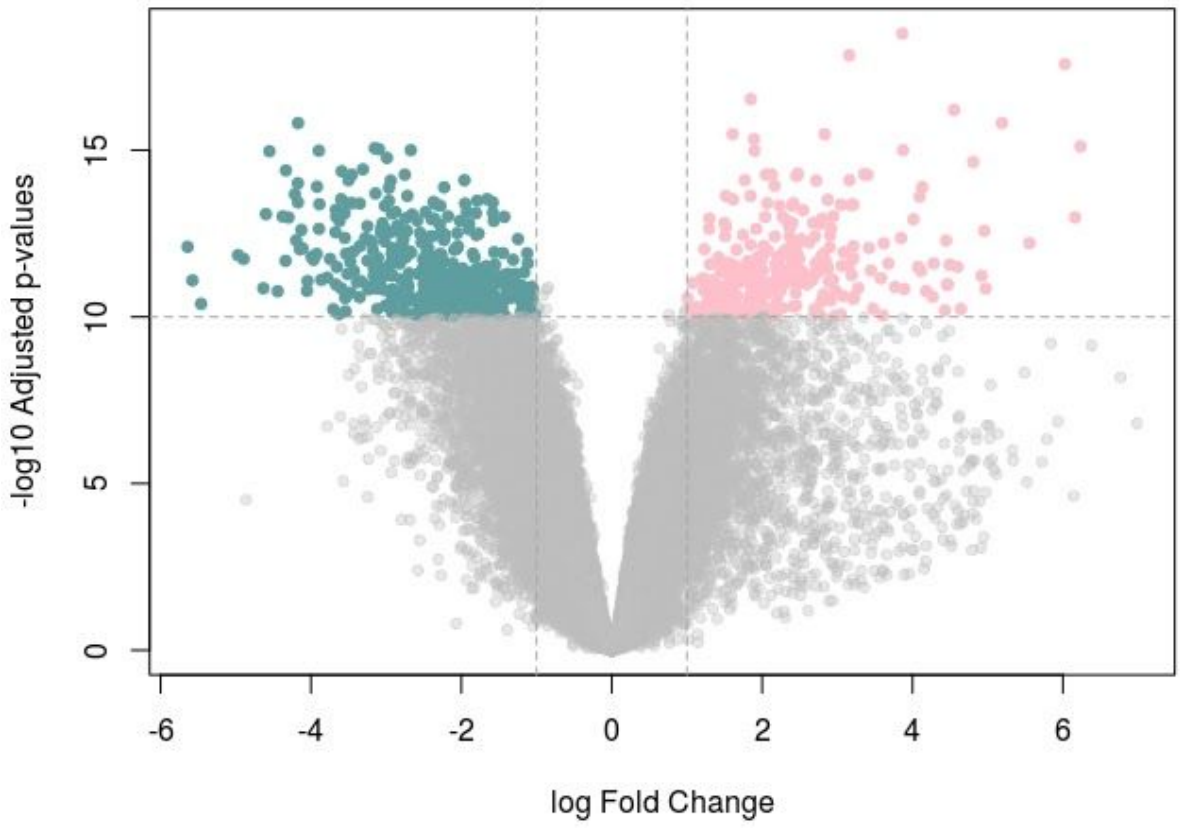

Fig 3: Volcano plot representation of genes with respect to their log of fold changes and adjusted p values

