Oncology Bioinformatics Analysis

Denver Ncube

2024-02-16

SECTION 1

Background and Experimental Design

The purpose of this mini-report is to present one of the ways we can use a simple transcriptomic analysis to gain insight into possible biomarkers for disease presence or progression.

The analysis laid out here is not meant to readily translate into a clinical outcome without necessary follow-up studies and exhaustive review.

The dataset used in this presentation is based on an actual study: "Kim SK, Kim SY, Kim JH, Roh SA, Cho DH et al. (2014) A nineteen gene-based risk score classifier predicts prognosis of colorectal cancer patients".

The dataset contains RNA-seq data from 18 subjects, but sampled (and this is very important) **from different tissues of these subjects, normal large intestine tissue, primary colon cancer and progressive liver metastasis (colon originating)**. The final number of samples is actually 54 but for this brief report 37 relevant samples, 18 normal and 19 primary colon cancer samples will be used.

The basic logic for any potential bio-marker discovery is to compare the cancer samples to non-cancer samples and observe the differences in their gene expressions.

For any potential biomarker discovery the goal is to compare the cancer samples to non-cancer samples and observe the differences in their gene expressions.

Overview of data

Run	Sample.Characteristic.biopsy.site.	Sample.Characteristic.Ontology.Term.biopsy.site.	Sample.Characteristic.disease.	Sample.Characteri
SRR975551	primary tumor	http://www.ebi.ac.uk/efo/EFO_0000616 (http://www.ebi.ac.uk/efo/EFO_0000616)	colorectal cancer	http://www.ebi.ac.u (http://www.ebi.ac.u
SRR975552	primary tumor	http://www.ebi.ac.uk/efo/EFO_0000616 (http://www.ebi.ac.uk/efo/EFO_0000616)	colorectal cancer	http://www.ebi.ac.u (http://www.ebi.ac.u
SRR975553	primary tumor	http://www.ebi.ac.uk/efo/EFO_0000616 (http://www.ebi.ac.uk/efo/EFO_0000616)	colorectal cancer	http://www.ebi.ac.u (http://www.ebi.ac.u
SRR975554	primary tumor	http://www.ebi.ac.uk/efo/EFO_0000616 (http://www.ebi.ac.uk/efo/EFO_0000616)	colorectal cancer	http://www.ebi.ac.u (http://www.ebi.ac.u
SRR975555	primary tumor	http://www.ebi.ac.uk/efo/EFO_0000616 (http://www.ebi.ac.uk/efo/EFO_0000616)	colorectal cancer	http://www.ebi.ac.u (http://www.ebi.ac.u
SRR975556	primary tumor	http://www.ebi.ac.uk/efo/EFO_0000616 (http://www.ebi.ac.uk/efo/EFO_0000616)	colorectal cancer	http://www.ebi.ac.u (http://www.ebi.ac.u
SRR975557	primary tumor	http://www.ebi.ac.uk/efo/EFO_0000616 (http://www.ebi.ac.uk/efo/EFO_0000616)	colorectal cancer	http://www.ebi.ac.u (http://www.ebi.ac.u
SRR975558	primary tumor	http://www.ebi.ac.uk/efo/EFO_0000616 (http://www.ebi.ac.uk/efo/EFO_0000616)	colorectal cancer	http://www.ebi.ac.u (http://www.ebi.ac.u
SRR975559	primary tumor	http://www.ebi.ac.uk/efo/EFO_0000616 (http://www.ebi.ac.uk/efo/EFO_0000616)	colorectal cancer	http://www.ebi.ac.u (http://www.ebi.ac.u
SRR975560	primary tumor	http://www.ebi.ac.uk/efo/EFO_0000616 (http://www.ebi.ac.uk/efo/EFO_0000616)	colorectal cancer	http://www.ebi.ac.u (http://www.ebi.ac.u
SRR975561	primary tumor	http://www.ebi.ac.uk/efo/EFO_0000616 (http://www.ebi.ac.uk/efo/EFO_0000616)	colorectal cancer	http://www.ebi.ac.u (http://www.ebi.ac.u
SRR975562	primary tumor	http://www.ebi.ac.uk/efo/EFO_0000616 (http://www.ebi.ac.uk/efo/EFO_0000616)	colorectal cancer	http://www.ebi.ac.u (http://www.ebi.ac.u
SRR975563	primary tumor	http://www.ebi.ac.uk/efo/EFO_0000616 (http://www.ebi.ac.uk/efo/EFO_0000616)	colorectal cancer	http://www.ebi.ac.u (http://www.ebi.ac.u
SRR975564	primary tumor	http://www.ebi.ac.uk/efo/EFO_0000616 (http://www.ebi.ac.uk/efo/EFO_0000616)	colorectal cancer	http://www.ebi.ac.u (http://www.ebi.ac.u
SRR975565	primary tumor	http://www.ebi.ac.uk/efo/EFO_0000616 (http://www.ebi.ac.uk/efo/EFO_0000616)	colorectal cancer	http://www.ebi.ac.u (http://www.ebi.ac.u
SRR975566	primary tumor	http://www.ebi.ac.uk/efo/EFO_0000616	colorectal cancer	http://www.ebi.ac.u

