BIOINFORMATICS - Course 2019/ 2019

OMICS DATA ANALYSIS

EXERCISE ON MICROARRAY DATA ANALYSIS **GSE113664**

Alejandro Sanchez Pla

Made by:
Ana Sanchez

Contents

1 Abstract

In this study, it has been examined the involvement of UHRF1, a major regulator of epigenetic mechanism, in abnormal DNA methylation in colorectal cancer (CRC).

In CRC cells, the temporary UHRF1 augmeny induced DNA demethylation across entire genomic regions, including CpG islands, gene bodies and repetitive elements.
UHRF1 reduction make a minimal effect on gene silencing. But the combination of the recuction of the UHRF1 and the inhibition of histone deacetylase (HDAC) causes a strong supression on CRC cell proliferantion by silencing the genes.

2 Objectives

In this experiment we want to know the effect of UHRF1 in CRC cells.

3 Material and methods

3.1 Data type, type of experiment, experimental design, type of —microarrays— used...

Data type that we have used is about Targets and GSE data.
It is an non-experimental experiment and it's done this way:
First we collect the data and we make plots about this row data, then, we normalize it and obtain other plots.
We want to compare the the different the raw and the normalized data and we can do so by comparing the different plots.

3.2 Methods that have been used in the analysis

3.2.1 General procedure of analysis (steps, "workflow" or "pipeline" that you

The steps that I have followed for this experiment are:

1 Installation of packages needed.

2 Load Data.

3 Quality control of arrays: Raw Data.

4 Boxplot.

5 Hierarchical clustering.

6 Principal component analysis

7 Data Normalization

8 Boxplot.

9 Hierarchical clustering.

10 Principal component analysis.

3.2.2 Targets table used

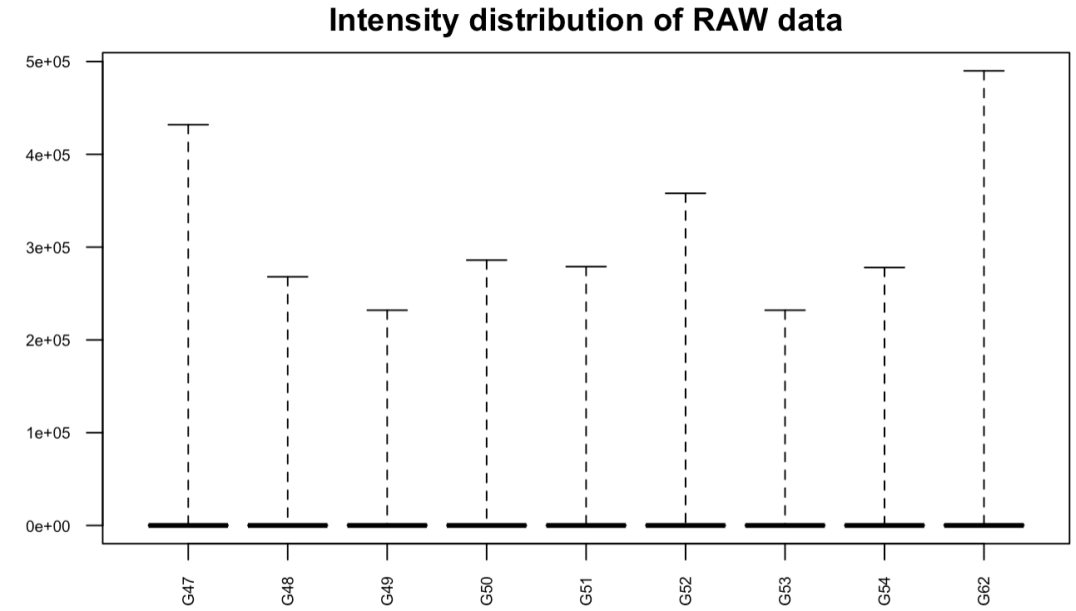
Identifier	Group	Color	ShortName	Type	Drug
GSM3110847	RKO_siCONT_r	Red	G47	control	mock
GSM3110848	RKO_siCONT_r	Red	G48	control	mock
GSM3110849	RKO_siUHRF1	Blue	G49	UHRF1	mock
GSM3110850	RKO_siUHRF1	Blue	G50	UHRF1	mock
GSM3110851	RKO_siCONT_pr	Green	G51	control	Trichostatin_A
GSM3110852	RKO_siCONT_pr	Green	G52	control	Trichostatin_A
GSM3110853	RKO_siCONT_pt	Yellow	G53	UHRF1	Trichostatin_A
GSM3110854	RKO_siCONT_pt	Yellow	G54	UHRF1	Trichostatin_A
GSM3110862	RKO_w_AZA	Black	G62	None	5aza2deoxycytidine

