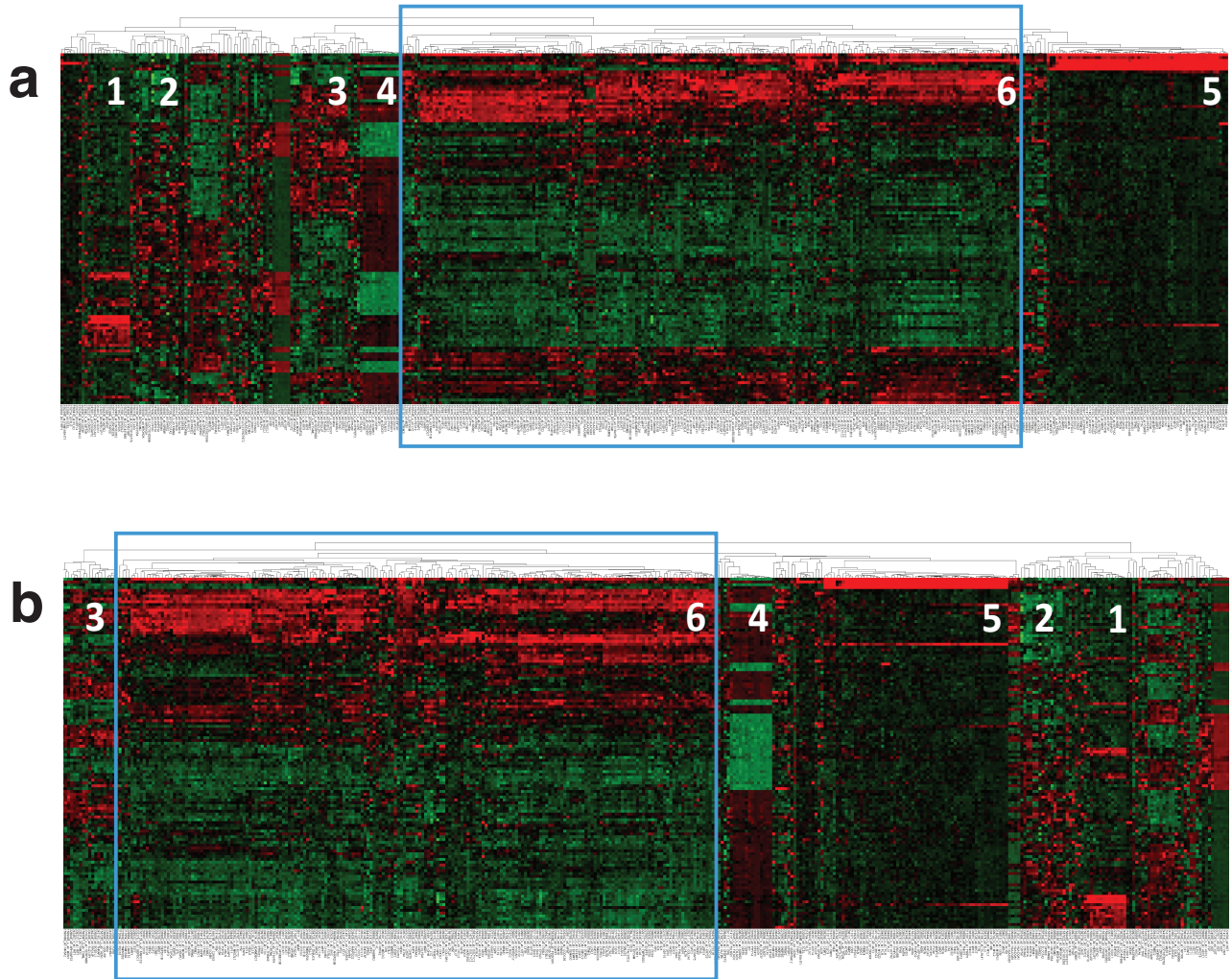