		(http://www.ebi.ac.uk/efo/EFO_0000616)		(http://www.ebi.ac.u
SRR975567	primary tumor	http://www.ebi.ac.uk/efo/EFO_0000616 (http://www.ebi.ac.uk/efo/EFO_0000616)	colorectal cancer	http://www.ebi.ac.u (http://www.ebi.ac.u
SRR975568	primary tumor	http://www.ebi.ac.uk/efo/EFO_0000616 (http://www.ebi.ac.uk/efo/EFO_0000616)	colorectal cancer	http://www.ebi.ac.u (http://www.ebi.ac.u
SRR975569	normal	http://purl.obolibrary.org/obo/PATO_0000461 (http://purl.obolibrary.org/obo/PATO_0000461)	colorectal cancer	http://www.ebi.ac.u (http://www.ebi.ac.u
SRR975570	normal	http://purl.obolibrary.org/obo/PATO_0000461 (http://purl.obolibrary.org/obo/PATO_0000461)	colorectal cancer	http://www.ebi.ac.u (http://www.ebi.ac.u
SRR975571	normal	http://purl.obolibrary.org/obo/PATO_0000461 (http://purl.obolibrary.org/obo/PATO_0000461)	colorectal cancer	http://www.ebi.ac.u (http://www.ebi.ac.u
SRR975572	normal	http://purl.obolibrary.org/obo/PATO_0000461 (http://purl.obolibrary.org/obo/PATO_0000461)	colorectal cancer	http://www.ebi.ac.u (http://www.ebi.ac.u
SRR975573	normal	http://purl.obolibrary.org/obo/PATO_0000461 (http://purl.obolibrary.org/obo/PATO_0000461)	colorectal cancer	http://www.ebi.ac.u (http://www.ebi.ac.u
SRR975574	normal	http://purl.obolibrary.org/obo/PATO_0000461 (http://purl.obolibrary.org/obo/PATO_0000461)	colorectal cancer	http://www.ebi.ac.u (http://www.ebi.ac.u
SRR975575	normal	http://purl.obolibrary.org/obo/PATO_0000461 (http://purl.obolibrary.org/obo/PATO_0000461)	colorectal cancer	http://www.ebi.ac.u (http://www.ebi.ac.u
SRR975576	normal	http://purl.obolibrary.org/obo/PATO_0000461 (http://purl.obolibrary.org/obo/PATO_0000461)	colorectal cancer	http://www.ebi.ac.u (http://www.ebi.ac.u
SRR975577	normal	http://purl.obolibrary.org/obo/PATO_0000461 (http://purl.obolibrary.org/obo/PATO_0000461)	colorectal cancer	http://www.ebi.ac.u (http://www.ebi.ac.u
SRR975578	normal	http://purl.obolibrary.org/obo/PATO_0000461 (http://purl.obolibrary.org/obo/PATO_0000461)	colorectal cancer	http://www.ebi.ac.u (http://www.ebi.ac.u
SRR975579	normal	http://purl.obolibrary.org/obo/PATO_0000461 (http://purl.obolibrary.org/obo/PATO_0000461)	colorectal cancer	http://www.ebi.ac.u (http://www.ebi.ac.u
SRR975580	normal	http://purl.obolibrary.org/obo/PATO_0000461 (http://purl.obolibrary.org/obo/PATO_0000461)	colorectal cancer	http://www.ebi.ac.u (http://www.ebi.ac.u
SRR975581	normal	http://purl.obolibrary.org/obo/PATO_0000461 (http://purl.obolibrary.org/obo/PATO_0000461)	colorectal cancer	http://www.ebi.ac.u (http://www.ebi.ac.u
SRR975582	normal	http://purl.obolibrary.org/obo/PATO_0000461 (http://purl.obolibrary.org/obo/PATO_0000461)	colorectal cancer	http://www.ebi.ac.u (http://www.ebi.ac.u
SRR975583	normal	http://purl.obolibrary.org/obo/PATO_0000461 (http://purl.obolibrary.org/obo/PATO_0000461)	colorectal cancer	http://www.ebi.ac.u (http://www.ebi.ac.u
SRR975584	normal	http://purl.obolibrary.org/obo/PATO_0000461 (http://purl.obolibrary.org/obo/PATO_0000461)	colorectal cancer	http://www.ebi.ac.u (http://www.ebi.ac.u
SRR975585	normal	http://purl.obolibrary.org/obo/PATO_0000461 (http://purl.obolibrary.org/obo/PATO_0000461)	colorectal cancer	http://www.ebi.ac.u (http://www.ebi.ac.u
SRR975586	normal	http://purl.obolibrary.org/obo/PATO_0000461 (http://purl.obolibrary.org/obo/PATO_0000461)	colorectal cancer	http://www.ebi.ac.u (http://www.ebi.ac.u
SRR975587	normal	http://purl.obolibrary.org/obo/PATO_0000461 (http://purl.obolibrary.org/obo/PATO_0000461)	colorectal cancer	http://www.ebi.ac.u (http://www.ebi.ac.u
SRR975588	colorectal cancer metastatic in the liver		colorectal cancer	http://www.ebi.ac.u (http://www.ebi.ac.u
SRR975589	colorectal cancer metastatic in the liver		colorectal cancer	http://www.ebi.ac.u (http://www.ebi.ac.u
SRR975590	colorectal cancer metastatic in the liver		colorectal cancer	http://www.ebi.ac.u (http://www.ebi.ac.u
SRR975591	colorectal cancer metastatic in the liver		colorectal cancer	http://www.ebi.ac.u (http://www.ebi.ac.u

