

Identification and Prioritization of Personalized Cancer Drivers

Cesim Erten, Antalya Bilim University, Antalya Turkey; Aissa Houdjedj, Antalya Bilim University, Antalya Turkey; Hilal Kazan, Antalya Bilim University, Antalya Turkey; Amine Taleb, Antalya Bilim University, Antalya Turkey

Motivation

A major challenge in cancer genomics is to distinguish the driver mutations that are causally linked to cancer from passenger mutations that do not contribute to cancer development. The majority of existing methods provide a single driver gene ranking for the entire cohort of patients. However, since mutation profiles of patients from the same cancer type show a high degree of heterogeneity, a more ideal approach is to identify patient-specific drivers.

Methods

We propose a novel method called PersonaDrive that integrates genomic data, biological pathways, and protein connectivity information for personalized identification of driver genes. The method is formulated on a personalized bipartite network (PBN) for each patient. One partition of the graph consists of the set of mutated genes of the specific patient under consideration and the other partition contains the set of differentially expressed genes (DEGs) of all the patients. An edge (i,j) is introduced in the PBN of a patient p_a if i is a mutated gene of the patient p_a , i and j interact in the input interaction network, and i is further mutated in patient p_b , where j is designated as a DEG in p_b . Our approach provides a personalized ranking of the mutated genes of a sample p_a based on the sum of normalized pairwise pathway coverage (PPC) scores across all the samples. Each edge (i,j) in the PBN of p_a is assigned a weight corresponding to its PPC, the number of pathways containing both i and j . The idea behind the PPC scoring is that the more the pathways shared by a mutated gene and DEG pair of a sample, the larger the potential for the mutated gene to be a cancer driver in that sample. The weight of each edge (i,j) is then normalized with respect to the pairwise patient similarity (PPS) of the sample pair p_a, p_b , where PPS measures the amount of overlap between the DEG sets of the pair.

Results

We compare our method against five state-of-the-art patient-specific cancer gene prioritization methods. The comparisons are with respect to a novel evaluation method that takes into account the personalized nature of the problem. We show that

our approach outperforms the existing alternatives for both the TCGA and the cell line data; the results of the former are depicted in Figure-1. Additionally, we show that the KEGG/Reactome pathways enriched in our ranked genes and those that are enriched in cell lines' reference sets overlap significantly when compared to the overlaps achieved by the rankings of the alternative methods. Our findings can provide valuable information towards the development of personalized treatments and therapies.

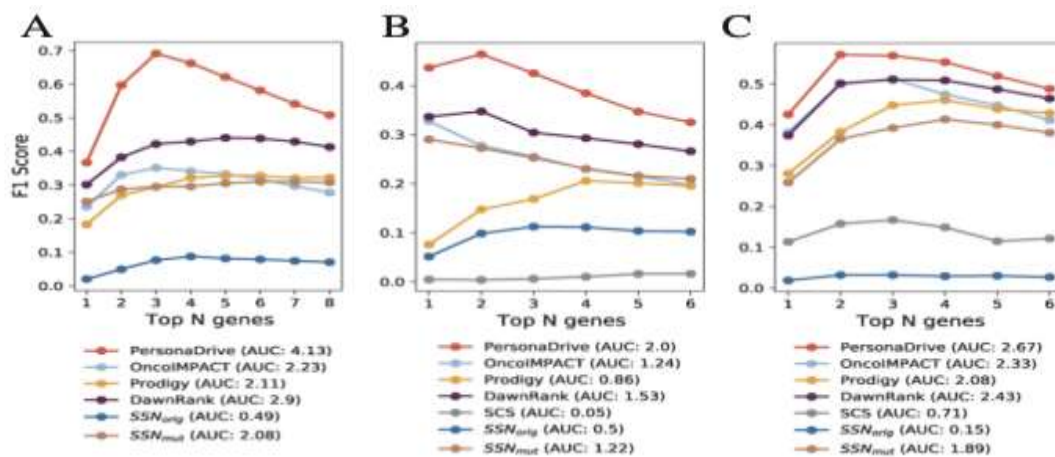


Figure-1

Availability

All the code and data are available at <https://github.com/abu-compbio/PersonaDrive>

Acknowledgements

This work has been supported by the Scientific and Technological Research Council of Turkey [117E879 to H.K. and C.E.] and Health Institutes of Turkey [TÜSEB 2019-TA-01-4069 to H.K and C.E].