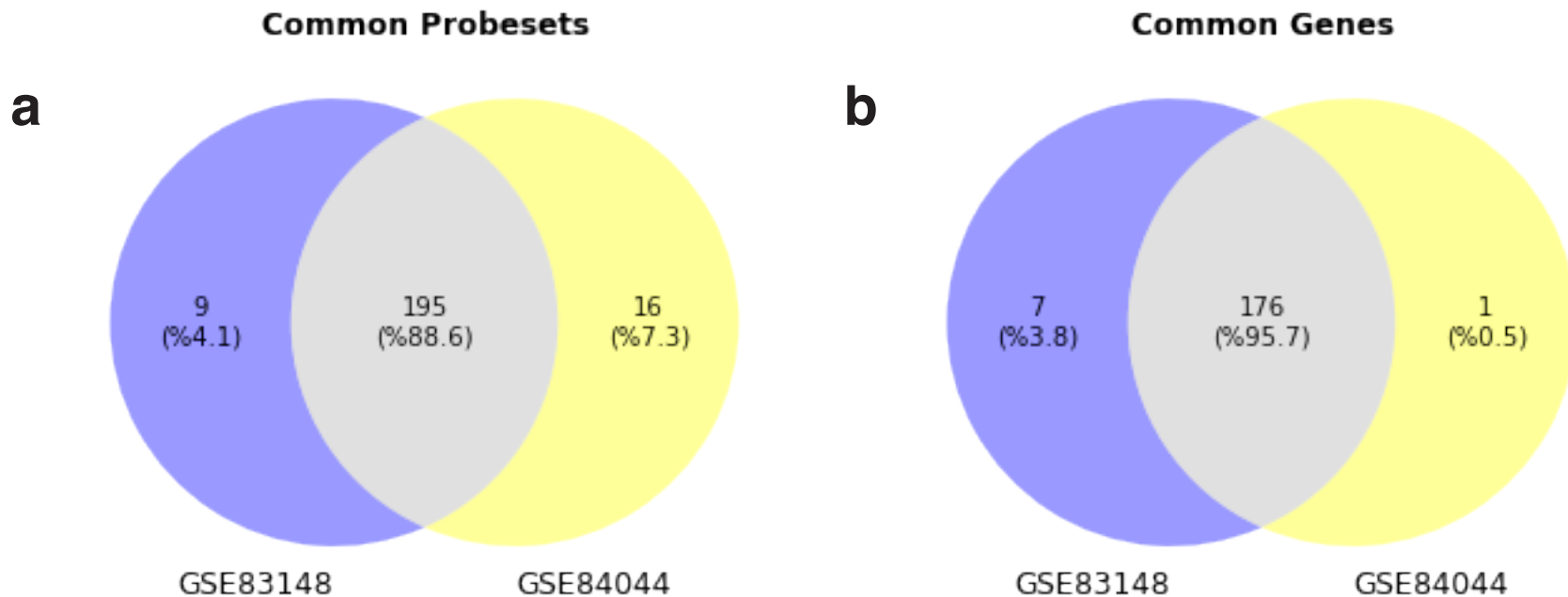


