Supplementary manuscript of

Multi-objective optimization based network control principles for identifying personalized drug targets of individual patients with cancer

Jing Liang[1], Zhuo Hu[1], Zong-Wei Li[1], Kang-Jia Qiao[1], Wei-Feng Guo[1,2]*

[1] School of Electrical Engineering, Zhengzhou University, Zhengzhou 450001, China
[2] State Key Laboratory of Oncology in South China, Collaborative Innovation Center for Cancer Medicine, Sun Yat-sen University Cancer Center, Guangzhou 510060, China
* Corresponding author(s).

Email: guowf@zzu.edu.cn

Section S-I: Constructing PGIN by Paired-SSN method

For the paired-SSN method [1], the first step is building the co-expression network based on the tumor sample and the normal sample of an individual patient [2]. 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 ΔPCC and then counting its Z-value of ΔPCC. The ΔPCC of an edge between gene i and gene j and its Z-score can be calculated :

$$\Delta PCC_{(ij,k)} = \left| PCC_{(ij,k)}^{n+1} - PCC_{ij}^{n} \right|$$

$$Z_{(ij,k)} = \frac{\Delta PCC_{(ij,k)}}{\left(1-(PCC_{ij}^{n})^2\right) \Big/ (N-1)}$$

where n represents the number of reference samples and k represents the k-th patient

in the perturbed network. PCC_{ij}^{n} represents the PCC of an edge between genes i and j in the reference network; and $PCC_{(ij,k)}^{n+1}$ 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,

$$e_{ij}^{k} = Index(comutation_{i,j}) * Index(comutation_{i,j})$$

$$Index(comutation_{i,j}) = \begin{cases} 1, if\ comutation_{i,j} > D_{10}(comutation) \\ 0, comutation_{i,j} < D_{10}(comutation) \end{cases}$$

$$Index(coexpression_{i,j}) = \begin{cases} 1, if\ coexpression_{i,j} > D_{10}(coexpression) \\ 0, coexpression_{i,j} < D_{10}(coexpression) \end{cases}$$

$$comutation_{i,j} = \frac{|S(i) \cap S(j)|}{|S(i) \cup S(j)|}$$

$$coexpression_{i,j,k} = \left| \log_2 \left| \frac{\Delta PCC_{(ij,k)}^{tumor}}{\Delta PCC_{(ij,k)}^{normal}} \right| \right|$$

where Norm represents the min-max normalized function. $S(i)$ and $S(j)$ 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; D_{10} indicates that 10% of the data falls under D_{10} after sorting a set of data in ascending order.

Section S-II: The the particular parameter setting of all CMOEAs

In this work, NSGA-II-CDP [3], CMME [4], CCMO [5], c-DPEA[6], MTCMO[7] and LSCV-MCEA adopt the simulated binary crossover [8]and the polynomial mutation[9] to generate offsprings, while CCMODE adopt the differential evolution [10]and the polynomial mutation to generate offsprings. The general parameters of the algorithms are set as follows:

1. Simulated binary crossover operators: the crossover probability p_c =1 and the distribution index η_c = 20;

2. Differential evolution operators: the crossover rate CR = 0.9 and the scaling factor F = 0.5;
3. Polynomial mutation operators: the mutation probability p_m = 1/n and the distribution index η_m = 20;

Section S-III: The biologic significance of patients on three cancer datasets.

To further verify the effectiveness of target genes and corresponding drugs, we queried the drug response datasets (GDSC) as shown in Table S1. For BRCA data, we found that three target genes under the framework of MDS, two target genes under the framework of NCUA and two target genes under the framework of DFVS. For instance, **Figure S1A** shows that the sensitivity of the drug afuresertib, which acts on the drug target AKT1, is significantly correlated with the PTEN mutation cell line in BRCA cancer tissues under the framework of MDS. Furthermore, **Figure S1B** shows that BRCA cancer cells with the PTEN mutation were significantly inhibited by afuresertib compared with the wildtype cell line, which was consistent with the findings of a previous study [11]. Therefore, afuresertib can be a candidate drug for BRCA patients with PTEN mutation. The sensitivity analysis of other drugs acting on AKT1, SMARCA2 and KIT is also shown in **Figure S1**. Similar results for NCUA and DFVS are shown in **Figure S2** and **Figure S3**.

