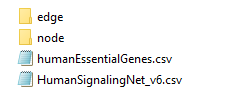
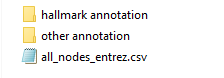
**Common files folder**

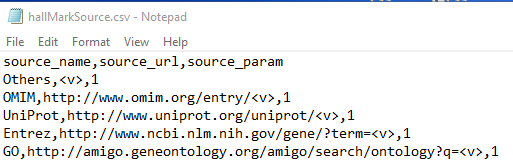


* HumanSignalingNet\_v6.csv: the baseline network from which views are created.
* humanEssentialGenes.csv: contains list of essential genes

**node subfolder**



* all\_nodes\_entrez.csv: list of unique nodes in HumanSignalingNet\_v6.csv
* hallmark annotation subfolder: contains all files associated with hallmark annotation of nodes



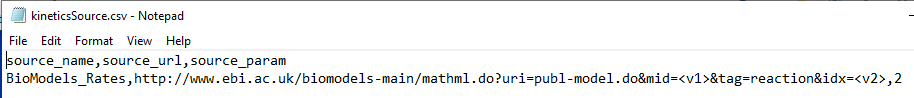
Note that for hallMarkSource.csv, source\_param indicates the number of parameter to be replaced in source\_url. For instance in the first second row (i.e., OMIM), there is one parameter denoted as <v> in source\_url. The exact hyperlink related to the annotation can be recreated by setting <v> with the corresponding source id in hallMarkAnnotation.csv

* other annotation subfolder: contains various annotation such as OMIM and GO of the nodes.

**edge subfolder**

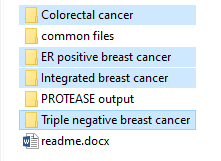


* kinetic annotation subfolder: contains all files associated with kinetic annotation of edges



Similar to hallMarkSource.csv, source\_param in kineticsSource.csv indicates number of parameters (e.g., <v1>, <v2>) in source\_url that should be replaced to retrieve the hyperlink. The exact values to replace the parameters can be found in rate id of kineticsRateAnnotation.csv

**View folders**



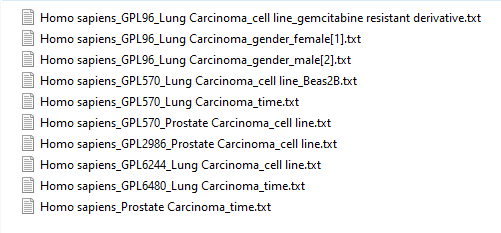
Each view folder contains file specific to that particular view which can be related to a specific cancer subtype, a specific treatment, or organism or microarray platform etc. Note that the files in Colorectal cancer, ER positive breast cancer and Triple negative breast cancer were manually extracted earlier for TROVE using the steps mentioned in Section 3.3 of TROVE User Guide (Win).pdf

There are 2 files in each of these folders:

* mutation frequency: COSMIC\_xxxx.csv
* gene expression fold change: GEO2R\_xxxx.csv

Note that it is not necessary for every view to have both files. For install, Integrate breast cancer folder only has integratedBC.csv which integrates data from multiple subtypes of breast cancer. In this case, the fold changes column in integratedBC.csv.

**PROTEASE output subfolder**



Each file is generated by PROTEASE and similar to GEO2R\_xxxx.csv in the above mentioned.

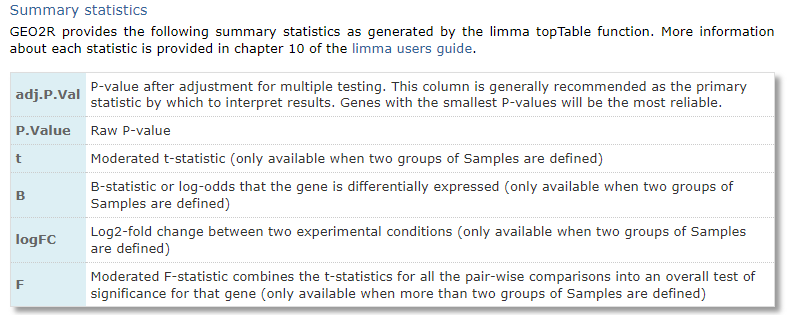
The file name or title provides context to the view. Each title is a concatenation of the most representative terms in the set of GSM records in the Gene Omnibus Repository that is used to generate this particular set of DEG. The representative terms are joined using underscores. See PROTEASE Readme.docx (pages 9 and 10) in Protease\_online git repo for more details of how the representative terms are obtained.

E.g. “Homo sapiens\_GPL96\_Lung Carcinoma\_cell line\_gemcitabine resistant derivative.txt” (in RTest folder) have representative terms “Homo sapiens”, “GPL96”, “Lung Carcinoma”, “cell line” and “gemcitabine resistant derivative” which correspond to organism, platform, associated disease, experimental type and specific description of the experimental type.

In the .txt, it consists of 8 columns: “ID”, “adj.P.Val”, “P.Value”, “t”, “B”, “logFC”, “Gene.symbol”, “Gene.title”.

Note that some .txt files (e.g., Homo sapiens\_Prostate Carcinoma\_time.R in RTest folder) are generated as 6 columns: “ID”, “adj.P.Val”, “P.Value”, “F”, “Gene.symbol”, “Gene.title”. The different formatting of the R files are in accordance to the GEO2R feature in GEO Browser. Details of generation of the R files is found in page 10 of PROTEASE\_Readme.docx in PROTEASE\_online git repo.

The definition of “adj.p.Val”, “P.Value”, “t”, “B”, “logFC” and “F” can be found in <https://www.ncbi.nlm.nih.gov/geo/info/geo2r.html>



“ID”: Probe ID corresponding to the specific platform.

“Gene.symbol”: Entrez official symbol

“Gene.title”: Entrez official full name