Recognizing Differentially Expressed Genes in HCC

Javad Razvian

Aug 7, 2019

1 Introduction

The study's goal is to recognize special key genes in hepatocellular carcinoma (HCC) in response to the Question 1 available at `http://oncinfo.org/r_test`. We have 8 control samples and 8 tissue samples given from Gene Expression Omnibus with access number GSE59259. As the question asked, the differentially expressed genes (DEGs) between the HCCs and control samples were identified with the aid of limma package in R with cutoff adj.p.value < 0.01.

2 Method

The DEGs between HCC tumors and Liver-non neoplastic control were identified with limma [2] package in R software. A small p-value (typically $\leq$ 0.05) indicates strong evidence against the null hypothesis, concluding that a significant difference does exist in expressed gene. The cutoff criterion was adjusted p-value < 0.01.

We also repeated the analysing stage with another cutoff as other studies like [1] suggest. This measure called log2 fold change (logFC) describing how much a quantity changes going from an initial to a final value. Postive value shows increment and negative value indicates decrement. Therefore, we chose "adjusted p-value < 0.01 and (logFC)| > 2.0" as another cutoff. While with the former cutoff we found 1570 genes, with the latter 143 genes
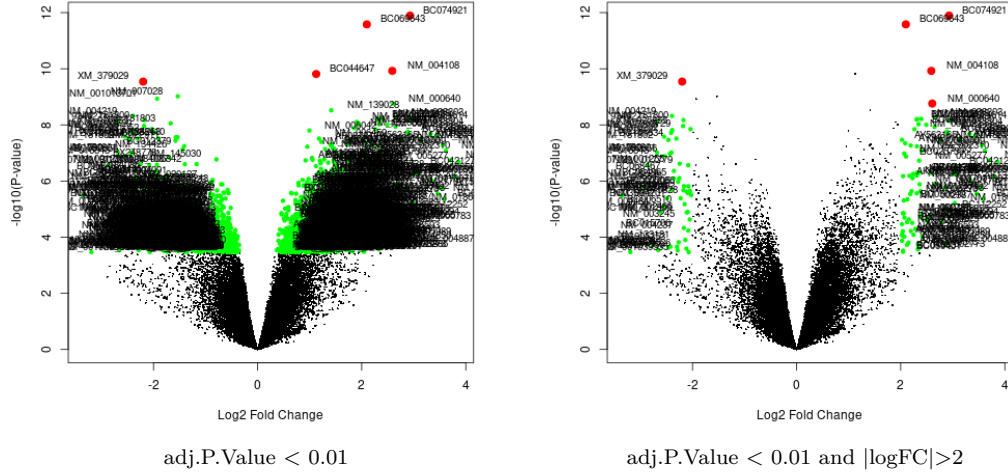
!htbp

Figure 1: Volcano Plot

were identified. However, for top five DE genes, they are similar except for just one gene (see table 1).

Finally we used the Volcano plot to justify our work. The farther the genes go the to right or left, the more dysregulated they are. However, genes appear higher on the volcano plot show highly significant changes. Green points depict the genes that have desired criteria with respect to adjusted p-values and logFC. The top 5 DE genes are illustrated in red.

3   Results

The table 1 depicts genes found as most frequent differentially expressed genes.

Figures 2 and 3 show heatmap for both cutoff criteria. The maps show how the genes correlate with each others, and also cluster the genes based on their similarity in term of their expressions.

