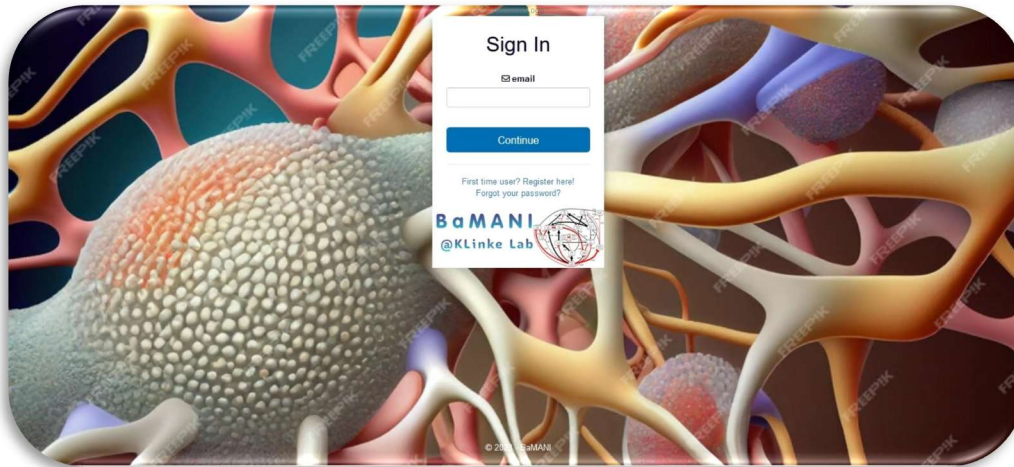


# User Guide for BaMANI: Bayesian Multi-Algorithm Network Inference

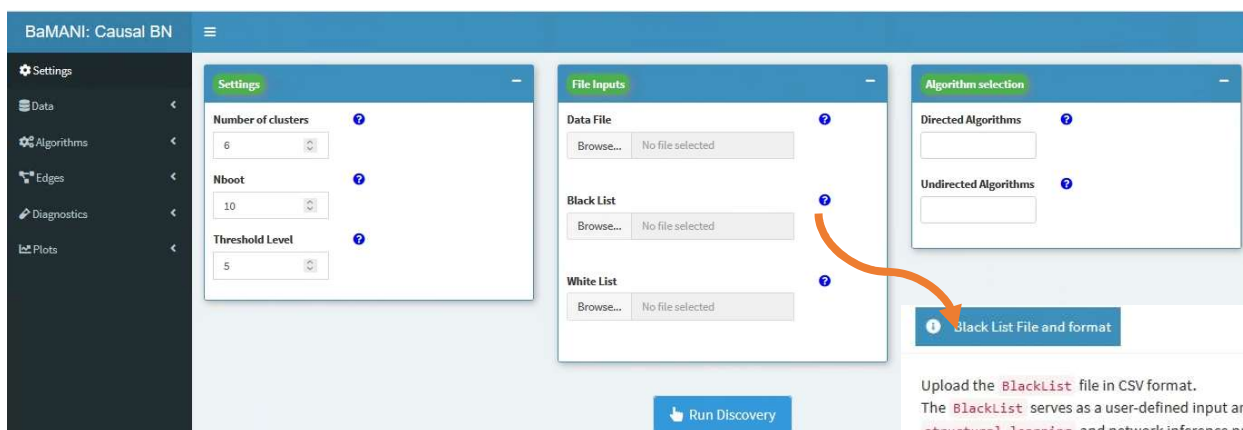
## Introduction

BaMANI is a powerful tool for Bayesian network inference, utilizing a range of algorithms to deduce structure and causality in networks. The following guide walks you through the process of using BaMANI, from creating blacklists based on domain knowledge to learning the final network structure.



## Step 1: Creating and Specifying a Blacklist

- **Objective:** To input a blacklist that outlines prior knowledge and removes implausible arcs. Ensure the blacklist aligns with your domain knowledge and data characteristics
- **Process:**
  - Remove arcs inconsistent with oncogenesis.
  - Specify nodes with majority zero values as leaf nodes.