SRR975592	colorectal cancer metastatic in the liver		colorectal cancer	http://www.ebi.ac.u (http://www.ebi.ac.u
SRR975593	colorectal cancer metastatic in the liver		colorectal cancer	http://www.ebi.ac.u (http://www.ebi.ac.u
SRR975594	colorectal cancer metastatic in the liver		colorectal cancer	http://www.ebi.ac.u (http://www.ebi.ac.u
SRR975595	colorectal cancer metastatic in the liver		colorectal cancer	http://www.ebi.ac.u (http://www.ebi.ac.u
SRR975596	colorectal cancer metastatic in the liver		colorectal cancer	http://www.ebi.ac.u (http://www.ebi.ac.u
SRR975597	colorectal cancer metastatic in the liver		colorectal cancer	http://www.ebi.ac.u (http://www.ebi.ac.u
SRR975598	colorectal cancer metastatic in the liver		colorectal cancer	http://www.ebi.ac.u (http://www.ebi.ac.u
SRR975599	colorectal cancer metastatic in the liver		colorectal cancer	http://www.ebi.ac.u (http://www.ebi.ac.u
SRR975600	colorectal cancer metastatic in the liver		colorectal cancer	http://www.ebi.ac.u (http://www.ebi.ac.u
SRR975601	colorectal cancer metastatic in the liver		colorectal cancer	http://www.ebi.ac.u (http://www.ebi.ac.u
SRR975602	colorectal cancer metastatic in the liver		colorectal cancer	http://www.ebi.ac.u (http://www.ebi.ac.u
SRR975603	colorectal cancer metastatic in the liver		colorectal cancer	http://www.ebi.ac.u (http://www.ebi.ac.u
SRR975604	colorectal cancer metastatic in the liver		colorectal cancer	http://www.ebi.ac.u (http://www.ebi.ac.u

As it can be seen the first 18 samples are primary tumors (colorectal cancer) and the next 19 are normal and then the rest are liver metastasis samples. For this tutorial we will use only the first 37, so primary tumor and normal samples.

