

Enrichment analysis with clusterProfiler

Rene Welch

February, 2018

Contents

1	Introduction	2
2	Preparing the data	2
2.1	Hallmark gene list V5.1	2
2.2	Diff expressed genes	2
3	Pathway analysis	3
3.1	GSEA analysis	3
3.2	Over-representation test	5
4	Bibliography	7

1 Introduction

We are going to perform a pathway analysis, I am using the data from Figure 3, i.e. genes that were differentially expressed in specific EBV or NOKS cell lines after applying the MC treatment:

```
## [1] "alignment"      "diff_genes_EBV"  "diff_genes_NOKS" "fdr"
## [5] "rlogmat"        "rsem_data"
```

Mark already analyzed the data using the TPM matrix and GSEA, so I am going to use *clusterProfiler* as an alternative method using the data already processed by *DESeq2*.

As far as I can tell, the advantages of this package is that:

1. It can perform pathway analysis using the Hallmark gene data set from **MolSigDB**
2. It contains several visualization methods, including some based on GSEA.
3. This may be minor, but for ChIP-seq annotation we have used *ChIPseeker* which was created by the same author.

2 Preparing the data

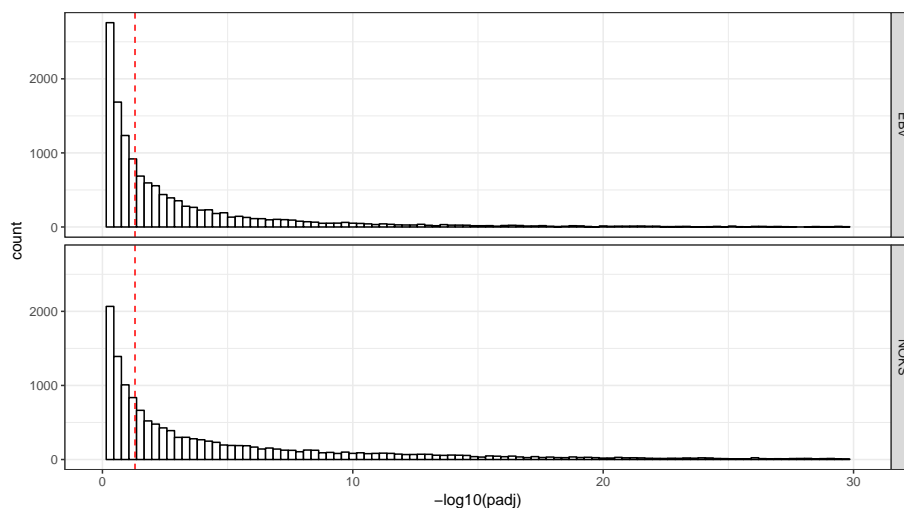
2.1 Hallmark gene list V5.1

We downloaded the **MolSigDB** curated gene sets from [the GSEA website](#), and then we load the hallmark gene set from **MolSigDB**, using the 5.1 version:

```
## [1] "h.all.v5.1.entrez.gmt" "h.all.v5.1.orig.gmt"
## [3] "h.all.v5.1.symbols.gmt"
```

2.2 Diff expressed genes

We define a gene to be diff. expressed if the adjusted p.value is ≤ 0.05



3 Pathway analysis

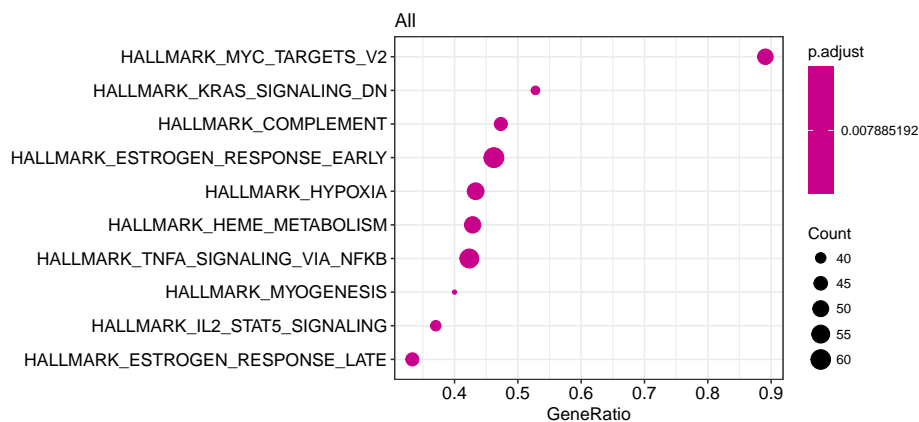
This package appears to be contains several methods:

- An R made version of GSEA, we are using the GSEA model for a lists of genes with a signal-to-noise measure for each diff. expressed gene (which in this case is calculated by [DESeq2](#)).
- An over-representation test, which is a hyper-geometric test of the gene lists vs the genes in a certain pathway.

3.1 GSEA analysis

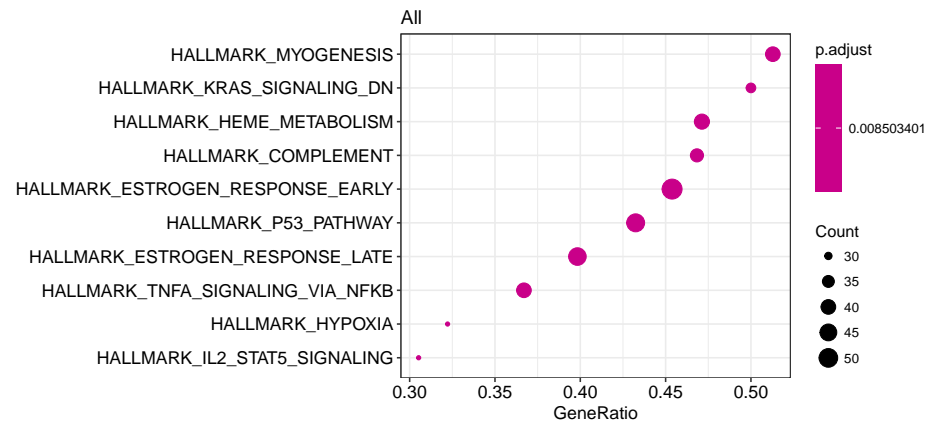
