

Part 1: Differential Analysis (DESeq2)

1. Download R (<https://cran.cnr.berkeley.edu/>)
2. In R, install the DESeq2 Library

```
if (!requireNamespace("BiocManager", quietly = TRUE))  
  install.packages("BiocManager")  
BiocManager::install("DESeq2", version = "3.8")
```

3. Run the Library

```
> Library("DESeq2")
```

4. Make sure that all of your files are in the same directory. Ex:

```
> dir <- getwd()  
> dir
```

5. Store the path to your sample table

```
> csvfile <- file.path(dir, "sampleTable.csv")
```

6. Store this sample table

```
> sampleTable <- read.csv(csvfile, row.names=1)
```

7. Once you have the path to the sample table, we can run DESeq2

```
> dds <- DESeqDataSetFromHTSeqCount(sampleTable = sampleTable, directory =  
dir, ~ celltype)  
> dds = DESeq(dds)  
> res = results(dds)  
> resOrdered <- res[order(res$padj),]  
> write.table(resOrdered, file = "GM12878_K562_out.txt", sep = "\t", quote  
= F)
```

8. The output file "GM12878_K562_out.txt" will be used for future analysis. Open it and look how the genes are formatted.

Part 2: Gene Ontology Analysis (GO)

1. Locate output file from DESeq2 (GM12878_K562_out.txt)
2. Run the following command in Terminal (unless you are running Windows):

```
cut -f 1 -d "." GM12878_K562_out.txt | head -501 | tail -500 >
GM12878_K562_500_genes.txt
```

3. The output file of the previous command is GM12878_K562_500_genes.txt
 - a. This file is provided for Windows users
4. Navigate to <https://biit.cs.ut.ee/gprofiler/gost>
5. Check the box that says “Ordered query”
6. Copy and paste the contents of GM12878_K562_500_genes.txt into the text box and run the query
7. Go to the “Detailed Results” tab
8. Download the results for the significantly enriched gene sets as a CSV

Part 3: Gene Set Enrichment Analysis (GSEA)

1. Locate output file from DESeq2 (GM12878_K562_out.txt)
2. Run the following command in Terminal (unless you are running Windows):

```
cut -f 1,5 GM12878_K562_out.txt | sed '1d' | sed 's/\. [0-9]\t\t/g' | sed
's/\.1[0-9]\t\t/g' | sort -k 2n > GM12878_K562_genes_values.txt
```

3. The output file of the previous command is GM12878_K562_genes_values.txt
 - a. This file is provided for Windows users
4. Navigate to <http://www.webgestalt.org/option.php>
5. Specify the following parameters:
 - a. Organism of Interest: hsapiens
 - b. Method of Interest: Gene Set Enrichment Analysis (GSEA)
 - c. Functional Database Class: pathway
 - d. Functional Database Name: KEGG
 - e. Gene ID Type: ensembl_gene_id
 - f. Ranked Gene List: copy and paste the contents of GM12878_K562_genes_values.txt into the text box
6. Click “Submit”