Upload the **BlackList** file in CSV format.

The **BlackList** serves as a user-defined input and plays a pivotal role in the **structural learning** and network inference processes within Bayesian networks, which are illustrated as Directed Acyclic Graphs (DAGs). This component is instrumental in molding the network structure, relying primarily on **prior knowledge** regarding the causal relationships amongst the data features. This list encompasses specific arcs, or **directed edges**, that are deliberately excluded from the proposed network, serving as constraints based on established or acknowledged information. This ensures the resultant structure and its causal interactions or arcs amongst nodes are in harmony with existing knowledge and logical constraints, thereby maintaining the integrity and accuracy of the network's representation of causal relationships.

BaMANI: Causal BN

Settings

Data

Data characteristics

Black List

Algorithms

Edges

Diagnostics

Plots

Black List

The **blackList** plays a pivotal role in the **structural learning** and network inference processes within Bayesian networks, which are illustrated as Directed Acyclic Graphs (DAGs). This component is instrumental in molding the network structure, relying primarily on **prior knowledge** regarding the causal relationships amongst the data features. This list encompasses specific arcs, or **directed edges**, that are deliberately excluded from the proposed network, serving as constraints based on established or acknowledged information. This ensures the resultant structure and its causal interactions or arcs amongst nodes are in harmony with existing knowledge and logical constraints, thereby maintaining the integrity and accuracy of the network's representation of causal relationships.

Black List

CSV Excel

Search:

	from	to
1	Endothelial	Cancer
2	CAF	Cancer
3	CD8	Cancer
4	NKactive	Cancer
5	NKrest	Cancer
6	Macrophages	Cancer
7	CD4	Cancer
8	B.cel	Cancer
9	Neutrophils	Cancer
10	proliferation	Cancer
11	Epithelial	Cancer
12	Mesenchymal	Cancer
13	CCN4	Cancer
14	pM0	Cancer

Showing 1 to 14 of 73 entries

## Step 2: Generating an Ensemble of Potential Arcs

- **Objective:** To generate a seed edge list using an ensemble of structural learning algorithms.
- **Process:**