3.1.1 Diff. expressed genes

- NOKS



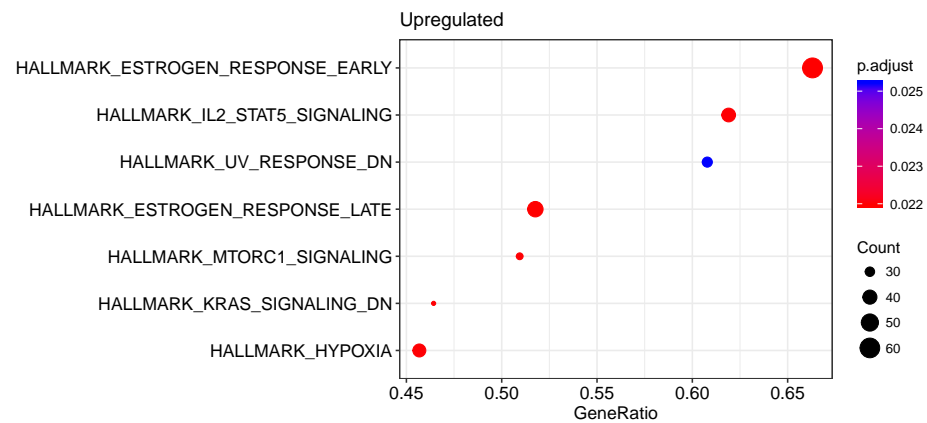
- EBV

Enrichment analysis with clusterProfiler

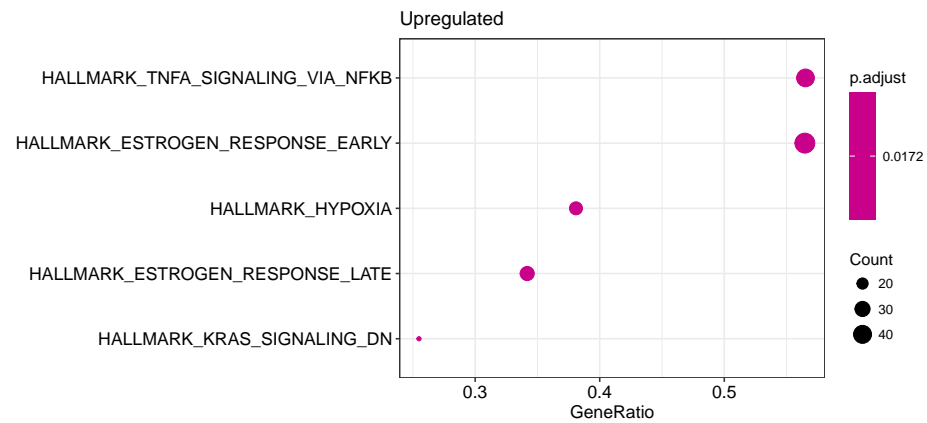


3.1.2 Upregulated genes

■ NOKS



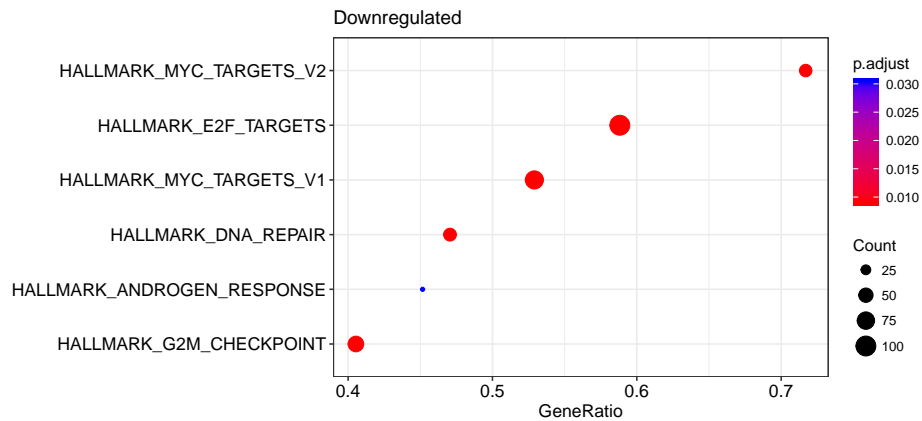
■ EBV



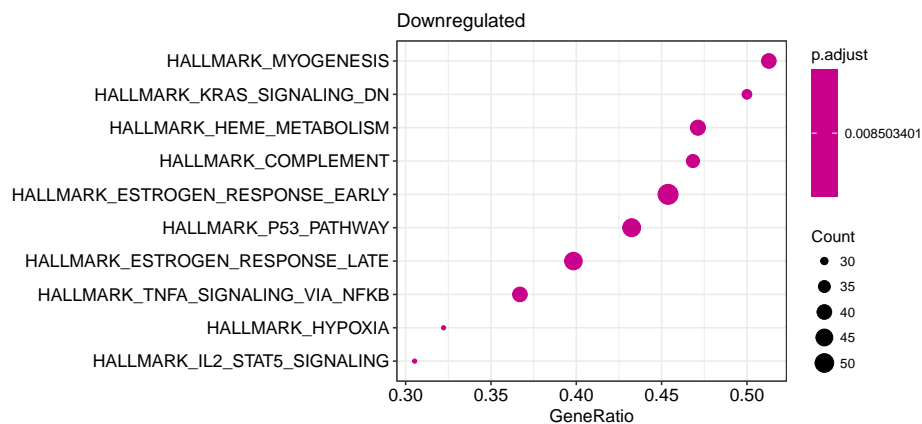
3.1.3 Downregulated genes

■ NOKS

Enrichment analysis with clusterProfiler



