ABSTRACT

Total RNA was collected from human cancer associated fibroblasts expressing the indicated shCtrl or shNNMT constructs or from normal WI-38 fibroblasts expressing Ctrl or NNMT constructs and expression analyzed with Affymetrix microarray.

OBJECTIVES

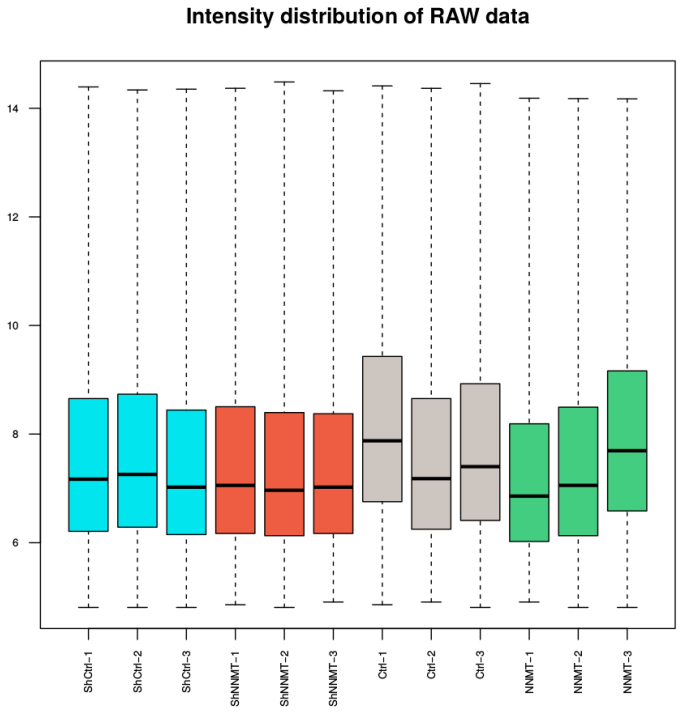
This analysis was performed in order to assess the differential gene expression in human CAFs with control (shCtrl) or shRNA against NNMT (shNNMT) expression or normal fibroblasts expressing control (Ctrl) or NNMT overexpression (NNMT) constructs.