Unsupervised clustering of CHB liver samples based on the most variant probesets



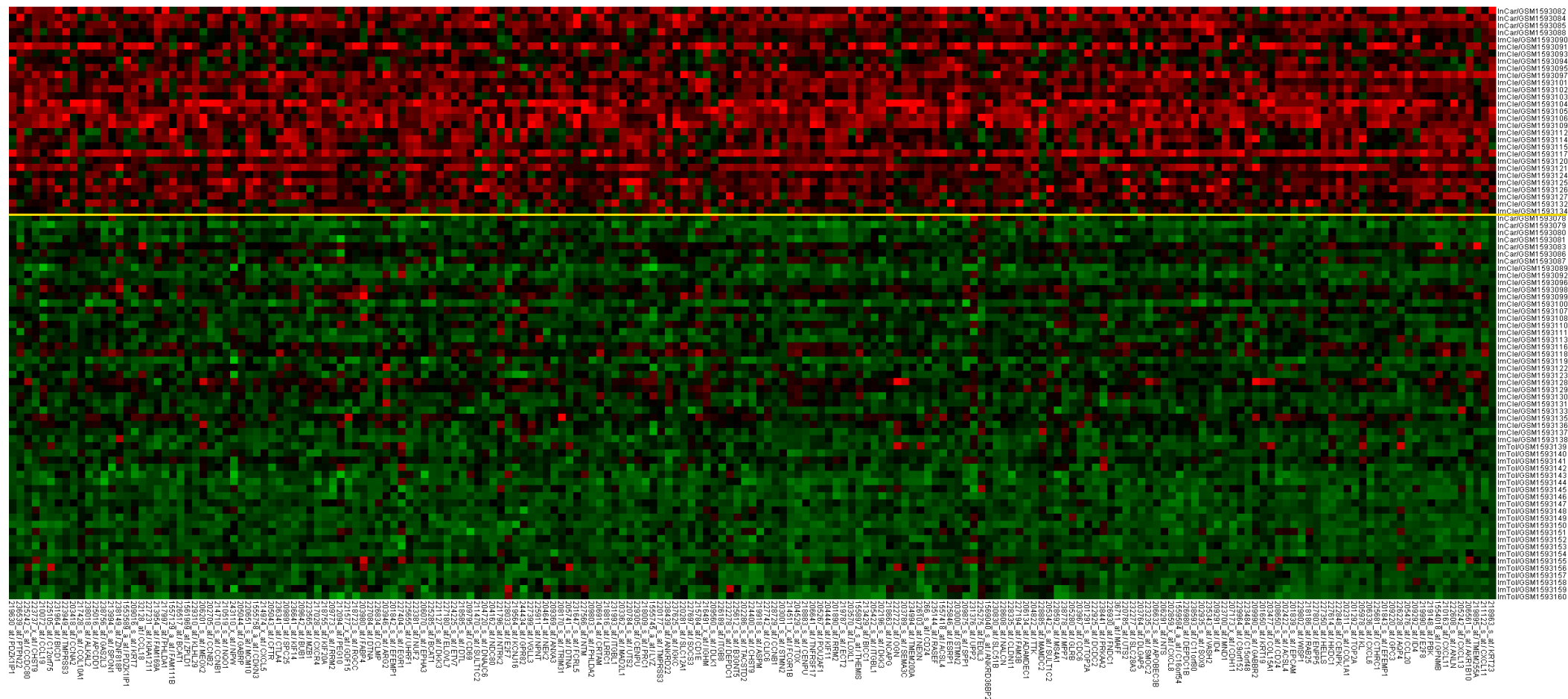
Supplementary Figure 1: Hierarchical clustering of CHB liver samples with the common most variant probesets in GSE83148 (a), GSE84044 (b). Expression of 476 probesets (columns) in 122 and 124 livers from CHB patients (rows) are shown. Red, black and green colors indicate high, moderate and low expression, respectively. Majority of the genes in the same numbered groups in two datasets overlapped. 6th group of genes shown in the blue box included 204 and 197 probesets that generated subgroups with high and low expression.

