
CGSEA User's Guide

Version 1.0

May 2019

Developers: Feng Zhang

License: GPL

URL: <https://github.com/ChengSQXJTU/CGSEA>

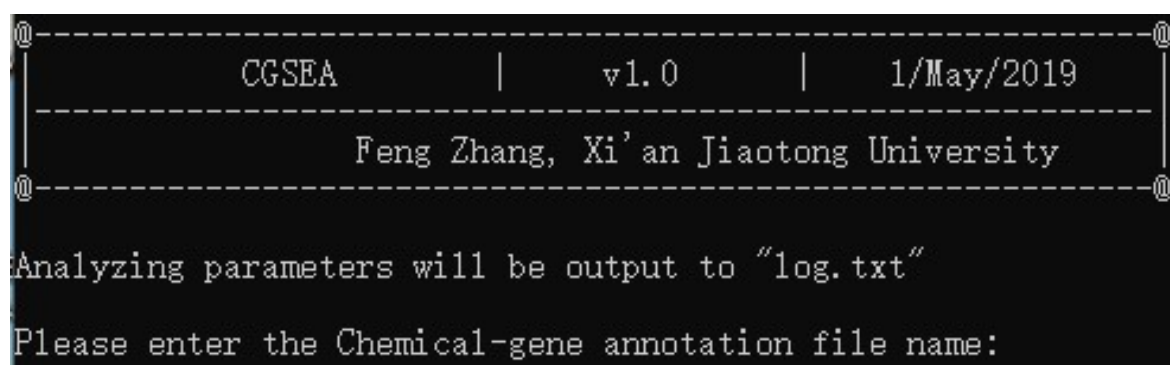
Email: fzhxjtu@xjtu.edu.cn

1. Installing and Running CGSEA

CGSEA is developed by C to interface with R for efficient data analysis. Please make sure that R (<http://www.r-project.org/>) has been installed on your system. CGSEA is a command line based program. Unzip the downloaded "CGSEA" package and run "CGSEA" program at your terminal window to start analysis.

Example:

`./CGSEA`



```
CGSEA      |    v1.0    |    1/May/2019
-----
Feng Zhang, Xi'an Jiaotong University
-----
Analyzing parameters will be output to "log.txt"
Please enter the Chemical-gene annotation file name:
```

2. Parameter Setting

CGSEA will ask users to input a set of parameters before starting analysis:

- [1] Chemical-gene annotation file name:** Please input the storage path and name of Chemical-gene annotation file. We have provided a Chemical-gene annotation file (CRGset.txt) in the CGSEA package. Users can also use the Chemical-gene annotation file prepared by themselves. Please make your Chemical-gene annotation files using following format:

Example:

```
grandidone A    MRPL20
methyl 5-aminolevulinate  TCEB3
fluvalinate     IQCC
```

....

Note: Each line of this file records a chemical (the first column) and corresponding gene (the second column).

- [2] Genome-wide gene expression association testing statistic file name:** Please

input the storage path and file name of genome-wide gene expression association testing statistics of target traits. The gene expression association testing statistics can be driven from transcriptome-wide association studies or gene expression profile studies. Please make your GWAS summary data files using following formats:

Example:

```
HES4  0.041
ISG15  3.276
SDF4   0.011
PER3   0.001
....
```

Note: Each line of this file record the gene (the first column) and corresponding expression association testing statistics (the second column).

- [3] **Permutation times:** Permutations are used for P value calculation in CGSEA. More than 1,000 permutations are recommended for obtaining accurate P values. Note, too large permutation times will make CGSEA taking a long time to complete data analysis.
- [4] **Maximum sizes of chemical response geneset:** Users need to define the maximum (fault value = 1000 genes) sizes of chemical response geneset analyzed by CGSEA.
- [5] **Minimum sizes of chemical response geneset:** Users need to define the minimum (fault value = 5 genes) sizes of chemical response geneset analyzed by CGSEA.

3. Output files

CGSEA will output two result files:

- [1] **CGSEA_result.txt:** Each line of this file records a chemical and corresponding enrichment analysis results.

Example:

Chemical name	NES	P
grandidone A	0.09	0.52
streptobiosamine	1.57	0.06

...

Note: NES denotes the normalized enrichment score statistics of CGSEA.

- [2] **NESplot.pdf:** Plot of CGSEA analysis results of target trait. In generated figures, each point denotes a chemical. X-axis presents the total number of chemicals analyzed by CGSEA. Y-axis shows $-\log_{10}(P \text{ values})$ calculated by CGSEA.
-

