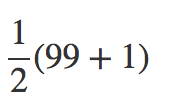
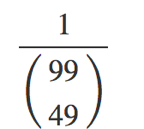
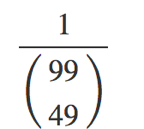
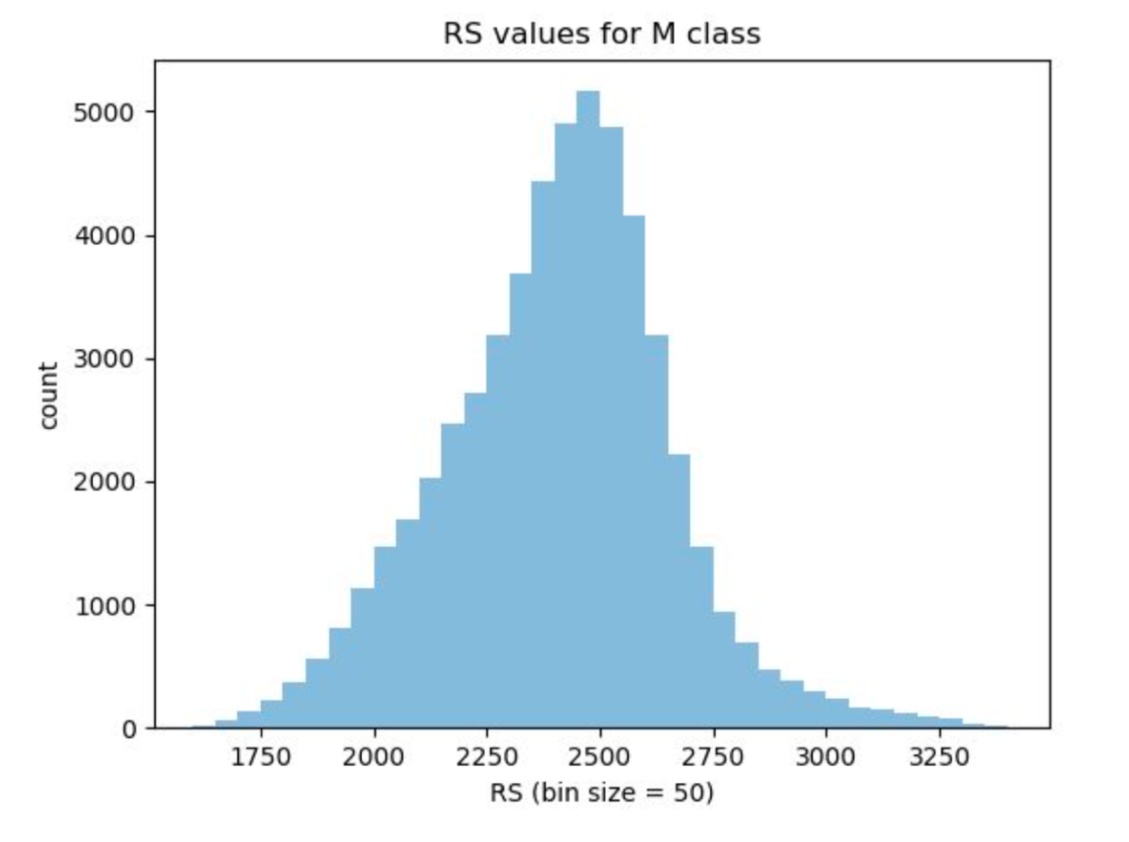
**DFE report**

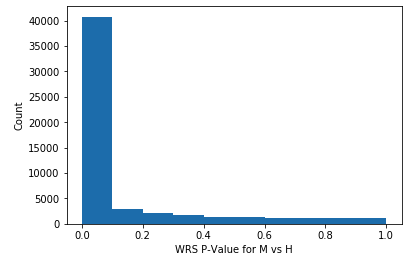
1. Describing the data:
   1. There are 54675 genes overall
   2. There are 99 patients in total
   3. The samples in each class are:
      1. 49 samples for M (acute myocardial infarction)
      2. 50 samples for H (healthy)
   4. After removing 47 genes we are left with 54628
2. WRS:
   1. Under the null model there is equal probability for getting + or – sample => p = 0.5. Which means that the value for M will be the average. 

