

t-SNE Analysis Tool – JAVA

For Visualization and Analysis of gene expression data

This takes advantage of the t-SNE algorithm developed by L.J.P. van der Maaten and G.E. Hinton.

Full details and code are available from <http://homepage.tudelft.nl/19j49/t-SNE.html>

L. van der Maaten and G. Hinton, Journal of Machine Learning Research 9, 2579 (2008).

The JAVA implementation of t-SNE allows command line controlled process of data using the t-SNE algorithm. Use requires Java SE 6. For Mac users this can be specified as the preferred versions using Java Preferences in the Applications/Utilities folder.

Download the “visgenex_” folder. The program is run from a terminal window. Open a terminal window and change the directory to /visgenex/. (e.g. \$ cd /Users/bob/visgenex/). To run the program:

```
$ Java -jar visgenex.jar [OPTIONS]
```

The following options can be specified:

```
-i <filename> - set the csv input filename, default: 'input.csv'  
-h <y/n> - is the first row of the input file column headers instead of data? default: y  
-o <filename> - set the csv output filename, default: '<inputfilename>_output.csv'  
-p <integer> - perplexity value, default: 30  
-gs <integer> - number of replicates per condition, this option will group adjacent columns  
into groups of <integer> and takes their mean, default: 1 (no replicates)  
ip <text> - pre-processing of gene expression values to be passed to t-SNE:  
    'raw' - unprocessed data is passed to t-SNE  
    'z-score' - standardized (z) scores are calculated and passed to t-SNE  
    'zero-means' - subtracts the row average from each row before t-SNE  
    'fold-induction' - divides entire row by its lowest value  
    'percent-induction' - subtract entire row by lowest value in the row and then divides by  
    the highest value (scales entire row (0=<n=<1)  
    default: 'z-score'
```

The input file should be a csv file containing a first column of gene names followed by columns containing gene expression values for each sample assayed. The first row can contain a header indicating the sample name in which case specify option “-h y” otherwise specify “-h n”.

The output will be a csv file containing three columns. The first column is the gene name the second and third column are the X and Y coordinates of the gene in the visualization space. These can be plotted with the graphing program of your choice.

Example:

```
visgenex bob$ java -jar visgenex.jar -i data.csv -o tSNEdata -h n -gs 3 -ip z-score
```

Will take a file “data.csv” which contains gene expression data from experiments performed in triplicate and use the z-scores from these data to generate a t-SNE output.

FootNote

Created for the Francis Crick Institute

Programmed by James N Smith and Chris Watkins, Royal Holloway College, London University.

Developed with James Briscoe, and Natascha Bushati

t-SNE Algorithm developed by: L.J.P van der Maaten and G.E. Hinton

more info available at <http://homepage.tudelft.nl/19j49/t-SNE.html>

selectdata.m - John d'Errico - <http://www.mathworks.com/matlabcentral/fileexchange/13857>

Last update – 8 June 2015