Common probesets and genes identified in GSE83148 and GSE84044



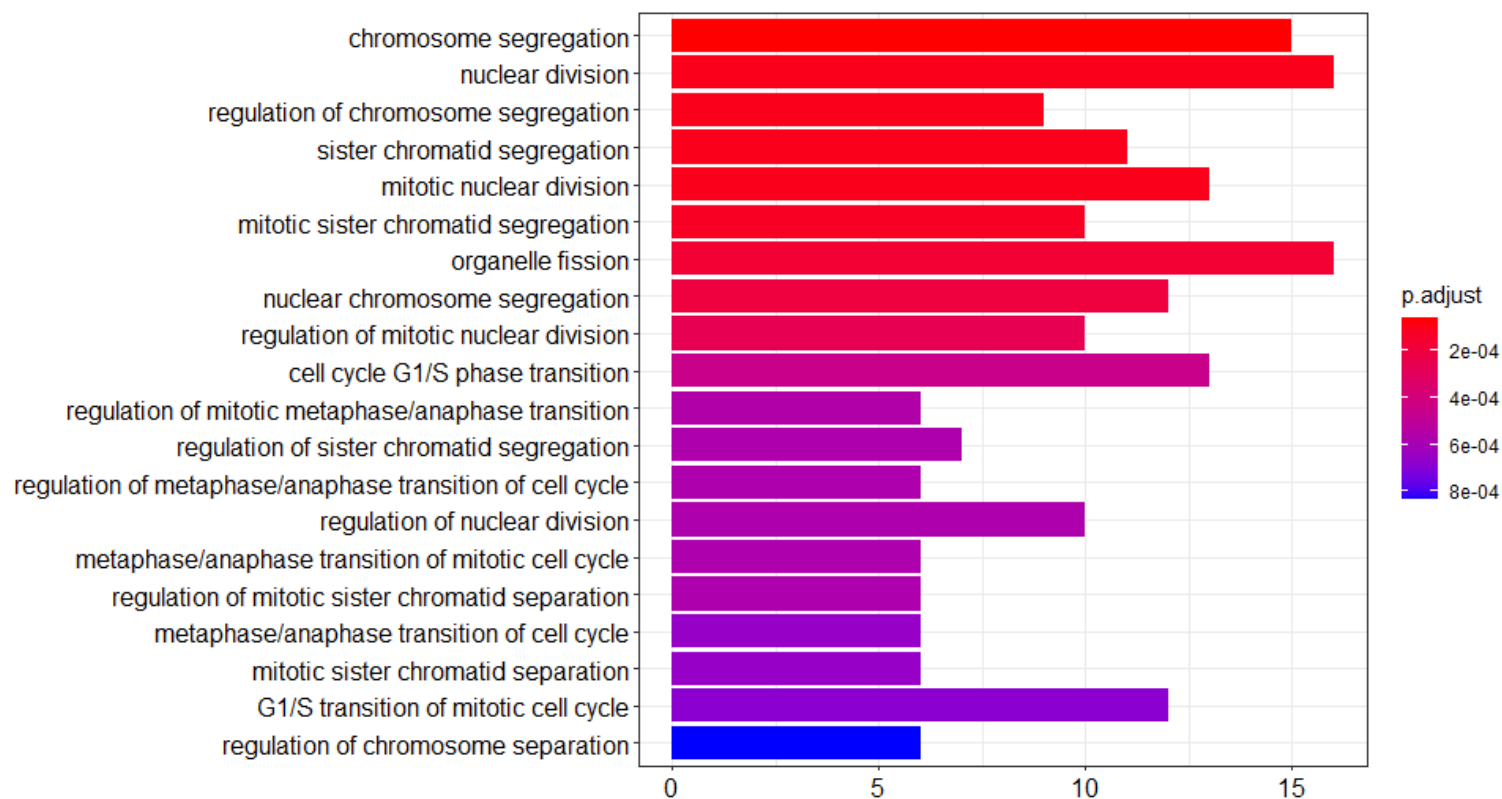
Supplementary Figure 2: Venn diagrams showing overlap of probesets (a) and genes (b) identified based on independent datasets (PCI).