Running DESEQ2

Creating DESeq results object using Benjamini-Hochberg correction

	baseMean	log2FoldChange	lfcSE	stat	pvalue	padj
ENSG00000000003	1006.252417	0.4171339	0.2540846	1.641713	0.1006496	0.2084175
ENSG00000000005	6.572053	-0.0476418	0.4225178	-0.112757	0.9102232	0.9473385
ENSG00000000419	370.323717	0.7674698	0.1742659	4.404016	0.0000106	0.0002247
ENSG00000000457	179.743933	-0.1868380	0.1148662	-1.626571	0.1038282	0.2134028
ENSG00000000460	58.131389	0.9189451	0.2353201	3.905086	0.0000942	0.0010838
ENSG00000000938	68.277653	0.4785311	0.2157950	2.217526	0.0265872	0.0766179

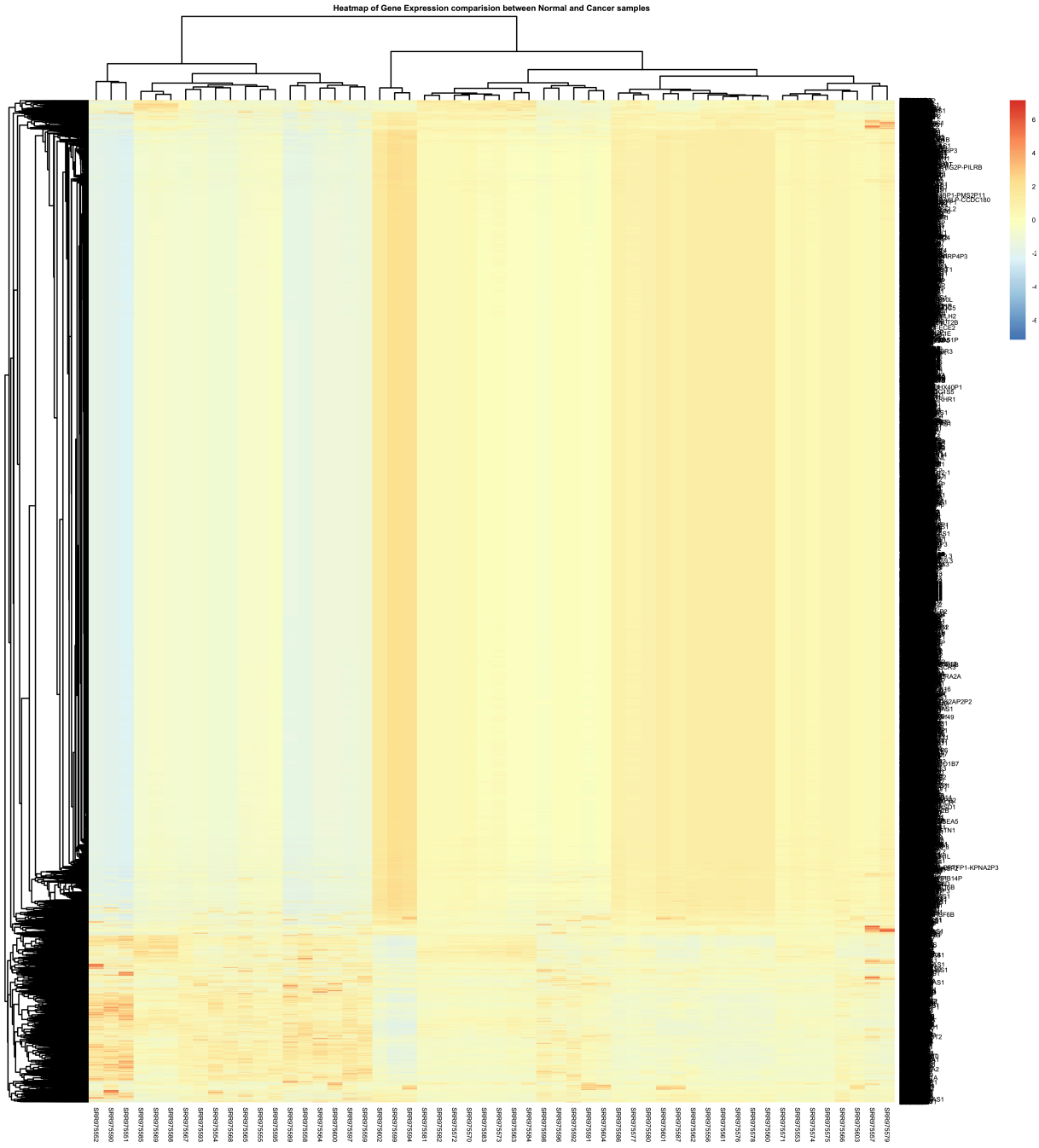
Out of 40166 non-zero gene counts there are *121* or *0.3%* or upregulated genes and *30 or 0.075%* of *down regulated genes*. So that would be around 151 genes of interest out of over 50 000. First conclusions is that there are much more upregulated genes of interest as potential causal factors for this type of cancer, but the 30 downregulated are not something to ignore either.

