

Enrichment analysis with clusterProfiler

Rene Welch

February, 2018

Contents

1	Introduction	2
2	Preparing the data	2
2.1	Gene Ontology list V5.1.	2
2.2	Keratinocytes pathways.	2
2.3	Diff expressed genes	3
3	Pathway analysis on the Biological processes ontology	3
3.1	Quick summary.	3
3.2	GSEA analysis with pval_threshold = 0.1	4
3.3	Extra analysis: Seached <code>kera</code> in MolSigDB	7
4	Bibliography	10

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"
```

2 Preparing the data

2.1 Gene Ontology 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:

```
##                  all                  bp
## "c5.all.v5.1.symbols.gmt" "c5.bp.v5.1.symbols.gmt"
##                  cc                  mf
## "c5.cc.v5.1.symbols.gmt" "c5.mf.v5.1.symbols.gmt"
```

There is one list of genes that contains all, and there are three separated subsets:

- **bp**: Biological processes
- **cc**: Cellular component
- **mf**: Molecular function

2.2 Keratinocytes pathways

First, we search for ontologies with **KERA** in the name, and notice that those gene list are only present in the biological processes. So, we are going to focus on that gene list.

```
onto_keratinocyte = ontologies %>%
  map(filter, str_detect(ont, "KERA"))

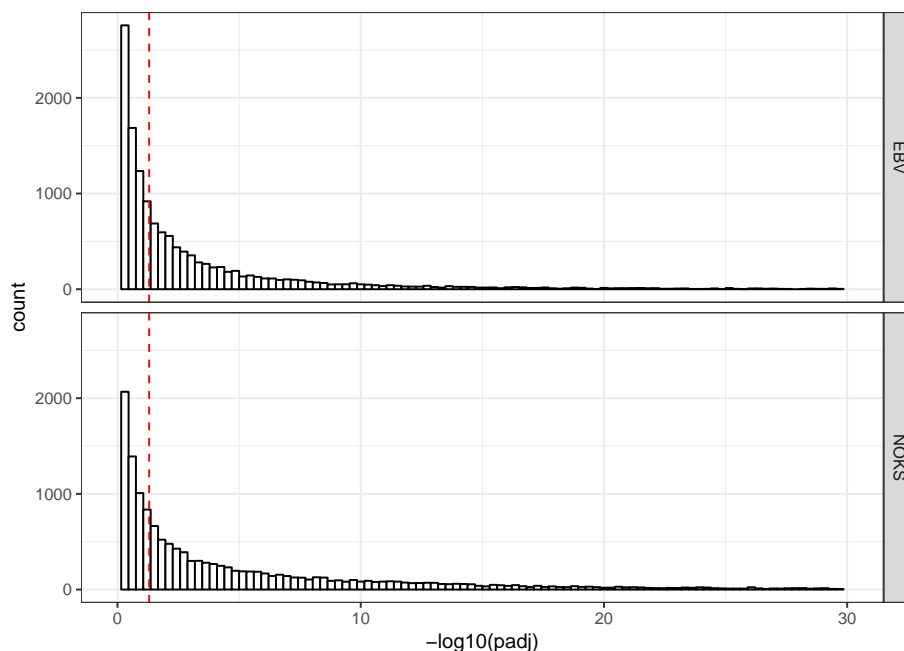
onto_keratinocyte %>%
  map_int(nrow)
## all bp cc mf
## 15 15 0 0

onto_keratinocyte[["bp"]]
##      ont      gene
## 1 KERATINOCYTE_DIFFERENTIATION LOR
## 2 KERATINOCYTE_DIFFERENTIATION TXNIP
## 3 KERATINOCYTE_DIFFERENTIATION ANXA1
## 4 KERATINOCYTE_DIFFERENTIATION SCEL
## 5 KERATINOCYTE_DIFFERENTIATION IL20
```

