

Introduction: My work was based on the GEO dataset GSE50737. This dataset was constructed for researching key genes associated with benzene hematotoxicity. Peripheral blood samples were randomly obtained from four patients of chronic benzene poisoning, three benzene-exposed workers and three health controls without benzene exposure. Total RNA was extracted and an integrated analysis of lncRNA and mRNA expression profiles was performed by GeneChip Human Gene 2.0 ST Array (*Bai W et al., 2014*)

Analysis and Results Discussion: After obtained GEO processed file and raw data from NCBI, I first find the interest variable “treatment:ch1”, which including different types of individuals: benzene exposed, benzene poisoned and control group. Then I extract these lists with geo accessions and performed rma for normalization. As there is only one variable, I just need to build a 2D matrix for comparison.

GSM1227407 - exprs

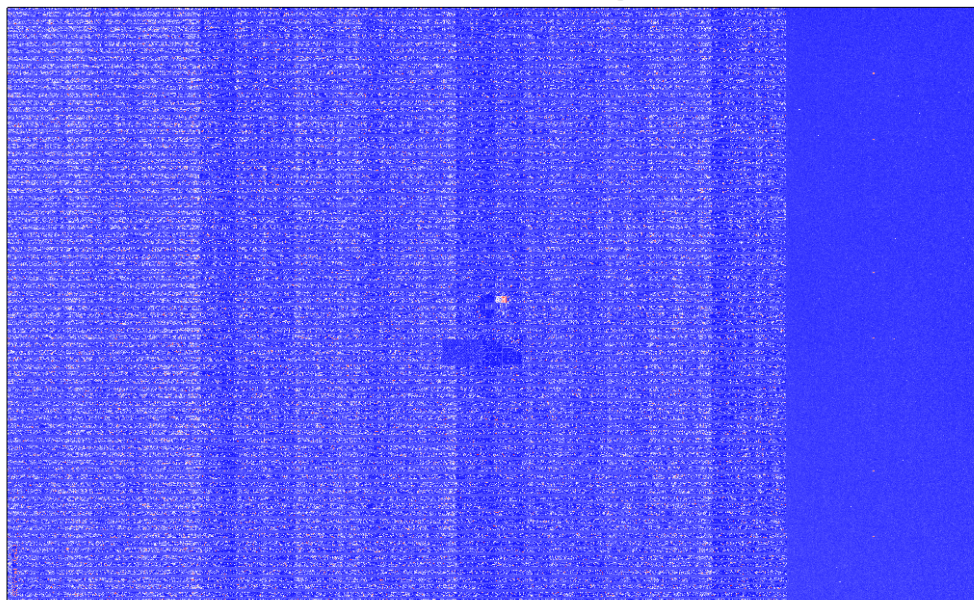


Fig 1. Chip intensities

Then I extracted differentially expressed genes, got interested genes and finally draw a heatmap showing differential expression. The heatmap shows the correlation of genes' expression level in different individuals. As we can see from the pic that there are many genes show significant relation between types and different individuals have different expression level. Many genes are up-regulated under benzene poison or exposure, indicates a possible stress response to benzene in gene expression in vivo.

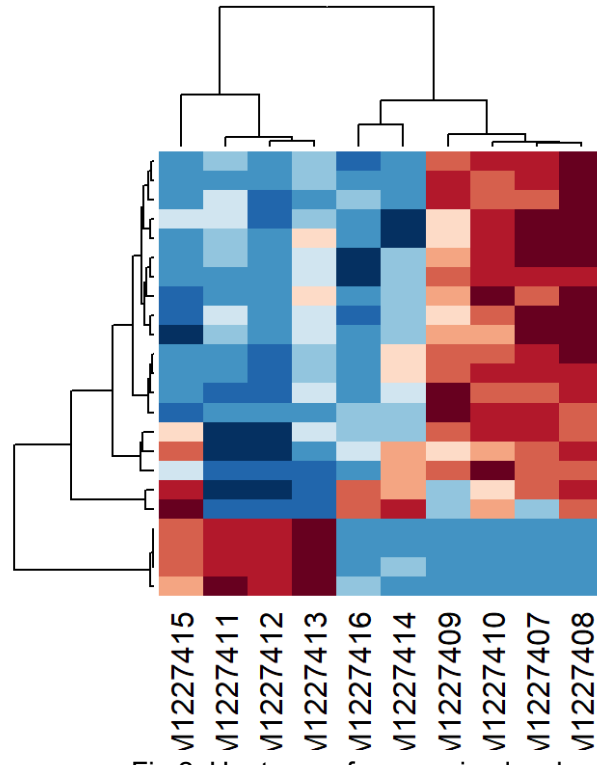


Fig 2. Heatmap of expression level