Unsupervised clustering of an independent dataset based on PCI signature



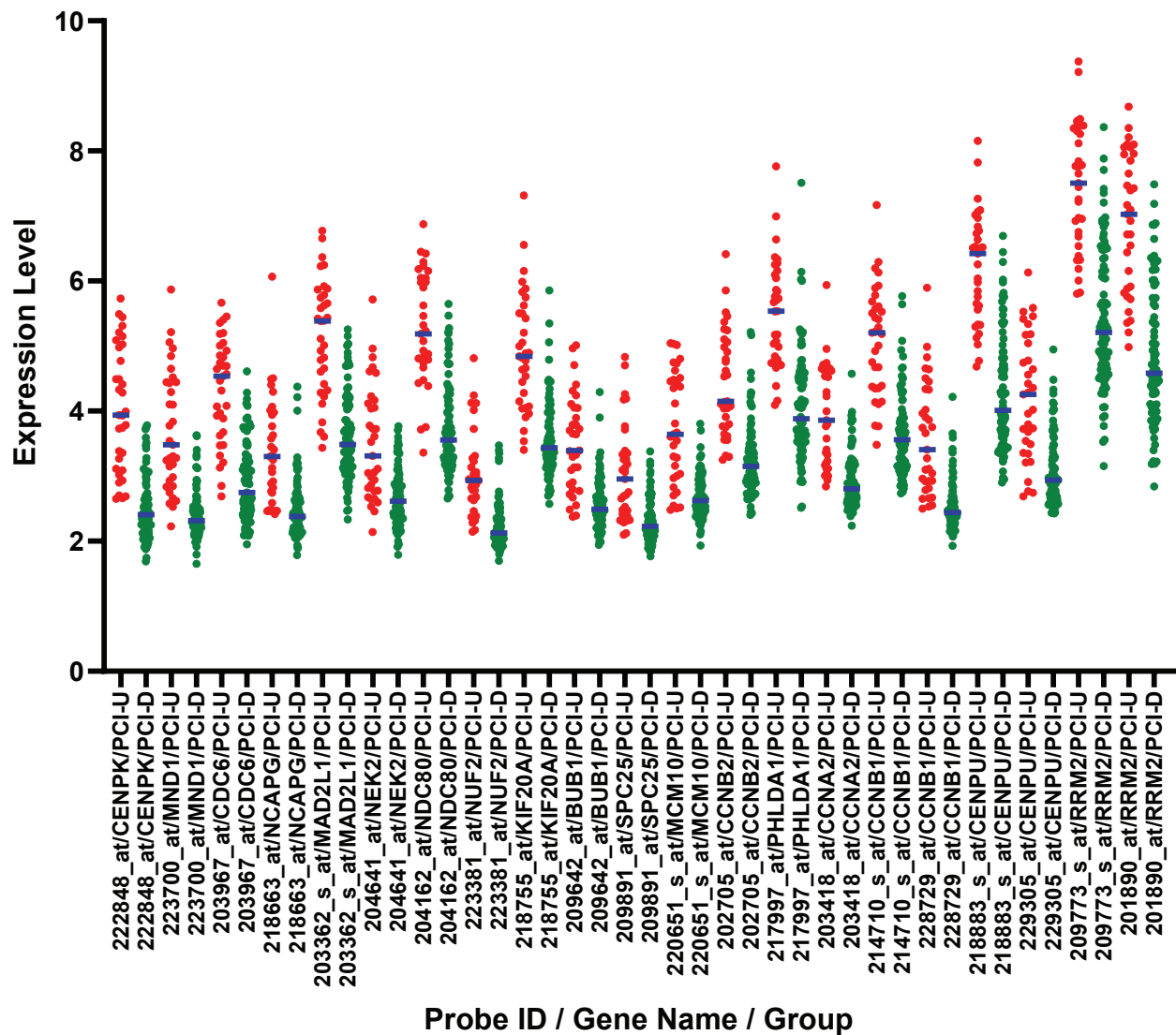
Supplementary figure 3: Hierarchical clustering of GSE65359 dataset. K-means clustering was performed using expression of 195 probes/176 genes from 83 CHB liver samples. Red, black and green colors indicate high, moderate and low expression, respectively. 29, 54 patients were assigned to groups with high and low expression, respectively. Groups are separated by the yellow line.

Functional analysis of genes included in PCI



Supplementary Figure 4: Functional over-representation of 176 gene panel. Top 20 mechanisms with lowest enrichR based adjusted p value are shown.