```
## 6 KERATINOCYTE_DIFFERENTIATION NME2
## 7 KERATINOCYTE_DIFFERENTIATION EVPL
## 8 KERATINOCYTE_DIFFERENTIATION EREG
## 9 KERATINOCYTE_DIFFERENTIATION SPRR1A
## 10 KERATINOCYTE_DIFFERENTIATION SPRR1B
## 11 KERATINOCYTE_DIFFERENTIATION TGM1
## 12 KERATINOCYTE_DIFFERENTIATION TGM3
## 13 KERATINOCYTE_DIFFERENTIATION DSP
## 14 KERATINOCYTE_DIFFERENTIATION CSTA
## 15 KERATINOCYTE_DIFFERENTIATION IVL
```

2.3 Diff expressed genes

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



3 Pathway analysis on the Biological processes ontology

3.1 Quick summary

For both methods `GSEA` and `enricher`, we performed the pathway analysis using the biological process subset of MolSigDB C5: gene ontologies. In total, for each cell and diff. expression category the following number of ontologies were deemed significant:

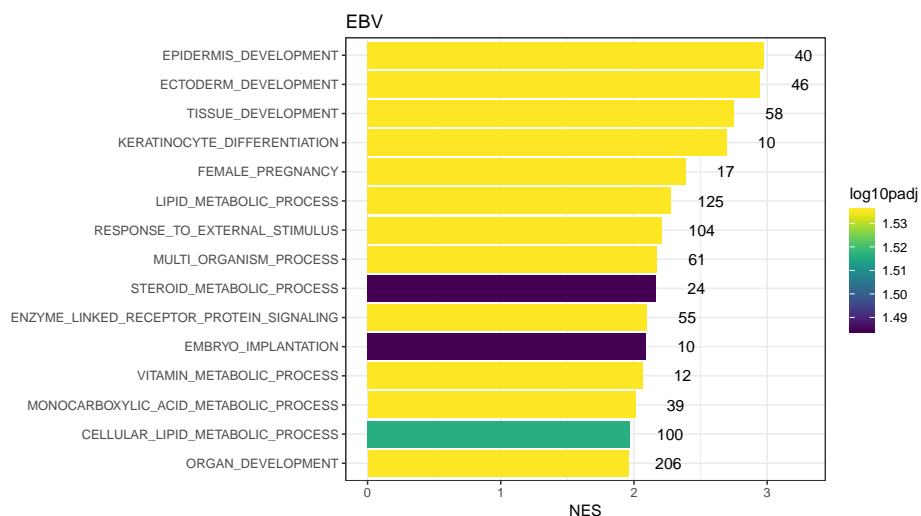
Enrichment analysis with clusterProfiler

