

Accepted Manuscript

A novel unsupervised learning model for detecting driver genes from pan-cancer data through matrix tri-factorization framework with pairwise similarities constraints

Jianing Xi, Ao Li, Minghui Wang

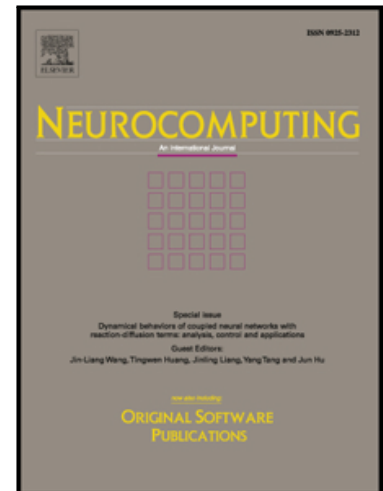
PII: S0925-2312(18)30318-7
DOI: [10.1016/j.neucom.2018.03.026](https://doi.org/10.1016/j.neucom.2018.03.026)
Reference: NEUCOM 19428

To appear in: *Neurocomputing*

Received date: 11 October 2017
Revised date: 22 January 2018
Accepted date: 14 March 2018

Please cite this article as: Jianing Xi, Ao Li, Minghui Wang, A novel unsupervised learning model for detecting driver genes from pan-cancer data through matrix tri-factorization framework with pairwise similarities constraints, *Neurocomputing* (2018), doi: [10.1016/j.neucom.2018.03.026](https://doi.org/10.1016/j.neucom.2018.03.026)

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.



A novel unsupervised learning model for detecting
driver genes from pan-cancer data through matrix
tri-factorization framework with pairwise similarities
constraints

Jianing Xi^a, Ao Li^{a,b,*}, Minghui Wang^{a,b,*}

^a*School of Information Science and Technology, University of Science and Technology of
China, Hefei 230027, PR China*

^b*Centers for Biomedical Engineering, University of Science and Technology of China,
Hefei 230027, PR China*

Abstract

Identifying cancer-causing mutated driver genes from passenger mutations is crucial to enhance the development of cancer diagnostics and therapeutics, and many previous efforts have been undertaken to identify cancer driver genes from somatic mutation data of specific types of cancers. However, many driver genes are underestimated when the mutation data of only specific cancers are investigated, which complicates the understanding of tumorigenesis. According to recent studies, cancers of disparate organs have many shared genomic mutations, and some driver genes that are not highly frequently mutated in patients of one cancer type may display considerable mutation frequencies across patients of multiple cancer types. By taking into account both the similarities of mutation profiles of different cancer types and the information of gene interaction network, we propose a novel unsupervised learning model based on matrix tri-factorization by learning the similarities from pairwise constraints to detect driver genes from pan-cancer data. In the evaluation of known benchmarking genes, our model achieves better performance than those of the existing matrix factorization based methods which do not consider the pairwise similarities between cancers. Furthermore, the detection performance of our model is also largely increased (area

*Corresponding authors.

Email addresses: aoli@ustc.edu.cn (Ao Li), mhwang@ustc.edu.cn (Minghui Wang)

Download English Version:

<https://daneshyari.com/en/article/6863974>

Download Persian Version:

<https://daneshyari.com/article/6863974>

[Daneshyari.com](https://daneshyari.com)