BaMANI: Causal BN

Settings

Data

Algorithms

Edges

Augmented Edge List

Augmented Threshold Columns

Possible Seed Arcs Filter

Diagnostics

Plots

Settings

Number of clusters

6

Nboot

10

Threshold Level

5

File Inputs

Data File

Browse... No file selected

Black List

Browse... No file selected

White List

Browse... No file selected

Algorithm selection

Directed Algorithms

iamb iamb.fdr pc.stable

hc tabu

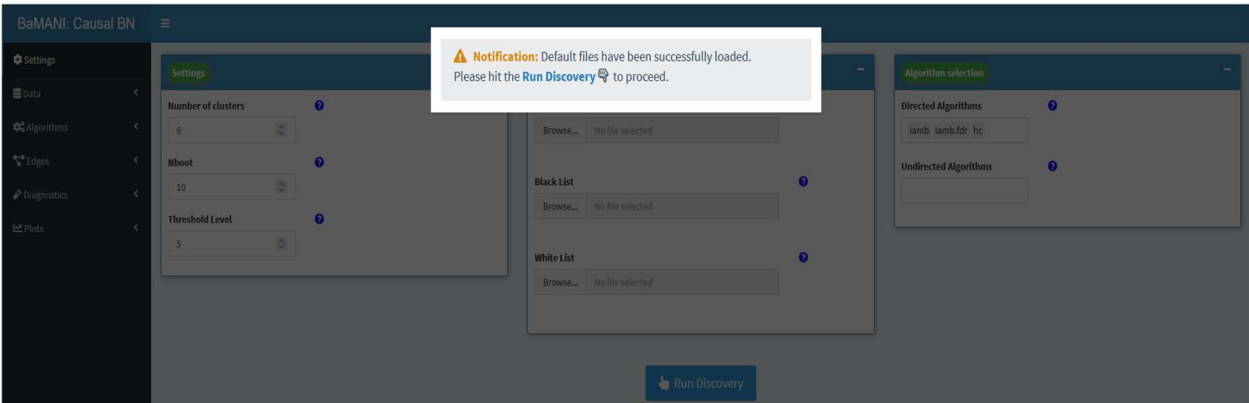
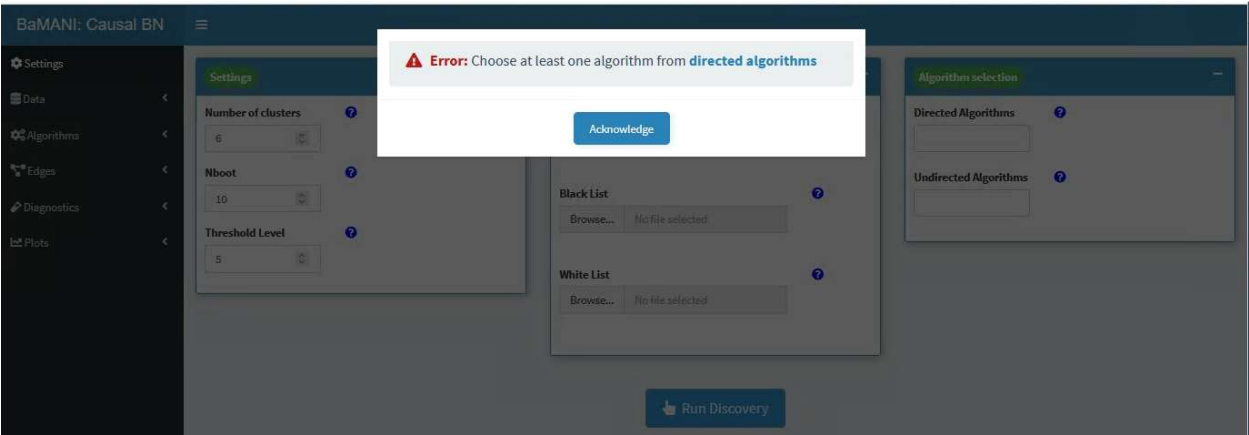
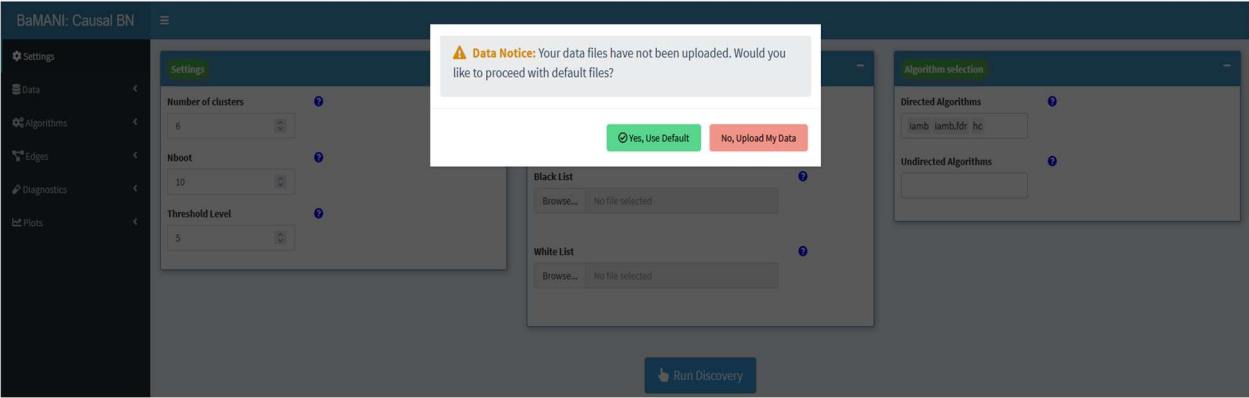
mmhc

rsmax2

gs

Run Discovery

- Use the predefined blacklist to learn the network structure.
- Employ various structur learning algorithms (e.g., IAMB.FDR, PC.STABLE) with bootstrapped arc strength calculations.



