Description:

The ASTRIX algorithm takes in as input transcriptomics data and a list of transcription factors. The algorithm outputs a transcriptional regulatory network (TRN) that contains interactions between transcription factors and transcripts in the transcriptomics data. The attached algorithm is implemented in MATLAB and requires the MATLAB bioinformatics toolbox and JAVA.

Syntax:

**>>[TRN, correl, rmsdval] = function\_astrix\_2015(data, probeids, tfprobes, N, z, RMSDTHRESHOLD)**

**Inputs:**

**data – This is the transcriptomics data with rows corresponding to transcripts and columns corresponding to different conditions.**

**probeids – This contains the name of the genes or probes (row labels) in the transcriptomics data. probeids should have the same length as the number of rows in data**

**tfprobes – This contains the names of the transcription factors in the transcriptomics data. The name/IDs used should match the corresponding names in probeids.**

**Optional Inputs:**

**N** – This represents the type of cross validation performed to test the fit of the TRN to the transcriptomics data. Default value is 10 (10-fold cross validation).

**Z** – This indicates if the algorithm is run to test the fit of the TRN to the transcriptomics data (Z = 0) or to infer a TRN from the entire transcriptomics data (Z = 1). If Z is set to be 1, then this sets N to be 1 and overrides the value of N set earlier.

**RMSDTHRESHOLD** – This value is the RMSD (Root Mean Square Deviation) cut off for finding significant interactions (default = 0.5). The lower the value, the more stringent the criterion for finding significant interactions.

**Outputs:**

**TRN –** This contains the transcriptional regulatory network. This *structure* variable contains the list of TFs, corresponding targets, the RMSD and Pearson’s correlation value for each interaction. This output is only provided if running ASTRIX for the entire dataset (Z = 1). The TRN is empty when running in cross validation mode (Z = 0).

**correl –** This containsthePearson’s correlation values in the training and test set for all the probes in the transcriptomics data. This contains only the training set correlation values if ASTRIX is applied to the entire dataset (Z = 1).

**rmsdval –** This containstheRMSD values in the training and test set for all the probes in the transcriptomics data. This contains only the training set RMSD values if ASTRIX is applied to the entire dataset (Z = 1).

**Notes:**

1. Make sure that the associated files (LARS MATLAB file, ARACNE JAVA file) are in the same folder and search path as the ASTRIX code.
2. The ARACNE algorithm used by ASTRIX requires JAVA to execute. For OSX update the environmental variables (DYLD\_LIBRARY\_PATH, DYLD\_FRAMEWORK\_PATH) so that Java is in the system path.
3. Download the Java Runtime Environment (<http://www.oracle.com/technetwork/java/javase/downloads/jre8-downloads-2133155.html>) and note the installation directory.
4. For PC, if you install it in the C drive, the path should be C:\ Java\jre1.8.0\_121\bin. Update the system call (line 170) to the right path if otherwise.
5. Its recommended that you use a higher RMSDTHRESHOLD (~ 1) for the inference and then prune out less significant interactions as a post processing step.

Example:

**>> load bee\_data\_challenge Expression probeids tfprobes**

**>> [TRN, correl,rmsdval] = function\_astrix\_2015(Expression, probeids ,tfprobes,10,0,1);**

This runs ASTRIX on the entire honeybee gene expression dataset and outputs interactions with RMSD less than 1. A high RMSD threshold is recommended so that the entire set of interactions are outputted by the algorithm. These interactions can be prioritized using Pearson’s correlation or a more stringent RMSD threshold

**>> [TRN, correl,rmsdval] = function\_astrix\_2015(Expression,probeids ,tfprobes,1,1,1);**

This runs 10-fold cross validation on the entire honeybee gene expression dataset and outputs the training and test set correlation and RMSD values for each round of cross validation.