Supplementary manuscript of

Knowledge-Embedded Constrained Multiobjective Evolutionary Algorithm Based on Structural Network Control Principles for Personalized Drug Targets Recognition in Cancer

## Constructing PGIN by Paired-SSN method

For the paired-SSN method [S1], the first step is building the co-expression network based on the tumor sample and the normal sample of an individual patient [S2]. Then, we needed to determine whether this edge is used to construct the PGIN according to the P-value of the edge between gene *i* and gene *j* in the normal sample network and tumor sample network. The specific conditions are as follows: If the P-value is lower than 0.05 in the tumor sample network (the coexpression relationship between the interaction of two genes is significant) and larger than 0.05 in the normal sample network (not significant), or vice versa, this edge is retained to constitute the PGIN. In addition, we can get P-value of an edge by calculating  and then counting its Z-value of . The  of an edge between gene *i* and gene *j* and its Z-score can be calculated:



where *n* represents the number of reference samples and k represents the *k*-th patient in the perturbed network.  represents the PCC of an edge between genes *i* and *j* in the reference network; and represents the PCC of the edge between genes *i* and gene *j* in the perturbed network. Here, we calculated a measure to score the pPCC of edges in the PGIN by integrating gene mutation data across cancer type-specific data into the PGIN as follows:

## 

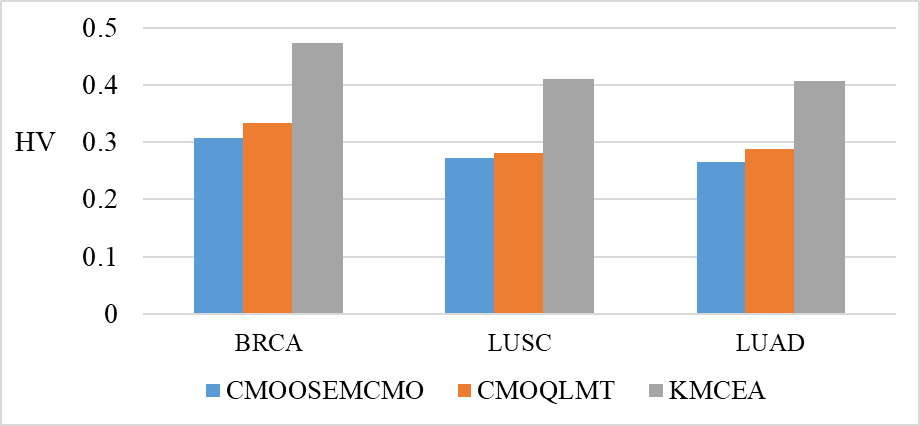
where norm represents the min-max normalized function.  and  respectively is the collection of tumors that exist mutated genes *i* and gene *j* after checking for somatic mutations in a given cancer data set;  indicates that 10% of the data falls under  after sorting a set of data in ascending order.