Table S1 DRUG TARGETS AND EFFECTIVE DRUGS PROVIDED BY LSCV-MCEA FOR THREE CANCER DATASETS UNDER THE FRAMEWORK OF MDS, NCUA AND DFVS.

Methods	BRCA			LUSC			LUAD		
	Drug targets	Drug(GDSC)	Drug(iGMDC)	Drug targets	Drug(GDSC)	Drug(iGMDC)	Drug targets	Drug(GDSC)	Drug(iGMDC)
MDS	AKT1	Afuresertib	allosteric AKT inhibitor	STAT5A		WT-161	SVEP1		GSK-650394
		AZD5363				ABT-737			Thapsigargin
		Ipatasertib				RITA			XAV 939
	KIT	Tivozanib		PAK1		indisulam			
	SMARCA2	PFI-3		ITPR1		etoposide			
NCUA	AKT1	Afuresertib	allosteric AKT inhibitor	RASGRF2		RITA	SVEP1		GSK-650394
		AZD5363		TNN		kinetin riboside			Thapsigargin
		Ipatasertib				ABT-737			XAV 939
	BRAF	PLX-4720				RITA	ESR2		navitoclax
									GW-405833
DFVS	AR	OSU-03012	Bicalutamide	TNN		kinetin riboside	TTN		CHEMBL1222381
		Niraparib	Enzalutamide			ABT-737	TPR		piperlongumine
		Taselisib	Orterone			RITA			
		PFI-3	4OHtestosterone						
	PTPN6	NSC-87877							

