

Predicting COVID-19 Severity Using Bayesian Networks in wiseR

group 6- Mahi, Aakarsh, Muthuraj, Kanishka, Sahas, Anikaith, Abhishant

Agenda

Topics Covered

DATA & PREPROCESSING

ASSOCIATION NETWORK

STRUCTURE LEARNING SETUP

NETWORK STRUCTURE
VISUALIZATION

INFERENCE LEARNING &
ANALYSIS

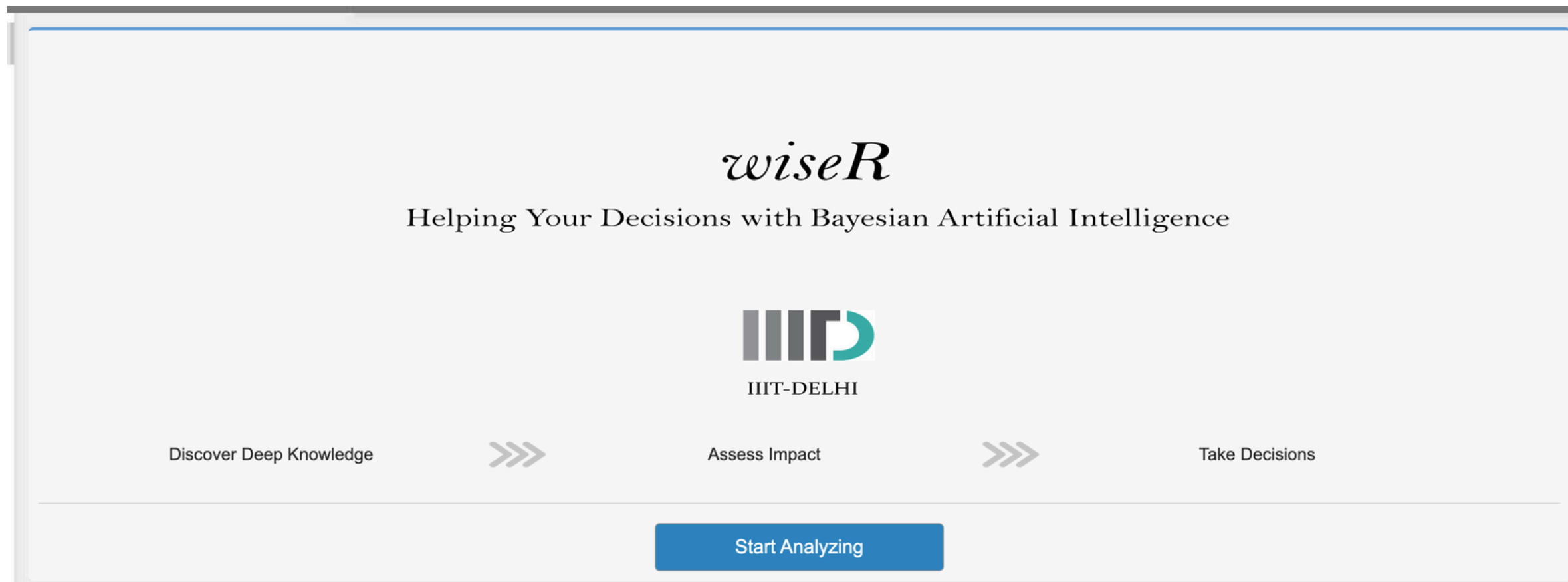
RESULTS & CONCLUSION

MODIFICATIONS

FUTURE WORK

OBJECTIVE

The aim of this project is to utilize Bayesian networks within the wiseR tool to analyze and predict COVID-19 severity in patients. By replicating and enhancing an existing research study, we explore relationships among various biomarkers that could indicate the likelihood of severe COVID-19 outcomes by using wiseR dashboard.



TOPIC SELECTION

COVID-19 severity prediction was chosen due to its high relevance to public health. Identifying key biomarkers linked to severe outcomes can inform better clinical decision-making, resource allocation, and patient management strategies during pandemics.

In this project, we aim to validate the hypothesis that biomarkers like S100A12 are significantly associated with an increased risk of severe COVID-19 outcomes, as suggested in the referenced studies by Shubham Maheshwari et al. (2021) and Gallo Marin et al. (2021).

DATA & PREPROCESSING

DATA & PREPROCESSING

Dataset Overview:

We used "data.csv," which contains COVID-19 patient data including demographic information, protein markers, and clinical variables relevant to COVID-19 severity. The dataset allows us to explore how specific biomarkers correlate with disease outcomes.

Preprocessing Steps:

- **Merging and Dimensionality Reduction:** Protein and metabolome variables were processed using PCA to reduce dimensionality and focus on significant features.
- **Discretization:** Variables were discretized to simplify the analysis, making it compatible with Bayesian network modelling.

wiseR						
App Settings	Data	Association Network	Bayesian Network	Decision Network	Publish your dashboard	App Tutorial
Dataset	Explore					
upload	Pre-Process	Download				
Show 10 entries		Search: <input type="text"/>				
	P_interleukin_6_interferon_beta_2	P_fibroblast_growth_factor_21	P_pleiotrophin	P_troponin_I_type_3_cardiac	P_natriuretic_peptide_B	P_int
1	[2.35,5.98)	[9.74,14.4]	[4.48,6.8)	[2.02,4.83)	[0.847,3.29)	[1.17,3
2	[2.35,5.98)	[6.69,9.74)	[4.48,6.8)	[-1.64,2.02)	[0.847,3.29)	[-3.13,
3	[9.63,14.7]	[6.69,9.74)	[4.48,6.8)	[2.02,4.83)	[0.847,3.29)	[1.17,3
4	[9.63,14.7]	[3.47,6.69)	[4.48,6.8)	[2.02,4.83)	[0.847,3.29)	[3.68,8

DATA & PREPROCESSING

wiseR

App Settings Data Association Network Bayesian Network Publish your dashboard

Dataset Explore

upload Pre-Process Download

Show 10 entries Search:

	CVP	PCWP	HIST	TPR	BP	CO	HRBP	HREK	HRSA	PAP	SAO2	FIO2	PRSS	ECO2	MINV	MVS	HYP	LVF	APL	ANES	PMB	INT
1	NORMAL	NORMAL	FALSE	LOW	NORMAL	HIGH	HIGH	HIGH	HIGH	NORMAL	NORMAL	LOW	HIGH	ZERO	HIGH	NORMAL	FALSE	FALSE	FALSE	FALSE	FALSE	NOR
2	NORMAL	NORMAL	FALSE	NORMAL	LOW	LOW	HIGH	HIGH	HIGH	NORMAL	LOW	NORMAL	HIGH	ZERO	ZERO	NORMAL	FALSE	FALSE	FALSE	FALSE	FALSE	NOR
3	NORMAL	HIGH	FALSE	NORMAL	NORMAL	HIGH	HIGH	HIGH	HIGH	NORMAL	LOW	NORMAL	NORMAL	ZERO	ZERO	NORMAL	FALSE	FALSE	FALSE	FALSE	FALSE	NOR
4	NORMAL	NORMAL	FALSE	LOW	LOW	HIGH	HIGH	HIGH	HIGH	NORMAL	NORMAL	NORMAL	HIGH	ZERO	ZERO	NORMAL	FALSE	FALSE	FALSE	FALSE	FALSE	NOR
5	NORMAL	NORMAL	FALSE	LOW	LOW	NORMAL	HIGH	HIGH	HIGH	NORMAL	LOW	NORMAL	LOW	ZERO	ZERO	NORMAL	FALSE	FALSE	FALSE	FALSE	FALSE	NOR
6	NORMAL	NORMAL	FALSE	LOW	NORMAL	HIGH	HIGH	HIGH	HIGH	NORMAL	LOW	NORMAL	HIGH	HIGH	ZERO	NORMAL	FALSE	FALSE	FALSE	TRUE	FALSE	NOR
7	NORMAL	NORMAL	FALSE	LOW	LOW	HIGH	HIGH	HIGH	HIGH	NORMAL	LOW	LOW	LOW	ZERO	ZERO	NORMAL	FALSE	FALSE	FALSE	FALSE	FALSE	NOR
8	NORMAL	NORMAL	FALSE	NORMAL	LOW	LOW	HIGH	NORMAL	NORMAL	LOW	LOW	NORMAL	HIGH	ZERO	ZERO	NORMAL	FALSE	FALSE	FALSE	FALSE	FALSE	NOR
9	NORMAL	LOW	FALSE	LOW	LOW	HIGH	HIGH	HIGH	HIGH	NORMAL	LOW	NORMAL	HIGH	ZERO	ZERO	NORMAL	FALSE	FALSE	FALSE	FALSE	FALSE	NOR
10	NORMAL	NORMAL	FALSE	NORMAL	NORMAL	HIGH	HIGH	HIGH	HIGH	LOW	LOW	NORMAL	HIGH	LOW	ZERO	NORMAL	FALSE	FALSE	FALSE	FALSE	FALSE	ESO

< >

Showing 1 to 10 of 20,000 entries

Previous 1 2 3 4 5 ... 2000 Next

Fundina Support: The Wellcome Trust/DBT India Alliance grant IA/CPHE/14/1/501504

Figure S7. Users can download the pre-processed data as CSV files.

ASSOCIATION NETWORK

ASSOCIATION NETWORK

Before going to Bayesian Network analysis, it is instructive to visualize the association graph constructed upon the data. Learning association networks on the data before Bayesian learning can give useful insights.

