RhoE isoform RNA-Seq Data analysis

Wencai Ma, Jing Wang

3/25/2019

1. Executive Summary

1.1 Introduction

This analysis is for Dr. Yuan Dai from Dr. Jiang Chang's lab.

The RNA seq data is for 2 cell lines, each having 4 samples. In each cell line, there are wild-typ control, KO control, KO+RhoE treatment, KO+RhoEa treatment.

Experimental design:

	Sample #	Experimental design
Cell line A	S1	Wild-type
	S2	KO
	S3	KO + RhoE expression
	S4	KO + RhoEα expression
Cell line B	S5	Wild-type
	S6	KO
	S7	KO + RhoE expression
	S8	KO + RhoEα expression

Comparisons:

Control group	Experimental group
S1	S2
S2	S3, S4 (Compare 3 groups together)
S5	S6
S6	S7, S8 (Compare 3 groups together)

sample info

The first comparison is to compare wild-typ control with KO control.

The second comparison is to compare among KO control, KO+RhoE treatment and KO+RhoEa treatment.

RNA-Seq fastq files were provided by other company. The files are located in following directorie (password required):

ftp://P202SC18110008-01_20190220_QDJwxk@hwftp.novogene.com/C202SC18110008 (ftp://P202SC18110008-01_20190220_QDJwxk@hwftp.novogene.com/C202SC18110008)

Fastqc software was used to check the quality of the fastq sequencing files.

We aligned the reads to hg38 using Tophat, producing intial BAM files.

We ran the bam files through samtools to sort the reads for easier access and indexing, producing revised BAM files.

We used HTSeq to count the numbers of reads mapping to each of the 56269 genes/locations, producing a quantification matrix.

We examined the matrix to see how many genes are receiving zero or few reads for each sample, to see if any samples appear grossly different.

We used box plot, hierarchical clustering and PCA plot to view the data for the purpose of quality control.

Because for each condition, there is only one sample. For comparisons with only two samples, only fold change was calculated to find differentially expressed genes. Fold change is defined as the ratio between the two quantities; for quantities A and B, then the fold change of B with respect to A is B/A. Fold change is not a statistical method that can generate p values. A fold change threshold is chosen as the cutoff the define differentially expressed genes. We usually use 1.5, 2 or 3 as thresholds.

IPA was used to analyze the differentially expressed gene with fold change >=3.