## The description of network and differential expression genes (DEG)-based methods

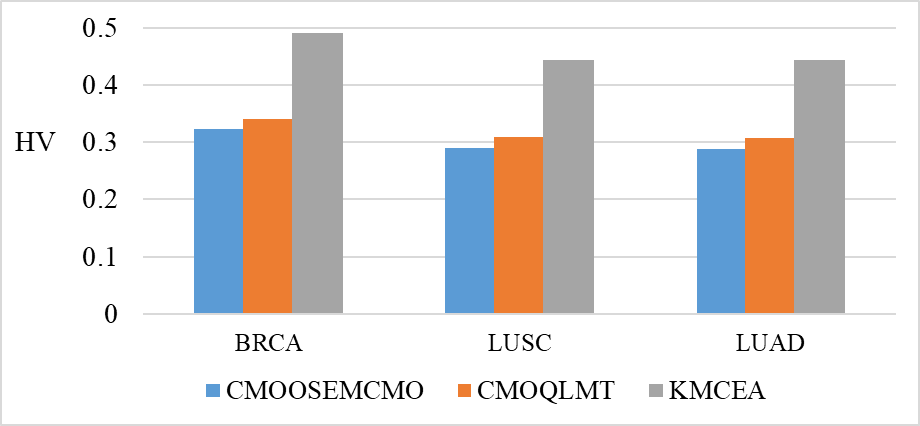
The network-based methods for identifying personalized drug targets (PDTs) were also taken for comparison in this paper, including CPGD [S2], ActiveDriver [S3], OncoDriveFM [S4], DriverML [S5] and Hub-genes. The cancer driver genes of CPGD and DriverMLwere obtained from their provided list of driver genes. Meanwhile, the driver genes in ActiveDriver and OncoDriveFM were obtained from the DriverDBv2 database [S6]. The hub gene selection method regards the hub genes in the constructed network as cancer driver genes. After the degree distribution of all genes T in the PGIN was obtained, a threshold was used in the following formula to obtain the hub genes:, where  and where  are the mean and standard variance of the degree distribution *T* of all genes, respectively. The DEG-based methods consist of DEG-Folchange, DEG-p-value, and DEF-FDR. Specifically, DEG-FoldChange selects the PDGs by calculating the fold-change between normal samples and tumor samples (log2(fold-change)| > 1). The DEG-p-value and DEG-FDR select the PDGs by calculating the p-value and FDR (<0.05) between a cancer tumor sample and a group of control samples, respectively.

## Compared results with two reinforcement learning-based CMOEAs

Two reinforcement learning-based CMOEAs are CMOOSEMCMO [S7] and CMOQLMT [S8]. CMOOSEMCMO uses deep reinforcement learning method to dynamically adjust the evolutionary operators of the population. CMOQLMT uses Q-learning and deep Q-learning methods to choose the most suitable auxiliary task during the evolutionary process. The compared results regarding HV and AUC results are provided in Figs. S-1 and S-2 respectively. Based on HV results, two compared algorithms are significantly worse than KMCEA on all three datasets under two frameworks. This is because two compared algorithms are designed for benchmark test functions. In addition, the AUC results also show that KMCEA is better than two compared algorithms. Therefore, the superiority of KMCEA is demonstrated.

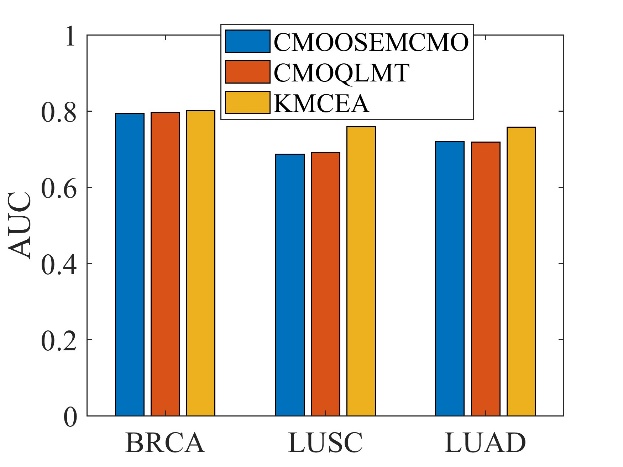
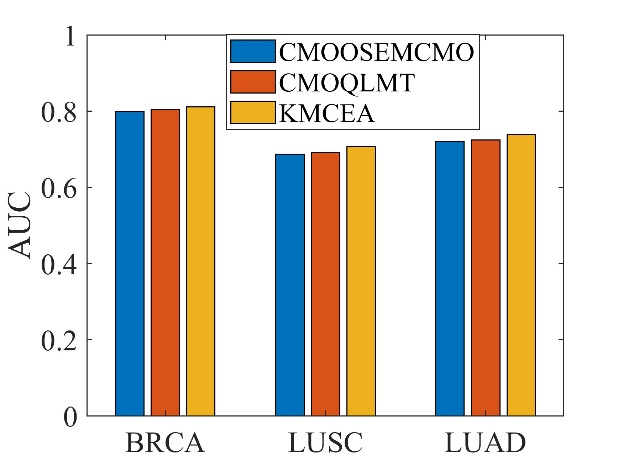


1. MMDS



1. MNCUA

Fig. S-1. Mean HV results of KMCEA and two compared algorithms on (a) MMDS and (b) MNCUA models.



|  |  |
| --- | --- |
| 1. MMDS | (b )MNCUA |

Fig. S-2. Mean AUC results of KMCEA and two compared algorithms on (a) MMDS and (b) MNCUA models.

