

# Using the PTM-GSEA R Package for Pathway Analysis

## R Markdown

This is a step-by-step description of how I am using the **PTM Signature Enrichment Analysis (PTM-SEA)** analysis developed by Karsten Krug at the Broad Institute. You can find the tool and documentation at: <https://github.com/broadinstitute/ssGSEA2.0> (<https://github.com/broadinstitute/ssGSEA2.0>). You can find the paper describing the methodology and application of the tool here <http://doi.org/10.1074/mcp.TIR118.000943> (<http://doi.org/10.1074/mcp.TIR118.000943>)

This analysis takes ratios of phosphoprotein abundance and returns the pathways that are enriched in proteins that have high (or low) ratios. The paper linked above describes how the enrichment score is calculated, and also how the p-value that describes significance is calculated. Briefly a running sum statistic for each signature is calculated by going through the ranked ratios (highest to lowest) and adding 1 when the phosphoprotein is in the signature and subtracting one when it is not. The enrichment score is the maximum of this statistic. The significance of the enrichment scores is evaluated by a p-value which is a measure of the probability that the enrichment score could be achieved by chance.

## Files Needed to run PTM-SEA

Download the following from <https://github.com/broadinstitute/ssGSEA2.0> (<https://github.com/broadinstitute/ssGSEA2.0>)

1. `ssgea-gui.r`
2. `ssGSEA2.0.R`
3. `ptm.sig.db.all.uniprot.human.v1.8.1.gmt` in the 'db' subfolder

## Input

The input to PTM-SEA is

1. A single site-centric data matrix, `m`, stored in GCT v1.2 or GCT v1.3 format. You can find more details on the GCT file format here [https://software.broadinstitute.org/cancer/software/gsea/wiki/index.php/Data\\_formats#GCT:\\_Gene\\_Cluster\\_Text\\_file\\_format\\_.28.2A.gct.29](https://software.broadinstitute.org/cancer/software/gsea/wiki/index.php/Data_formats#GCT:_Gene_Cluster_Text_file_format_.28.2A.gct.29) ([https://software.broadinstitute.org/cancer/software/gsea/wiki/index.php/Data\\_formats#GCT:\\_Gene\\_Cluster\\_Text\\_file\\_format\\_.28.2A.gct.29](https://software.broadinstitute.org/cancer/software/gsea/wiki/index.php/Data_formats#GCT:_Gene_Cluster_Text_file_format_.28.2A.gct.29)).
2. PTM signatures database (PTMsigDB), eg. `ptm.sig.db.all.uniprot.human.v1.8.1.gmt`

## Instructions:

- In RStudio: open `ssgea-gui.R`
- Press 'Source' in the upper right part of the editor window

- The script will invoke two dialog boxes to specify your input file(s) and databases
- The 1st dialog lets you choose a folder containing input files in GTC v1.2 format.
- The script will loop over all gct files in this directory and run ssGSEA on each file separately.
- The 2nd dialog window lets you choose a gene set database eg.ptm.sig.db.all.uniprot.human.v1.8.1.gmt.
- MAC users: XQuartz is required for invoking file dialogs

## Output

The analysis will return a number of files.

### datanew8\_ssGSEA-combined.gct

this is a gct file that can be opened in excel. It contains the enrichment scores, p-values, and false discovery rate corrected p-values.

### datanew8\_ssGSEA-scores.gct

this is a gct file that can be opened in excel. It contains the enrichment scores.

```
ES=read.table("datanew8_ssGSEA-scores.csv",sep=" ",header=T)
ES[1:6,17:22]
```

```
##          X.12          X.13          X.14          X.15          X.16          X.17
## 1 1min_0min 5min_0min 30min_0min 5min_1min 30min_1min 30min_5min
## 2  -0.9226  -1.1576  -0.9953  -0.3954  -0.3408  0.2314
## 3   1.8412  -0.4871   0.4032  -2.2243  -1.2537  0.7479
## 4   0.0316   1.7904   1.4483   2.1135   1.0258  0.3235
## 5  -1.8804  -0.2882   1.5416   1.5881   2.8848  2.5032
## 6  -1.9593  -0.3543  -1.1681   2.3371   0.1598  -1.0228
```

### datanew8\_ssGSEA-pvalues.gct

This is a gct file that can be opened in excel. It contains the p-values.

