# TREMSUCS

***Release 1.0***

**Gabor Balogh**

**Jun 20, 2024**

# CONTENTS:

"TREMSUCS" a tool to choose, harvest and analyse expression and methylation data of the TCGA-projects for revealing Biomarkers which indicate treatment success.

CHAPTER

ONE

# EXAMPLE REPORT:

An example report can be downloaded here. Be aware that this report has a size of about 300 MB.

CHAPTER

TWO

# INSTALLING FROM GITHUB.COM:

```
$ git clone https://github.com/dendemayer/TREMSUCS-TCGA.git
$ cd TREMSUCS-TCGA
$ pip install .
```

To start the analysis with help of the interactive mode, call the pipeline without any argument:

```
$ TREMSUCS
```

Calling the help or the manual page:

```
$ TREMSUCS --help
$ man TREMSUCS
```

CHAPTER

# THREE

# HELP PAGE OF THE PIPELINE:

## 3.1 TREMSUCS

"TREMSUCS" a tool to choose, harvest and analyse expression and methylation data of the TCGA-projects for revealing Biomarkers which indicate threapy specific treatment success predictions.

Calling the pipeline without any argument starts the interactive mode to help setting all needed parameters for the analysis.

```
TREMSUCS [OPTIONS]
```

### Options

**-o, --out_path** <out_path>

path to save the result files

**Default**

'/home/dende/TREMSUCS'

**-p, --project** <project>

TCGA project(s) to be applied. Any TCGA project can be chosen, like: -p TCGA-CESC -p TCGA-HNSC ...

**-d, --drugs** <drugs>

drug(s), like: -d drug1 -d drug2 or drugcombination(s), like: -d drug1,drug2

**-c, --cores** <cores>

number of cores provided to snakemake

**Default**

1

**-C, --cutoff** <cutoff>

Cut-off parameter

**Default**

0

**-t, --threshold** <threshold>

threshold parameter

**Default**

0

**-e, --execute** <execute>

choose which pipeline shall be executed

**Default**
'DESeq2', 'metilene'

**-N, --dryrun**

snakemake dryrun

**Default**
False

**-D, --download**

if set, just download raw and meta data for given projects and analysis types, revise them, link them, but do not run any analysis

**Default**
False

**-u, --unlock**

in case the analysis crashs, snakemake locks the output directory, run with -u to unlock, then repeat the analysis

**Default**
False

**-v, --version**

printing out version information: Version 1.0

CHAPTER

FOUR

# SHORT TUTORIAL:

## 4.1 Usage of the interactive mode:

The following example composition of projects, drugs and parameters creates the configuration given in the example report here.

The same configuration can be applied by issuing the following command (the number of cores hereby can be adjusted and would also give the same results):

```
$ TREMSUCS -p TCGA-CESC -p TCGA-HNSC -p TCGA-LUSC -d cisplatin -d carboplatin,paclitaxel␣
↪\
-d carboplatin -o /scr/TREMSUCS_out -c 40 -t 5 -t 10 -t 20 -C 5 -C 8
```

Calling the pipeline without any argument starts the interactive mode:

```
$ TREMSUCS

 OUTPUT_PATH:              /homes/biertruck/gabor/TREMSUCS
 SCRIPT_PATH:              /homes/biertruck/gabor/phd/test_git_doc/TREMSUCS/src/shared/
↪modules
 PIPELINES executed:       ['DESeq2', 'metilene']

 which projects do you want to include in your analysis:

  0:     TCGA-CESC           Cervical Squamous Cell Carcinoma and Endocervical␣
↪Adenocarcinoma
  1:     TCGA-HNSC           Head and Neck Squamous Cell Carcinoma
  2:     TCGA-LUSC           Lung Squamous Cell Carcinoma
  3:     TCGA-ESCA           Esophageal Carcinoma
  4:     TCGA-BRCA           Breast Invasive Carcinoma
  5:     TCGA-GBM            Glioblastoma Multiforme
  6:     TCGA-OV             Ovarian Serous Cystadenocarcinaoma
  7:     TCGA-LUAD           Lung Adenocarcinoma
  8:     TCGA-UCEC           Uterine Corpus Endometrial Carinoma
  9:     TCGA-KIRC           kindney renal clear cell carcinoma
 10:     TCGA-LGG            brain lower grade glioma
 11:     TCGA-THCA           thyroid carcinoma
 12:     TCGA-PRAD           prostate adenocarcinoma
 13:     TCGA-SKCM           skin cutaneous melanoma
 14:     TCGA-COAD           colon adenocarcinoma
 15:     TCGA-STAD           stomach adenocarcinoma
```