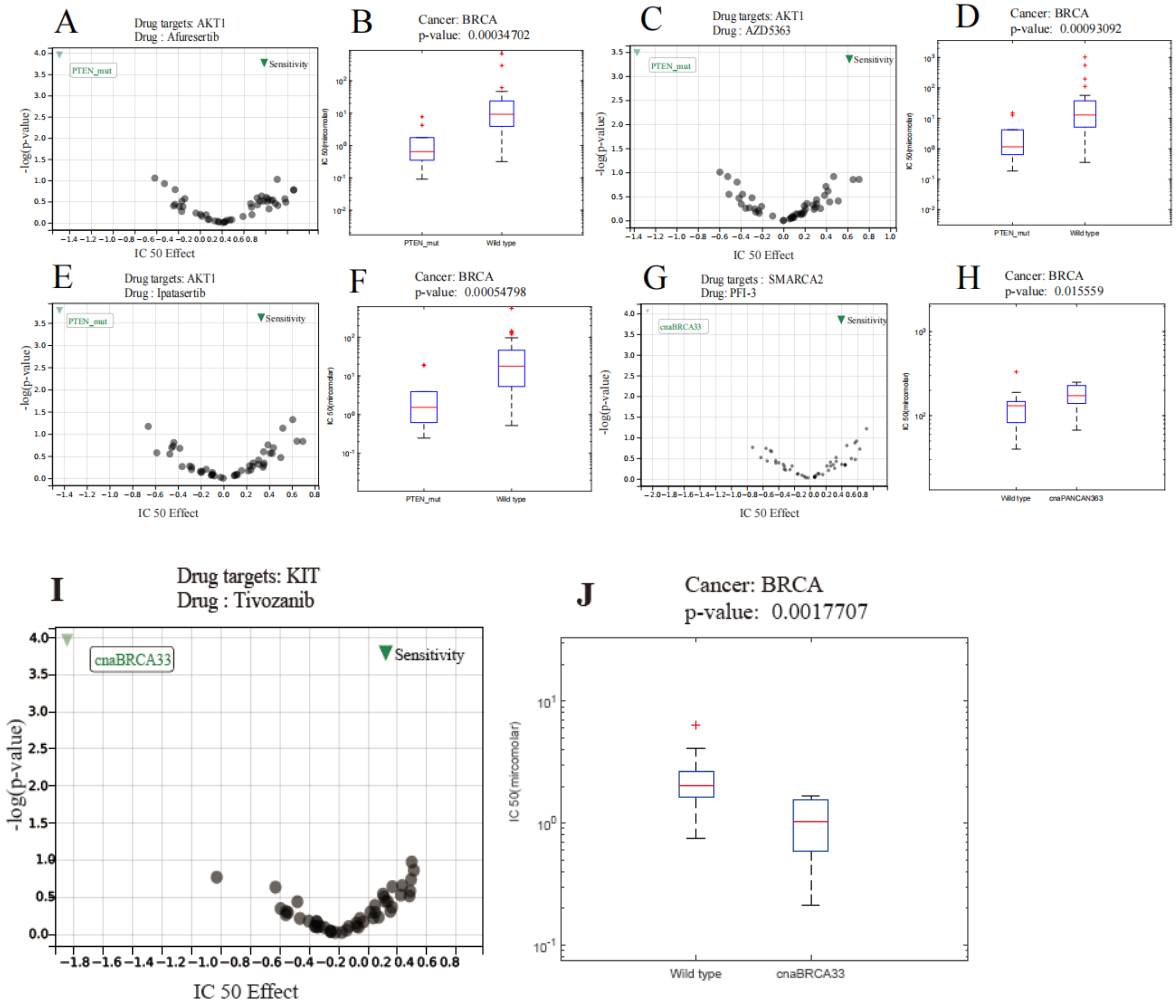

Figure S1 The sensitivity of drugs acting on drug targets of BRCA under the framework of MDS. (**A,C,E,G,I**) The volcano plot of drugs acting on drug targets . (**B,D,F,H,J**) The box-plots of IC50 on specific genomic changes cell line and wild type cell line.