One thing is very important and that is the multiple testing correction procedure. Over 40 000 nonzero genes means that the analysis will include over 40 000 hypothesis tests. FWER or family wise error rates occurs at high rates in these situations, finding positive results by random is almost certain if multiple correction procedure is implemented. That's why DESeq always has some multiple correction procedures.

I added the pAdjustMethod = 'BH', and this means using Benjamini-Hochberg correction, I did this on purpose so you can see how to implement the multiple testing procedures, even tough its implemented by default. But multiple testing procedure is very important in this case and I want a more robust method implemented. Therefore I will use the Holm multiple testing adjustment, which is going to be more robust on FWER.

Heatmap

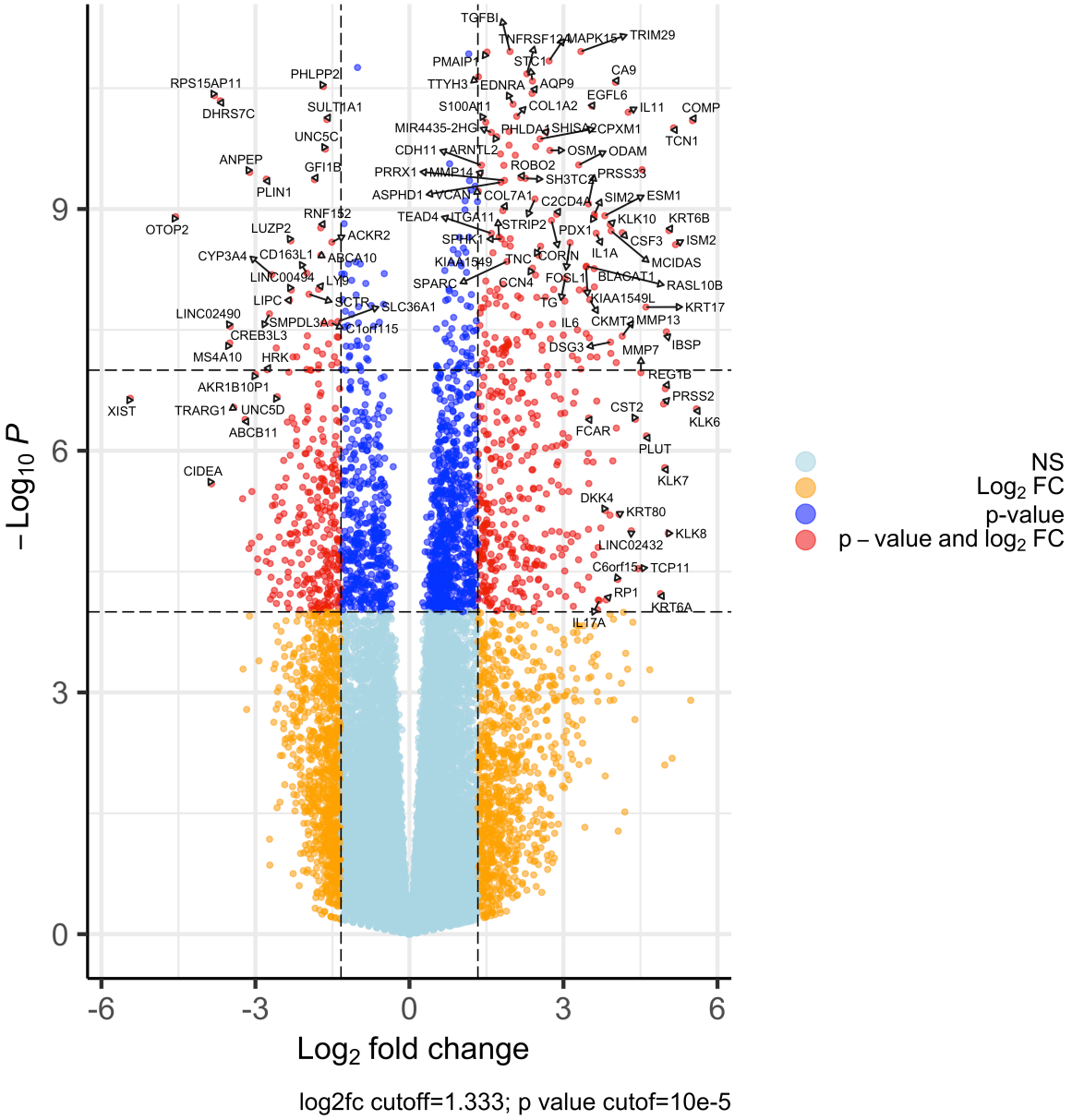
The heatmap in this section is merely an overview of the data though it might not be as informative since we have a lot of genes in this dataset so it becomes really hard to track them. However this is good a snapshot of relative expression differnces



