

# Package ‘BICORN’

May 6, 2026

**Title** Integrative Inference of De Novo Cis-Regulatory Modules

**Version** 0.1.0

**Description** Prior transcription factor binding knowledge and target gene expression data are integrated in a Bayesian framework for functional cis-regulatory module inference. Using Gibbs sampling, we iteratively estimate transcription factor associations for each gene, regulation strength for each binding event and the hidden activity for each transcription factor.

**Depends** R (>= 3.4)

**License** GPL-2

**Encoding** UTF-8

**LazyData** true

**RoxygenNote** 6.0.1

**Suggests** knitr, rmarkdown

**Imports** stats

**VignetteBuilder** knitr

**NeedsCompilation** no

**Author** Xi Chen [aut, cre],  
Jianhua Xuan [aut]

**Maintainer** Xi Chen <xichen86@vt.edu>

**Repository** CRAN

**Date/Publication** 2018-06-06 10:30:06 UTC

## Contents

A	2
alpha	3
A_old	3
A_sampling	3
baseline_sampling	4
base_line	5
base_line_old	5
beta	5

BICORN . . . . .	6
Binding_genes . . . . .	6
Binding_matrix . . . . .	7
Binding_TFs . . . . .	7
C . . . . .	7
C_old . . . . .	8
C_prior . . . . .	8
C_sampling_cluster . . . . .	9
data_integration . . . . .	9
Exp_data . . . . .	10
Exp_genes . . . . .	11
sigmanoise_sampling . . . . .	11
sigma_A . . . . .	12
sigma_baseline . . . . .	12
sigma_noise . . . . .	12
sigma_X . . . . .	13
X . . . . .	13
X_old . . . . .	13
X_sampling . . . . .	14
Y . . . . .	14

<b>Index</b>	<b>15</b>
--------------	-----------

---

A	<i>TF-gene regulation strength matrix</i>
---	---

---

### Description

A matrix of TF-gene regulation strength with genes as rows and TFs as columns.

### Usage

A

### Format

numeric matrix

---

alpha	<i>Inverse-gamma distribution hyper-parameter alpha</i>
-------	---

---

**Description**

Hyper-parameter alpha of inverse-gamma distribution.

**Usage**

alpha

**Format**

scalar

---

A_old	<i>TF-gene regulation strength matrix sampled from the previous round</i>
-------	---

---

**Description**

A matrix of TF-gene regulation strength with genes as rows and TFs as columns, sampled from the previous round. During the Gibbs sampling process, this matrix is used as prior for a new round of regulation strength sampling.

**Usage**

A\_old

**Format**

numeric matrix

---

A_sampling	<i>Regulation Strength Sampling Function</i>
------------	--

---

**Description**

Function 'A\_sampling' estimates a regulation strength for each sampled binding event in C, according to a posterior Gaussian distribution.

**Usage**

```
A_sampling(Y, C, A_old, X, base_line, C_prior, sigma_noise, sigma_A,
           sigma_baseline, sigma_X)
```

**Arguments**

Y	gene expression data matrix
C	sampled TF-gene binding network
A_old	regulatory strength sampled from the previous round, used as a prior in current function
X	sampled transcription factor activity matrix
base_line	sampled gene expression baseline activity
C_prior	prior TF-gene binding network
sigma_noise	variance of gene expression fitting residuals
sigma_A	variance of regulatory strength
sigma_baseline	variance of gene expression baseline activity
sigma_X	variance of transcription factor activity

---

baseline\_sampling      *Gene Baseline Expression Sampling Function*

---

**Description**

Function 'baseline\_sampling' estimates a baseline expression for each gene, according to a posterior Gaussian distribution.

**Usage**

```
baseline_sampling(Y, C, A, X, base_line_old, C_prior, sigma_noise, sigma_A,
                 sigma_baseline, sigma_X)
```

**Arguments**

Y	gene expression data matrix
C	sampled TF-gene binding network
A	sampled regulatory strength matrix
X	sampled transcription factor activity matrix
base_line_old	prior gene expression baseline activity
C_prior	prior TF-gene binding network
sigma_noise	variance of gene expression fitting residuals
sigma_A	variance of regulatory strength
sigma_baseline	variance of gene expression baseline activity
sigma_X	variance of transcription factor activity

---

base_line	<i>Gene baseline expression</i>
-----------	---------------------------------

---

**Description**

A vector of baseline expression for all genes.