1.2 Results

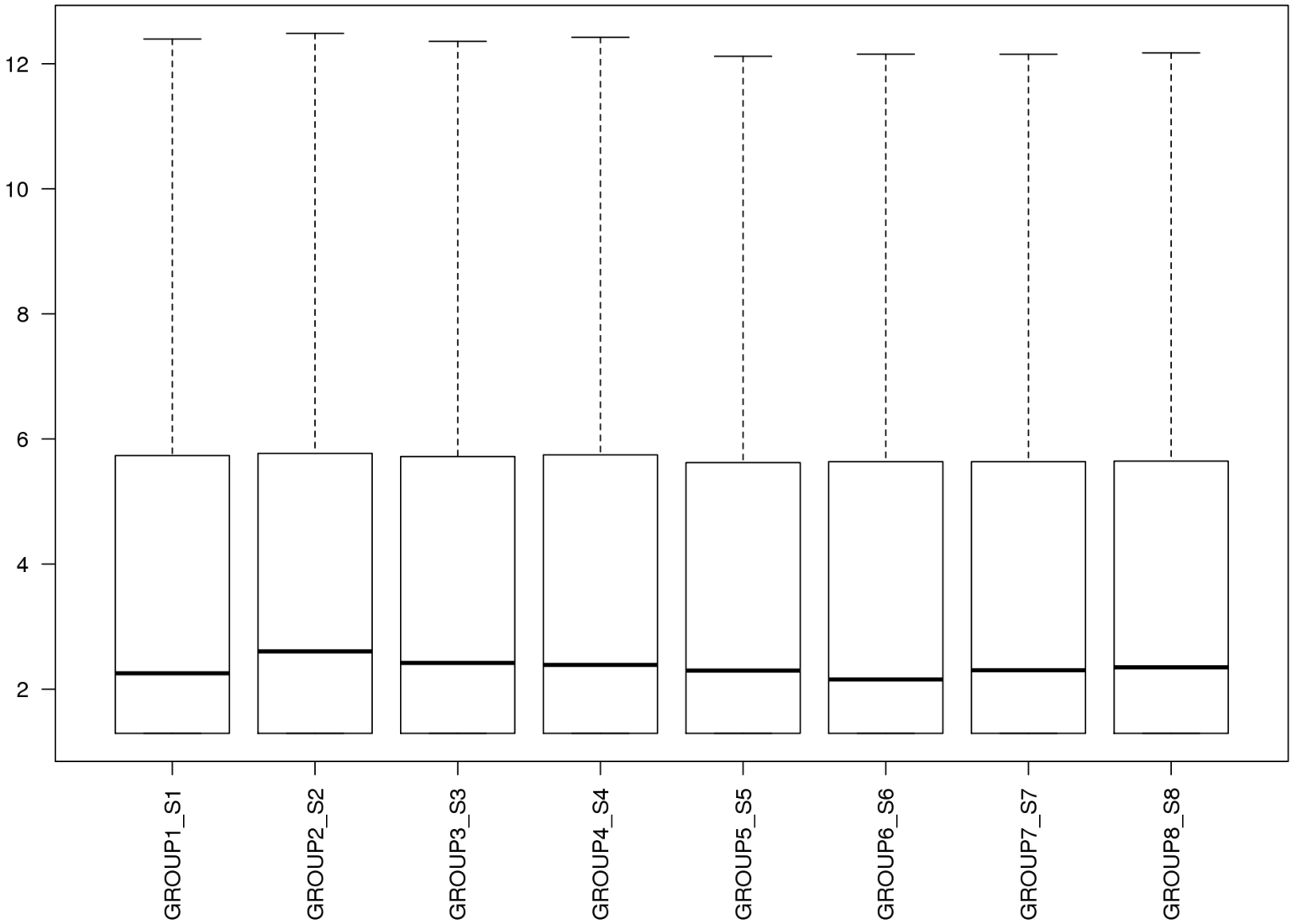
Fold changes for all genes in all required comparisons have been calculated.

IPA has been done for all required comparisons.