Volcano plot to display differentially expressed genes and their relative significance

Results

Differential expression analysis for Normal vs Cancer Samples

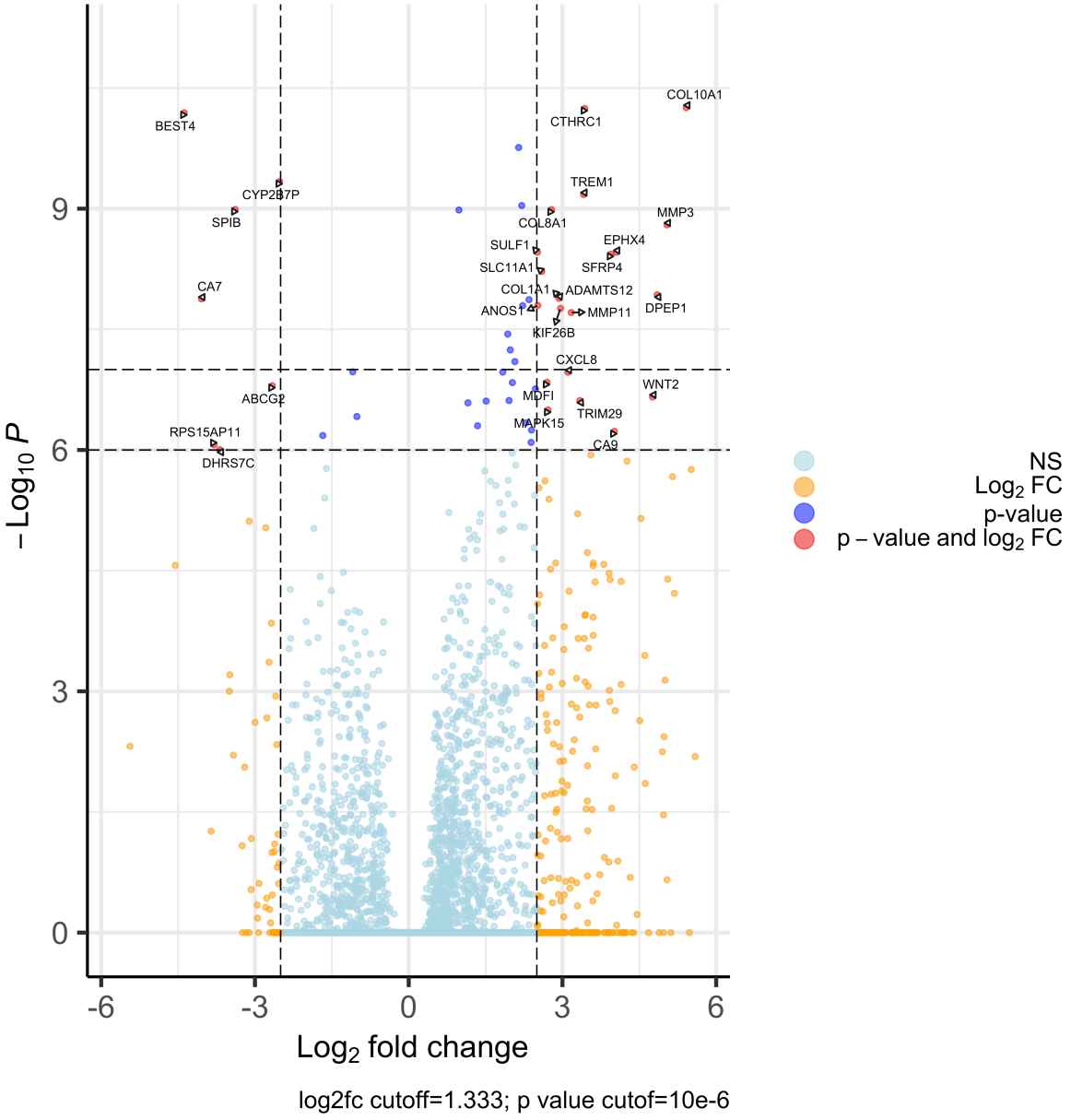


log2fc cutoff=1.333; p value cutof=10e-5

Selecting only marker genes of interest using the adjusted p-value.