Association Network

Export Table

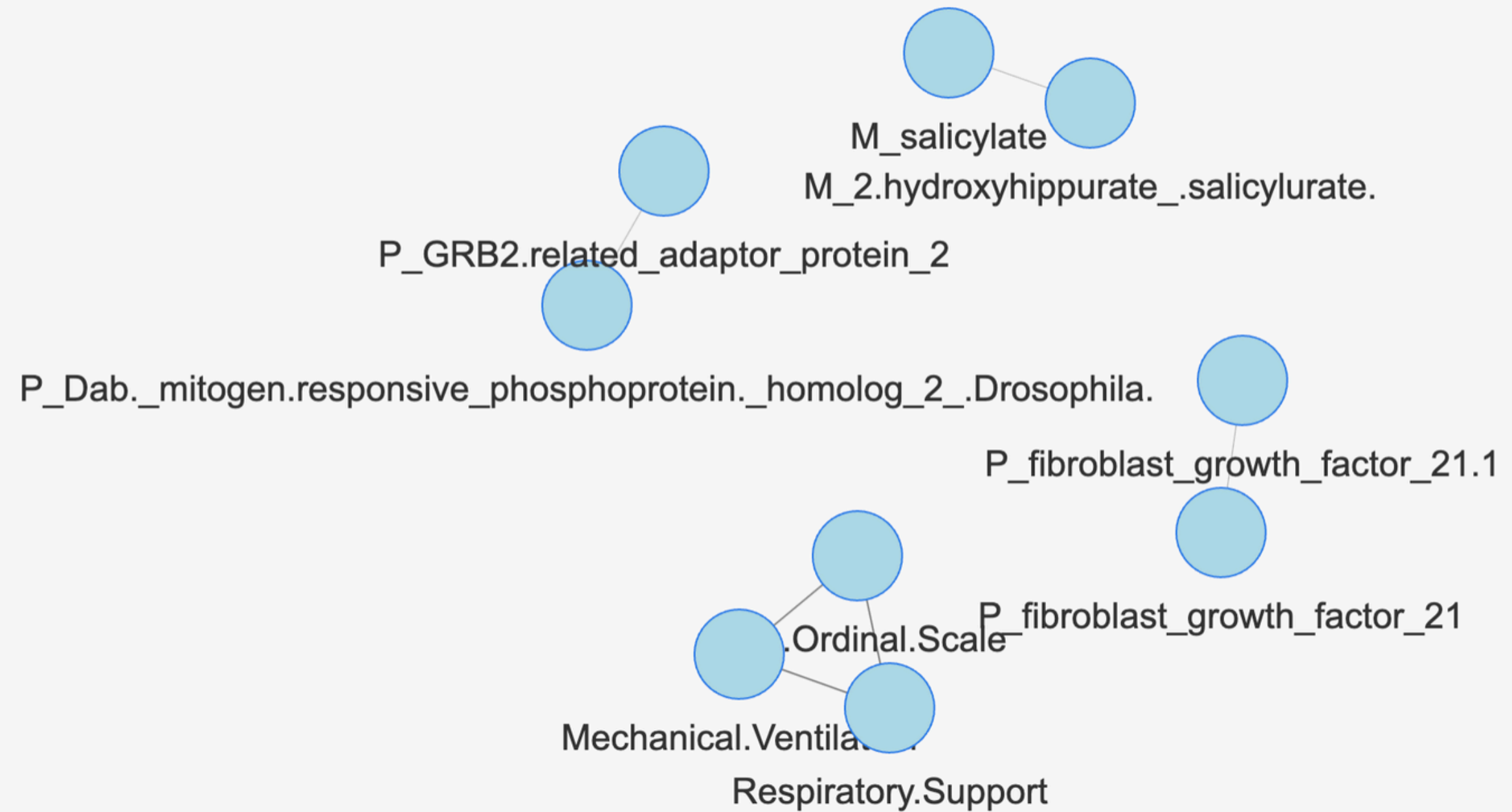
⚙ Visual Settings ▾

↻ Refresh

Nth neighbors(of selection):

graph ▾

Detect M



STRUCTURE LEARNING SETUP

STRUCTURE LEARNING SETUP

Bayesian Structure Learning was used to uncover probabilistic relationships between variables. We chose the Hill Climbing algorithm for its efficiency and applied Akeike Information Criterion (AIC) for network scoring to balance model complexity and accuracy.

Parameters:

Algorithm: Hill Climbing

Scoring Metric: Akeike Information Criterion (AIC)

Parameter Estimation: Bayesian parameter estimation with an imaginary sample size for robustness.

Bootstrap Replicates: 51, with edge strength set to 0.51.

Learning Algorithm

Hill Climbing (Recommended) ▼

Network Score

Bayesian Information Criterion ▼

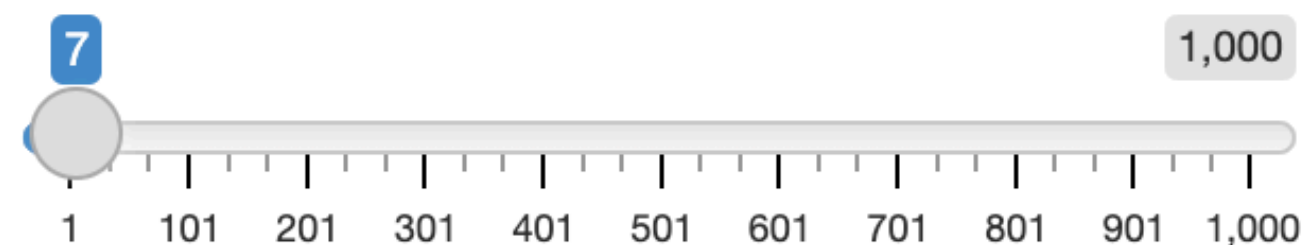
Use Expert Knowledge by Forcing/Prohibiting Edges

Blacklist ▼

Parameter fitting algorithm

Bayesian parameter estimation ▼

Imaginary sample size



Browse...