2. Analysis

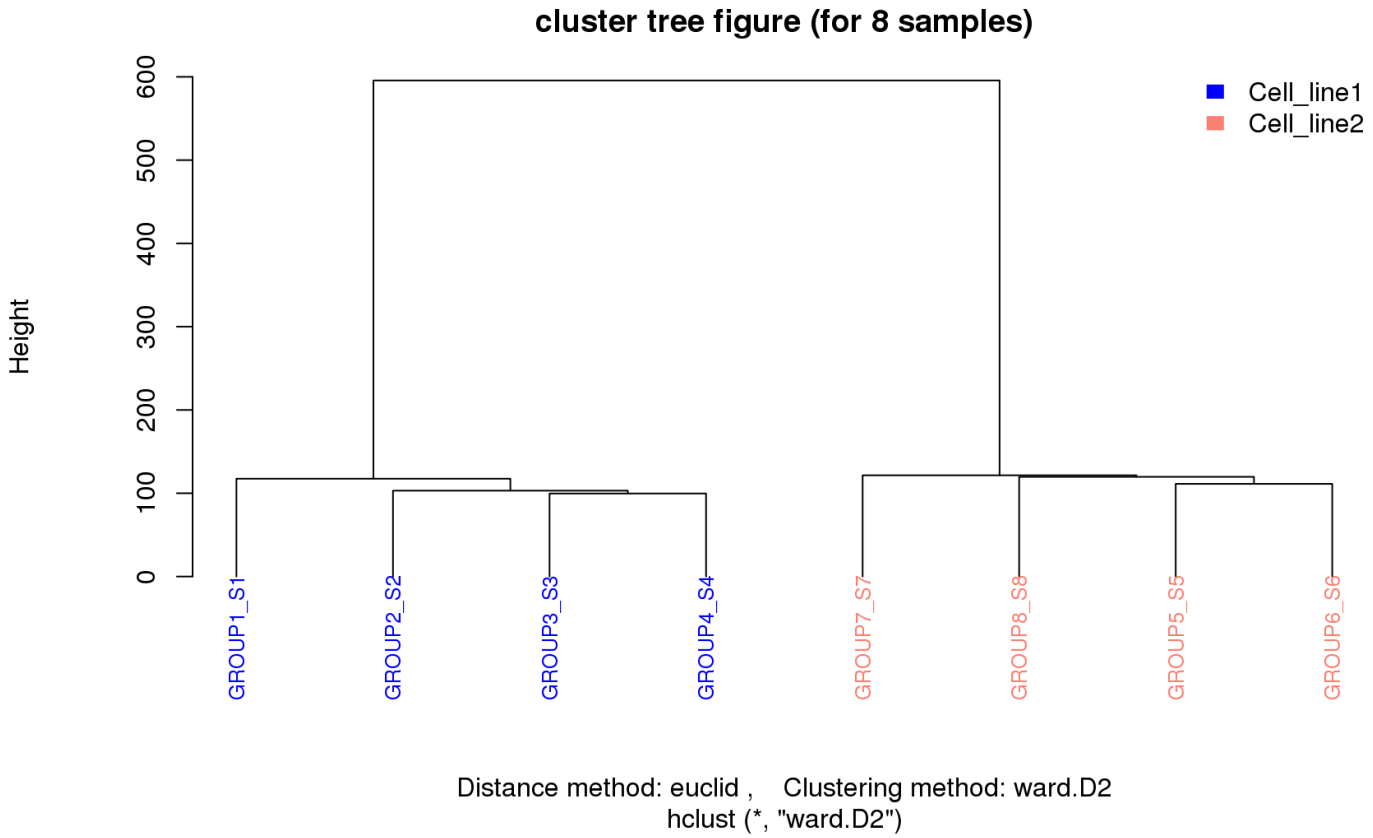
2.1 Process the RNA seq data and data quality control

```
## # Only the Variance Stabilizing Transformation was done, the result can be considered to be really normalized and be used for furhter analysis,such as correlation analysis, heatmap et al.
```