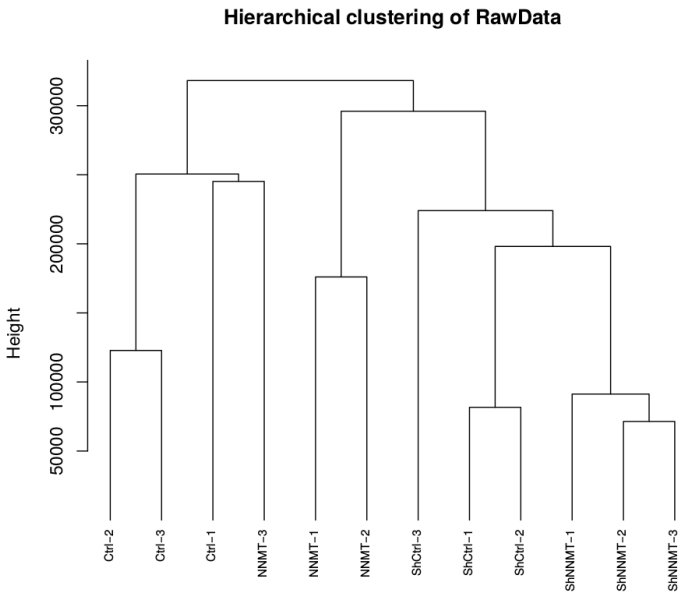
METHODS

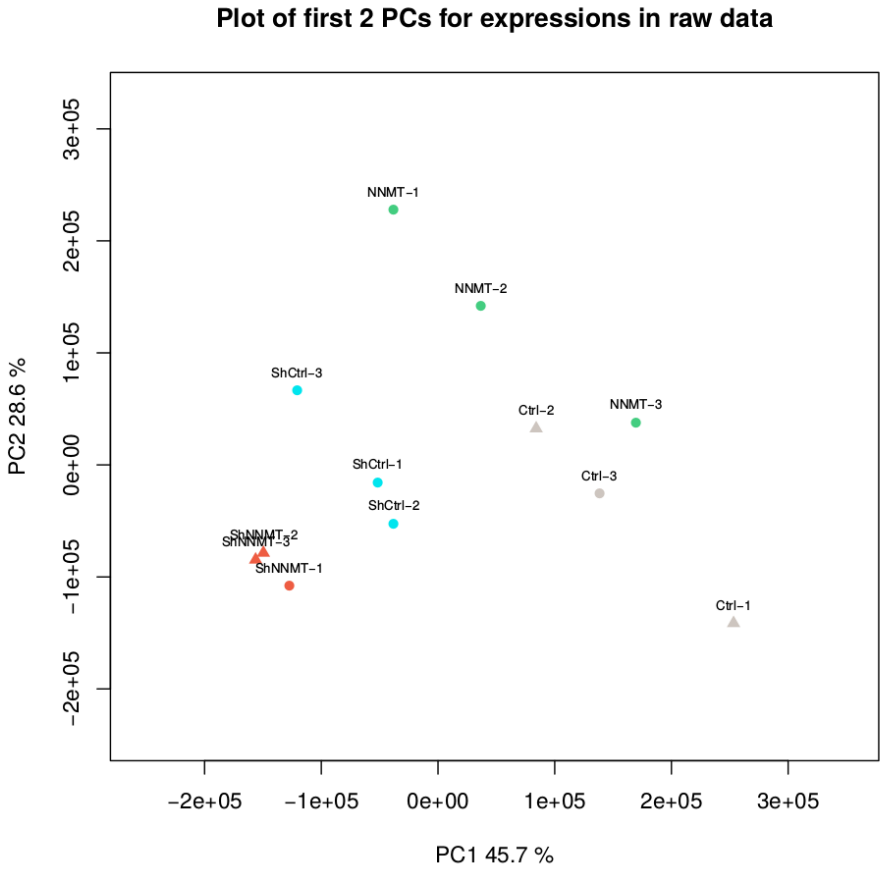
1. Boxplots for variance and mean analysis among groups.
2. PCA for data validation and sampling distribution.
3. Quality analysis with arrayQualityMetrics in order to look for reproduciblity, clean the data and erradicate noise.
4. Data normalization and filtering by means of Multi-Array Average Expression and nsFilter.
5. Repeating of methods 1,2 and 3 in order to detect new generated errors.
6. Fitting the linear model. By fixing 1 variable we can join the shCrtl, shNNMT, Ctrl, NNMT overexpression and control groups. We obtained a contrast matrix.
7. Genes deferentially expressed obtained by creating a top table with the best results of each table.
8. By means of a Volcano plot, visualize deferentially expressed genes.
9. Annotation of the differentially expressed genes.

RESULTS AND DISCUSSION

After a first analysis we can see that the data is pretty good evenly distributed without a major sign for artifacts. PCA validation backs up this first analysis.