### datanew8\_ssGSEA-fdr-pvalues.gct

This is a gct file that can be opened in excel. It contains the false discovery rate corrected p-values.

## Plot scores and p-values as heatmaps

You can also embed plots, for example: `####Plot enrichment scores`

```

library(NMF)
library(dplyr)
scores = read.table("enrichmentscore.csv", sep = ",", stringsAsFactors=F)
scores1 = scores [2:77,2:7]
rownames = scores[2:77,1]
colnames = scores[1,]
scores2 = apply(scores1,2,function(z){data.matrix(z) %>% as.numeric})
colnames(scores2) = scores[1,2:7]
rownames(scores2) = rownames
pdf("heatmap_enrichmentscores.pdf")
aheatmap(scores2,Rowv = TRUE, Colv = TRUE, cellwidth = 10, cellheight = NA)
dev.off()

```

```

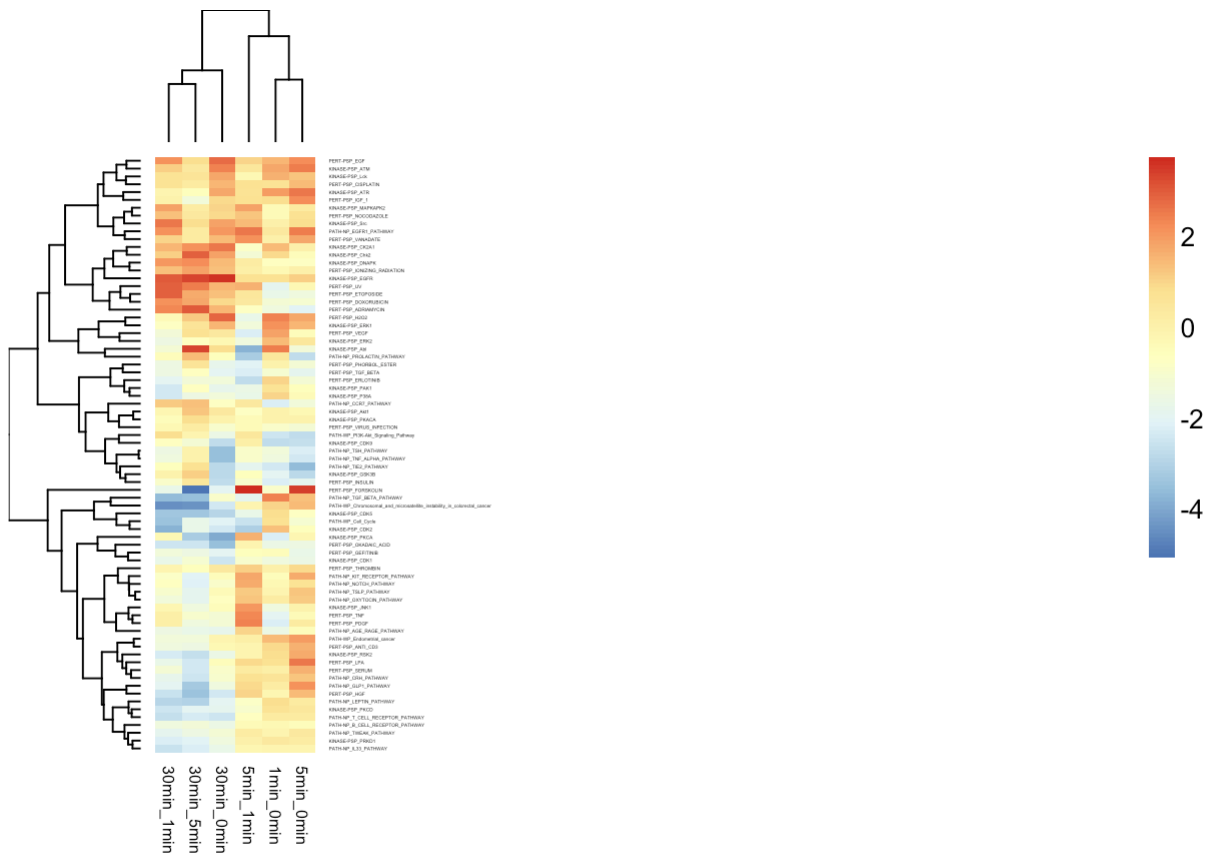
## quartz_off_screen
##                2