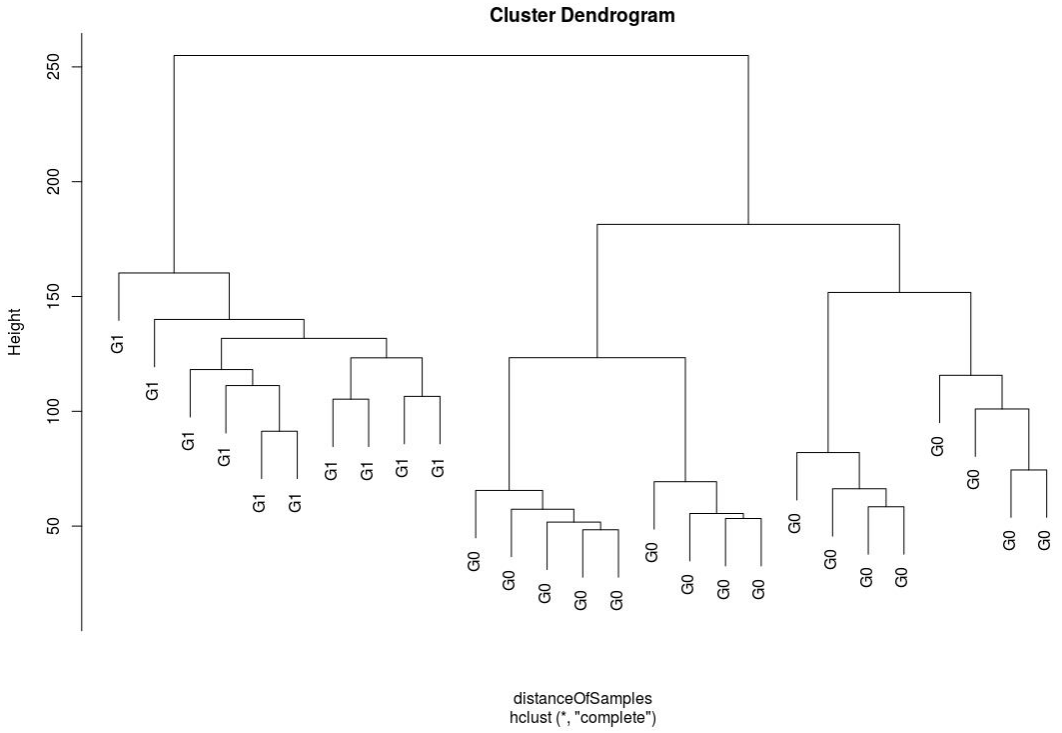

Fig 4: Dendrogram of hierarchical clustering of between the samples. G1 stands for healthy donors and G0 stands for ALF.