blacklist.csv

Upload complete

Bootstrap without resampling is available only for score-based learning

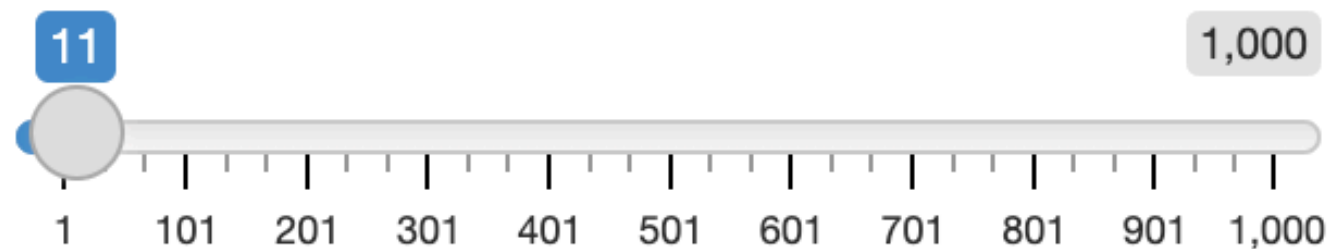
Disable resampling in bootstrap



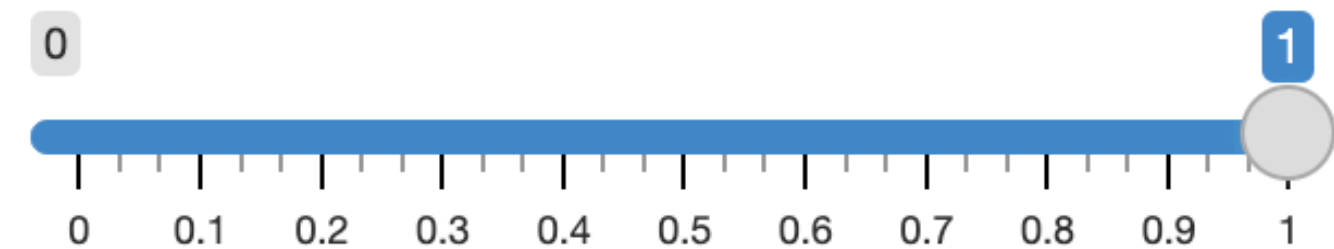
Disable resampling in bootstrap



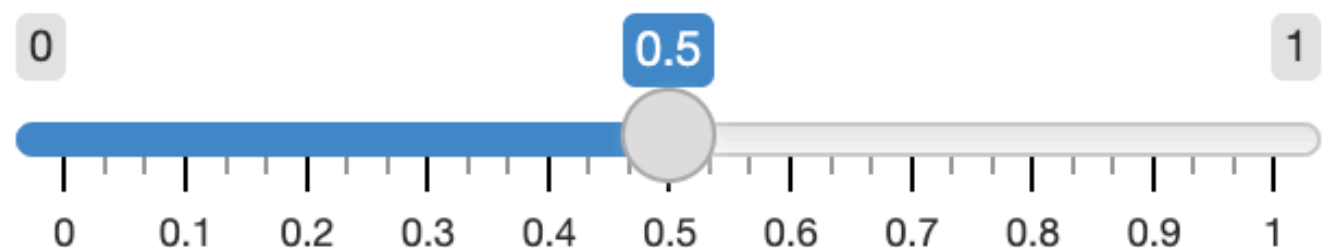
Bootstrap replicates



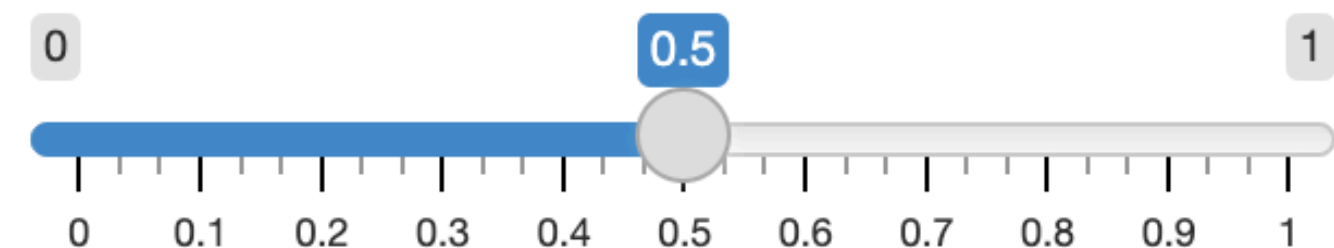
Proportion of sample for Bootstrap:



Edge Strength



Direction Confidence:



Bootstrap

One-time

Parameter Tuning

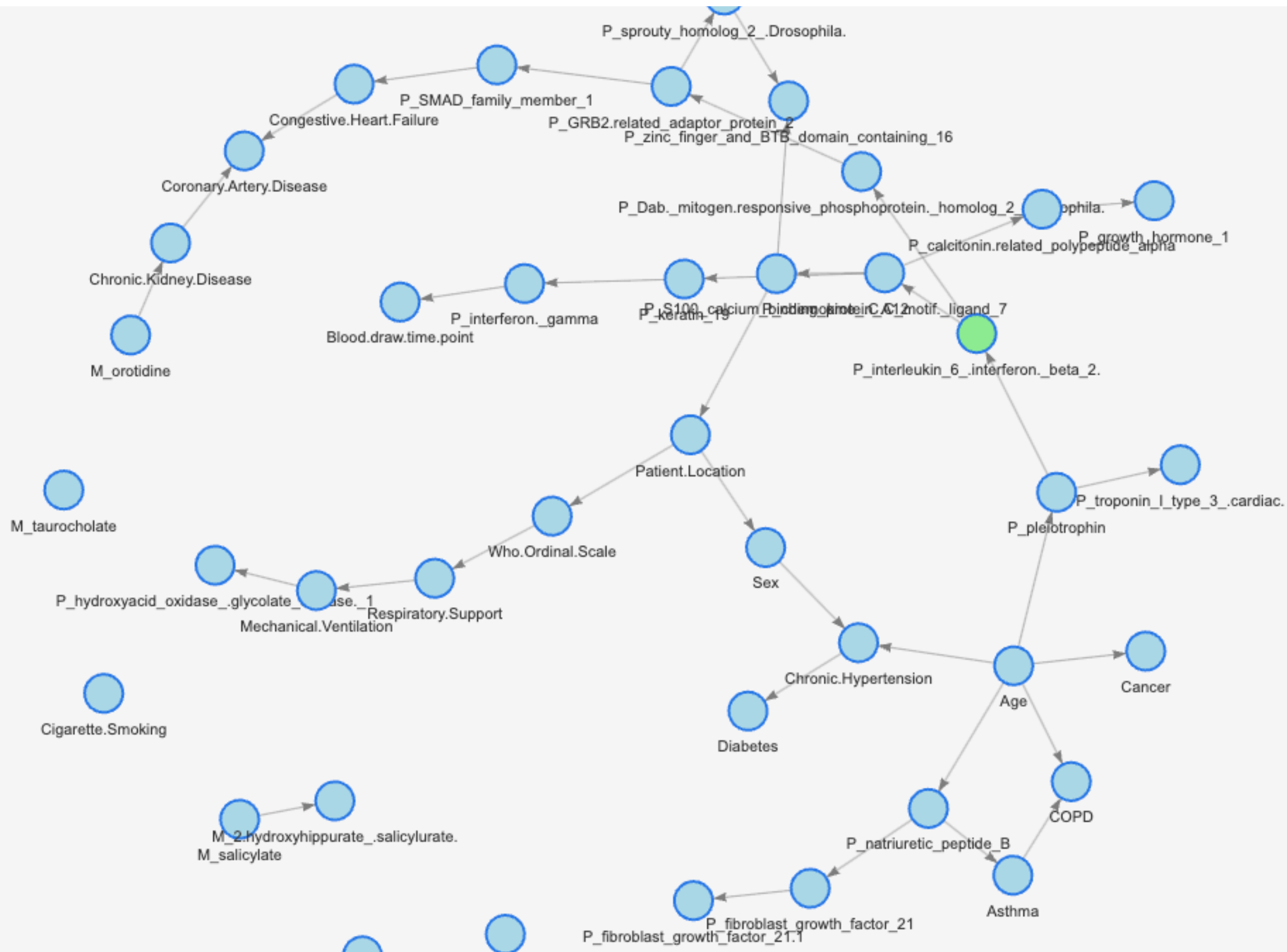
Blacklist Usage

Purpose:

The blacklist.csv file was utilized to restrict certain causal relationships within the network. This allowed us to prevent known irrelevant or non-causal pathways, focusing the network on more meaningful associations.

```
1 from,to
2 Who.Ordinal.Scale,Sex
3 Who.Ordinal.Scale,Age
4 Who.Ordinal.Scale,Ethnicity
5 Who.Ordinal.Scale,Race
6 Who.Ordinal.Scale,Cigarette.Smoking
7 P_interleukin_6_.interferon._beta_2.,Sex
8 P_fibroblast_growth_factor_21,Sex
9 P_pleiotrophin,Sex
10 P_troponin_I_type_3_.cardiac.,Sex
11 P_natriuretic_peptide_B,Sex
12 P_interferon._gamma,Sex
13 P_fibroblast_growth_factor_21.1,Sex
14 P_growth_hormone_1,Sex
15 P_Dab._mitogen.responsive_phosphoprotein._homolog_2_.Drosophila.,Sex
16 P_S100_calcium_binding_protein_A12,Sex
17 P_calcitonin.related_polypeptide_alpha,Sex
18 P_chemokine_.C.C.motif._ligand_7,Sex
19 P_keratin_19,Sex
20 P_zinc_finger_and_BTBDomain_containing_16,Sex
21 P_sprouty_homolog_2_.Drosophila.,Sex
22 P_GRB2.related_adaptor_protein_2,Sex
23 P_hydroxyacid_oxidase_.glycolate_oxidase._1,Sex
24 P_SMAD_family_member_1,Sex
25 P_apoptosis.inducing_factor._mitochondrion.associated._1,Sex
26 M_salicylate,Sex
27 M_orotidine,Sex
28 M_2.hydroxyhippurate_.salicylurate.,Sex
29 M_erythritol,Sex
30 M_tauricholate,Sex
31 Blood.draw.time.point,Sex
32 Asthma,Sex
```


NETWORK STRUCTURE VISUALIZATION



- Structure Learning ▾

Inference Learning ▾

Bayesian Network

Consensus Plot

Fitted Local Distributions

Infer Decisions

Export Tables

Nth Neighbors(of selection):

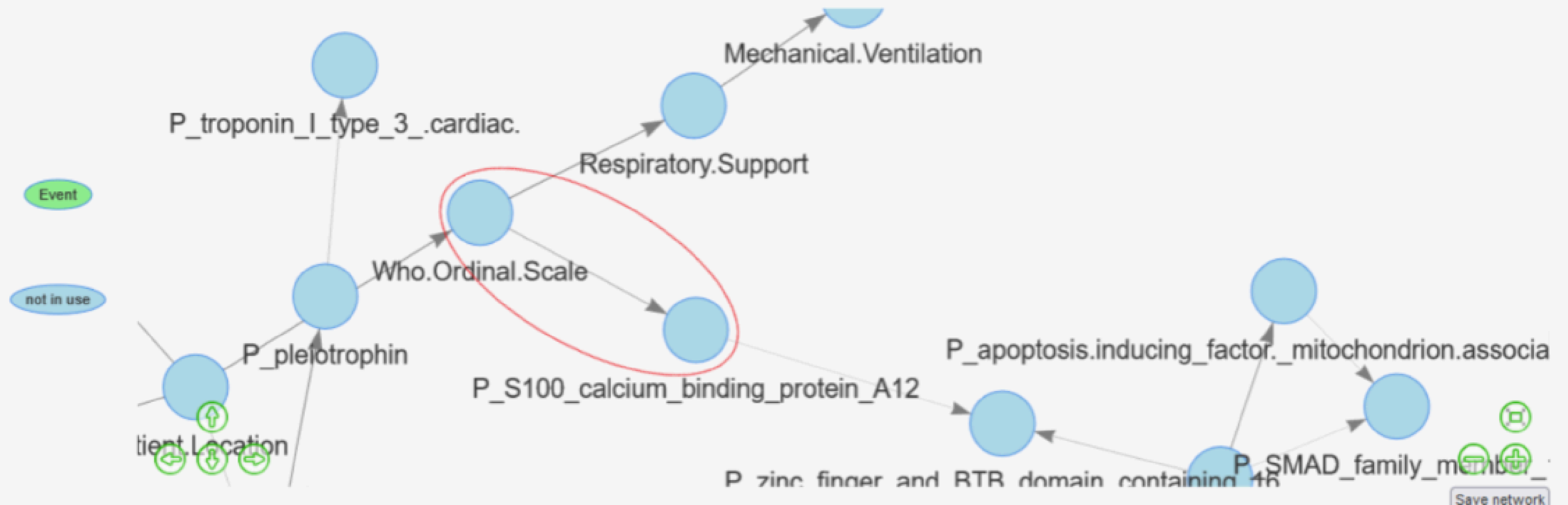
P_zinc_finger_and_BTb_domain_con

Modules:

graph

Detect Modules ▾

Visual Settings ▾

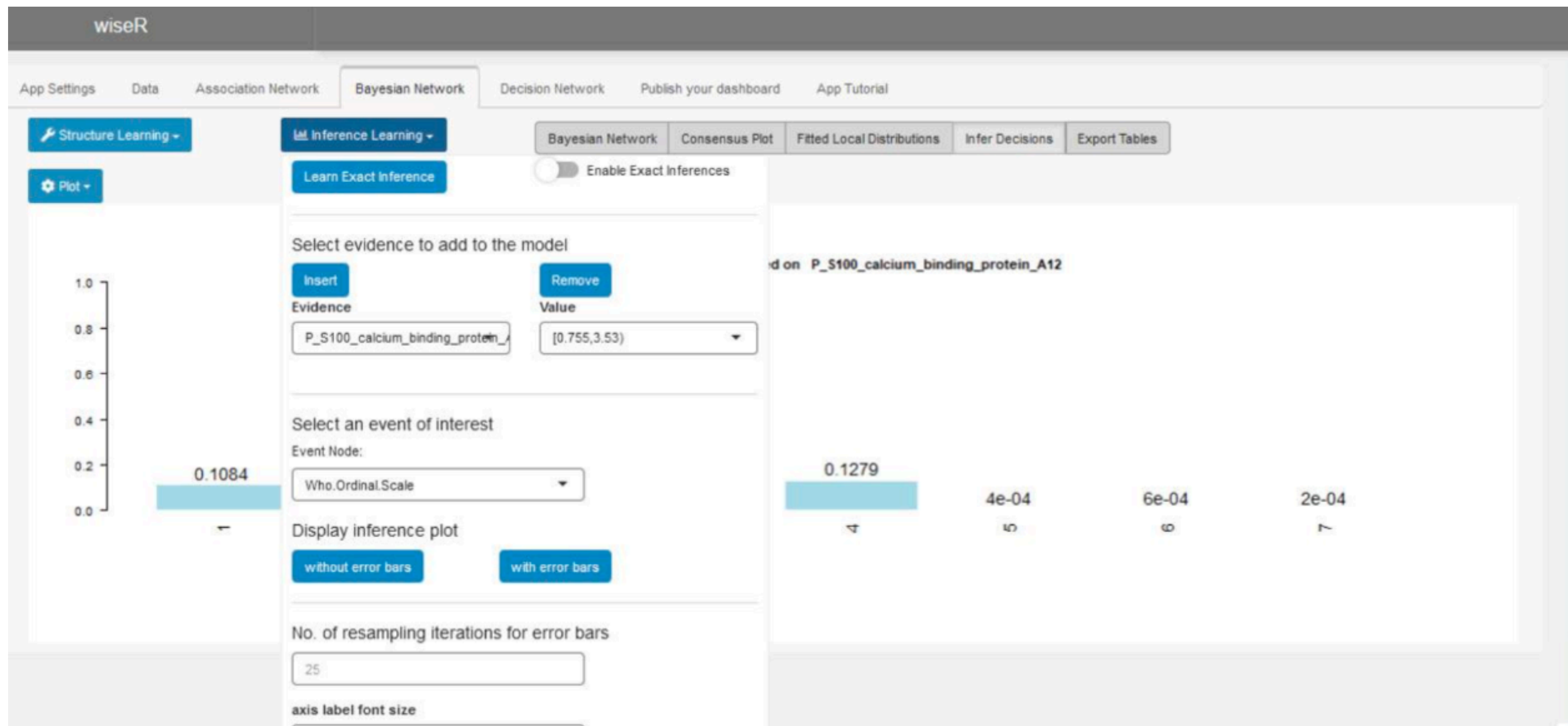
 RefreshSelect by id 

NETWORK STRUCTURE VISUALIZATION

Result:

The resulting Bayesian network illustrates connections among clinical and biomarker variables, highlighting key nodes such as "Who Ordinal Scale" (WOS) and "S100A12." These markers were identified as significant predictors of COVID-19 severity.

INFERENCE LEARNING & ANALYSIS



Inference Analysis of COVID Severity with S100A12 Markers

S100A12 is a protein marker studied for its correlation with COVID-19 severity. The Who Ordinal Scale (WOS) is used to classify COVID-19 severity, with levels ranging from 1 (mild) to 7 (severe).

wiseR's Bayesian inference feature allows us to set S100A12 levels as an evidence variable and predict COVID-19 severity.

Structure Learning

Inference Learning

Bayesian Network

Consensus Plot

Fitted Local Distributions

Infer Decisions

Export Tables

Plot

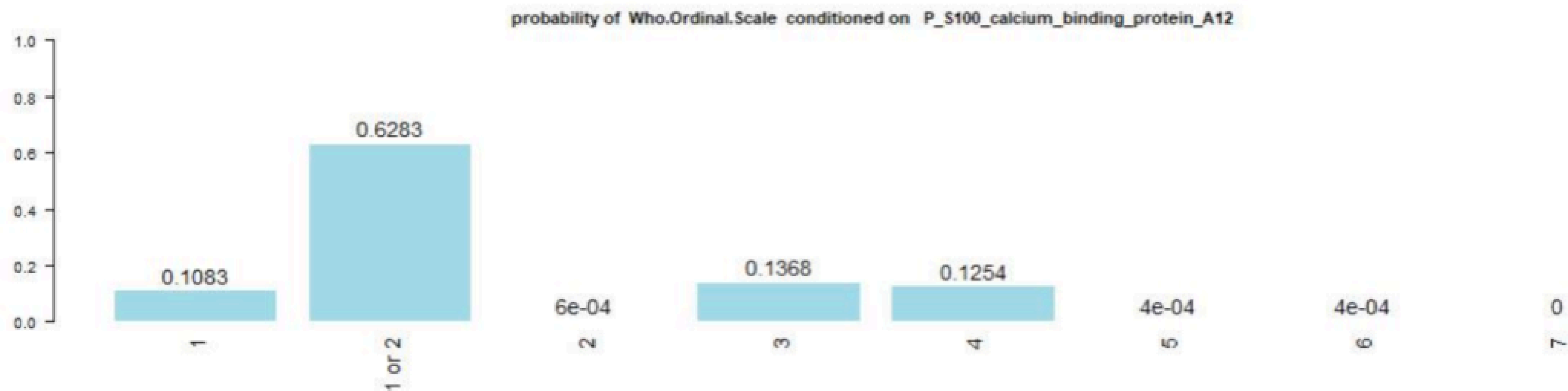


Fig. 7. Lowest Bracket for S100A12 markers indicate high chances of mild covid

S100A12 Protein as Evidence of COVID Severity

Low S100A12 Bracket: Indicate that patients with low levels of S100A12 have a high probability of mild cases (severity levels 1-2).