BaMANI: Causal BN

Settings

Data

Algorithms

Learning Process

Edges

Diagnostics

Plots

Learning process of directed & Undirected algorithms:

The message printed below displays the progression of the learning process as executed by various structural learning algorithms, encompassing both directed and undirected algorithms.

The **directed algorithm** in Bayesian network inference is crucial for deducing the structure and causality in a network, represented as a Directed Acyclic Graph (dag). Users employ an ensemble of different structural learning algorithms to accurately identify potential causal interactions or arcs among nodes and to estimate the **conditional probability distributions** from datasets. These algorithms contribute to constructing **blacklists** and/or **whitelists** of arcs, serving as a basis for exclusion and inclusion based on **prior knowledge** and **arc strength**. Some algorithms have been utilized for the local discovery of undirected graphs to furnish additional evidence substantiating the existence of an edge between two nodes, in conjunction with the **directed algorithms**.

The table displayed on the right panel enumerates the arcs learnt through using various structural learning algorithms. The column labeled **edge\_no** corresponds to the number of the arcs as listed in the table.

Algorithm Progression

List of Arc

Algorithm Progression

```

*****
Does the list Empty before directed algorithm <iamb> ? YES
Number of arcs: ==> before the algorithm: 0 || after algorithm: 40
*****
Does the list Empty before directed algorithm <iamb.fdr> ? NO
Number of arcs: ==> before the algorithm: 40 || after algorithm: 50
*****
Does the list Empty before directed algorithm <pc.stable> ? NO
Number of arcs: ==> before the algorithm: 50 || after algorithm: 68
*****
Does the list Empty before directed algorithm <hc> ? NO
Number of arcs: ==> before the algorithm: 68 || after algorithm: 108
*****
Does the list Empty before directed algorithm <mmhc> ? NO
Number of arcs: ==> before the algorithm: 108 || after algorithm: 108
-----This algorithm learned some existed Arcs but could not learn more!-----
*****
Does the list Empty before directed algorithm <gs> ? NO
Number of arcs: ==> before the algorithm: 108 || after algorithm: 110
*****

```

BaMANI: Causal BN

Settings

Data

Algorithms

Learning Process

Edges

Diagnostics

Plots

Learning process of directed & Undirected algorithms:

The message printed below displays the progression of the learning process as executed by various structural learning algorithms, encompassing both directed and undirected algorithms.

The **directed algorithm** in Bayesian network inference is crucial for deducing the structure and causality in a network, represented as a Directed Acyclic Graph (dag). Users employ an ensemble of different structural learning algorithms to accurately identify potential causal interactions or arcs among nodes and to estimate the **conditional probability distributions** from datasets. These algorithms contribute to constructing **blacklists** and/or **whitelists** of arcs, serving as a basis for exclusion and inclusion based on **prior knowledge** and **arc strength**. Some algorithms have been utilized for the local discovery of undirected graphs to furnish additional evidence substantiating the existence of an edge between two nodes, in conjunction with the **directed algorithms**.

The table displayed on the right panel enumerates the arcs learnt through using various structural learning algorithms. The column labeled **edge\_no** corresponds to the number of the arcs as listed in the table.

Algorithm Progression

List of Arc

Arcs Learnt from Algorithms

from

to

Edge\_No

Cancer	proliferation	1
proliferation	Cancer	1
Cancer	Epithelial	2
Epithelial	Cancer	2
Cancer	CCN4	3
CCN4	Cancer	3
Endothelial	CD4	4
CD4	Endothelial	4
CAF	CD8	5
CD8	CAF	5
CAF	NKActive	6
NKActive	CAF	6
NKActive	Macrophages	7
Macrophages	NKActive	7
Macrophages	CD4	8

Showing 1 to 18 of 110 entries

- Calculate the "arc.strength", "CorSign", and unclear\_direction" columns for each arc.