```

```

aheatmap(scores2,Rowv = TRUE, Colv = TRUE, cellwidth = 10, cellheight = NA)

```



Plot p-values



In the first column of the scores file, `datanew8_ssGSEA-scores.csv`, is a list of the signatures.

```
ES=read.table("datanew8_ssGSEA-scores.csv",sep=" ",header=T)
ES[2:10,1]
```

```
## [1] PERT-PSP_VIRUS_INFECTIO PERT-PSP_VEGF
## [3] PERT-PSP_VANADATE          PERT-PSP_UV
## [5] PERT-PSP_TNF                PERT-PSP_THROMBIN
## [7] PERT-PSP_TGF_BETA           PERT-PSP_SERUM
## [9] PERT-PSP_PHORBOL_ESTER
## 77 Levels: id KINASE-PSP_Abl KINASE-PSP_Akt1 ... PERT-PSP_VIRUS_INFECTIO
```

In the 10th column of the scores file, `datanew8_ssGSEA-scores.csv`, there is a list of the identifiers for the phosphoproteins included in each signature.

## Phosphoproteins in each signature

```
ES=read.table("datanew8_ssGSEA-scores.csv",sep=" ",header=T)
ES[2:10,10]
```

```
## [1] P04637;S15-p;u|P27361;Y204-p;u|P04637;S46-p;u|Q16539;Y182-p;u|P04637;S20-p;u|Q16539;T180-p;u|Q04206;S276-p;u|Q04206;S536-p;u|P47712;S505-p;u|P16104;S139-p;u|Q9UQM7;T286-p;u|P31749;S473-p;u|P27361;T202-p;u
## [2] Q9UQL6;S259-p;u|P35968;Y1059-p;u|P35968;Y951-p;u|P35968;Y1054-p;u|Q03135;Y14-p;u|Q05397;Y576-p;u|Q05397;Y861-p;u|P19174;Y783-p;u|Q05397;Y397-p;u|Q05397;Y407-p;u|Q05397;Y925-p;u|Q9UQL6;S498-p;u|P29474;S1177-p;u|P31749;S473-p;u
## [3] P27361;Y204-p;u|P16885;Y753-p;u|P16885;Y1217-p;u|Q05397;Y397-p;u|P31749;T308-p;u|Q16539;Y182-p;u|Q16539;T180-p;u|P15311;Y478-p;u|P16284;Y713-p;u|P00533;Y1172-p;u|P25963;Y42-p;u|P31749;S473-p;u|P27361;T202-p;u
## [4] P27361;Y204-p;u|O96017;T68-p;u|Q14683;S957-p;u|O60934;S343-p;u|P00533;Y1092-p;u|P15336;T71-p;u|P78527;T2647-p;u|P04637;S46-p;u|Q16539;Y182-p;u|P00533;Y1069-p;u|P04637;T18-p;u|P04637;S20-p;u|P00533;Y869-p;u|Q16539;T180-p;u|P04637;S37-p;u|P15336;T69-p;u|P04637;S315-p;u|P06493;Y15-p;u|P16220;S133-p;u|O75582;S376-p;u|P05412;S63-p;u|P30307;S216-p;u|P04637;S6-p;u|P40763;S727-p;u|P04637;S33-p;u|O14757;S317-p;u|P16220;S121-p;u|P31749;S473-p;u|P38398;S1524-p;u|P04637;S9-p;u|P04637;S392-p;u|P27361;T202-p;u