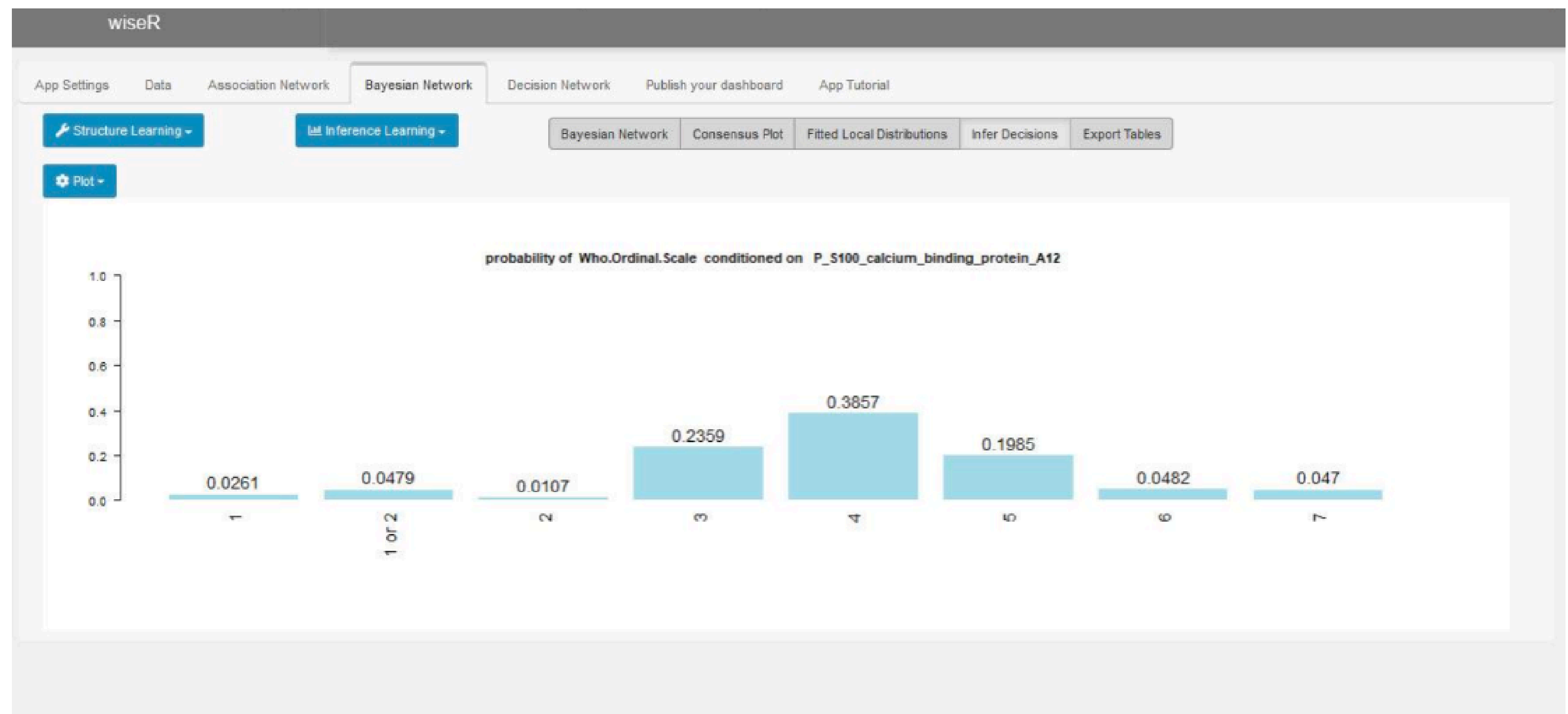


Fig. 8. Middle Bracket for S100A12 markers indicate high chances of moderate to severe covid

S100A12 Protein as Evidence of COVID Severity

Middle S100A12 Bracket: Higher protein levels show a shift toward moderate severity (severity levels 3-5).

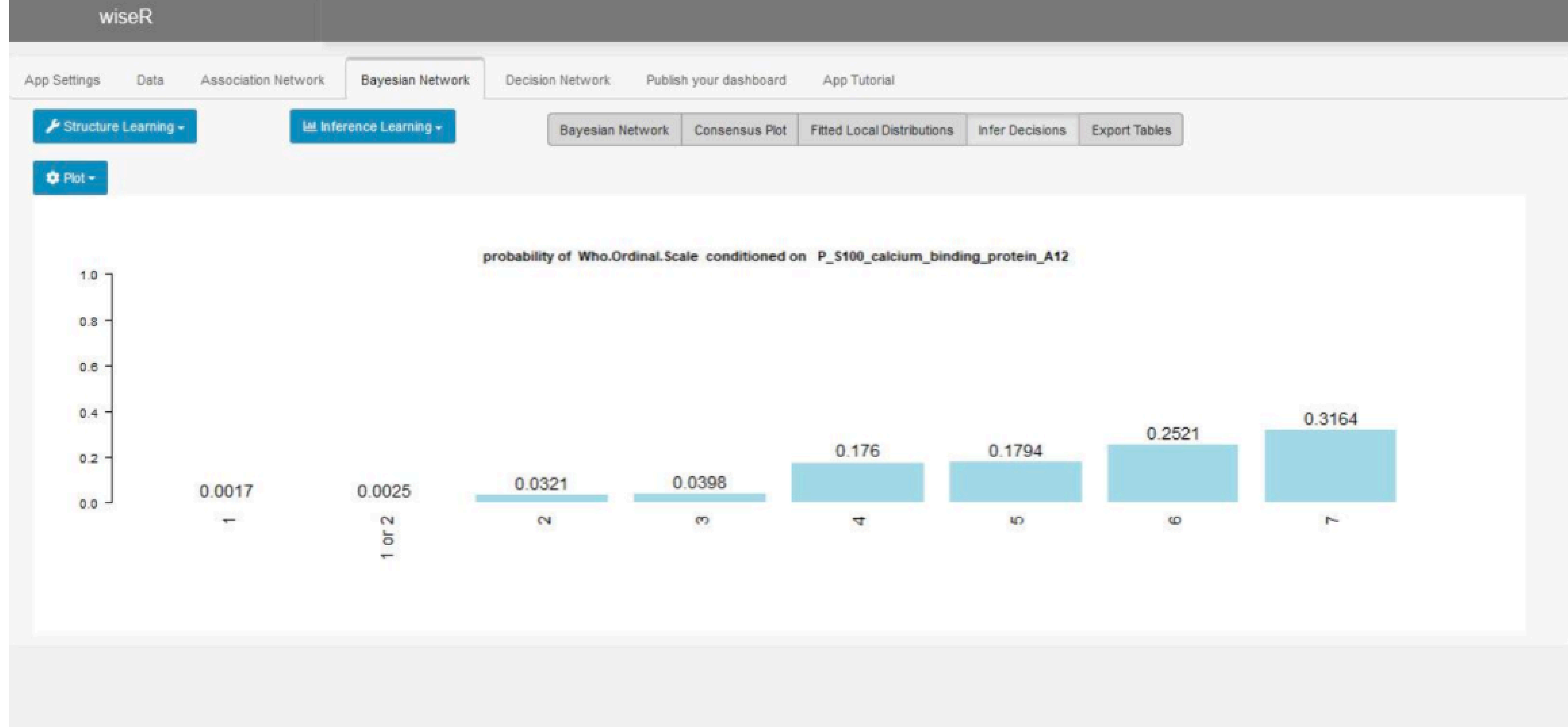


Fig. 9. Highest Bracket for S100A12 markers indicate high chances of severe covid

S100A12 Protein as Evidence of COVID Severity

High S100A12 Bracket: The highest bracket corresponds to an increased likelihood of severe COVID cases (severity levels 6-7).