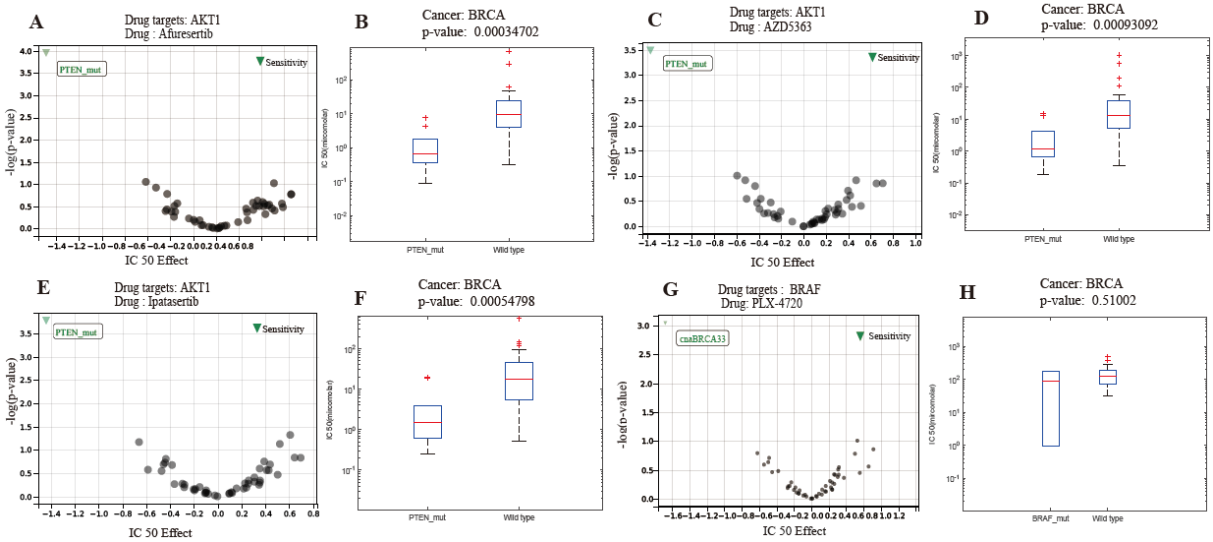

Figure S2 The sensitivity of drugs acting on drug targets of BRCA under the framework of NCUA. (**A,C,E,G**) The volcano plot of drugs acting on drug targets . (**B,D,F,H**) The box-plots of IC50 on specific genomic changes cell line and wild type cell line.