(continues on next page)

(continued from previous page)

```
 16:     TCGA-BLCA           bladder urothelial carcinoma
 17:     TCGA-LIHC           liver hepatocellular carcinoma
 18:     TCGA-KIRP           kidney renal papillary cell carcinoma
 19:     TCGA-SARC           sarcoma
 20:     TCGA-PAAD           pancreatic adenocarcinoma
 21:     TCGA-PCPG           pheochromocytoma and paraganglioma
 22:     TCGA-READ           rectum adenocarcinoma
 23:     TCGA-TGCT           testicular germcelltumors
 24:     TCGA-THYM           thymoma
 25:     TCGA-KICH           kidney chromophobe
 26:     TCGA-ACC            adrenochordical carcinoma
 27:     TCGA-MESO           mesothelioma
 28:     TCGA-UVM            uveal melanoma
 29:     TCGA-DLBC           lymphoid neoplasm diffuse large b-cell lymphoma
 30:     TCGA-UCS            uterine carcinoma
 31:     TCGA-CHOL           cholangiocarcinoma
 enter your choices one by one, when you are done, simply press "Enter":
```

As suggested, you can now, one by one include the projects you are interested in. A default OUTPUT_PATH is also already given together with the default analysis types “DESeq” and “metilene”. Those defaults can also be adjusted in next steps with help of the interactive mode.

To recreate the example set, the first three projects have to be selected, afterwards the following prompt is given:

```
you choose:
PROJECTS:        ['TCGA-CESC', 'TCGA-HNSC', 'TCGA-LUSC']

which therapy approach do you want to include in your analysis:

 0: cisplatin                                TCGA-CESC: 103 TCGA-HNSC: 64 TCGA-LUSC: 1
 1: carboplatin,paclitaxel                   TCGA-CESC: 5 TCGA-HNSC: 26 TCGA-LUSC: 14
 2: 5-fluorouracil,cisplatin                 TCGA-CESC: 5 TCGA-HNSC: 2 TCGA-LUSC: 0
 3: carboplatin                              TCGA-CESC: 3 TCGA-HNSC: 6 TCGA-LUSC: 3
 4: carboplatin,cisplatin,paclitaxel         TCGA-CESC: 3 TCGA-HNSC: 0 TCGA-LUSC: 1
 5: cisplatin,gemcitabine                    TCGA-CESC: 3 TCGA-HNSC: 0 TCGA-LUSC: 9
 6: paclitaxel                               TCGA-CESC: 2 TCGA-HNSC: 1 TCGA-LUSC: 0
 7: erbitux                                  TCGA-CESC: 1 TCGA-HNSC: 9 TCGA-LUSC: 0
 8: cisplatin,vectibix                       TCGA-CESC: 0 TCGA-HNSC: 5 TCGA-LUSC: 0
 9: carboplatin,erbitux,paclitaxel           TCGA-CESC: 0 TCGA-HNSC: 4 TCGA-LUSC: 0
10: cisplatin,erbitux                        TCGA-CESC: 0 TCGA-HNSC: 3 TCGA-LUSC: 0
11: carboplatin,cisplatin,erbitux,paclitaxel TCGA-CESC: 0 TCGA-HNSC: 3 TCGA-LUSC: 0
12: carboplatin,cisplatin                    TCGA-CESC: 0 TCGA-HNSC: 2 TCGA-LUSC: 0
13: docetaxel,erbitux                        TCGA-CESC: 0 TCGA-HNSC: 2 TCGA-LUSC: 0
14: cisplatin,docetaxel                      TCGA-CESC: 0 TCGA-HNSC: 1 TCGA-LUSC: 10
15: carboplatin,docetaxel                    TCGA-CESC: 0 TCGA-HNSC: 1 TCGA-LUSC: 3
16: cisplatin,vinorelbine                    TCGA-CESC: 0 TCGA-HNSC: 0 TCGA-LUSC: 21
17: carboplatin,vinorelbine                  TCGA-CESC: 0 TCGA-HNSC: 0 TCGA-LUSC: 8
18: cisplatin,etoposide                      TCGA-CESC: 0 TCGA-HNSC: 0 TCGA-LUSC: 7
19: carboplatin,gemcitabine                  TCGA-CESC: 0 TCGA-HNSC: 0 TCGA-LUSC: 5
20: cisplatin,pemetrexed                     TCGA-CESC: 0 TCGA-HNSC: 0 TCGA-LUSC: 3
21: cisplatin,docetaxel,gemcitabine          TCGA-CESC: 0 TCGA-HNSC: 0 TCGA-LUSC: 2
22: carboplatin,gemcitabine,paclitaxel       TCGA-CESC: 0 TCGA-HNSC: 0 TCGA-LUSC: 2
23: carboplatin,cisplatin,vinorelbine        TCGA-CESC: 0 TCGA-HNSC: 0 TCGA-LUSC: 2
```