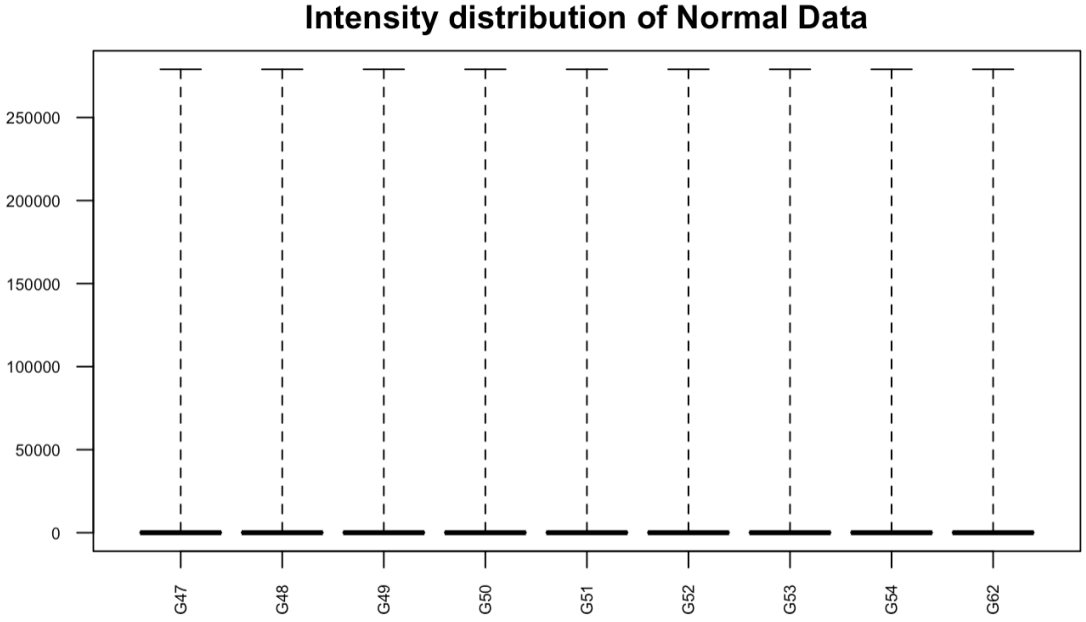
Table 1: 2 minutes for each example to get our α

In order to classify the different data, I have taken into account different classifications related to:
Group: I have taken into account that each Identifier has, there are 5 different groups: RKO_siCONT_r, RKO_siUHRF1, RKO_siCONT_pr, RKO_siCONT_pt and RKO_w_AZA.
Color: I have assigned different colors depending on the group each one pertains.
ShortName: I have assigned a short name to each one. The different short names I have assigned are : G47,G48,G49,G50,G51,G52,G53,G54,G62.
Type: By looking to the GEO i have assigned different types to each entry.
Drug By looking to the GEO accession number I have assigned also the different drugs used.