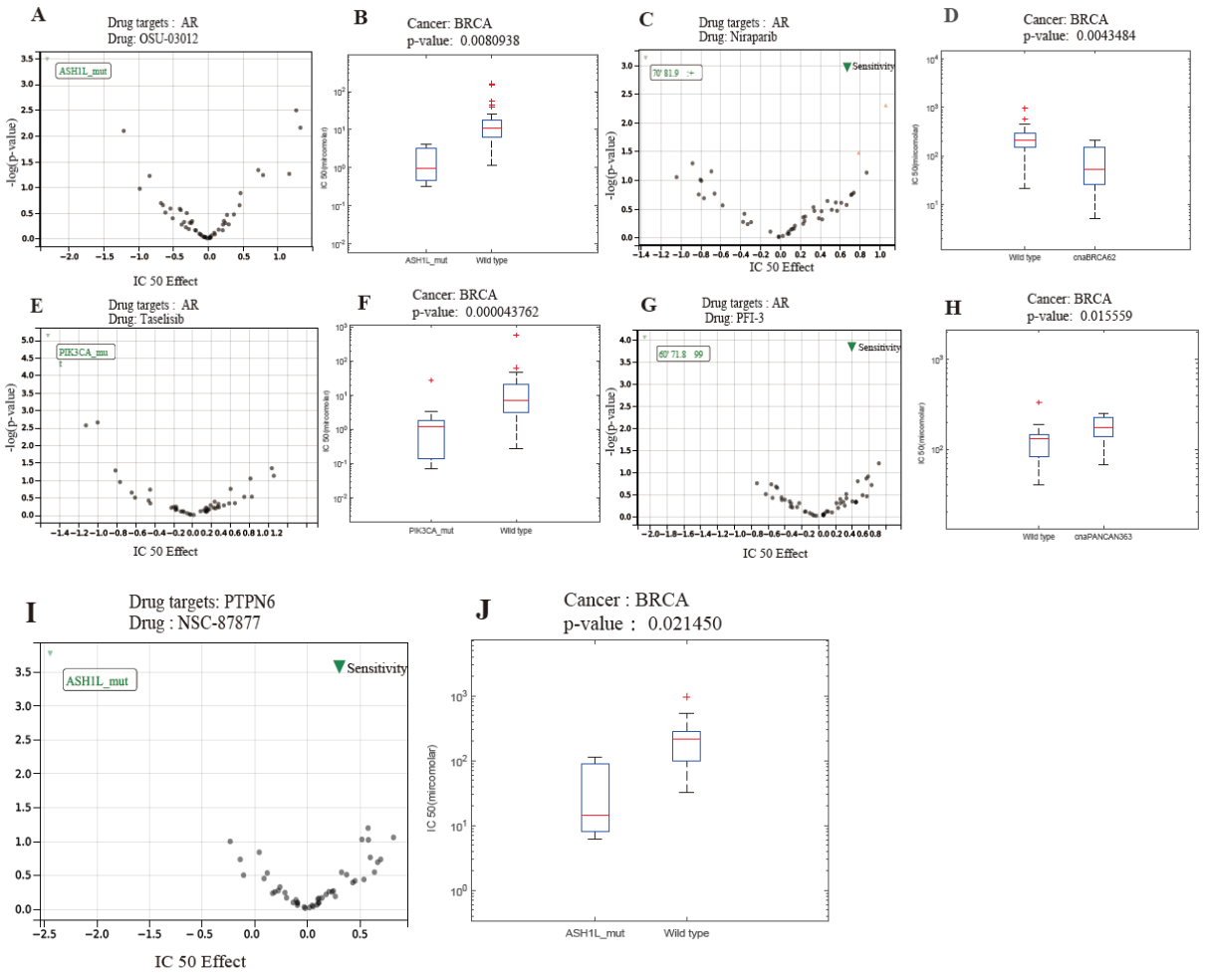

Figure S3 The sensitivity of drugs acting on drug targets of BRCA under the framework of DFVS. (**A,C,E,G,I**) The volcano plot of drugs acting on drug targets . (**B,D,F,H,J**) The box-plots of IC50 on specific genomic changes cell line and wild type cell line.

Section S-IV: 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 [12], PNC[13] , ActiveDriver[14] , OncoDriveFM[15], DriverML[16] and Hub-genes. The cancer driver genes of CPGD, DriverML, and PNC were obtained from their provided list of driver genes. Meanwhile, the driver genes in ActiveDriver and OncoDriveFM were obtained from the DriverDBv2 database[17]. 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: $w = \mu(T) + 2\sigma(T)$,where $\mu(T)$ and where $\sigma(T)$ 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 ($\log2(\text{fold-change})| > 1$). 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. All the above methods were executed on the same TCGA datasets (i.e., BRCA, LUSC, and LUAD) as our LSCV-MCEA according to their manuals.

Section S-V: The computational details for obtaining the probability of each combinatorial drug for AUC calculation

The drug combinations annotated in the CAC drugs were applied to obtain the AUC of the top-ranked/predicted anti-cancer drug combinations from different methods. This paper assigned a probability to each combinatorial drug for an individual patient according to its rank by using the following formula:

$$p(CD_i^j) = 1 - \frac{rank(CD_i^j)}{\left\|CD_i^j\right\|}$$

where CD^j_i denotes the predicted score of combinatorial drug j for patient i; CD^j_i denotes the number of all the combinatorial drugs for patient i; rank(CD^j_i) denotes the sequence number in descending order according to the predicted score. The AUC value of the predicted anti-cancer drug combinations was obtained based on the predicted probability and the true label in the CAC.

Section S-VI: The computational details of obtaining p-value for enriching in cancer gene census (CGC) dataset

We validated whether the PF of our LSCV-MCEA was significantly enriched in the Cancer Gene Census (CGC) dataset as well-established driver gene sets. To investigate whether the solutions (i.e., the set of driver genes) in the PF are enriched in the given CGC dataset, this paper first calculated the number of nodes in the CGC dataset for each solution in the PF on the three cancer datasets under the framework of MDS, NCUA, and DFVS. Then, random sets of driver genes were generated, each of which has the same number of driver genes for a solution in the PF, and the number of nodes in the CGC was re-calculated. Subsequently, the random data sets were used to obtain an empirical null distribution for the number of nodes in the CGC for each random set of driver genes. z-score is calculated as:

$$z_i = \frac{p_i - mean(SD_i)}{std(SD_i)}$$

where p_i is the number of nodes in the CGC for solution x_i . SD_i is the distribution of

the number of nodes in the CGC of random sets of driver genes. In this paper, the mean and std of SD_i were calculated from 1000 simulations of random sets of driver genes. Based on the z-score, the empirical p-value was obtained, i.e., the p_i (modeled as Gaussian distribution) of each solution in the PF. Finally, the fraction of solutions whose p-values are smaller than 0.05 was considered as the Enrichment Significance Score (ESS) for a given PF.

