

<sup>1</sup>Universidade Católica de Minas Gerais - PUCMINAS; <sup>2</sup>Universidade Federal de Minas Gerais - UFMG; <sup>3</sup>AMECI - Associação Mineira de Epidemiologia e Controle de Infecções, Belo Horizonte, MG, Brazil

# Background

"Every 3 seconds, someone in the world dies of sepsis. Globally, sepsis claims 11 million lives a year. Yet, for many patients, with treatable" diaanosis IS (https://sepsistrust.org/about/).

We used causal inference theory as a in silico method to identify biomarkers of sepsis. Causal Inference is a theory in Machine Learning that seeks for the root causes of an event.

The objective of our study is to identify mRNA expression profiles of sepsis and non-sepsis patients.

## Methods

In the study it was used transcription profile data downloaded from http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE12624. This dataset has 70 samples, being 34 sepsis and 36 non-sepsis samples. Data set contains 8,519 attributes: 7,672 genes obtained after preprocessing of the mRNA expression profile data.

The method applied in the dataset was a modification in the HEISA, a local learner of two stages algorithm (Figure 1). In the first stage HEISA identifies variables that compounds the set of parents, children, parents of parents and children of children of a target.

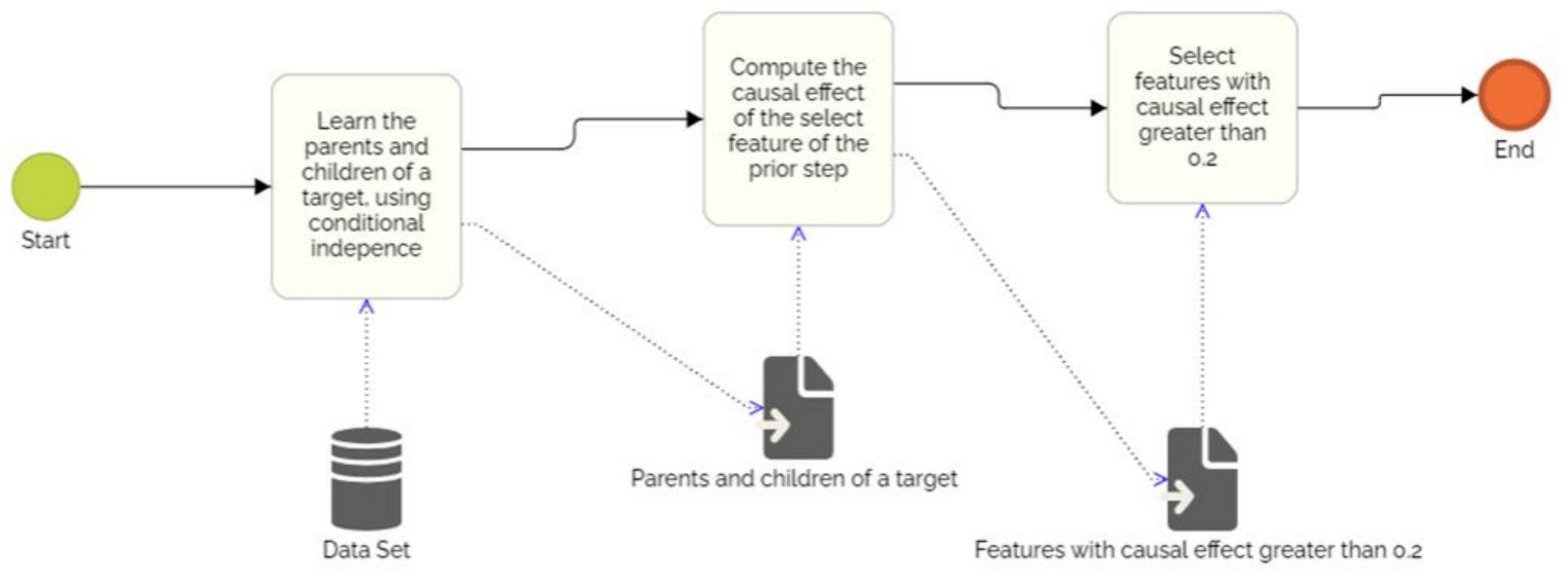
During the second stage is calculated the causal effect using do-calculus method, of the selected variables of the first stage in the target. At end of the second stage, features with causal effect greater than 0.2 is selected. After selecting the features mRNA expression, it was applied two algorithms of classification, Random Forest and K-means, in order to evaluate the ability of the selected variables of identifying the occurrence of sepsis.

# In Silico Modeling by Causal Inference for Identifying Biomarkers of Sepsis

#### Walisson Carvalho<sup>1</sup>, Rita Silvério-Machado<sup>2</sup>, Bráulio R.G.M Couto<sup>3</sup>, Luis Enrique Zarate<sup>1</sup>

# Methods

The method applied in the dataset was HEISA, this method is a local learner of three stages as presented in Figure above:



#### Figure 1 - Causal inference HEISA algorithm

## Conclusion

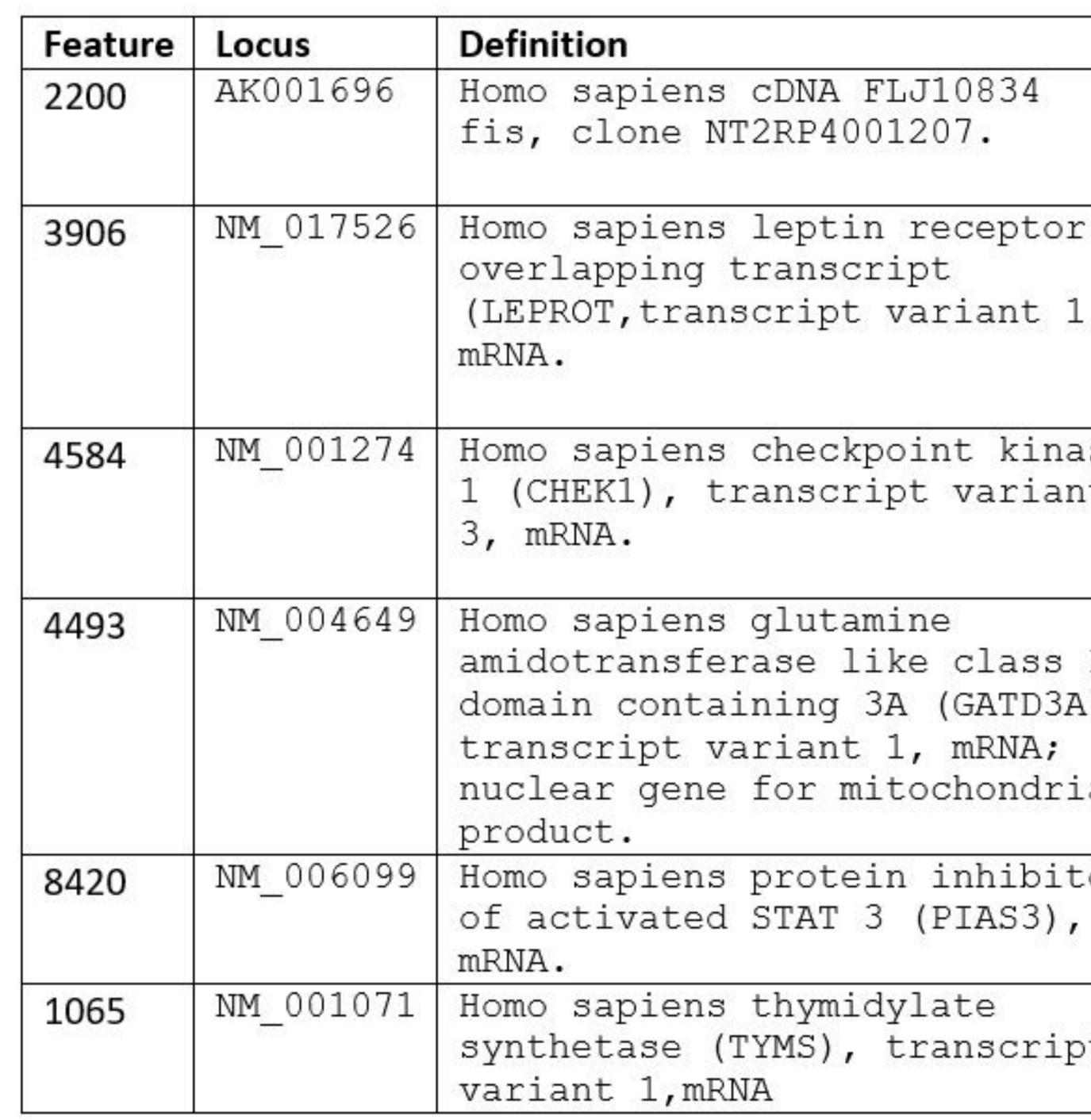
In a big set of 7,672 genes, only six were returned as sepsis biomarkers candidates by the causal inference method. This is a promising in silico discovery, but they need experimental validation for their effectiveness.

Regarding the classification task, using only those six mRNA expressions it was possible to identify all patients, in the data set, that had sepsis. In other words, the accuracy of the task of classification was 100%.

### Results

sepsis occurs, or not.

#### **Table I - Sepsis biomarkers candidates**



### References

de Carvalho, W. F. & Z´arate, L. E. (2020). A new local causal learning algorithm applied in learning analytics. The International Journal of Information and Learning Technology, 38(1), 103–115. Epskamp, S. & Fried, E. I. (2018). A tutorial on regularized partial correlation networks. Psychological Methods, 23(4), 617–634.

PEARL, J. Causality: Models, Reasoning, and Inference. 2. ed. [S.l.]: Cambridge University Press, 2009. Yu, K., Guo, X., Liu, L., Li, J., Wang, H., Ling, Z., & Wu, X. (2019). Causality-based feature selection: Methods and evaluations.





#### As shown above, after applying our algorithm, it was select six mRNA expressions that better explains whether

	link	Causal Effect
	https://www.ncbi.nlm.nih.gov/ nuccore/AK001696	-0.433910
r 1,	https://www.ncbi.nlm.nih.gov/ nuccore/NM_017526	-0.218894
ase nt	https://www.ncbi.nlm.nih.gov/ nuccore/NM_001274	-0.303484
1 A), ial	https://www.ncbi.nlm.nih.gov/ nuccore/NM_004649	0.302175
tor ,	https://www.ncbi.nlm.nih.gov/ nuccore/NM_006099	0.6340434
pt	https://www.ncbi.nlm.nih.gov/ nuccore/NM_001071	-0.256266

**CONTACT INFORMATION** 

coutobraulio@hotmail.com