CONCLUSIONS AND KEY INSIGHTS

Observations:

Key Findings: High levels of S100A12 and other protein markers correlated strongly with severe COVID-19 outcomes. This suggests that certain biomarkers can act as reliable predictors for managing patient risk.

Inference Plots: Show plots that illustrate the relationship between biomarkers (like S100A12) and severity likelihood.

Conclusion:

These insights demonstrate the potential for Bayesian networks to inform COVID-19 management by highlighting critical biomarkers associated with severity, supporting clinical triage and resource prioritization.

The results align with Lei (2021), underscoring S100A12's predictive power for COVID severity and the utility of wiseR in healthcare decision-making.”

Modifications

Modifications:

We significantly enhanced the original model by optimizing its structure with AIC scoring and the Hill Climbing algorithm, providing a balance between accuracy and complexity.

By using Bayesian parameter estimation, we improved the stability of the network's probability estimates, especially under data sparsity.

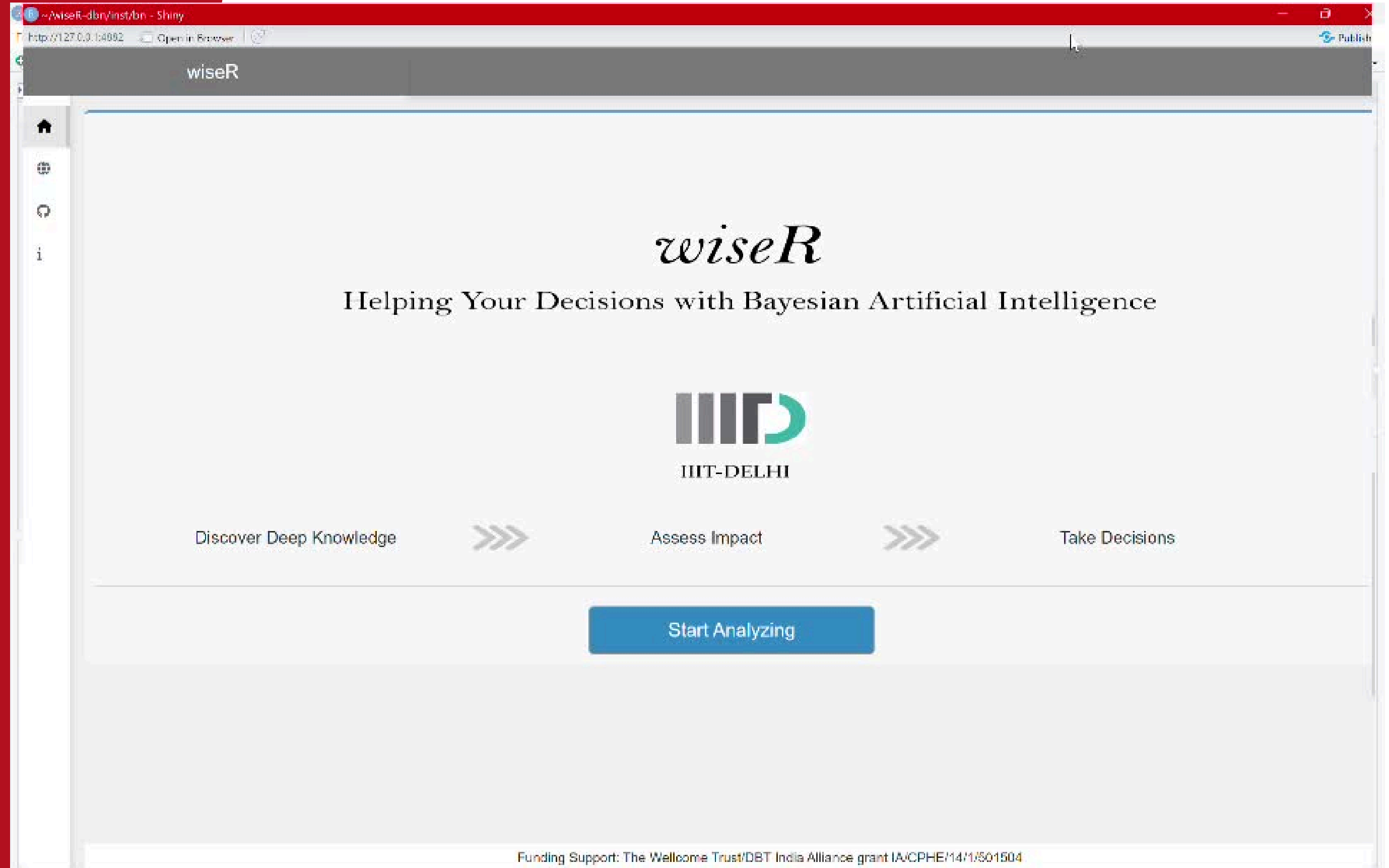
Our extensive evidence testing with S100A12 levels offered in-depth insights into severity prediction. These modifications illustrate wiseR's adaptability and value in healthcare decision-making.

Future Work

Future Improvements:

Future modifications could include expanding the network with additional biomarkers, optimizing the structure learning algorithm, and validating the model on different datasets to improve robustness and generalizability.

Demo Working Video



References:

- 1.Covid,2021-Gallo Marin B, Aghagoli G, Lavine K, Yang L, Siff EJ, Chiang SS, Salazar-Mather TP, Dumenco L, Savaria MC, Aung SN, Flanigan T, Michelow IC. Predictors of COVID-19 severity: A literature review. Rev Med Virol. 2021 Jan;31(1):1-10. doi: 10.1002/rmv.2146. Epub 2020 Jul 30. PMID: 32845042; PMCID: PMC7855377
- 2.WiseR- Shubham Maheshwari, Khushbu Pahwa, Tavpritesh Sethi, et al."WiseR: An end-to-end structure learning and deployment framework for causal graphical models" arXiv:2108.07046, 2021



Thank you!