Section S-VII: Visualization and gene/pathway enrichment analysis for molecular medicine insights with p-values.

First, we selected genes that appeared in the identified Pareto optimal solution set (PS) with a frequency equal to 1 in every BRCA patient by expecting to obtain the overlapping genes in PS. These selected genes were considered as candidate personalized driver genes. Then we integrated these personalized driver genes of patients at the same stage to obtain the stage specific driver genes of different stages during cancer progress. Finally, we did pathway enrichment analysis of KEGG and WikiPathway for these stage specific driver genes of different stages. The results of the stage specific pathways with the top 20 significant p-values are presented in **Figure S4** and **Figure S5**. The following findings can be derived from **Figure S4** and **Figure S5**:

(1) The different stages of cancer patients are associated with certain pathways that are related to cancer. For example, Pathways in cancer was detected from KEGG dataset while Retinoblastoma gene in cancer, Breast cancer pathway, ErbB signaling pathway were detected from WikiPathway dataset.

(2) We identified certain cancer pathways that are related to patients with relatively early and intermediate stages of cancer (I, II, III). From KEGG dataset we detected Apoptosis , Cell cycle, PD-L1 expression and PD-1 checkpoint pathway in cancer. From WikiPathway dataset, we detected Gastrin signaling pathway, Integrated cancer pathway.

(3) We also predicted cancer pathways associated with patients in the advanced stage of cancer (IV). From KEGG dataset we detected Focal adhesion, Glioma and Gastric cancer. From WikiPathway dataset, we detected Focal adhesion,

TGF-beta signaling pathway.

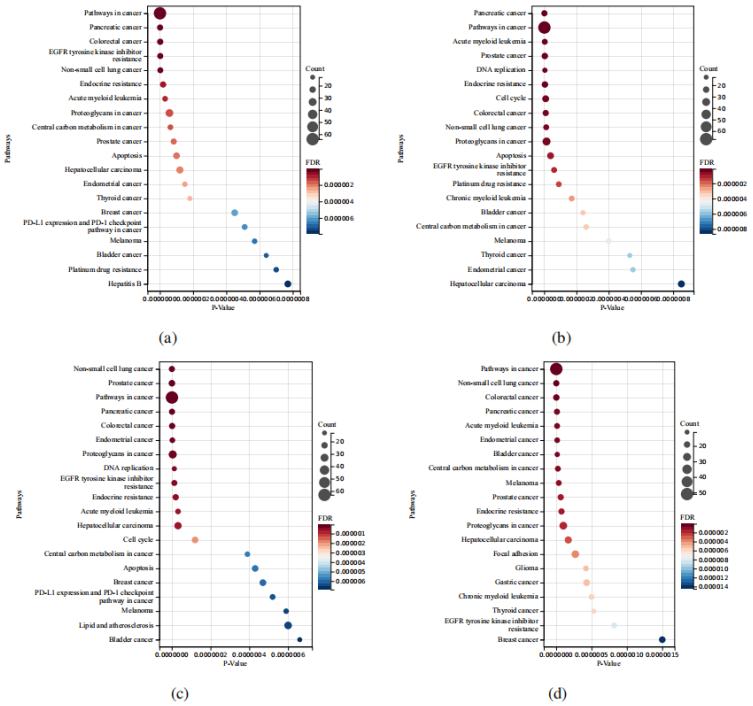

Figure S4 The KEGG pathway enrichment analysis for molecular medicine insights with p-values of BRCA . (A-D)pathway enrichment analysis for the patients of (A)stage I, (B)stage II, (C)stage III and (D)stage IV.