**Usage**

base\_line

**Format**

numeric vector

---

base_line_old	<i>Gene baseline expression sampled from the previous round.</i>
---------------	--

---

**Description**

A vector of baseline expression for all genes, sampled from the previous round. During the Gibbs Sampling process, this is used as a prior for a new round of gene baseline expression sampling.

**Usage**

base\_line\_old

**Format**

numeric vector

---

beta	<i>Inverse-gamma distribution hyper-parameter beta</i>
------	--

---

**Description**

Hyper-parameter beta of inverse-gamma distribution.

**Usage**

beta

**Format**

scalar

---

BICORN                      *BICORN Algorithm Function*

---

### Description

Function 'BICORN' infers a posterior module-gene regulatory network by iteratively sampling regulatory strength, transcription factor activity and several key model parameters.

### Usage

```
BICORN(BICORN_input = NULL, L = 100, output_threshold = 10)
```

### Arguments

BICORN\_input    this list structure contains TF symbols, gene symbols and candidate modules  
 L                      total rounds of Gibbs Sampling.  
 output\_threshold                      number of rounds after which we start to record results.

### Examples

```
# load in the sample data input
data("sample.input")

# Data initialization (Integrate prior binding network and gene expression data)
BICORN_input<-data_integration(Binding_matrix = Binding_matrix, Binding_TFs = Binding_TFs,
Binding_genes = Binding_genes, Exp_data = Exp_data, Exp_genes = Exp_genes,
Minimum_gene_per_module_regulate = 2)

# Infer cis-regulatory modules (TF combinations) and their target genes
BICORN_output<-BICORN(BICORN_input, L = 2, output_threshold = 1)
```

---

Binding\_genes                      *Genes in the prior binding network*

---

### Description

A list of official gene symbols in the binary binding network.

### Usage

```
Binding_genes
```

### Format

character vector

---

Binding_matrix	<i>Prior TF-gene binding network</i>
----------------	--------------------------------------

---

**Description**

A prior binary TF-gene regulatory network with each unit either 1 (binding) or 0 (non-binding).

**Usage**

Binding\_matrix

**Format**

numeric matrix

---

Binding_TFs	<i>TFs in the prior binding network</i>
-------------	---

---

**Description**

A list of transcription factors in the prior binding network.

**Usage**

Binding\_TFs

**Format**

character vector

---

C	<i>TF-gene binding network</i>
---	--------------------------------

---

**Description**

A matrix of TF-gene regulatory network with each unit either 1 (binding) or 0 (non-binding).

**Usage**

C

**Format**

numeric matrix

---

C_old	<i>TF-gene binding network sampled from the previous round</i>
-------	--

---

**Description**

A matrix of TF-gene binding network sampled from the previous round, with each unit either 1 (binding) or 0 (non-binding). During the Gibbs sampling process, this is used as a prior for a new round of binding network sampling.

**Usage**

C\_old

**Format**

numeric matrix

---

C_prior	<i>Prior TF-gene binding network</i>
---------	--------------------------------------

---

**Description**

A matrix of prior TF-gene binding events, with each unit either 1 (binding) or 0 (non-binding). Such a prior network can be obtained from TF-gene binding database, motif searching, ChIP-seq peaks or ATAC-seq peaks.

**Usage**

C\_prior

**Format**

numeric matrix

---

C\_sampling\_cluster      *cis-Regulatory Module Sampling Function*

---

### Description

Function 'C\_sampling\_cluster' samples a candidate cis-regulatory module for each gene, according to a discrete posterior probability distribution.