## Network schematic

Fig. S-3 plots the global and local topologies of four stages of patient data (the patient numbers are TCGA-BH-A0C3, TCGA-BH-A1EN, TCGA-BH-A0B5, and TCGA-BH-A1FH, respectively), in which the red points and green points indicate prior nodes and non-prior nodes respectively, and gray lines indicate the edges that connect two nodes. If one node is pointed by more gray lines, it has more neighbor nodes. Some observations are as follows:

1) For four data, the number of prior nodes is significantly larger than that of non-prior nodes.

2) For global topology, apart from two large circles, other points are plotted within the green circle. These points indicate the nodes with more than 100 neighbor nodes. At stage I, there are seven prior nodes within the green circle. While, there are four prior nodes within the green circle at other stages, respectively. For non-prior nodes, there are two points within the green circle at each stage, respectively. Therefore, most of nodes have fewer than 100 neighbor nodes.

3) For local topology, node TP53 (TP53 is a prior node, and it is marked yellow in figures) is selected as an example to plot its neighbor relationship. In four local topology figures, the nodes with more than 20 neighbor nodes are plotted within the circle, and the numbers of nodes within the circle are different at four stages. In addition, it can be seen that only a small portion of prior nodes have some neighbor nodes.

|  |  |
| --- | --- |
|  |  |
| (a) Stage I: TCGA-BH-A0C3 (Left: Global topology; Right: Local topology) | |
|  |  |
| (b) Stage II: TCGA-BH-A1EN (Left: Global topology; Right: Local topology) | |
|  |  |
| (c) Stage III: TCGA-BH-A0B5 (Left: Global topology; Right: Local topology) | |
|  |  |
| (d) Stage IV: TCGA-BH-A1FH (Left: Global topology; Right: Local topology) | |
| Fig. S-3. Global and local topologies of four stages of patient data, in which red points and green points indicate prior nodes and non-prior nodes respectively, and gray lines indicate the edges that connect two nodes. | |

**References**

[S1] W. F. Guo, S. W. Zhang, T. Zeng, Y. Li, J. Gao, and L. Chen, "A novel network control model for identifying personalized driver genes in cancer," *PLOS Computational Biology,* vol. 15, 2019.

[S2]Guo, W.F., et al. "Network controllability-based algorithm to target personalized driver genes for discovering combinatorial drugs of individual patients," 2021;49(7).

[S3]Reimand, J.r. and Bader, G.D.J.M.S.B., 9,1. "Systematic analysis of somatic mutations in phosphorylation signaling predicts novel cancer drivers," 2013;9(1).

[S4]Abel, G.P. and Nuria, L.B.J.N.A.R. "Functional impact bias reveals cancer drivers," 2012(21):e169.

[S5]Han, Y., et al. "DriverML: a machine learning algorithm for identifying driver genes in cancer sequencing studies," 2019(8):e45-e45.

[S6]I-Fang, et al. "DriverDBv2: a database for human cancer driver gene research," 2015;44(D1):D975-D979.

[S7] F. Ming, W. Gong, L. Wang, and Y. Jin, Constrained multi-objective optimization with deep reinforcement learning assisted operator selection, IEEE/CAA Journal of Automatica Sinica, 2024, 11(4): 1-13.

[S8] F. Ming, W. Gong, and L. Gao, Adaptive auxiliary task selection for multitasking-assisted constrained multi-objective optimization [feature], IEEE Computational Intelligence Magazine, 2023, 18(2): 18-30.