BaMANI: Causal BN

Settings

Data

Algorithms

Learning Process

Edges

Diagnostics

Plots

Arcs and thier corresponding info

CSV

Excel

Search:

	from	to	Edge_No	iamb_strength	iamb_CorSign	iamb.fdr_strength	iamb.fdr_CorSign	pc.stable_strength	pc.stable_CorSign	hc_strength	hc_CorSign	mmhc_strength	mmhc_CorSign	gs_strength	gs_CorSign
1	Cancer	proliferation	1	6.7e-23	+	6.7e-23	+	2e-8	+	2.9e-11	+	4e-10	+	6.7e-23	
2	proliferation	Cancer	1												
3	Cancer	Epithelial	2	1.7e-25	-	1.7e-25	-	2.2e-22	-	3.6e-29	-	2.2e-22	-	1.7e-25	
4	Epithelial	Cancer	2												
5	Cancer	CCN4	3	2.1e-55	+	2.1e-55	+	2.1e-55	+	2.1e-55	+	2.1e-55	+	2.1e-55	
6	CCN4	Cancer	3												
7	CD4	Endothelial	4												
8	Endothelial	CD4	4	2.9e-7	+	2.4e-9	+	7.8e-7	+	5.8e-10	+	2.9e-7	+	2.4e-9	
9	CAF	CD8	5	3.1e-7	-	3.1e-7	-	3.1e-7	-	1.7e-8	-	3.1e-7	-	3.1e-7	
10	CD8	CAF	5												
11	CAF	NKActive	6	4.2e-9	-					4.2e-9	-	4.2e-9	-	4.2e-9	
12	NKActive	CAF	6					9.2e-7	-						
13	Macrophages	NKActive	7												

Showing 1 to 14 of 110 entries

### Step 3: Filtering Potential Arcs

- **Objective:** To create a whitelist based on Bayesian Information Criterion (BIC) and network complexity.
- **Process:**
  - Set arc strength thresholds.
  - Create temporary whitelists and learn BN structures for each threshold.

**BaMANI: Causal BN**

Settings Data Algorithms Edges

» Augmented Edge List

» Augmented Threshold Columns

» Possible Seed Arcs Filter

Diagnostics

Plot

Arcs and their corresponding info in different algorithm

Algorithm count and arcs strength of each arc

List of Arcs and their corresponding arcs strength and their inhibitive or promotive impact/ correlation sign in different algorithm

Arcs and their corresponding info

CSV Excel

Search:

	from	to	Edge_No	lamb _strength	lamb _CorSign	lamb.fdr _strength	lamb.fdr _CorSign	pc.stable _strength	pc.stable _CorSign	hc _strength	hc _CorSign	mmhc _strength	mmhc _CorSign	gs _strength	gs _CorSign	Htt.Count	Min.strength	Max.strength
1	Cancer	proliferation	1	6.7e-23	+	6.7e-23	+	2e-08	+	2.9e-11	+	4e-10	+	6.7e-23	+	6	6.7e-23	6.7e-23
2	Cancer	Epithelial	2	1.7e-25	-	1.7e-25	-	2.2e-22	-	3.6e-29	-	2.2e-22	-	1.7e-25	-	6	3.6e-29	2
3	Cancer	CD4	3	2.1e-55	+	2.1e-55	+	2.1e-55	+	2.1e-55	+	2.1e-55	+	2.1e-55	+	6	2.1e-55	2
4	Endothelial	CD4	4	2.9e-07	+	2.4e-09	+	7.8e-07	+	5.8e-10	+	2.9e-07	+	2.4e-09	+	6	5.8e-10	7
5	CAF	CD8	5	3.1e-07	-	3.1e-07	-	3.1e-07	-	1.7e-08	-	3.1e-07	-	3.1e-07	-	6	1.7e-08	3
6	NK.active	Macrophages	7	7.9e-08	+	7.9e-08	+	1.2e-08	+	NA		6.8e-08	+	6.8e-08	+	5	1.2e-08	7
7	Macrophages	CD4	8	5.1e-16	+	7.6e-14	+	2e-13	+	6e-16	+	5.1e-16	+	7.6e-14	+	6	5.1e-16	1
8	Macrophages	Neutrophils	9	0.0016	+	NA		0.0018	+	0.00045	+	0.0016	+	0.0016	+	5	0.00045	0
9	Macrophages	CD8	10	2.4e-12	+	2.4e-12	+	2.4e-12	+	1.7e-11	+	2.4e-12	+	2.4e-12	+	6	2.4e-12	1
10	B.cel	CD4	11	1.9e-06	+	1.4e-07	+	7.3e-05	+	3e-08	+	1.9e-06	+	1.4e-07	+	6	3e-08	7
11	B.cel	CD8	12	2.9e-27	+	2.9e-27	+	2.9e-27	+	2.6e-17	+	2.9e-27	+	2.9e-27	+	6	2.9e-27	2
12	Epithelial	Neutrophils	15	0.0028	-	NA		0.00068	-	NA		0.00068	-	0.0028	-	4	0.00068	0
13	Mesenchymal	CAF	16	9.6e-73	+	6.8e-74	+	6.4e-85	+	1.6e-92	+	2.3e-86	+	9.6e-73	+	6	1.6e-92	9

Showing 1 to 14 of 49 entries

- Calculate and compare BIC for different thresholds.

**BaMANI: Causal BN**

Settings  
Data  
Algorithms  
Edges  
Diagnostics

» L1 Merged  
» Npar Merged  
» BIC Merged Table  
» BIC Min Table  
» WhiteList/ Check acyclicity  
Plots

### Loss value for each node in final DAG / threshold

Descriptin: Loss value for each node in final DAG / threshold

Loss value for each node / threshold

CSV Excel Search:

	L1_empty	L1_thresh_#0	L1_thresh_#1.1e-148	L1_thresh_#2.9e-27	L1_thresh_#3.5e-14	L1_thresh_#3e-08	L1_thresh_#4.3e-05	L1_thresh_#0.19	L1_thresh_all
B.cel	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
CAF	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
Cancer	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2
CCN4	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
CD4	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3
CD8	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
Endothelial	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2
Epithelial	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
Macrophages	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
Mesenchymal	0.2	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
Neutrophils	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2
NKactive	0	0	0	0	0	0	0	0	0
NKrest	0	0	0	0	0	0	0	0	0
pM0	0.2	0	0	0.1	0.1	0	0	0	0

Showing 1 to 14 of 17 entries

**BaMANI: Causal BN**

**Number of parents of each node in final DAG / threshold**  
 Descriptin: Number of parents of each node in final DAG / threshold

Number of parents of each node/ threshold

CSV Excel

	npar_empty_graph	npar_thresh_# 0	npar_thresh_# 1.1e-148	npar_thresh_# 2.9e-27	npar_thresh_# 3.5e-14	npar_thresh_# 3e-08	npar_thresh_# 4.3e-05	npar_thresh_# 0.19	npar_thresh_# all
B.cel	0	1	1	1	1	0	2	1	1
CAF	0	2	1	3	1	3	2	1	2
Cancer	0	0	0	0	0	0	0	0	0
CCN4	0	1	1	1	1	1	1	1	1
CD4	0	3	3	3	3	3	3	3	3
CD8	0	4	4	4	4	4	4	4	4
Endothelial	0	2	1	2	2	1	2	2	2
Epithelial	0	1	1	1	1	1	1	1	1
Macrophages	0	1	2	2	2	2	2	1	2
Mesenchymal	0	1	2	1	1	2	1	1	1
Neutrophils	0	2	2	3	2	3	2	3	3
NKactive	0	0	0	0	1	0	1	1	0
NKrest	0	0	0	0	0	0	0	0	0
pM0	0	2	2	1	1	2	2	2	2