### Usage

```
C_sampling_cluster(Y, C_old, A_old, X_old, base_line_old, C_prior, sigma_noise,
  sigma_A, sigma_baseline, sigma_X, BICORN_input)
```

### Arguments

Y	gene expression data matrix
C_old	TF-gene binding network sampled from the previous round
A_old	regulatory strength matrix sampled from the previous round
X_old	transcription factor activity matrix sampled from the previous round
base_line_old	gene expression baseline activity sampled from the previous round
C_prior	prior TF-gene binding network
sigma_noise	variance of gene expression fitting residuals
sigma_A	variance of regulatory strength
sigma_baseline	variance of gene expression baseline activity
sigma_X	variance of transcription factor activity
BICORN_input	this list structure contains TF symbols, Gene symbols and candidate modules

---

data\_integration      *Data Initialization for BICORN*

---

### Description

Function 'data\_integration' integrates the prior TF-gene binding network and gene expression data together. It will remove any genes missing either TF bindings or gene expression and identify a list of candidate cis-regulatory modules.

### Usage

```
data_integration(Binding_matrix = NULL, Binding_TFs = NULL,
  Binding_genes = NULL, Exp_data, Exp_genes = NULL,
  Minimum_gene_per_module_regulate = 2)
```

**Arguments**

Binding\_matrix loaded prior binding network  
Binding\_TFs loaded transcription factors  
Binding\_genes loaded genes in the prior binding network  
Exp\_data loaded properly normalized gene expression data  
Exp\_genes loaded genes in the gene expression data  
Minimum\_gene\_per\_module\_regulate  
the minimum number of genes regulated by each module, used for candidate module filtering.

**Examples**

```
# load in the sample data input
data("sample.input")

# Data initialization (Integrate prior binding network and gene expression data)
BICORN_input<-data_integration(Binding_matrix = Binding_matrix, Binding_TFs = Binding_TFs,
Binding_genes = Binding_genes, Exp_data = Exp_data, Exp_genes = Exp_genes,
Minimum_gene_per_module_regulate = 2)
```

---

Exp_data	<i>Gene expression data</i>
----------	-----------------------------

---

**Description**

A matrix of normalized gene expression data with genes as rows and samples as columns. The gene expression data can be either time-course data measured under multiple time points or steady state data generated from at least two different conditions.

**Usage**

```
Exp_data
```

**Format**

```
numeric matrix
```

---

Exp_genes	<i>Genes in the expression data</i>
-----------	-------------------------------------

---

**Description**

A list of official gene symbols in the gene expression data set.

**Usage**

Exp\_genes

**Format**

character vector

---

sigmanoise_sampling	<i>Fitting Residule Variance Sampling Function</i>
---------------------	--

---

**Description**

Function 'sigmanoise\_sampling' estimates the variance of overall gene expression fitting residuals, according to an inverse-gamma distribution.

**Usage**

```
sigmanoise_sampling(Y, C, A, X, base_line, C_prior, sigma_A, sigma_baseline,
  sigma_X, alpha, beta)
```

**Arguments**

Y	gene expression data matrix
C	sampled TF-gene binding network
A	sampled regulatory strength matrix
X	sampled transcription factor activity matrix
base_line	sampled gene expression baseline activity
C_prior	prior TF-gene binding network
sigma_A	variance of regulatory strength
sigma_baseline	variance of gene expression baseline activity
sigma_X	variance of transcription factor activity
alpha	hyper-parameter for inverse-gamma distribution
beta	hyper-parameter for inverse-gamma distribution

---

sigma_A	<i>Regulation strength variance</i>
---------	-------------------------------------

---

**Description**

Variance of regulation strength matrix A.

**Usage**

sigma\_A

**Format**

scalar

---

sigma_baseline	<i>Variance of baseline gene expression.</i>
----------------	--

---

**Description**

Variance of baseline gene expression.

**Usage**

sigma\_baseline

**Format**

scalar

---

sigma_noise	<i>Variance of gene expression fitting residuals.</i>
-------------	---

---

**Description**

Variance of gene expression fitting residuals.

**Usage**

sigma\_noise

**Format**

scalar

---

sigma_X	<i>Transcription factor activity variance</i>
---------	---

---

**Description**

Variance of transcription factor activity matrix X.