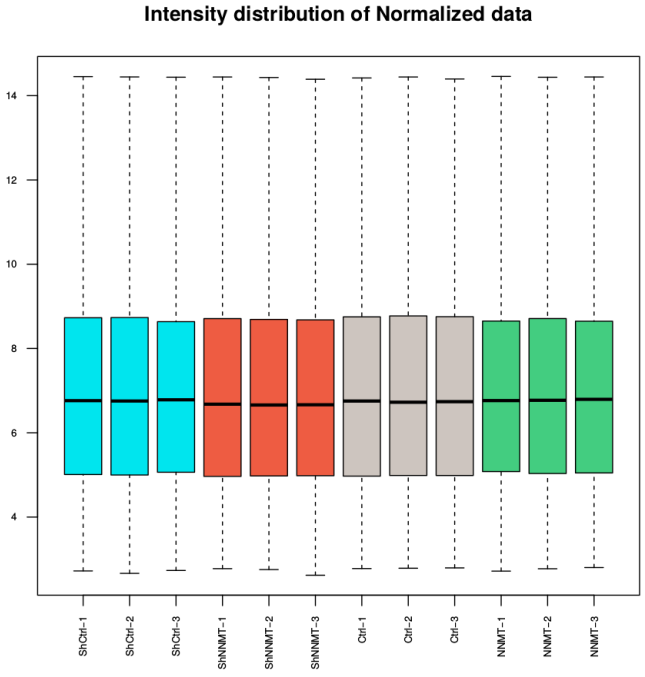


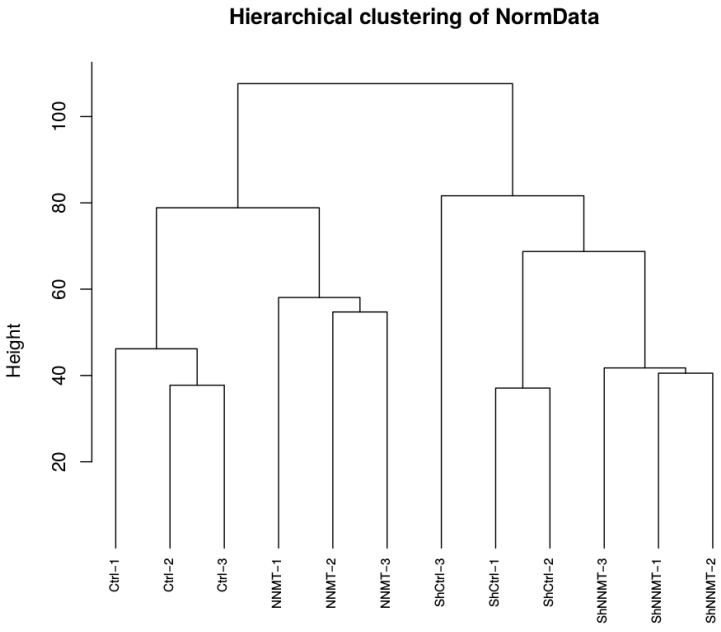


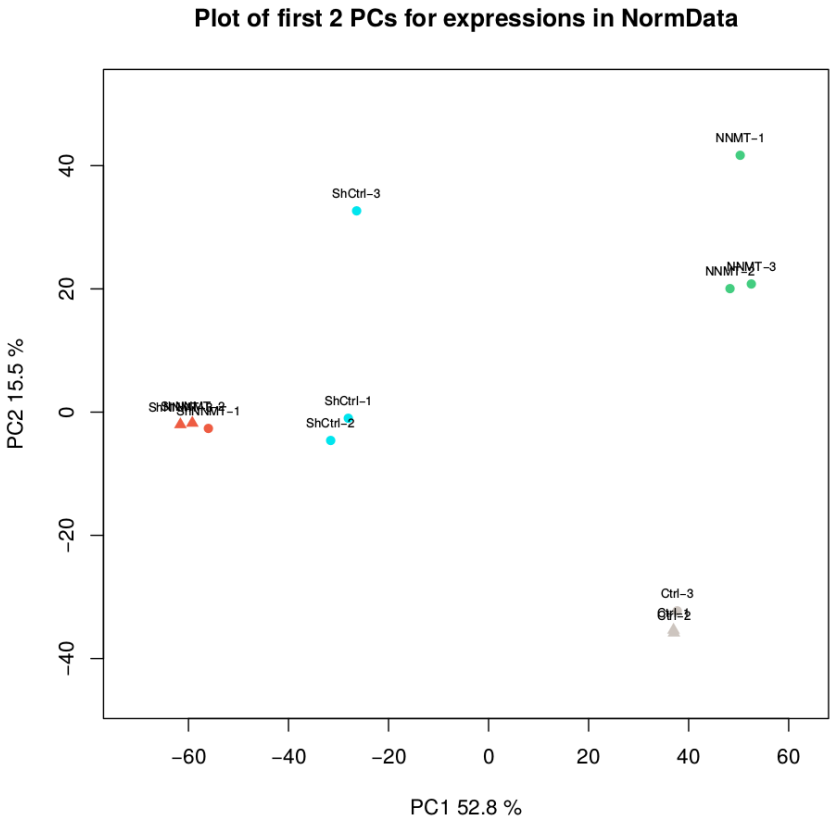


PCA shows that some groups are really clustered like SHNNMT and others are pretty distinct like Ctrl or NNMT.

Then we proceeded to normalize the data. After that, we see that now it presents almost no differences and is homogeneously distributed. This way we got rid of any possible artifact not seen in the initial data.