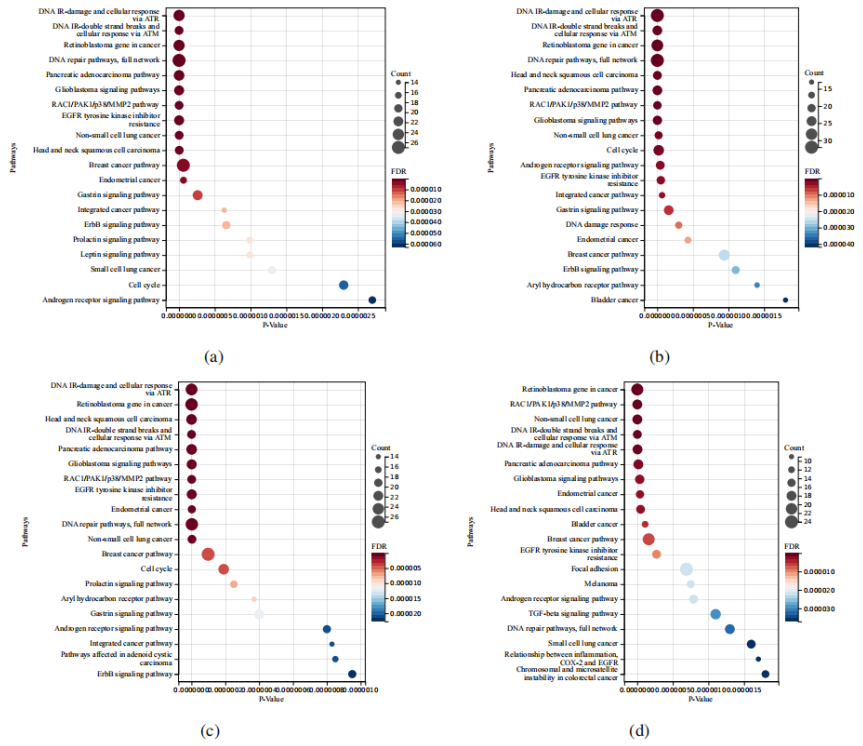

Figure S5 The WikiPathway enrichment analysis for molecular medicine insights with p-values of BRCA . (A-D)pathway enrichment analysis for the patients of (A)stage I, (B)stage II, (C)stage III and (D)stage IV.

Section S-VIII: The calculation times of of all methods on three cancer datasets under the framework of MDS

we have compared calculation time of different algorithms on three cancer data. All the experiments used a specialized computer with Intel(R) Xeno(R) Gold 6230 CPU and 640 GB RAM. Please note that all experiments were implemented on PlatEMO [*Y. Tian, R. Cheng, X. Zhang, Y. Jin, IEEE Computational Intelligence Magazine,2017,12(4):73–87.*]. The results of the comparison are presented in **Figure S6** , which leads to the following conclusions:

(1)As shown in **Figure S6**, under the framework of MDS, NCUA, and DFVS the calculation time of LSCV-MCEA is shorter than that of the other two CMOEAs (i.e., CCMO and MTCMO) and NSGA-II-CDP, CMME, and cDPEA have slight better performance in terms of calculation time compared with our LSCV-MCEA. In addition, LSCV-MCEA has better performance than other six CMOEAs in terms of algorithm performance (higher HV and lower IGD) and performance in identifying PDGs (higher AUC) under the framework of MDS, NCUA, and DFVS as shown in **Figure 5** and **Figure 3 (B-D)**.

(2) Although the calculation time of single objective based network controllability algorithms (i.e., MDS, NCUA, and DFVS) is generally shorter than that of LSCV-MCEA , these methods for identifying one minimum set of driver nodes cannot effectively identify drug targets (lower AUC as shown in **Figure 3 (A)**) compared with our LSCV-MCEA for discovering multiple sets of driver nodes.

