# ModulOmics Manual

Dana Silverbush[*†1], Simona Cristea[*† 2,3,4,5], Gali Yanovich[6], Tamar Geiger[6], Niko Beerenwinkel[‡2,3] and Roded Sharan[‡ 1]

[1]Blavatnik School of Computer Science, Tel Aviv University, Tel Aviv 69978, Israel
[2]Department of Biosystems Science and Engineering, ETH Zurich, Basel, Switzerland
[3]SIB Swiss Institute of Bioinformatics, Basel, Switzerland
[4]Dana-Farber Cancer Institute, Boston, Massachusetts, USA
[5]Harvard T.H. Chan School of Public Health, Boston, Massachusetts, USA
[6]Department of Human Molecular Genetics and Biochemistry, Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv 69978, Israel

**ModulOmics** is a method to identify cancer driver pathways de novo by integrating multiple data types (protein-protein interactions, mutual exclusivity of mutations or copy number alterations, transcriptional co-regulation, and RNA co-expression) into a single probabilistic model. ModulOmics uses patient data, and protein-protein interaction networks, based here on the Hippie database [1] and regulatory connections, based here on the Transcriptional Regulatory Relationships Unraveled by Sentence-based Text mining database (TRRUST) [2].

The following manual describes the R code for ModulOmics – including the input and output data, and how driver modules can be easily inferred. Example files are based on Her2 breast cancer subtype from TCGA [3].

**Dependencies**

```
library("TiMEx")
library("igraph")
library("cplexAPI")
library("gtools")
```

The R code uploaded here contains all the functions needed for ModulOmics and a wrapper function named *run_ModulOmics*.

**Libraries**

**Code** containing callPipeSP, funcsTiMExPPI.R, ilpSolver.R and librariesTiMExPPI.R, and an inner library data downloaded here.
**Input** mutation and expression input files
**output** for result files

**Input**

**mutInput** location of the binary genetic alterations file

**zScoreInput** location of the zScored expression file
PPIInput location of the shortest path score for each pair of proteins, computed based on the PPI network
**TRRUSTInput** full path for the regulatory network
**pathCode** full path for the library containing funcsTiMExPPI.R, ilpSolver.R and librariesTiMExPPI.R
**pathOutput** full path for the output directory
**quantNo** quantile to consider active transcription factor, default value = 3
**topResults** number of modules to retreive from each cluster, default value = 5
**noClusters** number of clusters to start to stochastic search in order to locate global maximum, default value = 10
**SETS** number of initial seed modules retreived from the ILP, default value = 200
**LOWER_LIMIT** lower K value (smallest modules), default value = 2, cannot drop below 2
**UPPER_LIMIT** upper K value (largest modules), default value = 5, cannot exceed 6

**Output**
**Return value** A matrix of the detected modules and their single and set scores.
**Files** The function saves (K-1) files named "ModuleOmicsK.txt" to the output library**,** containing the sets and scores of each K.

**Code example**

```
# format: column names are gene followed by either "-Mut" or "-CNA, rownames are patients. Data as downloaded from TCGA
mutInput<-"<full path>/input/Her2.txt"

# format: column names are genes, rownames are patients. Data as downloaded from TCGA
zScoreInput<-"<full path>/input/Her2_zScore.txt"
pathOutput<-"<full path>/output/"
pathCode<-"<full path>/code/"

source(paste(pathCode,"callPipeSP.R",sep=""))
# The function reads cancer cohort and computes ModuleOmics scores
results<-run_ModulOmics(mutInput, zScoreInput, pathCode, pathOutput)
```

**References**

[1] Martin H. Schaefer, Jean-Fred F. Fontaine, Arunachalam Vinayagam, Pablo Porras, Erich E. Wanker, and Miguel A. Andrade-Navarro. HIPPIE: Integrating protein interaction networks with experiment based quality scores. PLoS ONE, 7(2), 2012.

[2] Heonjong Han, Hongseok Shim, Donghyun Shin, Jung Eun Shim, Yunhee Ko, Junha Shin, Hanhae Kim, Ara Cho, Eiru Kim, Tak Lee, et al. Trrust: a reference database of human transcriptional regulatory interactions. Scientific Reports, 5:11432, 2015.

[3] Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. Nature, 490(7418):61–70, 2012.