Identification, classification, and prioritization of most influential players in normal biological processes and diseases

Abbas Salavaty[1,2], Mirana Ramialison[1,2], and Peter D Currie[1,3]

[1]Australian Regenerative Medicine Institute-, [2]Systems Biology Institute Australia-, [3]EMBL Australia-Monash University

ABSTRACT

High-throughput technologies have enabled the identification of thousands of genes and proteins at a time and across several conditions. However, one of the biggest challenges is the selection of right candidates amongst thousands of features for experimental validation. Currently, numerous models have been developed for candidate prioritization, most of which rely on external sources of information and are not able to classify features. On the other hand, we present ExIR, a one-stop and versatile model that extracts, classifies and prioritizes candidate features from any type of experimental data and is accessible on the Influential Software Package web portal at:

https://influential.erc.monash.edu/ExIR

CONTACT

Adrian (Abbas) Salavaty
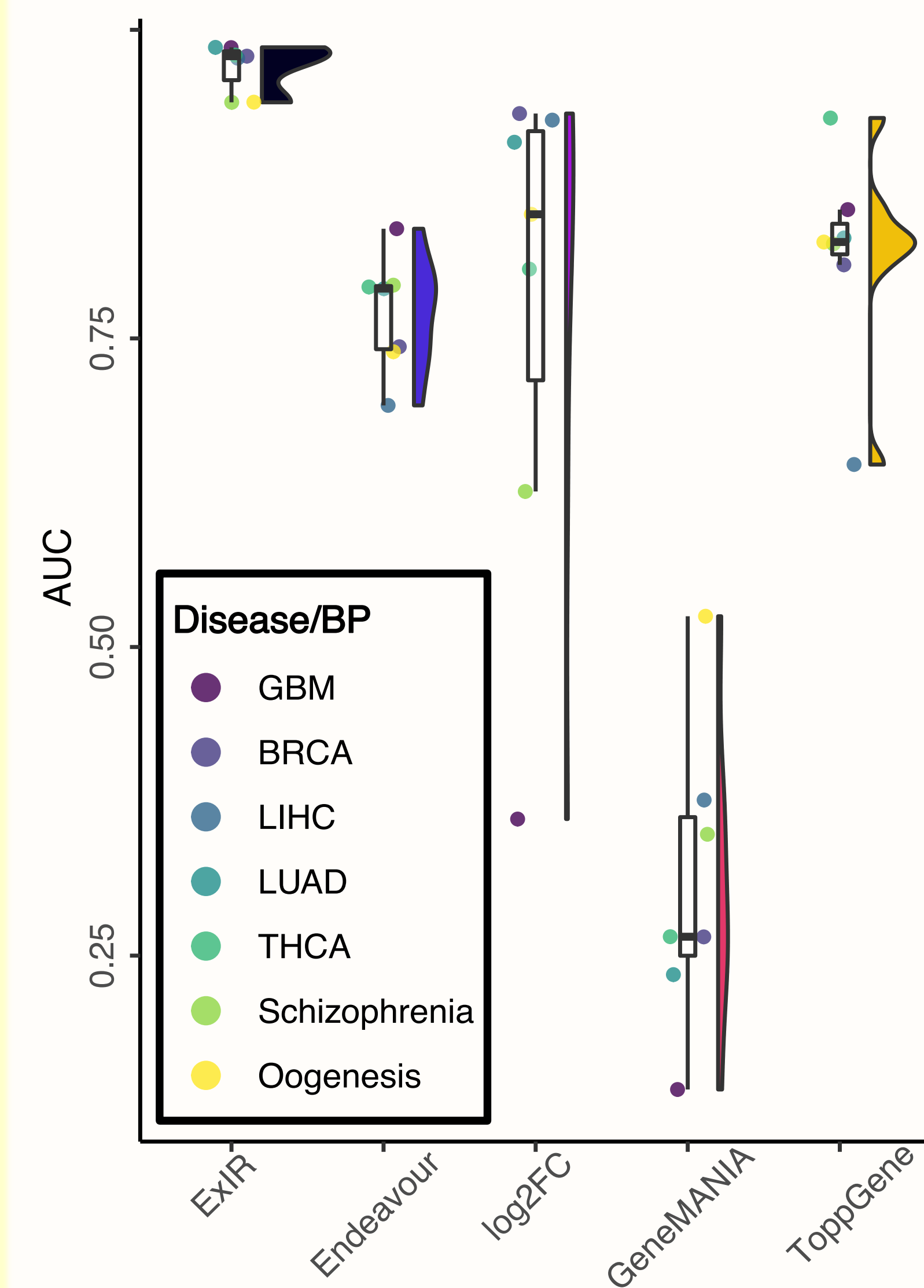ARMI, Monash University
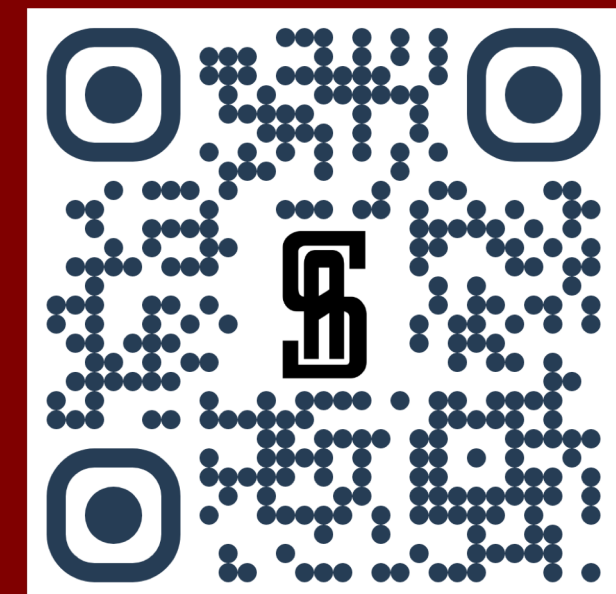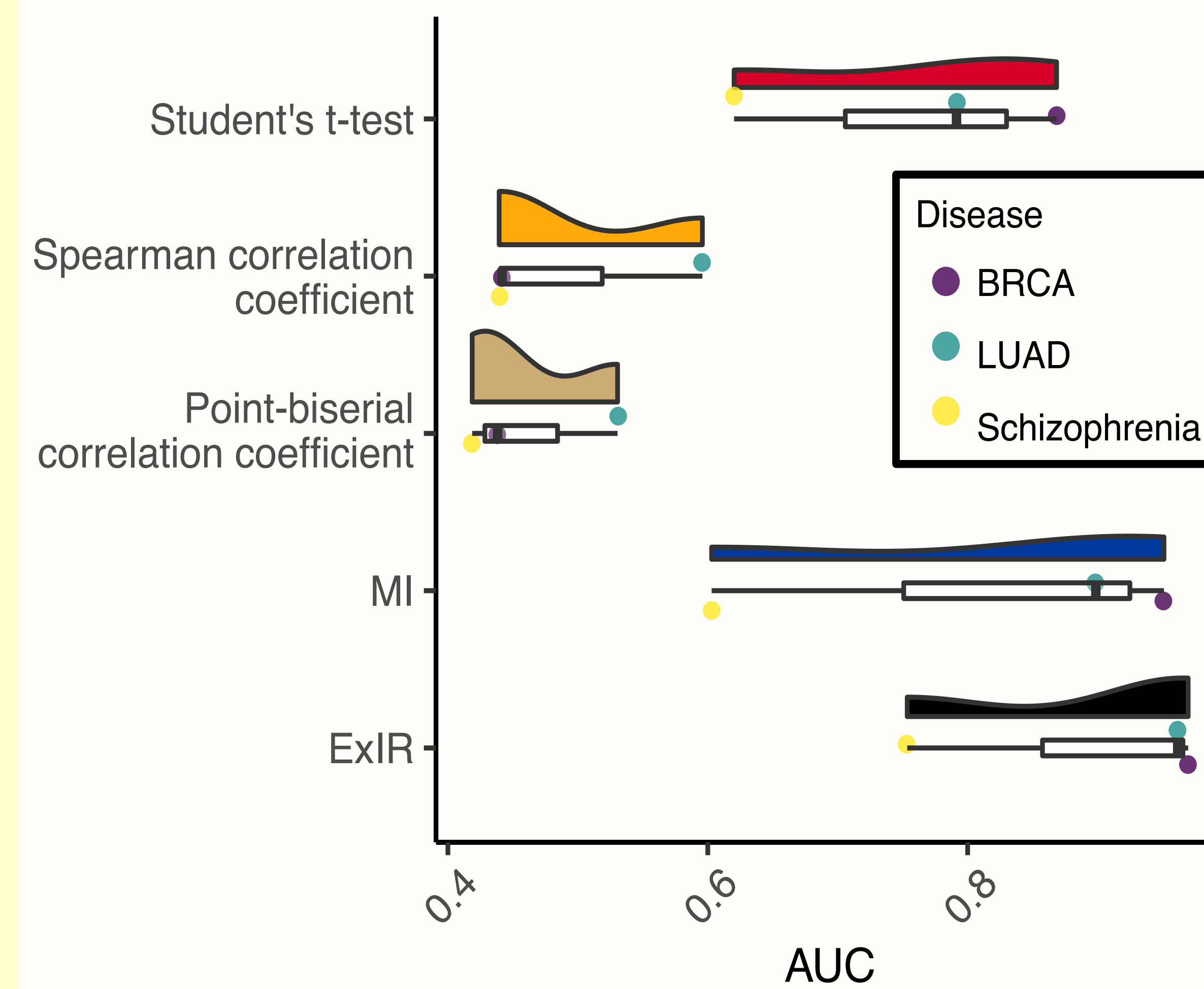
Personal Website