Showing 1 to 14 of 17 entries

**BaMANI: Causal BN**

**Bayesian Information Criterion (BIC) score for each node/ threshold**  
 Descriptin: Bayesian Information Criterion (BIC) score for each node/ threshold

BIC score/ threshold

CSV Excel

	BIC_thresh_empty_graph	BIC_thresh_0	BIC_thresh_1.1e-148	BIC_thresh_2.9e-27	BIC_thresh_3.5e-14	BIC_thresh_3e-08	BIC_thresh_4.3e-05	BIC_thresh_0.19	BIC_thresh_all
B.cel	58.2	64.57	64.57	64.57	64.57	58.2	70.93	64.57	64.57
CAF	58.2	70.93	64.57	77.3	64.57	77.3	70.93	64.57	70.93
Cancer	116.4	116.4	116.4	116.4	116.4	116.4	116.4	116.4	116.4
CCN4	58.2	64.57	64.57	64.57	64.57	64.57	64.57	64.57	64.57
CD4	174.6	193.7	193.7	193.7	193.7	193.7	193.7	193.7	193.7
CD8	58.2	83.67	83.67	83.67	83.67	83.67	83.67	83.67	83.67
Endothelial	116.4	129.13	122.77	129.13	129.13	122.77	129.13	129.13	129.13
Epithelial	58.2	64.57	64.57	64.57	64.57	64.57	64.57	64.57	64.57
Macrophages	58.2	64.57	70.93	70.93	70.93	70.93	70.93	64.57	70.93
Mesenchymal	116.4	64.57	70.93	64.57	64.57	70.93	64.57	64.57	64.57
Neutrophils	116.4	129.13	129.13	135.5	129.13	135.5	129.13	135.5	135.5
NKactive	0	0	0	0	6.37	0	6.37	6.37	0
NKrest	0	0	0	0	0	0	0	0	0
pM0	116.4	12.73	12.73	64.57	64.57	12.73	12.73	12.73	12.73

Showing 1 to 14 of 17 entries

- Look for the minimum BIC value while avoiding the creation of cycles. Remove arcs that create cycles.



**BaMANI: Causal BN**

**Min BIC score for a node with min number of parents in corresponding final DAG**  
 Descriptin: Min BIC score for a node with min number of parents in corresponding final DAG / threshold

Min BIC table    Possible White List

**Min BIC table**

CSV    Excel    Search:

	bic_min	bic_min_index
B.cel	58.2	1
CAF	58.2	1
Cancer	116.4	1
CCN4	58.2	1
CD4	174.6	1
CD8	58.2	1
Endothelial	116.4	1
Epithelial	58.2	1
Macrophages	58.2	1
Mesenchymal	64.57	2
Neutrophils	116.4	1
NK.active	0	1
NK.rest	0	1
pM0	12.73	2

Showing 1 to 14 of 17 entries

## Step 4: Learning the Network Structure

- **Objective:** To learn the final network structure using the blacklist and whitelist.
- **Process:**
  - Utilize the final whitelist and blacklist to learn the structure of the Bayesian Network (BN).

**BaMANI: Causal BN**

**Settings**

Number of clusters: 6

Nboot: 10

Threshold Level: 5

**File Inputs**

Data File:  No file selected

Black List:  No file selected

White List:

**Algorithm selection**

Directed Algorithms: iamb iamb.fdr hc

Undirected Algorithms:

**Proceed to Further Analysis:**  
 Navigate to the tab 'WhiteList/ Check acyclicity' [Click here.](#)

OK

- Observe how the network structure and parameters are determined based on these lists.
- **Note:** Review the final network structure for consistency with your domain knowledge.