3.3 Plots

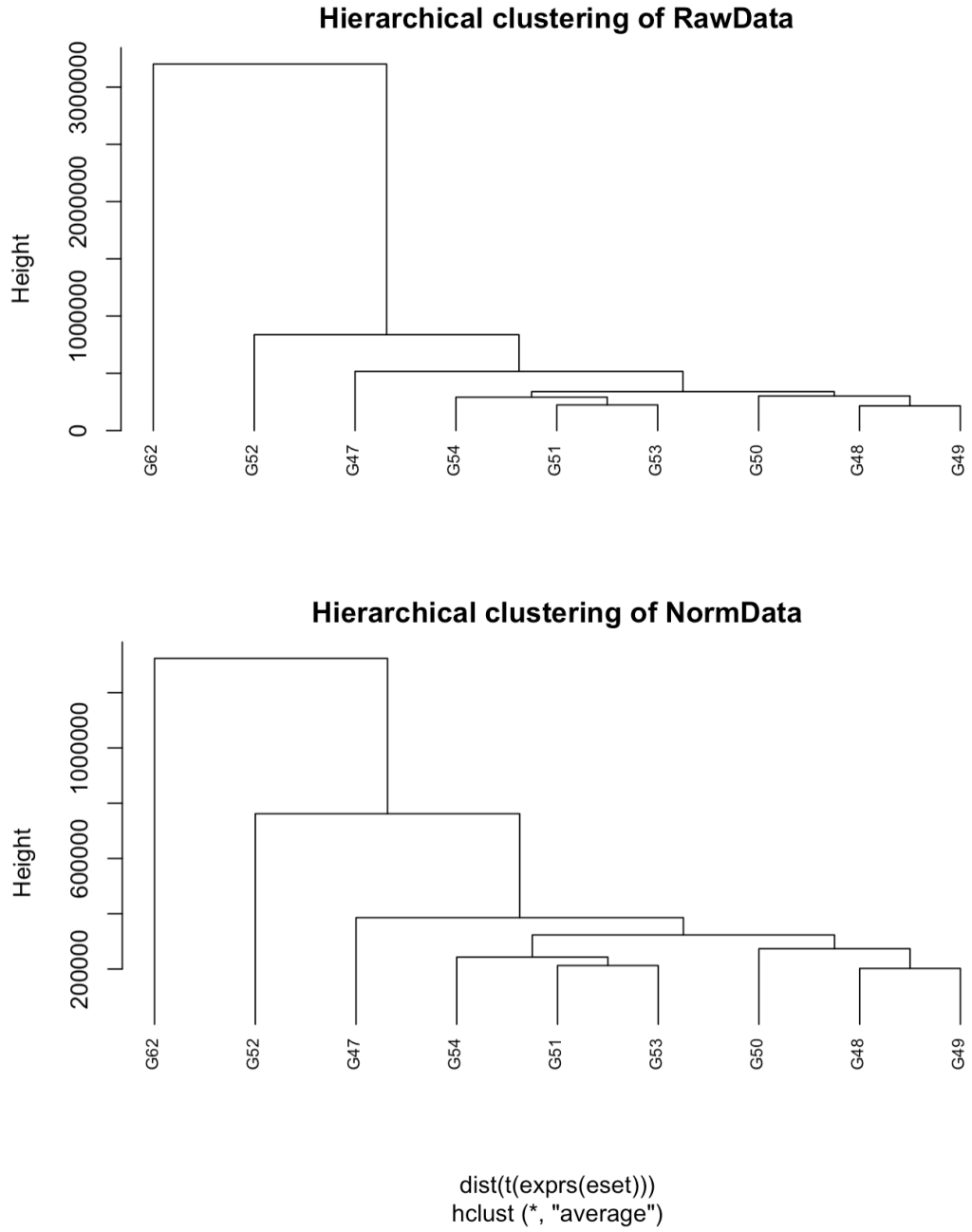
3.3.1 Plot of intensity distribution





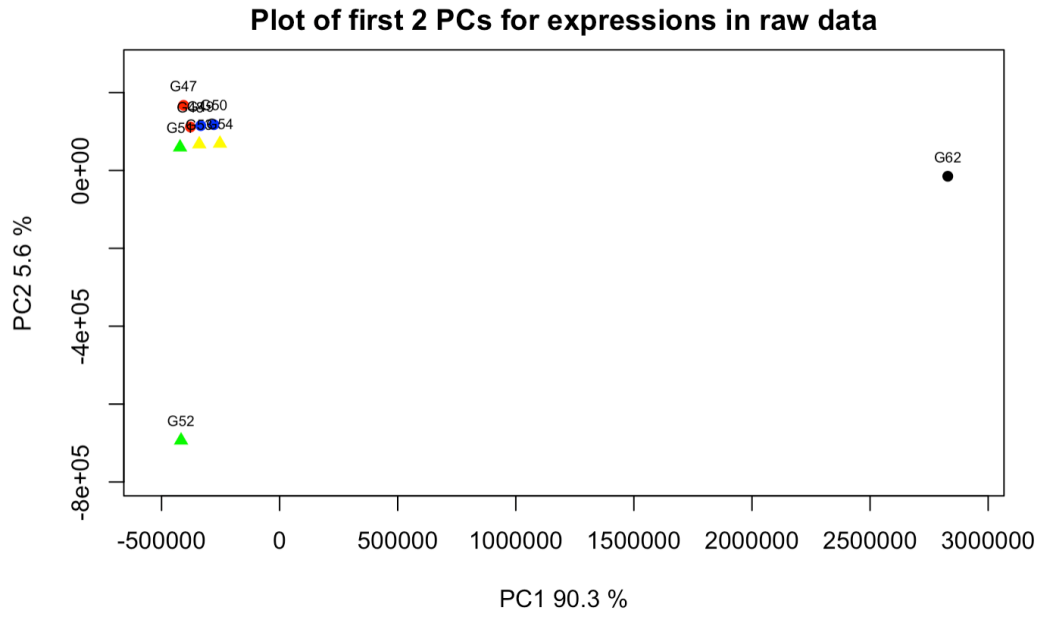
By comparing this two plots we can see that the y axis has higher values on the Normal data than on the Raw Data.

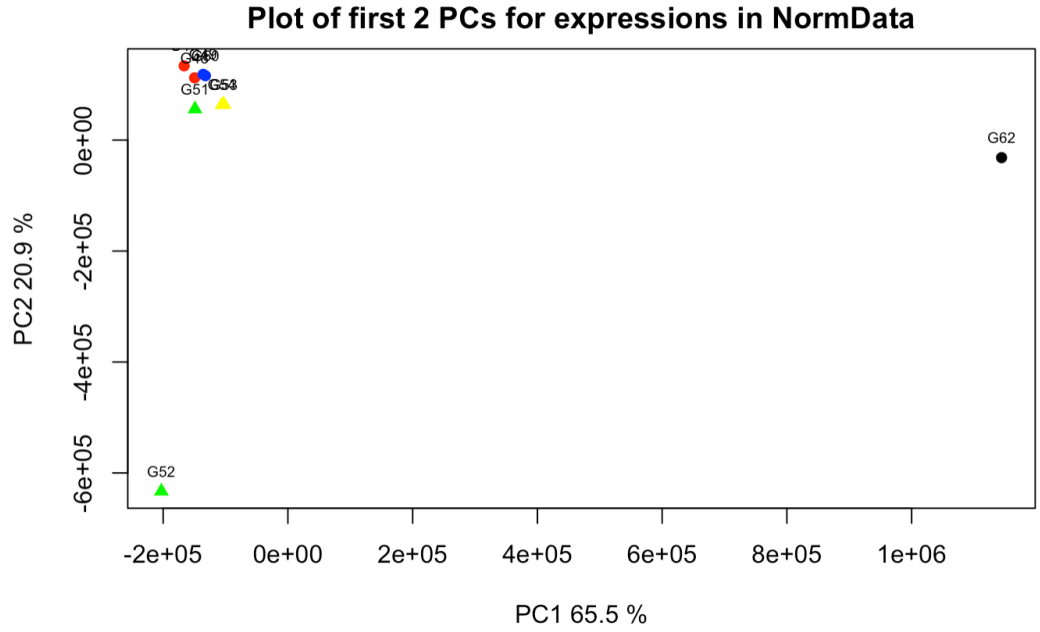
3.3.2 Plot of Hierarchical clustering



This two plots about Hierarchical clustering are very similar but the one that is about the Raw Data has bigger height than the one from the Normal data.

3.3.3 Plot of PCA expressions





This two plots about PCA and the difference that we can observe is that the % in PC1 of the Normal data is lower than the one from the Raw Data.