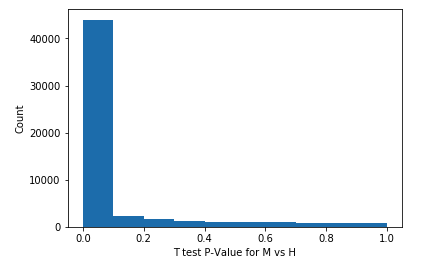
And the Total RS for M is 99\*50=2450.

* 1. The minimal value for RS(g) is when we will arrange all the M Samples in the lowest indices. Which means 1+2+…+49 = 1225.
  2. There is only one model which satisfies RS = m. Which means that the probability for getting the exact RS = m is like choosing one order from the whole ordering options. 
  3. Same as above there is only one ordering which can produce a situation where RS(g) = m+1. This ordering will be accepted when we will replace the last M rank with H, considering the RS(g) = m ordering as a baseline.
  4. For getting RS(g) = m+2 there are two ordering options:
     1. Switch between M48 and H50, taking the RS(g)=m ordering as a baseline.
     2. Switch between M49 and H50 and then switch between M48 and M49
     3. Taking these two ordering options to consideration we get P(RS(g)) = m+2 => 2\*

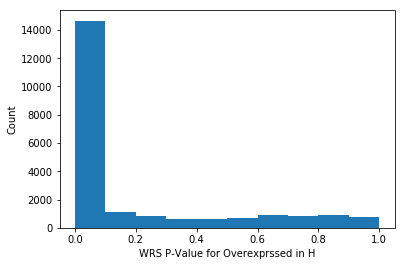


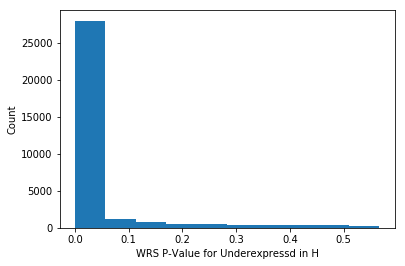
* 1. WRS – p values for M vs H:

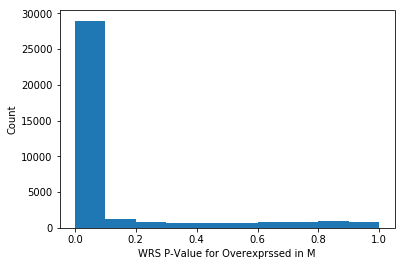


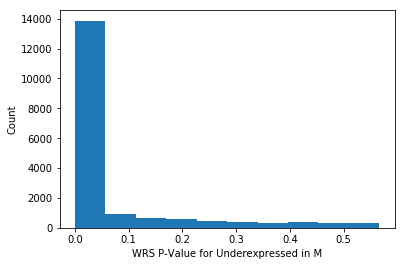


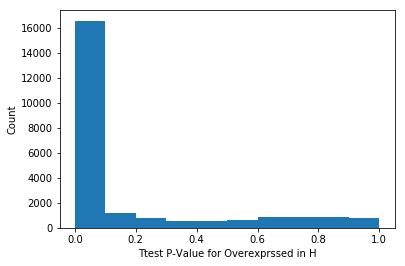
* 1. T test – p values for M vs H

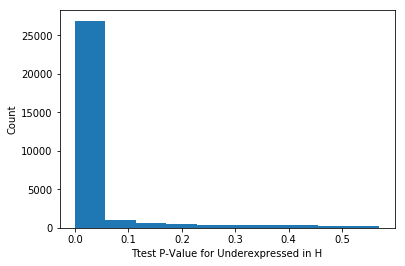


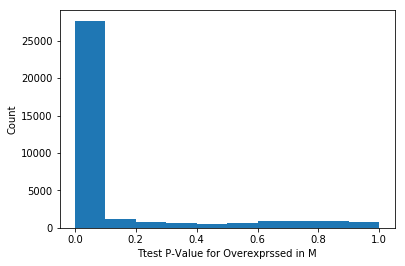


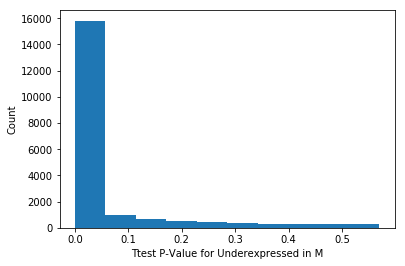




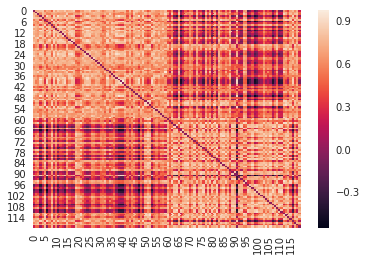






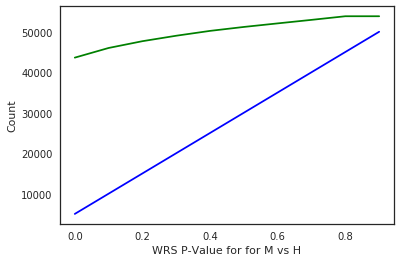


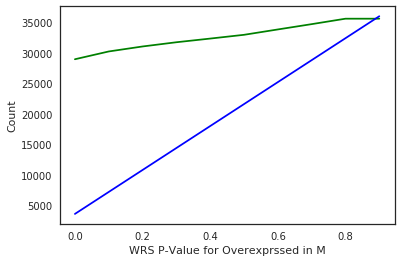
1. Correlations:
   1. Heatmap
      1. Heatmap attached.
      2. The heatmap describes correlation of 120X120 genes.
      3. The first 60 are overexpressed in class m.
      4. The last 60 are underexpressed in class m.
      5. From the heatmap is can be seen that there is a good correlation between genes which are overexpressed in class and it is the same for those which are underexpressed in the same class.

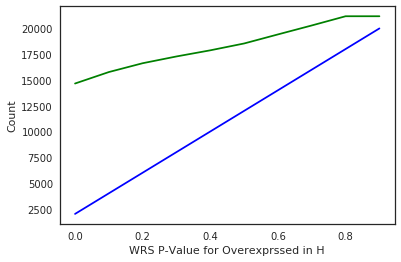


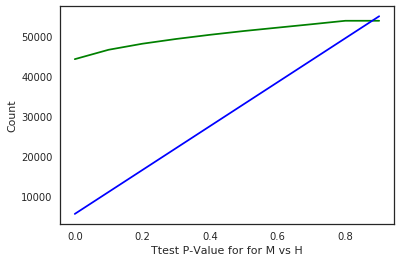
* 1. Calculating the whole data:
     1. The main advantage for working with the whole data set is zero data loss.
     2. when looking at partial set of the data, we can get an idea for how the whole data will act but it is still assumption.
     3. The ones labeled as H is in the right down corner, it can be seen that there is a good correlation grade for those genes.

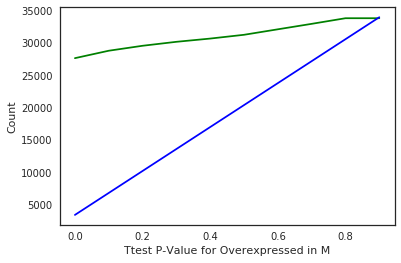
1. Plots and conclusions:

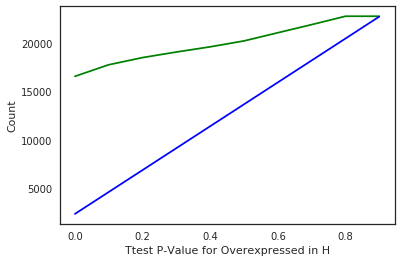




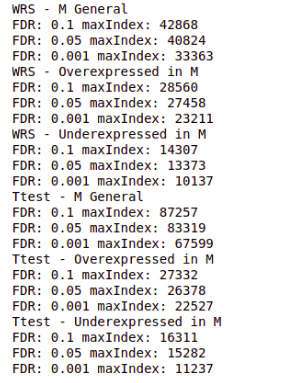








* 1. FDR:



* 1. It is shown that there are lower pvalues in WRS than Ttest which causes larger FDR result for WRS.
  2. 3 genes were selected:
     1. first two are overexpressed in M
     2. last one is underexpressed in M
     3. heatmap is attached:
     4. It can be seen that there is a good correlation for the overxpressed genes in the left side and it’s get less good as long as we are approaching the left side.
     5. for the underexpressed genes it is exactly the opposite.
  3. 