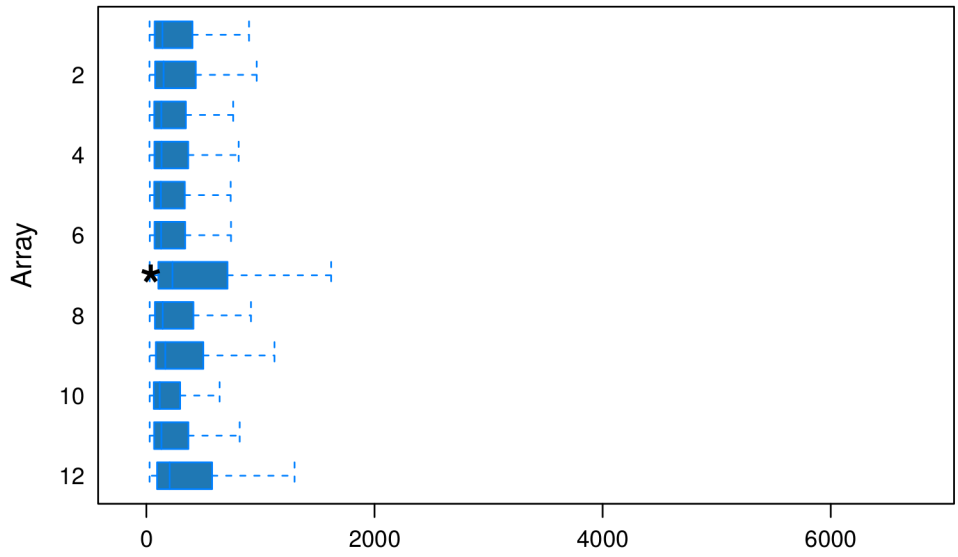


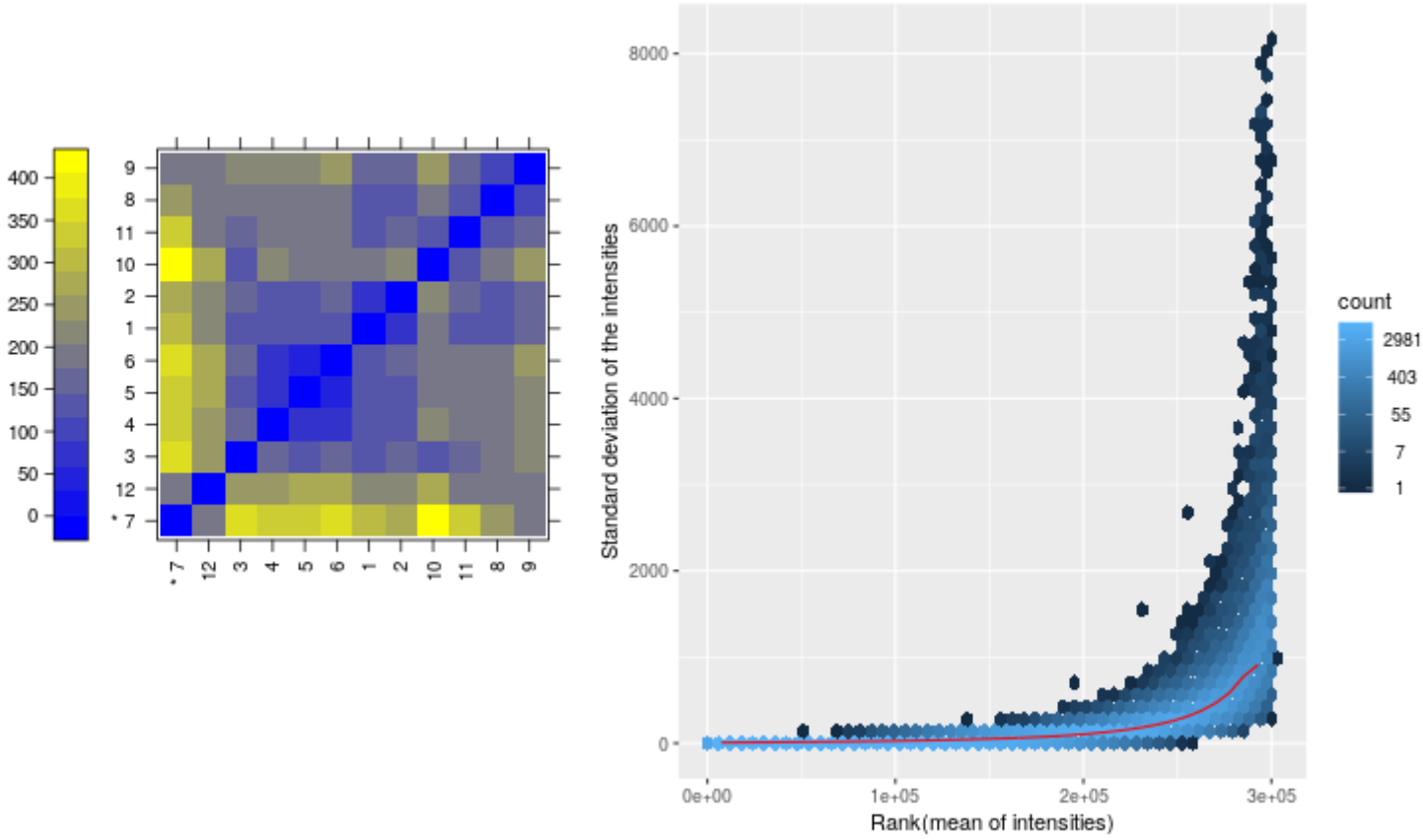


Now clusters are formed by different samples of the same observations, which means that the difference between the samples is not big, but between the groups is bigger.

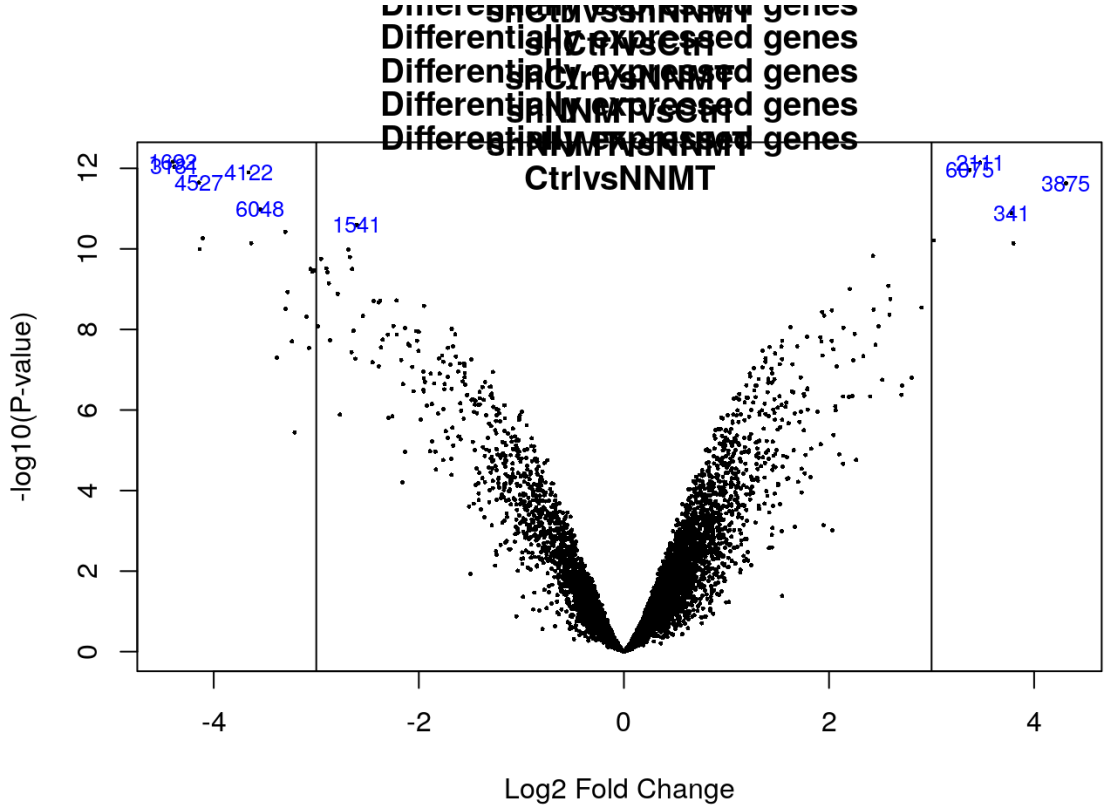
Samples of the same experiment are together which means that the expression of genes between the types of experiments which is good.

With the array quality matrices we checked the correctness of the data. Under this filter 20341 genes were discarded and 6780 remained in the dataset for subsequent use.





After the quality check, by means of the design and contrast matrices we fitted a linear model. The model was used to create a table with the best genes. Then a volcano plot was used to visualize the data.



Some of the most relevant annotations obtained were:

"ADAM metallopeptidase with thrombospondin type 1 motif 1"
"ADP ribosylation factor like GTPase 1"
"ATP binding cassette subfamily A member 1"
"bromodomain adjacent to zinc finger domain 1A"
"cytochrome c oxidase assembly factor 3"
"ubiquitin specific peptidase 2"
"zinc finger and BTB domain containing 25"