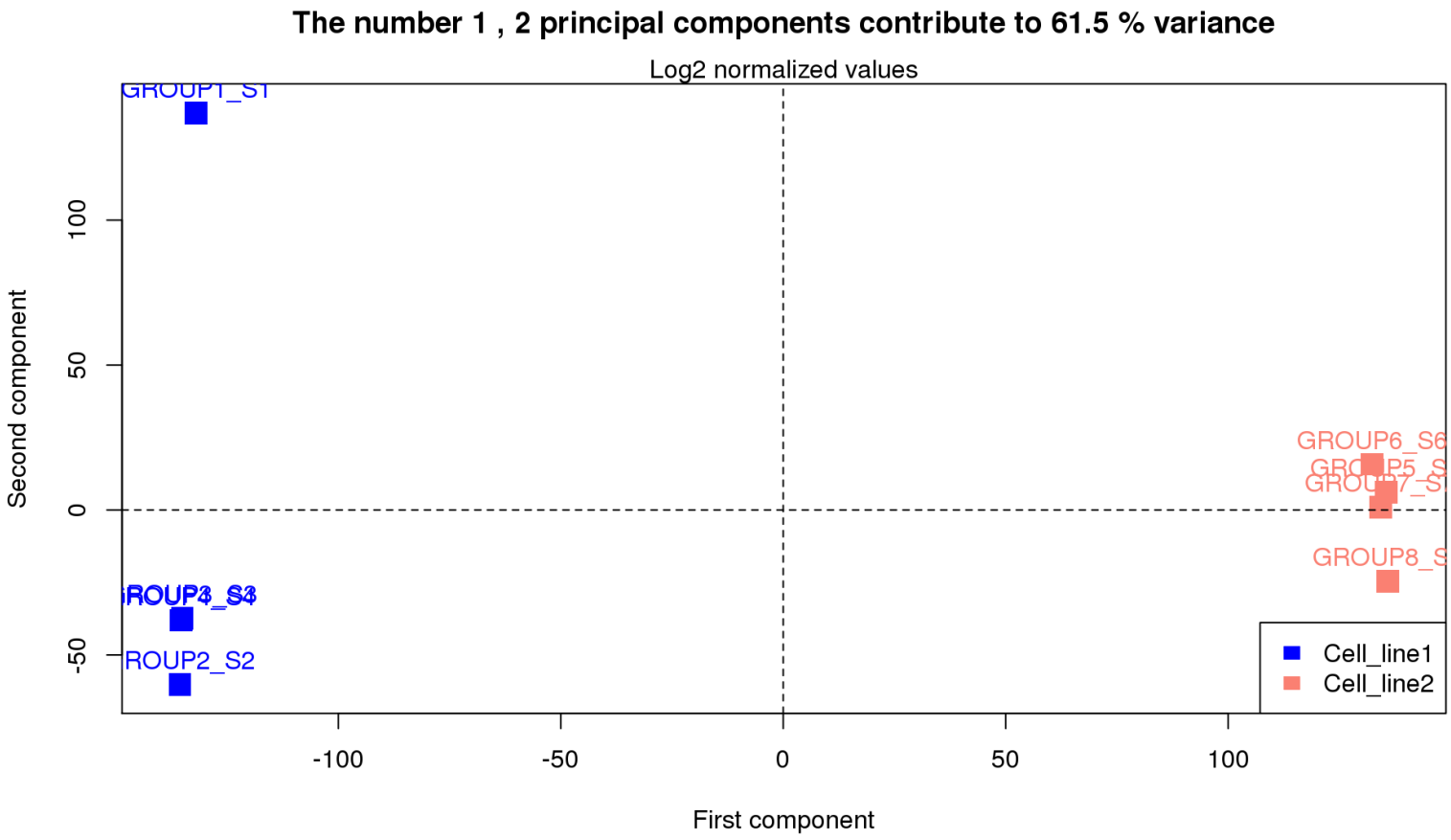
2.2 Data quality control using hierarchical cluster

```
## ## The whole data was used to cluster the samples.
```



2.3 Data quality control using PCA plot

```
## ## We used principal components (# 1 2 ) to examine the sample groupings.
## ## The number 1 2 principal components contribute to 61.5 % of whole Variance.
```



2.4 For S2 vs S1. Differential expression comparison using fold change. Making file for IPA.

```
## We kept 781 genes with fold change threshold>= 3 for IPA analysis.
```

2.5 For S6 vs S5. Differential expression comparison using fold change. Making file for IPA.

```
## We kept 903 genes with fold change threshold>= 3 for IPA analysis.
```

2.6 For S3 vs S2. Differential expression comparison using fold change. Making file for IPA.

```
## We kept 360 genes with fold change threshold>= 3 for IPA analysis.
```

2.7 For S4 vs S2. Differential expression comparison using fold change. Making file for IPA.

```
## We kept 335 genes with fold change threshold>= 3 for IPA analysis.
```

2.8 For S4 vs S3. Differential expression comparison using fold change. Making file for IPA.

```
## We kept 311 genes with fold change threshold>= 3 for IPA analysis.
```

2.9 For S7 vs S6. Differential expression comparison using fold change. Making file for IPA.

```
## We kept 1012 genes with fold change threshold>= 3 for IPA analysis.
```

2.10 For S8 vs S6. Differential expression comparison using fold change. Making file for IPA.

```
## We kept 968 genes with fold change threshold>= 3 for IPA analysis.
```

2.11 For S8 vs S7. Differential expression comparison using fold change. Making file for IPA.

```
## We kept 1328 genes with fold change threshold>= 3 for IPA analysis.
```