

Goal 1:

Disease Tissues

sample29
sample28
sample1
sample5
sample4
sample8
sample7
sample30
sample11
sample20
sample12
sample22
sample23
sample24
sample25
sample26
sample27

Healthy Tissues

sample19
sample18
sample17
sample21
sample3
sample10
sample2
sample15
sample9
sample16
sample13
sample14
sample6

m=17 samples are taken from disease tissues.

n=13 samples are taken from healthy tissues.

I upload the dataset to the site that you provided us as an example tool and I chose to cluster in the Kmeans method since there should be 2 clusters as healthy and disease. Then I see the class of samples and take extract them as my first goal.

Goal 2:

Top 10 Genes

('205725_at', 30437.0823529412),
('220542_s_at', 20045.491402715),
('204892_x_at', 18234.7262443439),
('203021_at', 10640.0538461538),
('210646_x_at', 8643.53619909497),
('206559_x_at', 8307.9723981901),
('213477_x_at', 6279.978280543),
('201257_x_at', 6041.47149321262),
('212869_x_at', 6000.143438914),
('201492_s_at', 5793.769683258)

Bottom 10 genes

('214076_at', 0.00407239819009),
('202354_s_at', 0.003619909502298),
('210570_x_at', 0.0031674208144),
('221210_s_at', 0.0027149321267),
('207650_x_at', 0.001809954751099),
('208832_at', 0.001809954750996),
('217443_at', 0.0013574660633),
('209754_s_at', 0.0013574660633),
('207599_at', 0.00090497737556),
('211483_x_at', 0.000904977375502)

First of all, I classify diseased tissues and healthy tissues in different excel files then I calculated the average expression values then I take them as healthy average expression value and diseased average expression values to find the difference between them, in excel again, then I write a python program to match the gene ids with their values as difference

average values, I kept the ids and values as two separate lists, then I merge them in a dictionary to sort easily without losing the ids of genes then I took the top ten genes at the top and the bottom ten genes.