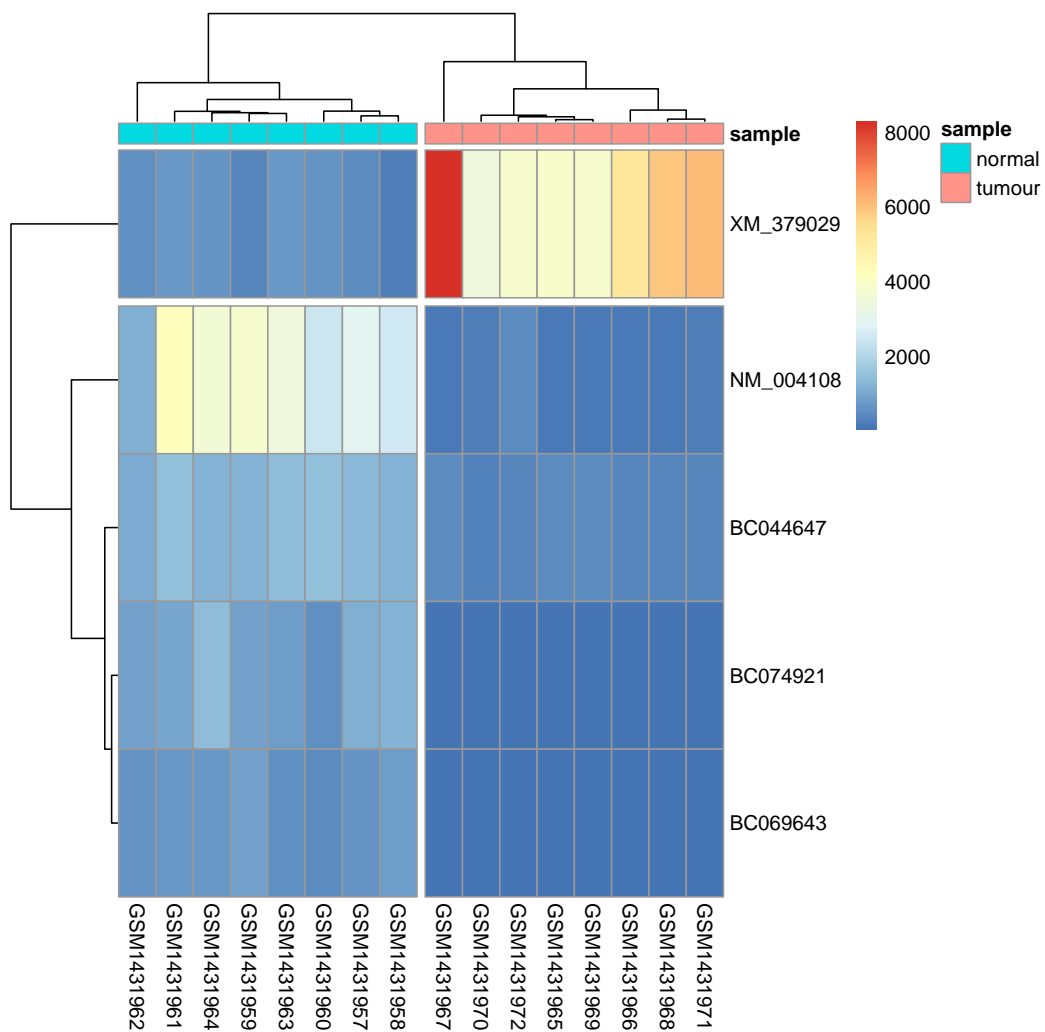| cutoff criteria	adj.P.Value < 0.01	adj.P.Value < 0.01 & |logFC|>2
genes	BC074921	BC074921
	C069643	BC069643
	NM_004108	NM_004108
	BC044647	XM_379029
	XM_379029	NM_000640