## [5] P04792;S82-p;u|P27361;Y204-p;u|Q15418;S380-p;u|Q04206;S468-p;u|P15336;T71-p;u|Q16539;Y182-p;u|P35568;S307-p;u|Q16539;T180-p;u|O14920;Y188-p;u|P16220;S133-p;u|P35568;S312-p;u|P45983;T183-p;u|P00533;S1070-p;u|P45983;Y185-p;u|P00533;T693-p;u|P19838;S932-p;u|P19838;S927-p;u|P27361;T202-p;u|O14920;Y199-p;u
## [6] P04792;S82-p;u|P27361;Y204-p;u|Q05397;Y397-p;u|P31749;T308-p;u|P04792;S78-p;u|Q16539;Y182-p;u|Q16539;T180-p;u|Q04206;S536-p;u|P05106;Y773-p;u|P47712;S505-p;u|P31749;S473-p;u|P27361;T202-p;u
## [7] P49841;S9-p;u|P84022;S204-p;u|Q05397;Y397-p;u|Q16539;Y182-p;u|Q16539;T180-p;u|P84022;S425-p;u|Q15796;S250-p;u|P31749;S473-p;u|Q15796;S467-p;u|P05556;T788-p;u
## [8] P49841;S9-p;u|P24941;T160-p;u|Q15418;S380-p;u|P06730;S209-p;u|P42345;S2448-p;u|P42345;S2481-p;u|P38936;T145-p;u|P06400;S795-p;u|P62753;S235-p;u|P04049;S338-p;u|P49840;S21-p;u|P06400;S780-p;u|P00533;T693-p;u|Q00987;S166-p;u
## [9] P49841;S9-p;u|P03372;S118-p;u|P27361;Y204-p;u|P05771;S661-p;u|P06239;S59-p;u|Q9BZL6;S876-p;u|O43521;S69-p;u|P00533;T678-p;u|P03372;S167-p;u|Q14289;Y402-p;u|Q15418;S380-p;u|P35611;S726-p;u|P06730;S209-p;u|P42345;S2448-p;u|Q05397;Y397-p;u|P31749;T308-p;u|Q16539;Y182-p;u|P35568;S307-p;u|Q16539;T180-p;u|O14920;Y188-p;u|P62753;S235-p;u|P04049;S338-p;u|P45983;T183-p;u|P55211;T125-p;u|Q96A00;T38-p;u|P40763;S727-p;u|Q05655;S645-p;u|Q9UQL6;S498-p;u|P45983;Y185-p;u|P03372;S104-p;u|P00533;T693-p;u|P04049;S43-p;u|P03372;S106-p;u|P27361;T202-p;u|O14920;Y199-p;u|P01730;S433-p;u
## 76 Levels: O60496;Y299-p;u|P16885;Y753-p;u|P19174;Y771-p;u|P29353;Y349-p;u|Q13094;Y128-p;u|P43403;Y292-p;u|P20963;Y142-p;u|O75582;T581-p;u|P19174;Y783-p;u|P15336;T71-p;u|Q16539;Y182-p;u|P42224;S727-p;u|Q16539;T180-p;u|P06127;Y453-p;u|P42229;Y694-p;u|P06493;T14-p;u|P06493;Y15-p;u|P16220;S133-p;u|Q04206;S536-p;u|P04049;Y341-p;u|P43403;Y493-p;u|P16284;Y713-p;u|P45983;T183-p;u|P06239;Y505-p;u|P06239;Y192-p;u|P31749;S473-p;u|P27361;Y204-p;d|P04049;S259-p;d|P27361;T202-p;d ...
```

## Converting identifiers to protein names

You can convert these Uniprot identifiers to protein names here: <https://www.uniprot.org/uploadlists/>  
(<https://www.uniprot.org/uploadlists/>)