@mania_abbas

ExIR Shiny App

INTRODUCTION

High-throughput technologies (*e.g.* next-generation sequencing) have enabled biologists to investigate the omics profiles of two or more biological conditions to examine hypotheses and identify drivers of a process. Although it is logistically impossible to evaluate all of the results in the lab and require filtration and selection of the right candidates. Accordingly, several computational models have been proposed for the prioritization of gene lists. However, most of them rely on external sources of information such as gene ontologies and pathways (1-3) which results in several deficits and makes the outcome biased towards well-studied genes (4). By contrast, a robust prioritization model that works based on experimental data could remove such deficits. The Experimental-data-based Integrative Ranking (ExIR) is a prioritization model we developed to solve these problems. Also, ExIR classifies the features into *Drivers*, *Biomarkers* and *Mediators*. ExIR was developed using network analysis (5) and machine learning (ML) techniques and recruits the *Addition* and *Multiplication* functions to integrate its generated scores.

Fig. 2. Superiority of ExIR in biomarker prioritization in comparison with other methods. This figure has been adapted from the ExIR manuscript.

Fig. 4. The percentage of biological pathways corresponding to ExIR-derived mediators that are associated with their respective disease.

- ExIR can identify the disease mediators which are not differentially expressed at all but might play mediatory and signaling roles between disease driver genes.
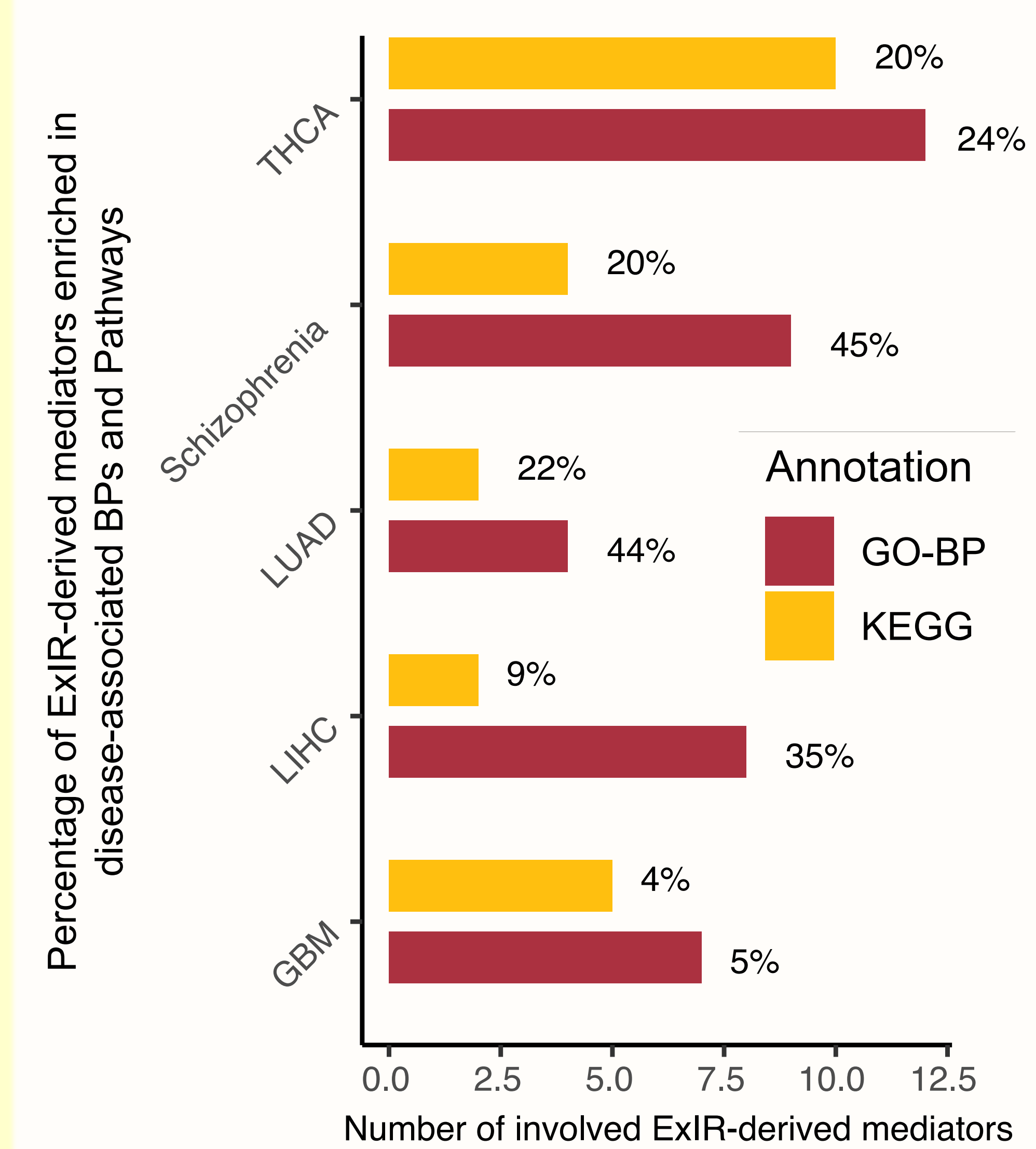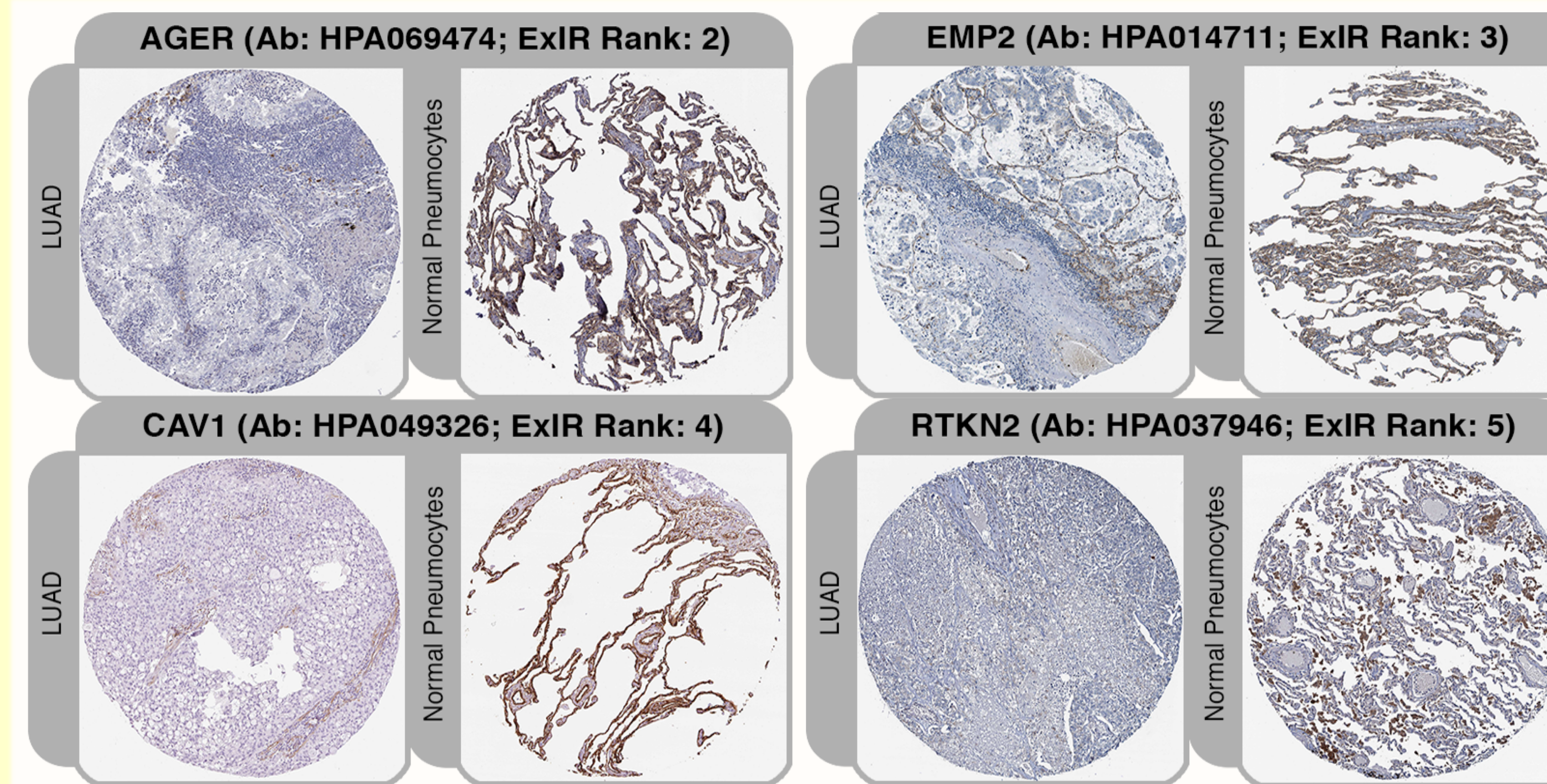