```
## # A tibble: 6 x 4
##   cell type      gsea_0.05 gsea_0.1
##   <chr> <chr>      <int>   <int>
## 1 EBV   all          75     93
## 2 NOKS  all          97    114
## 3 EBV   upreg         9     10
## 4 NOKS  upreg         0      9
## 5 EBV   downreg       22     29
## 6 NOKS  downreg       18     24
```

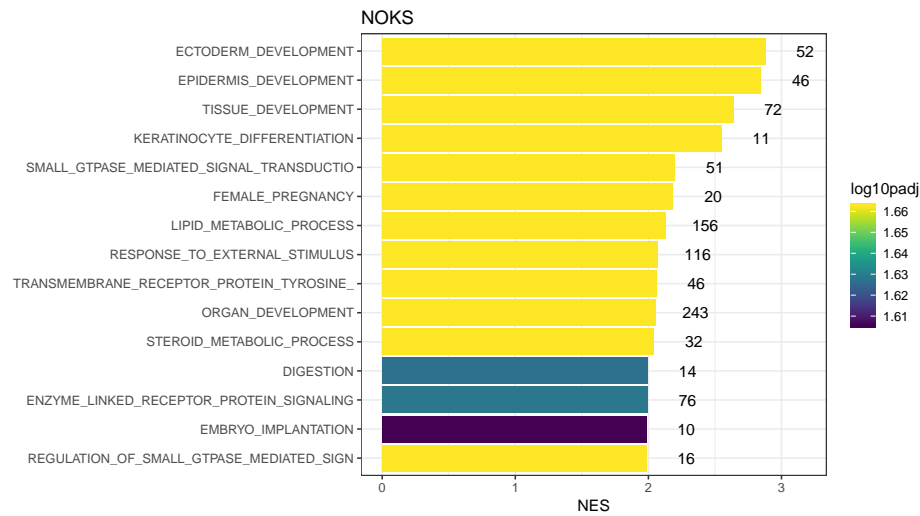
3.2 GSEA analysis with pval_threshold = 0.1

Quick note: In the figures below, I truncated the gene sets names at 40 character to get better visualizations.

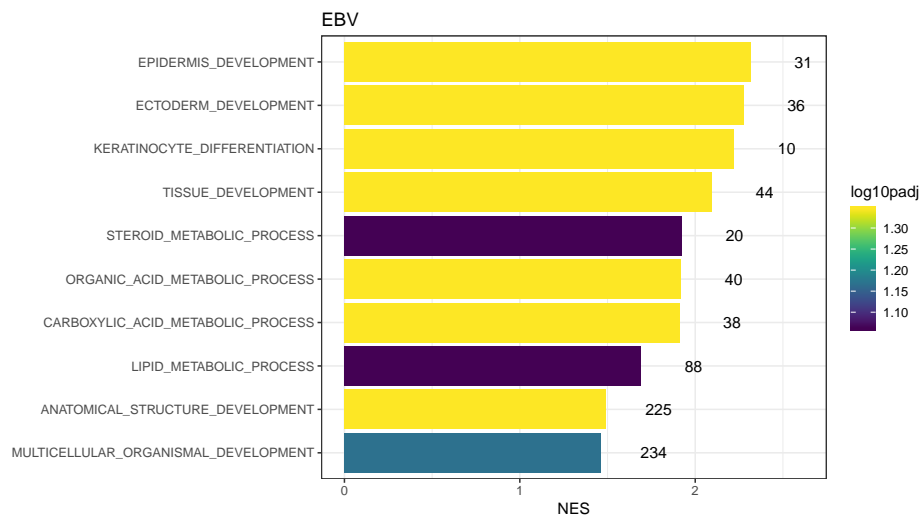