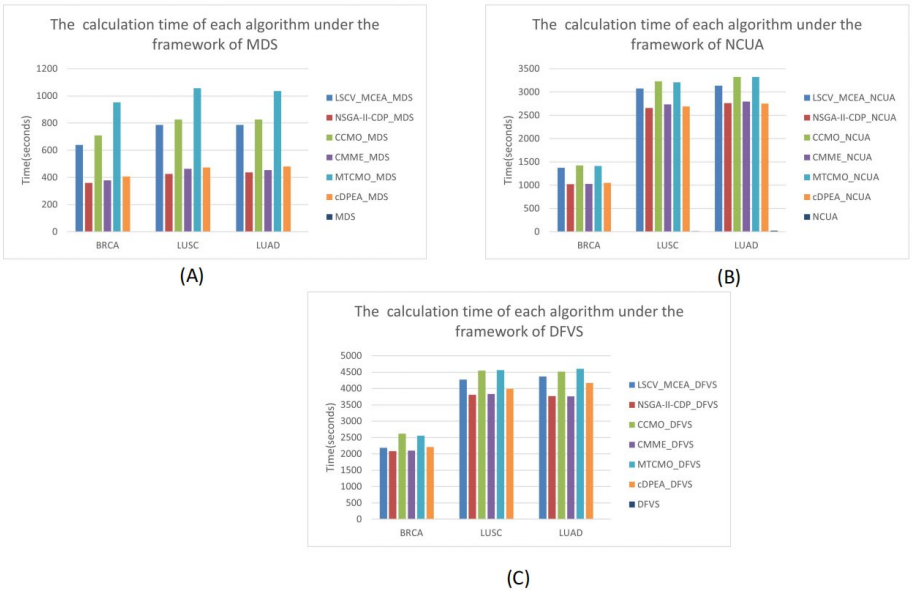

Figure S6 The average calculation time of different algorithms on three cancer data.

References

[1]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.

[2]X. Liu, Y. Wang, H. Ji, K. Aihara, and L. Chen, "Personalized characterization of diseases using sample-specific networks," *Nucleic Acids Research,* vol. 44, no. 22, p. gkw772, 2016.

[3]K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: Nsga-ii, " *IEEE Transactions on Evolutionary Computation*, 6(2):182 – 197, 2002.

[4]F. Ming, W. Y. Gong, L. Wang, and L. Gao, "A constrained many-objective optimization evolutionary algorithm with enhanced mating and environmental selections," *IEEE Transactions on Cybernetics*, 2022.

[5]Y. Tian, T. Zhang, J. H. Xiao, X. Y. Zhang, and Y. C. Jin, "A coevolutionary framework for constrained multiobjective optimization problems," *IEEE Transactions on Evolutionary Computation*, 25(1):102 – 116, 2020.

[6]M. J. Ming, A. Trivedi, R. Wang, D. Srinivasan, and T. Zhang, "A dual-population-based evolutionary algorithm for constrained multiobjective optimization," *IEEE Transactions on Evolutionary Computation*, 25(4):739 – 753, 2021.

[7]K. J. Qiao, K. J. Yu, B. Y. Qu, J. Liang, H. Song, C. T. Yue, H. Y. Lin, and K. C. Tan, "Dynamic auxiliary task-based evolutionary multitasking for constrained multiobjective optimization," *IEEE Transactions on Evolutionary Computation*, 2022.

[8]D. K, A. R B, "Simulated binary crossover for continuous search space[J]," *Complex systems*, 1995, 9(2): 115-148.

[9]D. K, G. M, "A combined genetic adaptive search (GeneAS) for engineering design[J]," *Computer Science and informatics*, 1996, 26: 30-45.

[10]Li. H, Z. Q, "Multiobjective optimization problems with complicated Pareto sets, MOEA/D and NSGA-II[J]," *IEEE Transactions on Evolutionary Computation*, 2008, 13(2): 284-302.

[11]L. Y, Z. J, N. P, et al, "The Critical Role of PTEN Mutation in Cellular Process and Drug Selection of Endometrial Cancer[J]," 2020.

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

[13]Guo, W.F., et al. "A novel network control model for identifying personalized driver genes in cancer," 2019;15.

[14]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).

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

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

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