(continues on next page)

(continued from previous page)

```
24: carboplatin,docetaxel,gemcitabine        TCGA-CESC: 0 TCGA-HNSC: 0 TCGA-LUSC: 2
25: carboplatin,docetaxel,paclitaxel         TCGA-CESC: 0 TCGA-HNSC: 0 TCGA-LUSC: 2
26: gemcitabine                              TCGA-CESC: 0 TCGA-HNSC: 0 TCGA-LUSC: 2

enter your choices one by one, when you are done, simply press "Enter":
```

Here are therapies listed where the maximum of a row is greater than 1. We apply row 0, 1 and 3 to include cisplatin, the combination of carboplatin and paclitaxel and cases which got solely treated with carboplatin. In the following, every other parameter is requested. With the next prompt, the default OUTPUT_PATH can be confirmed or replaced:

```
do you want to keep the default OUTPUT_PATH of:
/homes/biertruck/gabor/TREMSUCS
if so, press ENTER, if not, enter your custom output path:
```

**In this example, we confirm the suggested OUTPUT_PATH and are asked to confirm**
or set the number of cores which shall be invoked into the analyses:

```
do you want to keep the default number of cores invoked of 1?
if so, press ENTER, if not, enter the number of cores:
40
```

We set the cores to 40 and then can decide which analysis approaches shall be triggered, per default, DESeq2 and metilene based biomarker predictions are produced:

```
which pipeline do you want to include into your analysis
press ENTER if DESeq2 and metilene (default) or
1 for DESeq2 or
2 for metilene
```

We confirm the default of those two analyses and can set the cutoff values, if we want to add those at all:

```
do you want to add one or multiple cutoffs?
it is recommend to choose cutoff values between 5 and 10 years
if not, just press ENTER, if so enter the coutoffs one by one:
5
8
```

Like the example set, we add here a cutoff of 5 and 8. Then the thresholds are requested:

```
do you want to add one or multiple thresholds?
it is recommend to choose threshold values which do not exceed a value of 50
if not, just press ENTER, if so enter the thresholds one by one:
5
10
20
```

We apply thresholds of 5, 10 and 20. All mandatory and optional parameters are set with that and are finally listed before the whole approach is started:

```
OUTPUT_PATH:             /homes/biertruck/gabor/TREMSUCS
PROJECT:                 ['TCGA-CESC', 'TCGA-HNSC', 'TCGA-LUSC']
DRUGS:                   ['carboplatin', 'carboplatin,paclitaxel', 'cisplatin']
pipelines executed:      ['DESeq2', 'metilene']
```

(continues on next page)

(continued from previous page)

```
cores:                  40
cutoff:                 [0, 5, 8]
threshold:              [0, 5, 10, 20]
press ENTER to start or q to quit:
```

If something went wrong, you can quit now and start over, or of course start the analysis.

# INDEX

## Symbols



## T