3.2.1 All genes



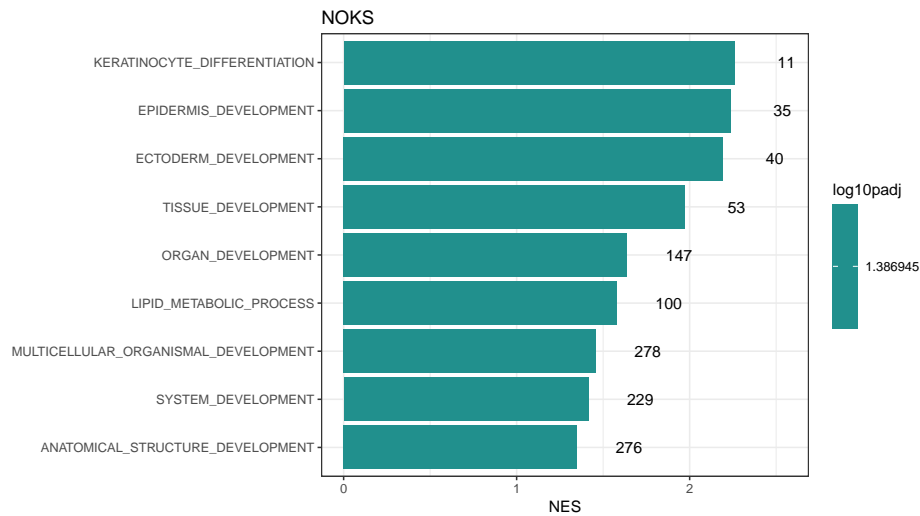
Enrichment analysis with clusterProfiler



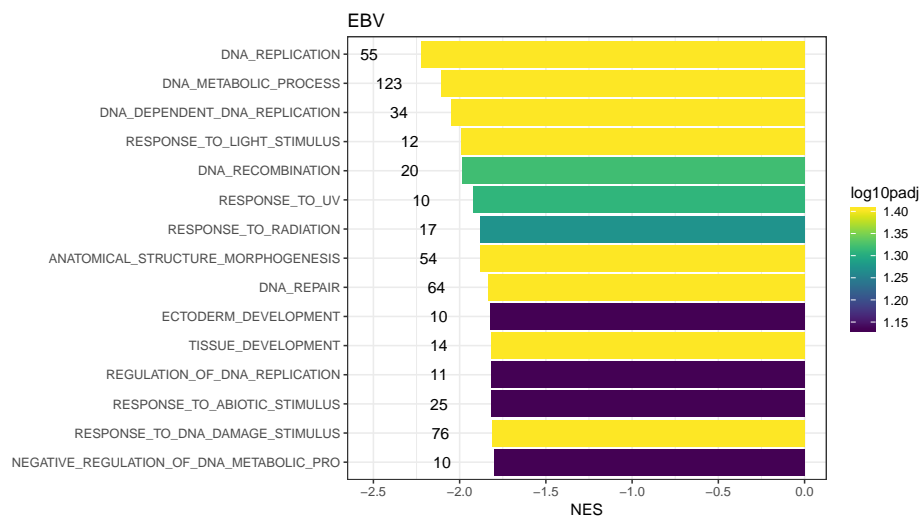
3.2.2 Upregulated



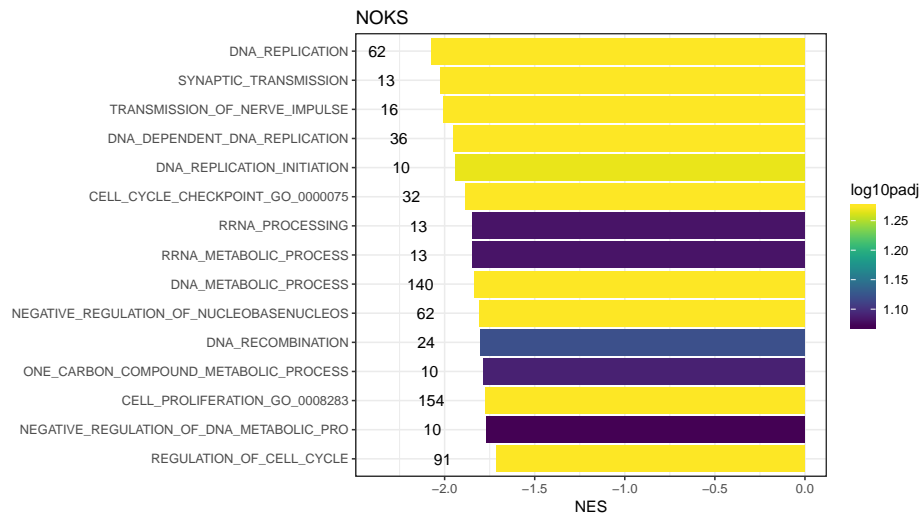
Enrichment analysis with clusterProfiler



3.2.3 Downregulated



Enrichment analysis with clusterProfiler



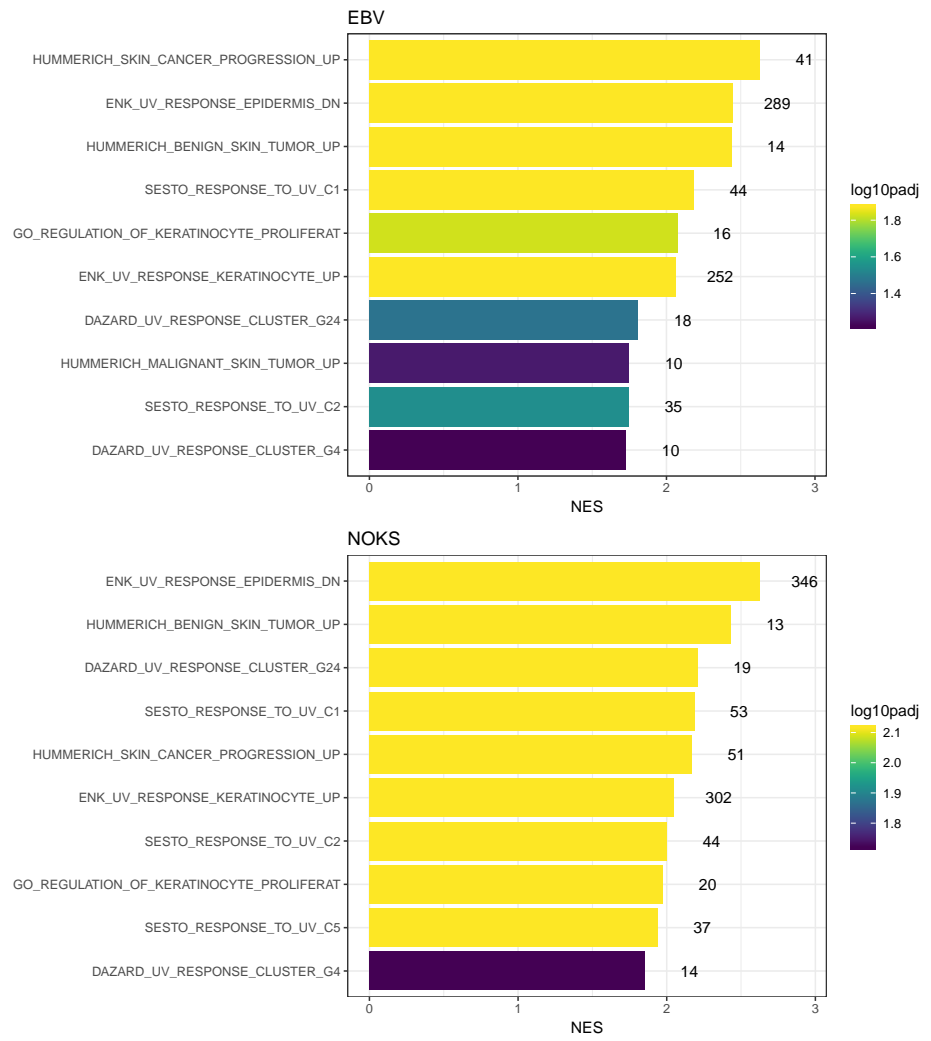
3.3 Extra analysis: Seached `kera` in MolSigDB

I constructed a gene list by searching in the GSEA website all the gene sets related to KERANOCYTES. We repeated the previous GSEA analysis:

```
## # A tibble: 6 x 4
##   cell type    gsea_0.05 gsea_0.1
##   <chr> <chr>      <int>    <int>
## 1 EBV   all         11      15
## 2 NOKS  all         15      19
## 3 EBV   upreg        9      11
## 4 NOKS  upreg        9      12
## 5 EBV   downreg       1       1
## 6 NOKS  downreg       1       1
```

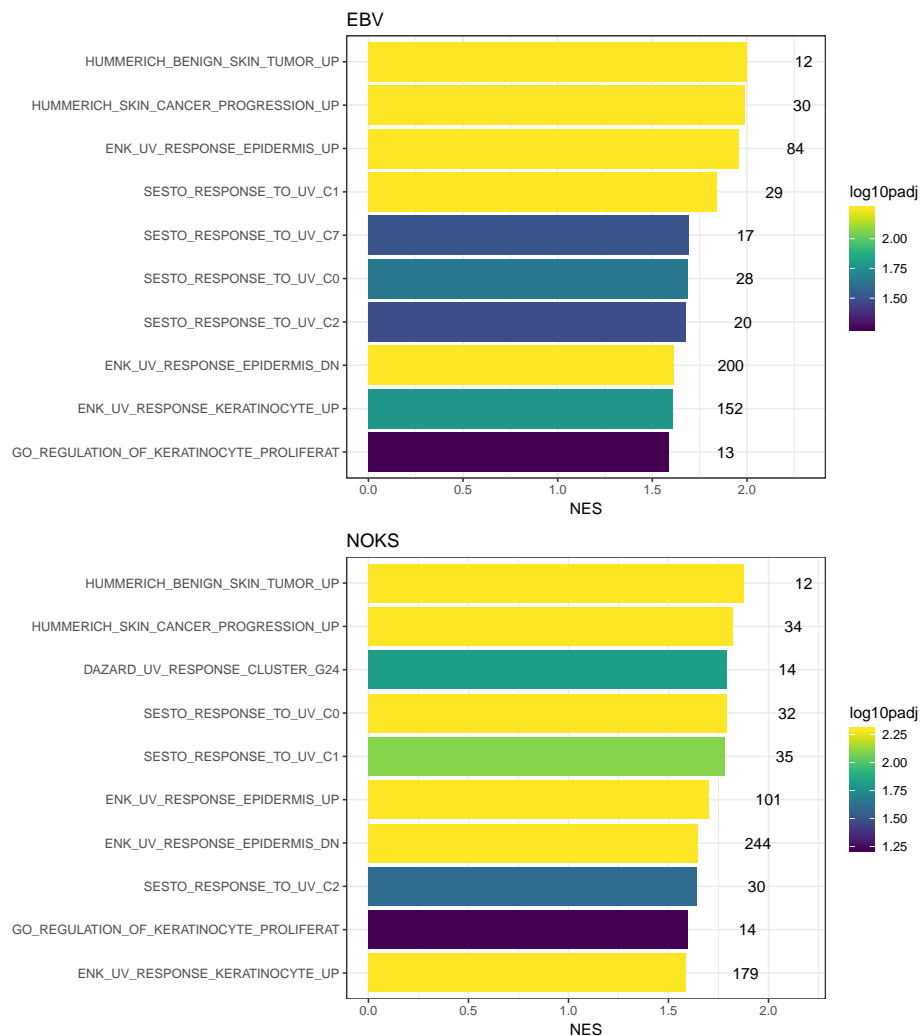
Enrichment analysis with clusterProfiler

3.3.1 All genes



Enrichment analysis with clusterProfiler

3.3.2 Upregulated



3.3.3 Downregulated

Note: Only the following, which can change due to the permutation:

```
kera_results %>%
  filter(type == "downreg") %>%
  pluck("gsea_0.1")
## [[1]]
## # A tibble: 1 x 13
##   ID      Description setSize enrichmentScore  NES  pvalue p.adjust qvalues
##   <chr>   <chr>         <int>          <dbl> <dbl>  <dbl>  <dbl>  <dbl>
## 1 ENK_UV~ ENK_UV_RES~    162          -0.307 -1.54 0.00200 0.0440 0.0400
## # ... with 5 more variables: rank <dbl>, leading_edge <chr>,
## #   core_enrichment <chr>, log10pval <dbl>, log10padj <dbl>
##
```

```
## [[2]]  
## # A tibble: 1 x 13  
##   ID      Description setSize enrichmentScore  NES  pvalue p.adjust qvalues  
##   <chr>   <chr>         <int>         <dbl> <dbl>  <dbl>   <dbl>  <dbl>  
## 1 ENK_UV~ ENK_UV_RES~    199         -0.331 -1.54 9.99e-4  0.0240  0.0200  
## # ... with 5 more variables: rank <dbl>, leading_edge <chr>,  
## #   core_enrichment <chr>, log10pval <dbl>, log10padj <dbl>
```

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](https://doi.org/10.1089/omi.2011.0118)(<http://dx.doi.org/10.1089/omi.2011.0118>)