**Usage**

sigma\_X

**Format**

scalar

---

X	<i>Transcription factr activity matrix</i>
---	--

---

**Description**

A matrix of hidden transcription factr activity estimated from gene expression data, with transcription factrs as rows and samples as columns.

**Usage**

X

**Format**

numeric matrix

---

X_old	<i>Transcription factr activity matrix sampled from the previous round</i>
-------	--

---

**Description**

A matrix of hidden transcription factr activity estimated from gene expression data, with transcription factrs as rows and samples as columns, sampled from the previous round. During the Gibbs sampling process, this is used as a prior for a new round of transcription factor activity sampling.

**Usage**

X\_old

**Format**

numeric matrix

---

X\_sampling                      *Transcription Factor Activity Sampling Function*

---

### Description

Function 'X\_sampling' estimates the hidden activities of each transcription factor, according to a posterior Gaussian random process.

### Usage

```
X_sampling(Y, C, A, X_old, base_line, C_prior, sigma_noise, sigma_A,
           sigma_baseline, sigma_X)
```

### Arguments

Y	gene expression data matrix
C	sampled TF-gene binding network
A	sampled regulatory strength matrix
X_old	sampled transcription factor activity matrix from the previous round
base_line	sampled gene expression baseline activity
C_prior	prior TF-gene binding network
sigma_noise	variance of gene expression fitting residuals
sigma_A	variance of regulatory strength
sigma_baseline	variance of gene expression baseline activity
sigma_X	variance of transcription factor activity

---

Y                                      *Gene expression data used for module inference*

---

### Description

A matrix of normalized gene expression for common genes of prior binding input and gene expression input, with genes as rows and samples as columns. Y is the matrix used for cis-regulatory module inference.

### Usage

```
Y
```

### Format

numeric matrix

# Index

- \* **BICORN**
  - BICORN, 6
- \* **Baseline**
  - base\_line, 5
  - base\_line\_old, 5
  - baseline\_sampling, 4
  - sigma\_baseline, 12
- \* **Binding**
  - Binding\_genes, 6
  - Binding\_matrix, 7
  - Binding\_TFs, 7
  - C, 7
  - C\_old, 8
  - C\_prior, 8
- \* **Gene\_expression**
  - Exp\_data, 10
  - Exp\_genes, 11
  - sigma\_noise, 12
  - Y, 14
- \* **Hyper-parameters**
  - alpha, 3
  - beta, 5
- \* **Initialization**
  - data\_integration, 9
- \* **Module**
  - C\_sampling\_cluster, 9
- \* **Noise**
  - alpha, 3
  - beta, 5
  - sigma\_A, 12
  - sigma\_baseline, 12
  - sigma\_noise, 12
  - sigma\_X, 13
  - sigmanoise\_sampling, 11
- \* **Regulation\_Strength**
  - A, 2
  - A\_old, 3
  - A\_sampling, 3
  - sigma\_A, 12
- \* **TFA**
  - sigma\_X, 13
  - X, 13
  - X\_old, 13
  - X\_sampling, 14
- \* **sampling**
  - baseline\_sampling, 4
  - BICORN, 6
  - C\_sampling\_cluster, 9
  - sigmanoise\_sampling, 11
  - X\_sampling, 14
- A, 2
- A\_old, 3
- A\_sampling, 3
- alpha, 3
- base\_line, 5
- base\_line\_old, 5
- baseline\_sampling, 4
- beta, 5
- BICORN, 6
- Binding\_genes, 6
- Binding\_matrix, 7
- Binding\_TFs, 7
- C, 7
- C\_old, 8
- C\_prior, 8
- C\_sampling\_cluster, 9
- data\_integration, 9
- Exp\_data, 10
- Exp\_genes, 11
- sigma\_A, 12
- sigma\_baseline, 12
- sigma\_noise, 12
- sigma\_X, 13
- sigmanoise\_sampling, 11

X, 13

X\_old, 13

X\_sampling, 14

Y, 14