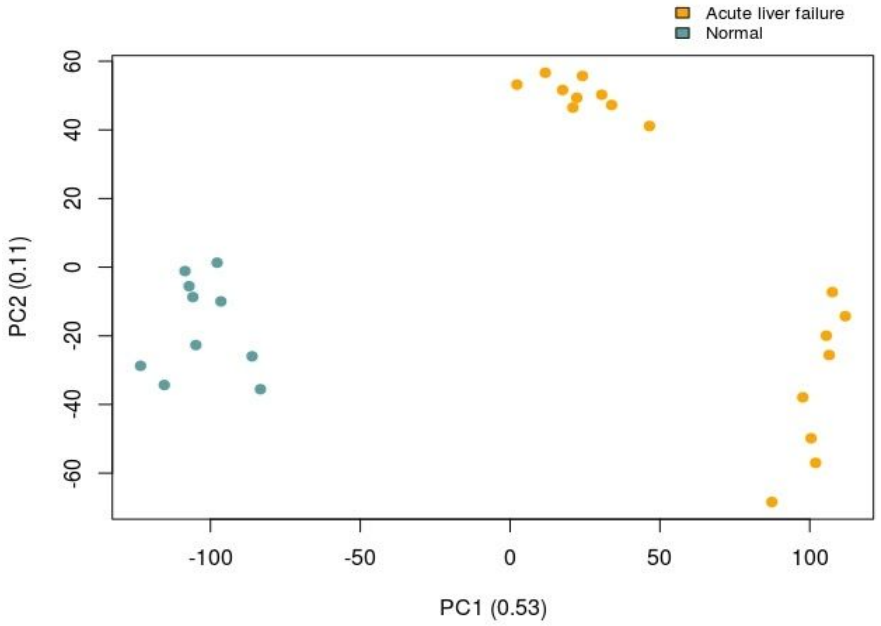

Fig 5: PCA plot of ALF and normal

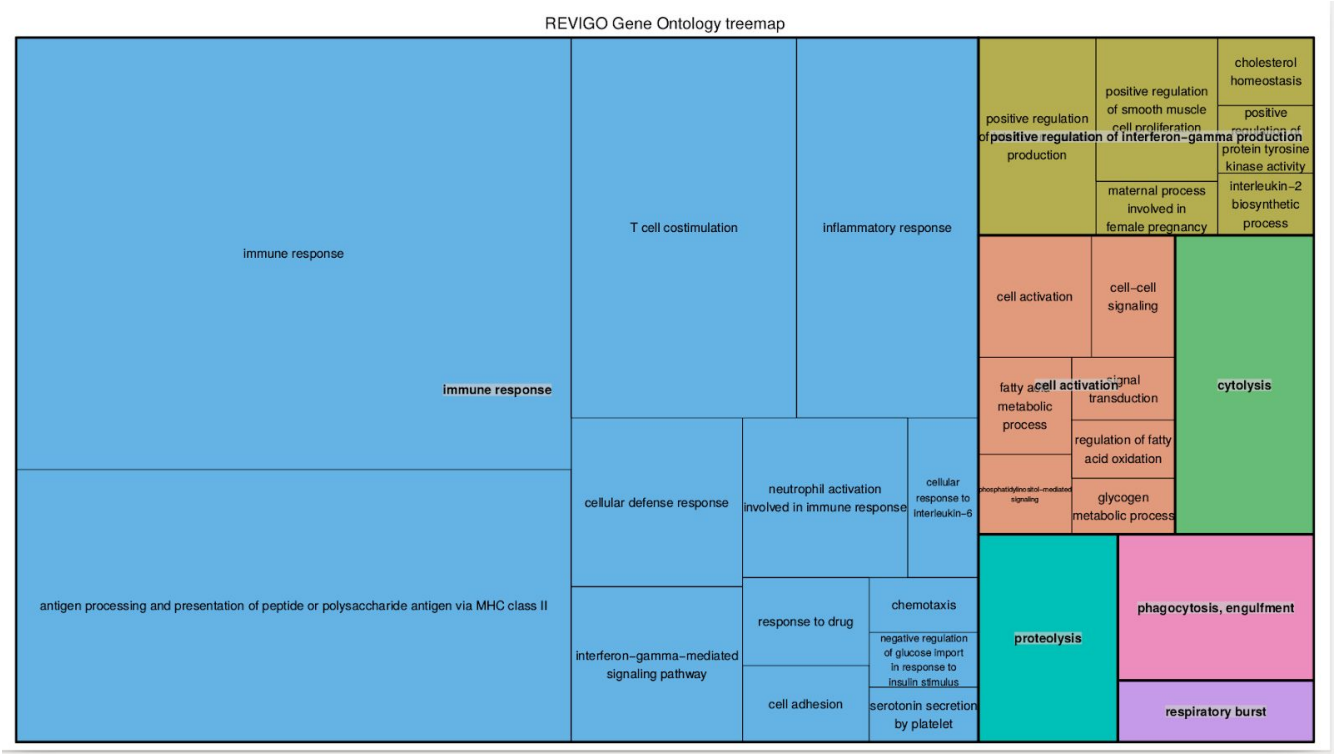

Fig 6: Tree map representation of enriched gene groups in ALF. Colors stand for a similar set of gene groups and the size of the squares stands for the significance of enrichment