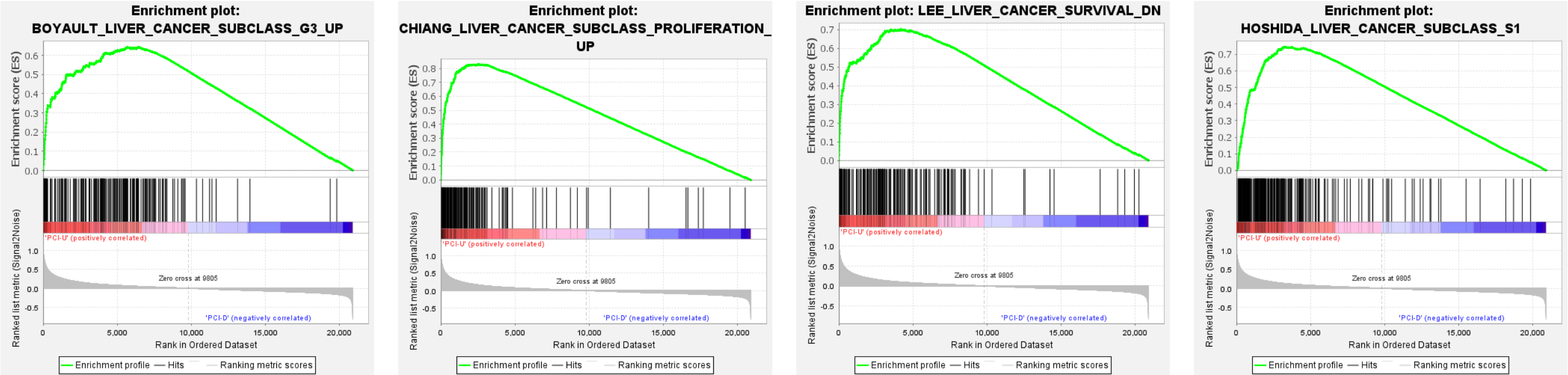
Expression of cell cycle related genes in PCI-U and PCI-D livers



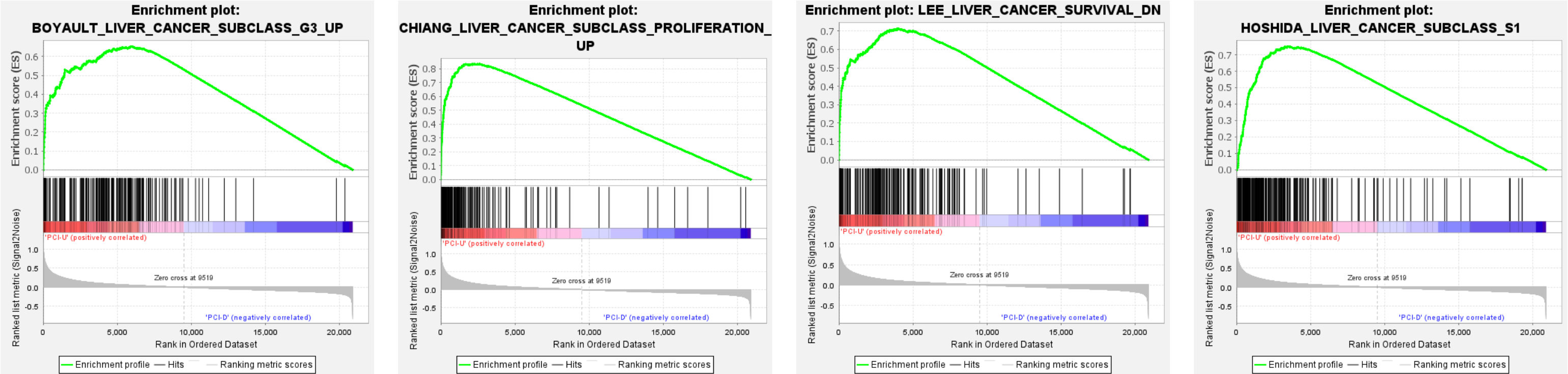
Supplementary Figure 5: Cell cycle related gene expression in PCI-U (red) and PCI-D (green) groups (GSE83148). Horizontal line indicates median.

