



ARACNE Documentation

Description: Runs the ARACNE algorithm
Author: Marc-Danie Nazaire (Broad Institute), gp-help@broad.mit.edu
Date: 06/01/07
Release: 1.0

Summary:

ARACNE (**A**lgorithm for the **R**econstruction of **A**ccurate **C**ellular **N**etworks) is an algorithm which reverse engineers a gene regulatory network from microarray gene expression data. ARACNE uses mutual information(MI), an information theoretical measure, to compute the correlation between pairs of genes and infer a best-fit network of probable interactions (I.E. an MI score of 0 between two genes implies they are independent of each other). The ARACNE algorithm first calculates the MI between each pair of all genes or a subset of genes and creates an adjacency matrix. Afterwards, MI thresholding is applied using a specified MI threshold value or p-value of the MI score. MI scores of gene pairs which fall below the MI threshold will be removed from the adjacency matrix. Finally, the DPI(Data Processing Inequality) tolerance is used to reduce the number of genes which do not interact directly but were missed in the MI thresholding step.

References:

- Margolin, A., et al., *ARACNE: An Algorithm for the Reconstruction of Gene Regulatory Networks in a Mammalian Cellular Context*. BMC Bioinformatics, 2006. **7**(Suppl 1): p. S7.
- Basso, K., et al., *Reverse engineering of regulatory networks in human B cells*. Nat Genet, 2005. **37**(4): p. 382-390.

Parameters:

Name	Description
dataset.file	Dataset - .res, .gct
hub.gene	The name of one gene whose network interactions you want to reconstruct. This is mutually exclusive with the hub.genes.file parameter.
hub.genes.file	A file containing a list of hub genes - .txt. This is a text file containing a subset of genes from the input dataset whose network interactions you are interested in reconstructing. This is mutually exclusive with the hub gene parameter.
transcription.factor.file	A file containing a list of all genes that encode transcription factors. Specifying transcription factors allows DPI to be applied more intuitively when reconstructing a transcriptional interaction network. As a result, indirect interactions of a transcription factor via another transcription factor will be removed
kernel.width	The kernel width (or window width) of the Gaussian Kernel Estimator (estimates the probability density function of the dataset). The kernel width affects the smoothness of the density function. Default value is 0.15.

GenePattern

mi.threshold	Threshold for a mutual info (MI) estimate to be considered statistical different from zero.
p.value	Significance level for a MI estimate to be considered statistically different from zero. This p-value is ignored if an MI threshold is specified. Default value is 1. This indicates no threshold is applied on the MI scores.
dpi.tolerance	The percentage of MI estimation considered as sampling error. For example if three genes A, B, C form a loop with gene pairs AB, BC, and AC. Then gene pair AB would be removed if: $(MI \text{ of gene AB}) \leq (1-e) (MI \text{ of gene AC})$ and $(MI \text{ of gene AB}) \leq (1-e) (MI \text{ of gene BC})$ where e is dpi.tolerance. Corresponding calculations would be done for gene pair BC and AC. The DPI tolerance is normally between 0 and 0.15 since values larger than 0.15 yields higher false positives.
mean.filter	Filter out non-informative genes whose mean expression value is smaller than mean.filter
cv.filter	Filter out non-informative genes whose coefficient of variance is smaller than cv.filter
output.file	The name of the output file - .adj

Output Files:

1. ARACNE (.adj) result file. The adjacency matrix file format is:

>	Input file	input_file.exp				
>	ADJ file					
>	Output file	output_file.adj				
>	Algorithm	accurate				
>	Kernel width	0.15				
>	No. bins	6				
>	MI threshold	0.065				
>	MI P-value	1				
>	DPI tolerance	0.15				
>	Correction	0				
>	Subnetwork file	hub_genes.txt				
>	Hub probe					
>	Control probe					
>	Condition					
>	Percentage	0.35				
>	TF annotation	transcription_factor_list.txt				
>	Filter mean	0.0				
>	Filter CV	0.0				
	GeneId1	GeneId2	0.07	GeneId5	0.16	...
	GeneId2	GeneId1	0.07	GeneId3	0.33	...
	GeneId3	GeneId2	0.33	GeneId9	0.09	

GenePattern

The first 18 lines of the .ADJ file start with “>” and contain all the parameters used to create the file. The remaining lines are formatted such that the first column in a row is the gene being reported on (or the hub gene) and the rest of the columns are gene and MI value pairs. In the example above, the MI between GeneId1 and GeneId2 is 0.07 and the MI between GeneId1 and GeneId5 is 0.16. The adjacency matrix is symmetric such the MI between gene A and B is the same as the MI between gene B and A.

Platform dependencies:

Module type:	Reverse Engineering
CPU type:	any
OS:	any
Java JVM level:	1.5
Language:	Java