|  |
| --- |
| Suplementary Data 2  **BADR: A method to infer Gene Regulatory**  **Networks from Gene Expression Time Series**  Luis F Iglesias-Martinez, Tapesh Santra\* and Walter Kolch |

**List of Contents**

|  |  |  |
| --- | --- | --- |
| **Content** | **Description** | **Page** |
| 1.BADR | The basic algorithm to infer gene regulatory networks from gene expression time series (e.g. DREAM4 networks) | 1 |
| 2.BADR\_Uneven | BADR version for scenarios where the number of time points between replicates differ (e.g. IRMA dataset) | 3 |
| 3.BADR\_Reduced | BADR version for very large systems (e.g. Breast cancer datasets) | 5 |

**1. BADR**

**Basic Usage**

**Input:**

Mandatory: Gene Expression Time Series, Time Differences, Number of Time Points, Number of Replicates

Optional: Transcription Factor Identifiers and Prior Probability Matrix

Gene Expression Time Series:

A matrix where each column represent a gene and each row a time series measurement.

Formula:

Time Differences:

A vector representing the differences between time points. The units are to be chosen by the user.

Formula:

Number of Time Points:

A scalar representing the number of time points. On our nomenclature it would be equal to .

Number of Replicates:

A scalar representing the number of replicates. On our nomenclature it would be equal to .

Transcription Factor Identifiers:

A vector with number representing the columns of the gene expression time series matrix where the transcription factors are located

Formula:

Prior Probability:

A matrix with the probability that each gene regulates the other.

Formula:

**Output:** Gene Regulatory Network and Coefficients

Gene Regulatory Network:

A matrix containing the posterior probability that genes regulate each other.

Coefficients:

A matrix containing the coefficients in the models after performing Bayesian Model Averaging. Can be used as an indicator of whether the interaction is activating or inhibiting.

**Basic Usage:**

[Adjacency\_Matrix, Coefficients] = BADR(Gene\_Exp, Dt, No\_Time\_Points, No\_Replicates);

**Transcription Factor IDs Usage:**

[Adjacency\_Matrix, Coefficients] = BADR(Gene\_Exp, Dt, No\_Time\_Points, No\_Replicates, Transcription\_Factors);

**Prior and Transcription Factors IDs Usage**

[Adjacency\_Matrix, Coefficients] = BADR(Gene\_Exp, Dt, No\_Time\_Points, No\_Replicates, Transcription\_Factors, Prior);

**Example Basic Usage: Dream 4 Network 1**

mRNA1 = dlmread('insilico\_size10\_1\_timeseries.tsv');

No\_Replicates = 5;

No\_Time\_Points = 21;

dt = 50/60;

dt = repmat(dt, No\_Time\_Points-1,1);

[Adjacency\_Matrix]=BADR(mRNA1, dt, No\_Time\_Points, No\_Replicates);

**2. BADR\_Uneven (e.g. IRMA)**

BADR can be used for a scenario where the number of time points differ between replicates. We just have to calculate the discretised differentials as follows:

**Input:**

Mandatory: Gene Expression Time Series, Time Points, Replicates

Gene Expression Time Series:

A matrix where each column represent a gene and each row a time series measurement.

Formula:

Time Points:

A vector with the time points per replicates that has the same order as the rows in the gene expression time series:

Replicates:

A vector with the replicate identifier that coincides with the order of the rows in the gene expression time series:

**Optional:**

Transcription Factor Identifiers:

A vector with number representing the columns of the gene expression time series matrix where the transcription factors are located

Formula:

Prior Probability:

A matrix with the probability that each gene regulates the other.

Formula:

**Output:** Gene Regulatory Network and Coefficients

Gene Regulatory Network:

A matrix containing the posterior probability that genes regulate each other.

Coefficients:

A matrix containing the coefficients in the models after performing Bayesian Model Averaging. Can be used as an indicator of whether the interaction is activating or inhibiting.

**Example: IRMA**

IRMA\_1 = dlmread('IRMA\_Switch\_ON.txt');

Times = IRMA\_1(:,1);

Dt = Times(2:end)-Times(1:end-1);

F = [find(Dt<0);length(Times)-1];

Replicates = zeros(size(Times));

S = 1;

for i = 1:length(F)

Replicates(S:F(i)+1) = i;

S = F(i)+1;

end

[Adjacency\_Matrix]=BADR\_Uneven(IRMA\_1(:,2:end), Times, Replicates);

**3. BADR\_Reduced**

For large gene systems (e.g. thousands of genes) the simple BADR will use too much memory primarily due to the dimensions of the matrixes. The same algorithm can be used but using a reduced representation of the adjacency matrix where only the known transcription factors are considered. This speeds running times by a substrantial margin.

**Basic Usage**

**Input:**

Mandatory: Gene Expression Time Series, Time Differences, Number of Time Points, Number of Replicates, Transcription Factors Identifiers

Optional: Prior Probability Matrix

Gene Expression Time Series:

A matrix where each column represent a gene and each row a time series measurement.

Formula:

Time Differences:

A vector representing the differences between time points. The units are to be chosen by the user.

Formula:

Number of Time Points:

A scalar representing the number of time points. On our nomenclature it would be equal to .

Number of Replicates:

A scalar representing the number of replicates. On our nomenclature it would be equal to .

Transcription Factor Identifiers:

A vector with number representing the columns of the gene expression time series matrix where the transcription factors are located

Formula:

Prior Probability:

A matrix with the probability that each gene regulates the other.

Formula:

**Output:** Gene Regulatory Network and Coefficients

Gene Regulatory Network:

A matrix containing the posterior probability that genes regulate each other.

Coefficients:

A matrix containing the coefficients in the models after performing Bayesian Model Averaging. Can be used as an indicator of whether the interaction is activating or inhibiting.

**Basic Usage:**

[Adjacency\_Matrix, Coefficients] = BADR\_Reduced(Gene\_Exp, Dt, No\_Time\_Points, No\_Replicates, Transcription\_Factors);

**Prior and Transcription Factors IDs Usage**

[Adjacency\_Matrix, Coefficients] = BADR(Gene\_Exp, Dt, No\_Time\_Points, No\_Replicates, Transcription\_Factors, Prior);

**Example Basic Usage: Dream 4 Network 1**

mRNA1 = dlmread('insilico\_size10\_1\_timeseries.tsv');

No\_Replicates = 5;

No\_Time\_Points = 21;

dt = 50/60;

dt = repmat(dt, No\_Time\_Points-1,1);

Transcription\_Factors = 1;

[Adjacency\_Matrix]=BADR\_Reduced(mRNA1, dt, No\_Time\_Points, No\_Replicates,Transcription\_Factors);