Enrichment of liver cancer related gene sets in PCI-U

a



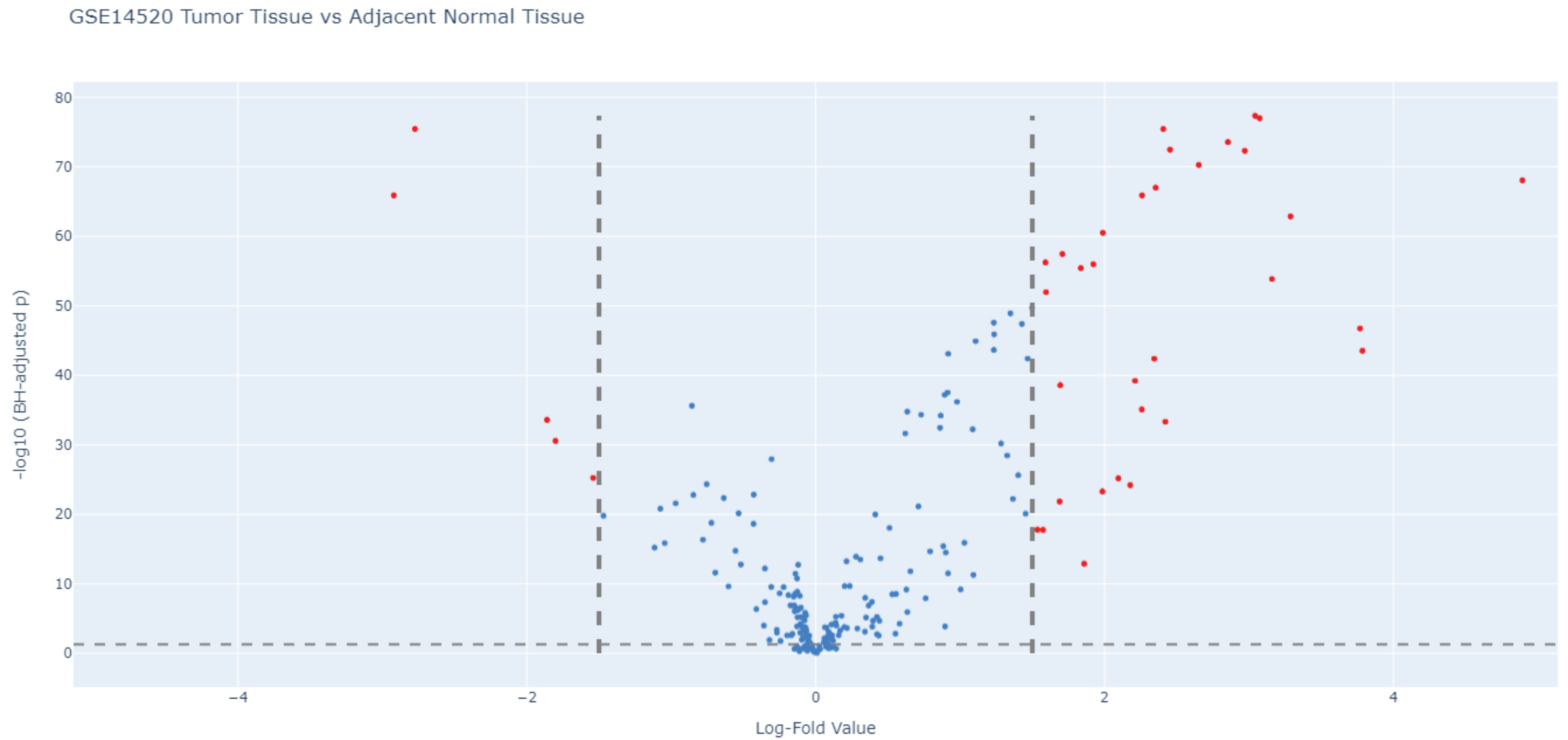
b



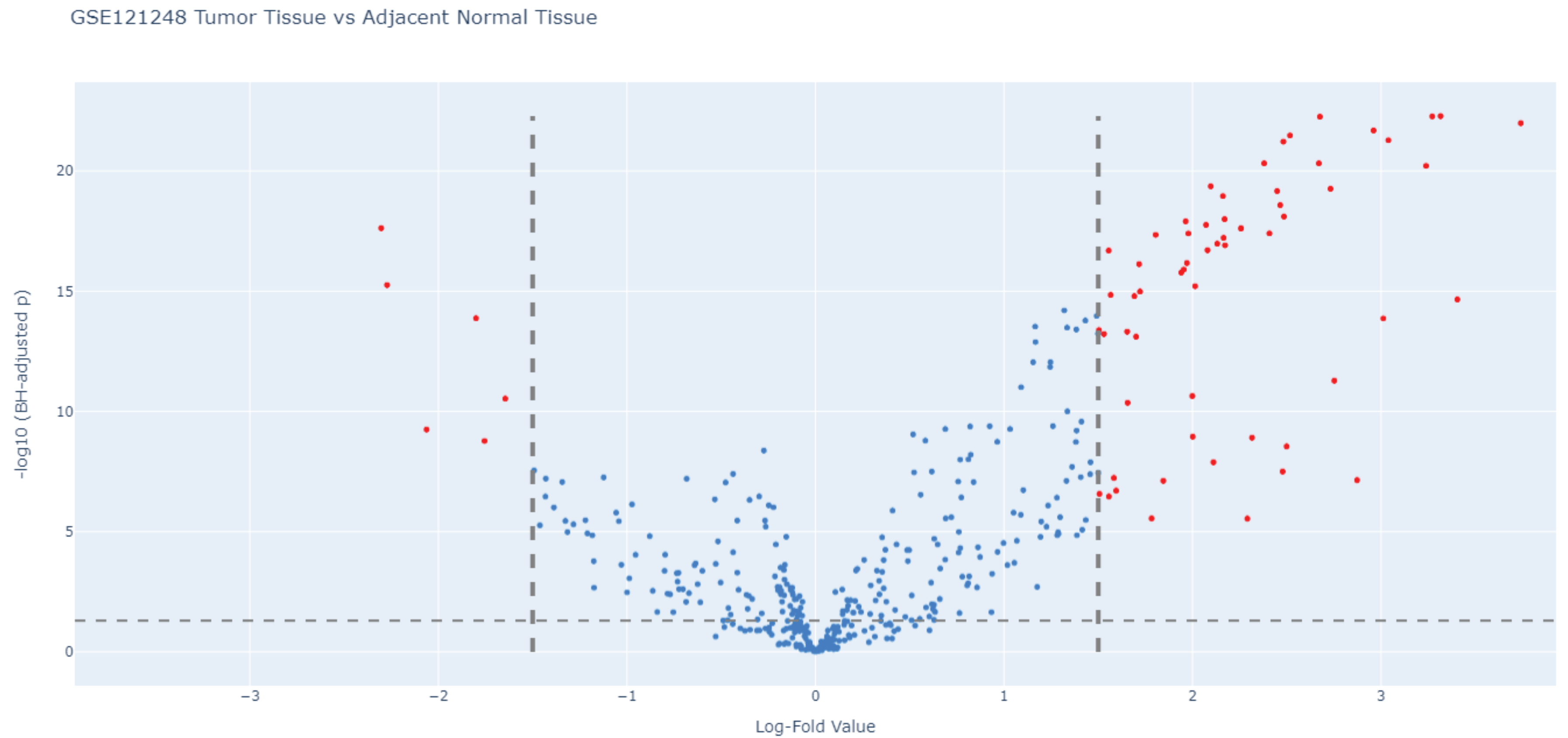
Supplementary figure 6: Enrichment plots of the curated liver cancer related gene sets in GSE83148 (a) and GSE84044 (b). PCI-U and PCI-D groups were compared.

PCI gene expression in HCC tumors vs. normal liver

a



b



Supplementary Figure 7: Volcano plot comparing HCC tumors vs normal liver for PCI genes. Unpaired and paired t-test were performed in GSE121248 and GSE14520 for each probeset, respectively. Each dot represents one probeset. Log fold changes and Benjamini Hochberg corrected p values are given in x and y axis, respectively.