Results

Differential expression analysis for Normal vs Cancer



log2fc cutoff=1.333; p value cutof=10e-6

SECTION 2

Gene Set Functional Enrichment analysis

Summary

To explore analyze the data more closely, I also did a short trial run and extracted some gene sets from the Kyoto Encylopedia of Genes and Genomes (KEGG) for colorectal cancer to check if there was significant enrichment of those specific gene sets. The result from the KEG colorectal cancer gene set shows that the specific gene sets in that dataset are not significantly enriched in our dataset as the p-ad = 0.73 and the NES (normalized expression) is -0.0879.

GSEA RESULT AND PLOT

```
## # A tibble: 19,190 × 2
##    Gene       stat
##    <chr>     <dbl>
##  1 A1BG-AS1 -4.70
##  2 A1CF     -1.56
##  3 A2M      -0.474
##  4 A2M-AS1  -0.886
##  5 A2MP1    -1.80
##  6 A4GALT    0.949
##  7 AAAS      1.21
##  8 AACS     -0.332
##  9 AADAC     1.60
## 10 AADACL2  -4.34
## # ℹ 19,180 more rows
```

Show 10 entries Search:

	pathway	padj	log2err	NES	size
1	KEGG_COLORECTAL_CANCER	0.7323076923076923	0.08175155825321447	-0.8791466556616183	61

Showing 1 to 1 of 1 entries Previous 1 Next

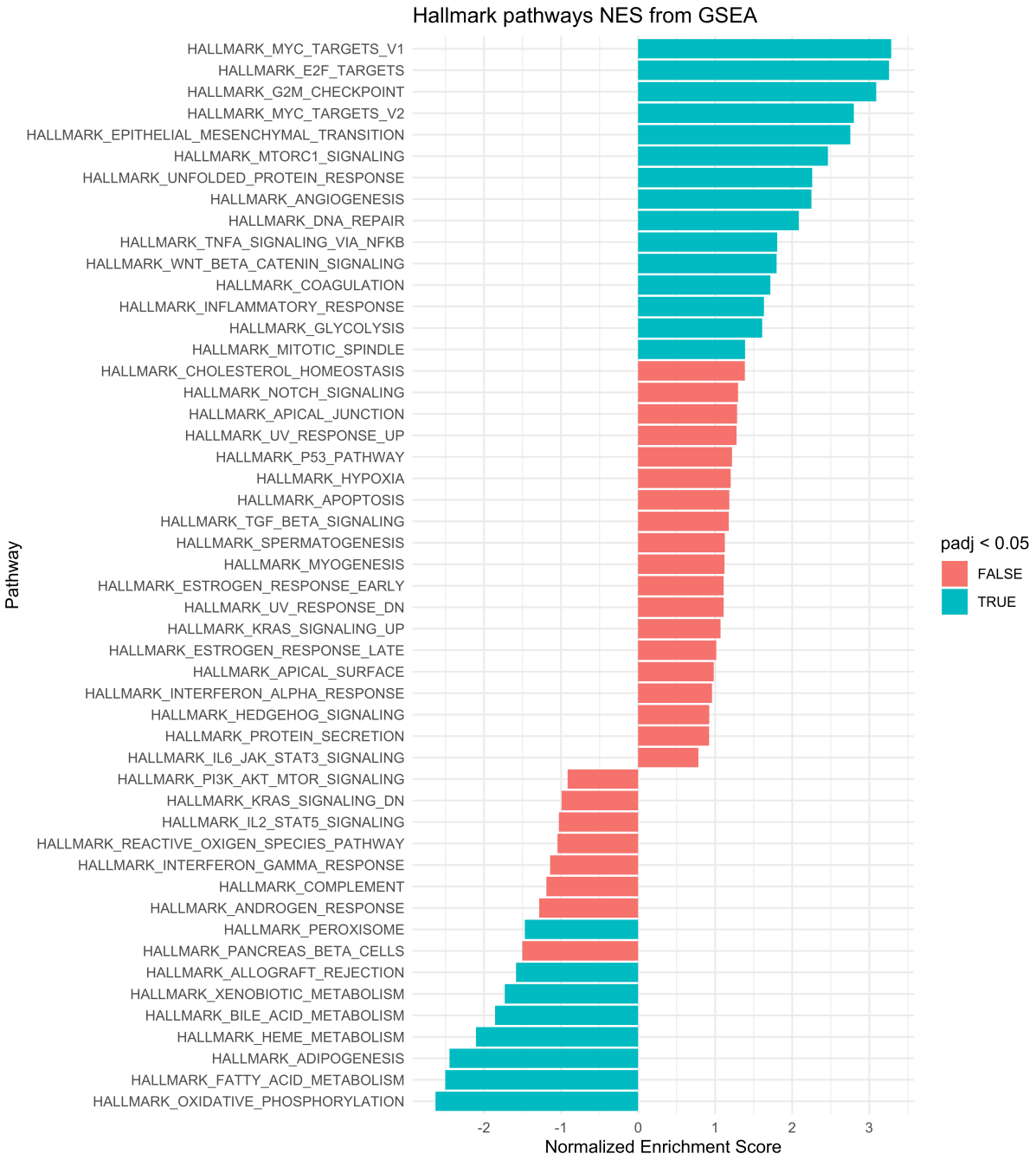
Show 10 entries Search:

	pathway	padj	log2err	NES	size
1	HALLMARK_MYC_TARGETS_V1	2.723924679263329e-40	1.690472378984485	3.285786563243036	193
2	HALLMARK_E2F_TARGETS	2.763536197016432e-39	1.659565341484397	3.255848283836536	194
3	HALLMARK_G2M_CHECKPOINT	4.260282178955118e-32	1.502386888814151	3.089470894939666	188
4	HALLMARK_EPITHELIAL_MESENCHYMAL_TRANSITION	5.209560250487714e-22	1.246232771916476	2.754502801104088	195
5	HALLMARK_OXIDATIVE_PHOSPHORYLATION	5.028678778873692e-15	1.027669870543582	-2.631656001364717	183
6	HALLMARK_MTORC1_SIGNALING	5.492086049063108e-15	1.027669870543582	2.463314263579093	194
7	HALLMARK_MYC_TARGETS_V2	9.300533038589517e-15	1.017544797309885	2.799082708273078	58
8	HALLMARK_ADIPOGENESIS	5.882973399166173e-13	0.9545416307257243	-2.447941636465511	186
9	HALLMARK_FATTY_ACID_METABOLISM	7.029570844593943e-12	0.9101197347744591	-2.503166967720976	147
10	HALLMARK_UNFOLDED_PROTEIN_RESPONSE	2.226968312633735e-8	0.7614608014455848	2.258748026948287	106

Showing 1 to 10 of 50 entries Previous 1 2 3 4 5 Next

GSEA Plot

This plot is generated using the normalized enriched values for



Conclusions

Increased expression of CCNB1, CLCA1, and PLK4 restrained proliferation as well as migration of cancer cells and induced apoptosis of cancer cells. CCNB1, KIF4A, TPX2, MT1F, PRC1, PLK4, CALD1, MMP9, CLCA1, and MMP1 were identified as hub genes and CCNB1, CLCA1, and PLK4 could inhibit the progression of colon cancer. These highly expressed genes are also used as prognostic indicators in colorectal cancer.

The top 10 pathways that are significantly enriched based on this analysis include the typical cancer pathways that we expect to have genes that exhibit high expression. The supplementary study to this analysis was a survival analysis based on expression of the key colorectal cancer genes and the relevant pathway expression modules. It is important to note that this analysis report is a tip of the iceberg as there are many possible additional analyses that would need to be done in order to have a more granular view of the data that will in turn inform clinical decision making.

SECTION 3 : Potential Future directions

1. Final interpretation would be that the interesting genes were identified starting with COL11A1, CEMIP, ADAM12, MMP1, OTOP3 (check volcano plot above in Section 2);
2. Further experimental research would be needed to conclude if any of these are potential cancer treatment targets;
3. Whole Exome Seq followed by variant curation to enahnce the insights we have obtained from this analysis;
4. CCNB1, KIF4A, TPX2, MT1F, PRC1, PLK4, CALD1, MMP9, CLCA1, and MMP1 were identified as some of the genes and CCNB1, CLCA1, and PLK4 could inhibit the progression of colon cancer.