Fig. 1. Outperformance of ExIR in driver gene prioritization in comparison with other methods. This figure has been adapted from the ExIR manuscript.

- The performance of ExIR was evaluated in comparison with other contemporary prioritization models and the ground truth genes for evaluations were obtained from databases of manually curated Drivers and Biomarkers

- Evaluations were done using the receiver operating characteristic (ROC) analyses and measuring the area under the ROC curves (AUCs).

Fig. 3. The immunostaining of top predicted LUAD biomarkers prioritized by ExIR. This figure has been adapted from the ExIR manuscript.

- ExIR is able to identify known as well as novel disease biomarkers. The accuracy of the novel biomarkers was validated by immunohistochemistry data obtained from the Human Protein Atlas.

CONCLUSIONS

- ExIR is a one-stop and versatile model for the classification and prioritization of biological entities
- ExIR can be applied on any type of experimental data
- ExIR proposes a new class of genes termed "mediators", which could enhance our understanding of the development of normal and disease processes

REFERENCES

1. Zolotareva O, Kleine M. A Survey of Gene Prioritization Tools for Mendelian and Complex Human Diseases. J Integr Bioinform. 2019;16(4):20180069.
2. Raj MR, Sreeja A. Analysis of Computational Gene Prioritization Approaches. Procedia Comput Sci. 2018;143:395-410.
3. Seyyedrazzagi E, Navimipour NJ. Disease genes prioritizing mechanisms: a comprehensive and systematic literature review. etw Model Anal Health Inform Bioinforma. 2017;6(1):13.
4. Piro RM, Di Cunto F. Computational approaches to disease-gene prediction: rationale, classification and successes. FEBS J. 2012;279(5):678-696.
5. Salavaty A, Ramialison M, Currie PD. Integrated Value of Influence: An Integrative Method for the Identification of the Most Influential Nodes within Networks. Patterns. 2020;1(5):100052.