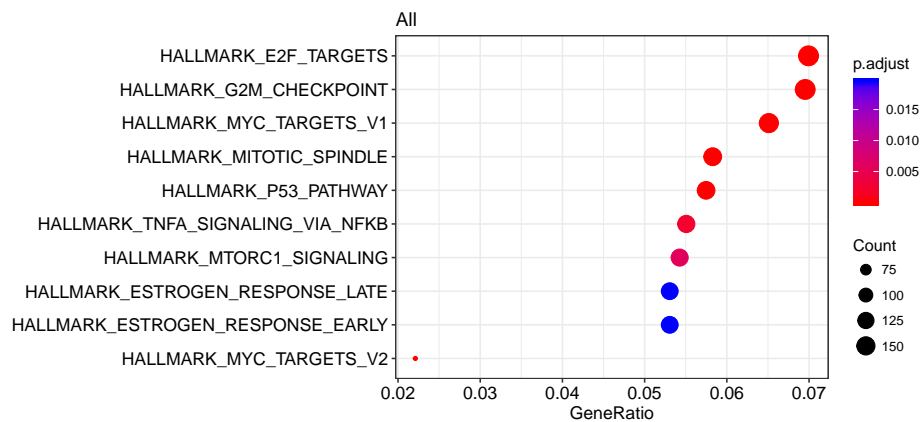
■ EBV



3.2 Over-representation test

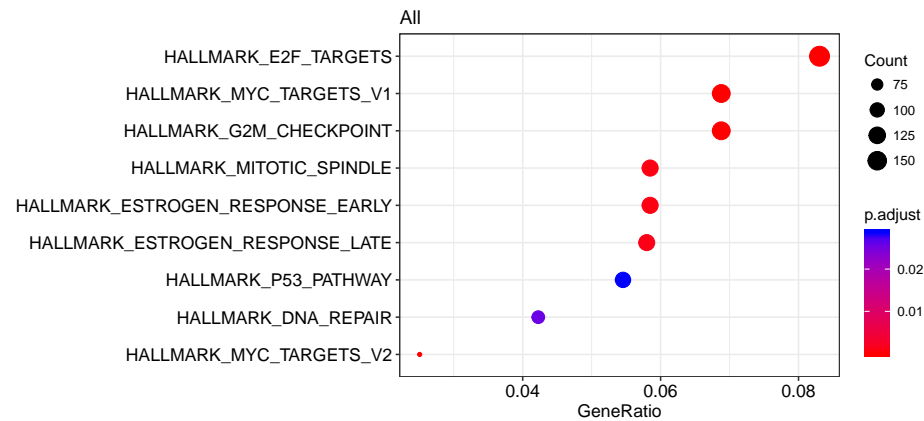
3.2.1 Diff. expressed genes

■ NOKS



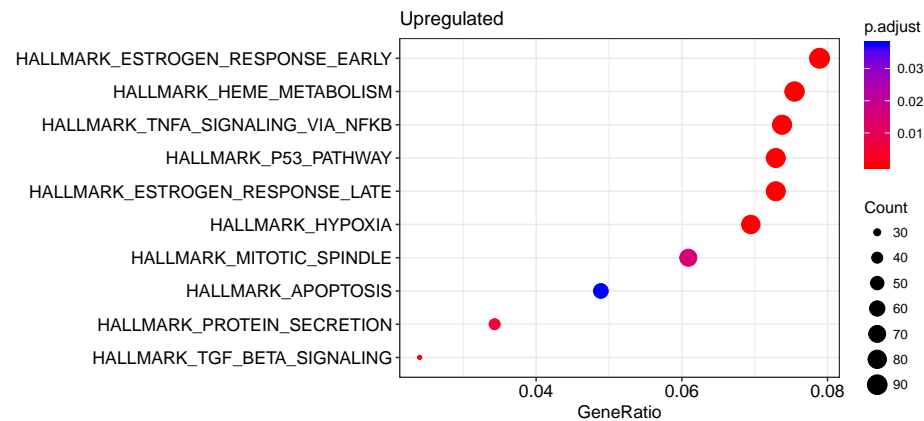
■ EBV

Enrichment analysis with clusterProfiler

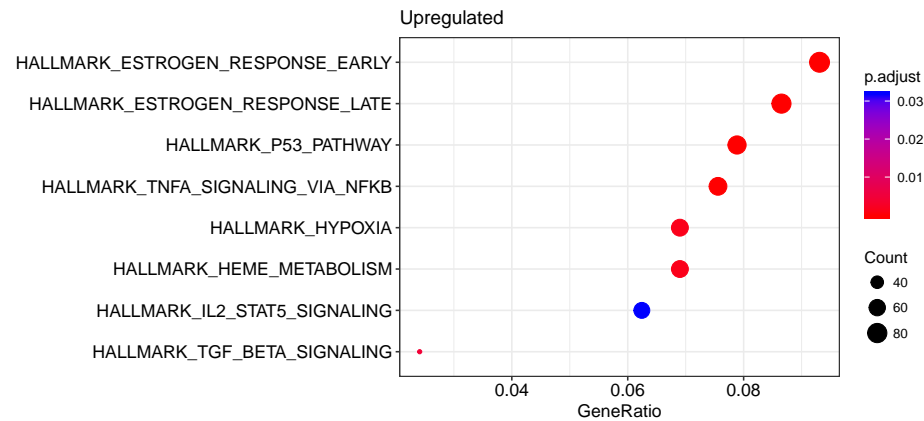


3.2.2 Upregulated genes

■ NOKS



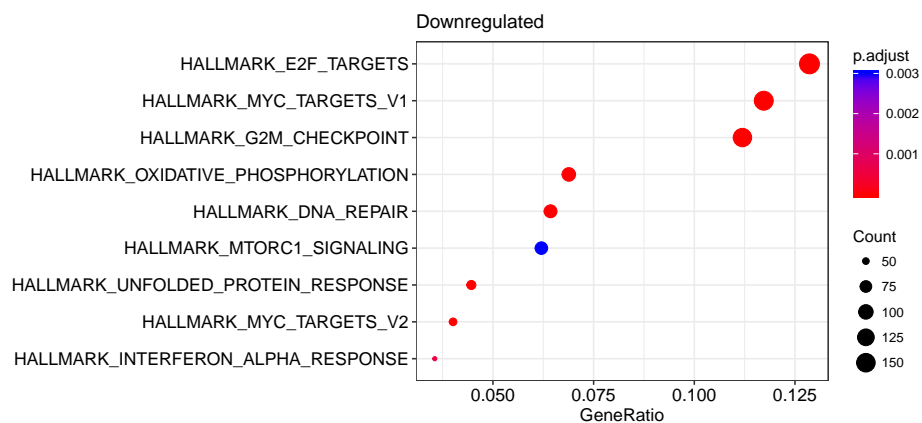