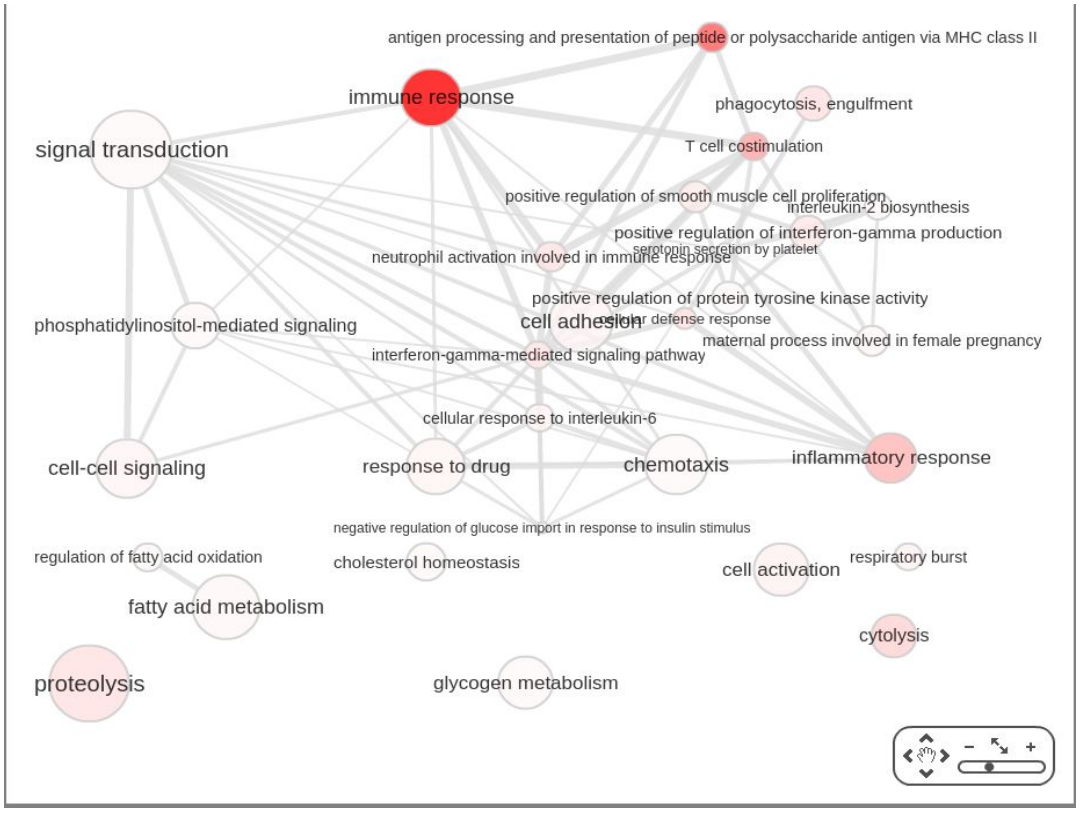

Fig 7: Network representation of enrichment. The color stands for the significance and size stands for the genes involved

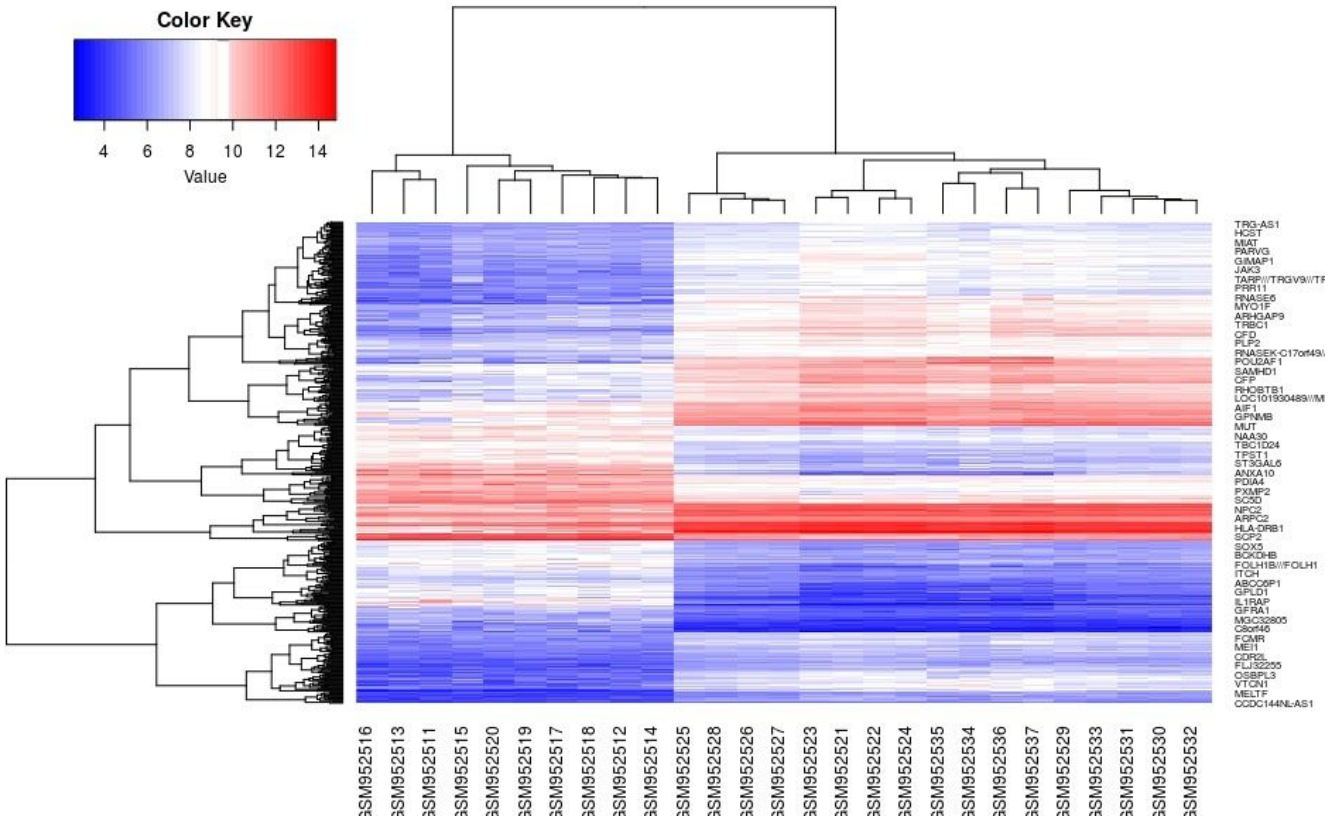

Fig 8: Heatmap representation of significant subset of differentially expressed genes between ALF and healthy donors. Sample names indicated most bottom, gene names are indicated on the right and blue color represents downregulation whereas red color indicates upregulation.