Table 1: Top 5 DE genes

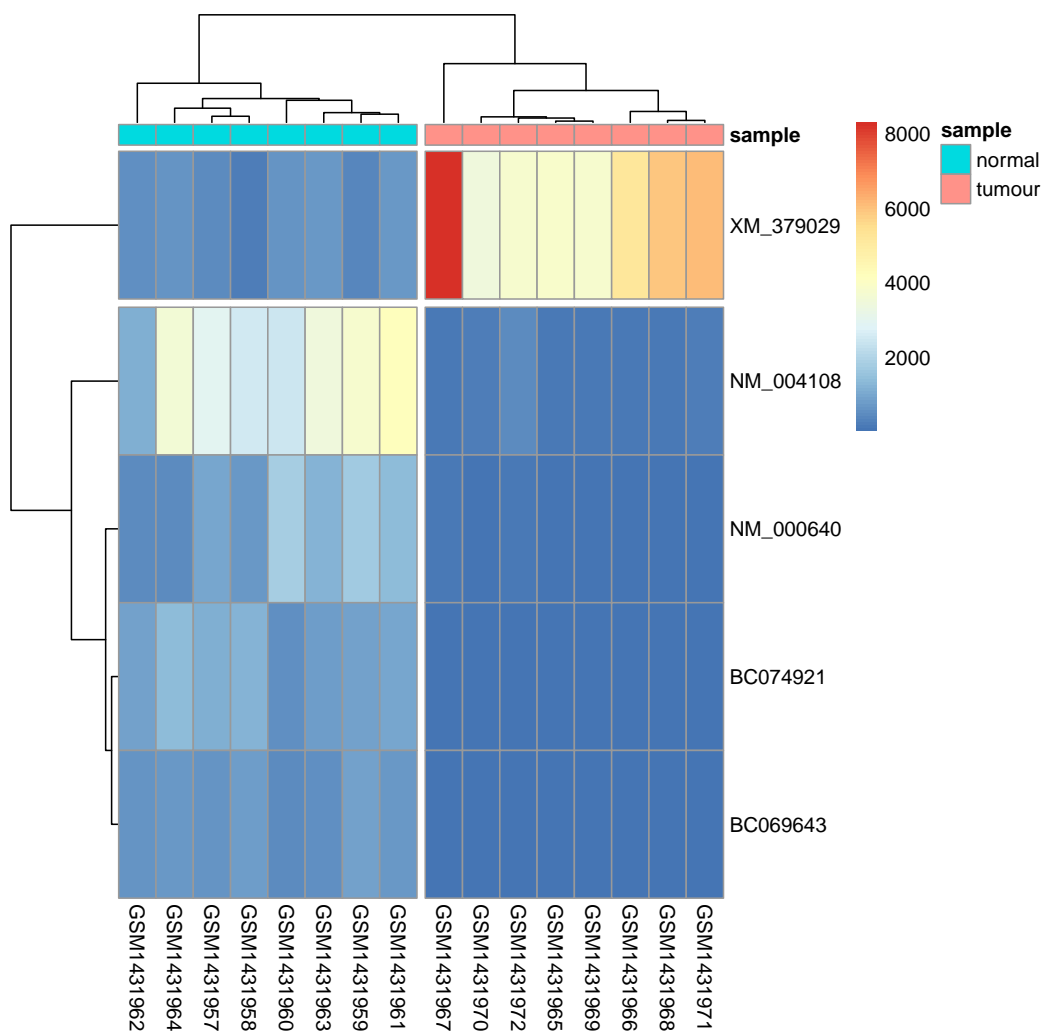Figure 2: Heatmap with cutoff "adj.P.Value < 0.01"

Figure 3: Heatmap with cutoff "adj.P.Value < 0.01 & |logFC|>2"

References

[1] Zhang, X., Kang, C., Li, N., Liu, X., Zhang, J., Gao, F. and Dai, L., 2019. Identification of special key genes for alcohol-related hepatocellular carcinoma through bioinformatic analysis. PeerJ, 7, p.e6375.

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

[3] Microarray Data Analysis `https://discover.nci.nih.gov/microarrayAnalysis/Statistical.Tests.jsp`

[4] Gene Expression Omnibus `https://www.ncbi.nlm.nih.gov/geo/geo2r/?acc=GSE59259`

[5] Limma Tutorial, Available at `https://www.youtube.com/watch?v=Tg_qttx2koI&t=563s`.

[6] Making a heatmap in R with the pheatmap package, Available at `https://davetang.org/muse/2018/05/15/making-a-heatmap-in-r-with-the-pheatmap-package/`.

[7] Using Volcano Plots in R to Visualize Microarray and RNA-seq Results, Available at `https://www.gettinggeneticsdone.com/2014/05/r-volcano-plots-to-visualize-rnaseq-microarray.html`