■ EBV



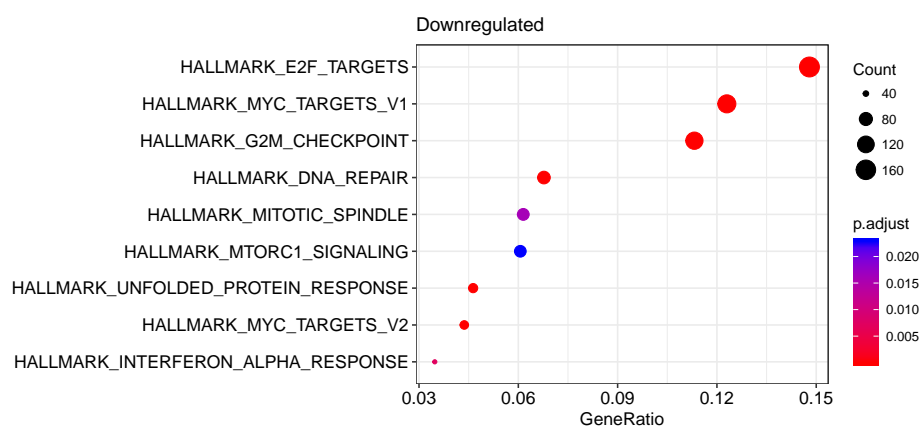
3.2.3 Downregulated genes

■ NOKS

Enrichment analysis with clusterProfiler



■ EBV



4 Bibliography

G Yu, LG Wang, Y Han, QY He. clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS: A Journal of Integrative Biology 2012, 16(5):284-287. doi:10.1089/omi.2011.0118 (<http://dx.doi.org/10.1089/omi.2011.0118>)