



OPEN

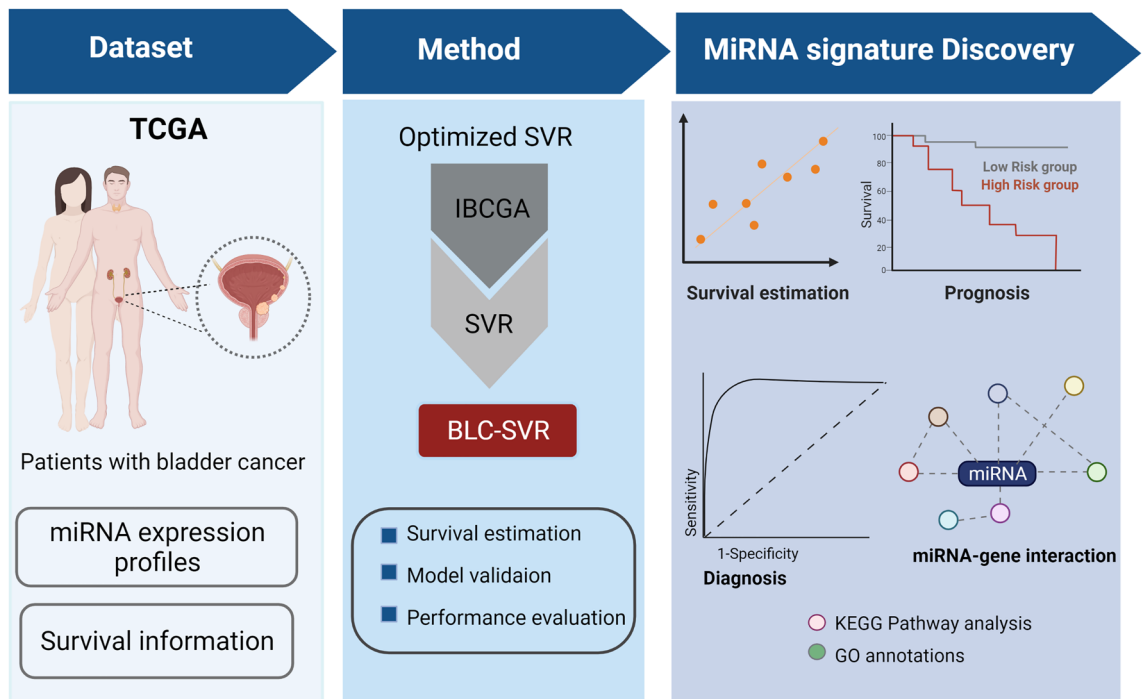
## MicroRNA signature for estimating the survival time in patients with bladder urothelial carcinoma

Srinivasulu Yerukala Sathipati<sup>1</sup>✉, Ming-Ju Tsai<sup>2,3</sup>, Sanjay K. Shukla<sup>1</sup>, Shinn-Ying Ho<sup>4,5</sup>, Yi Liu<sup>6</sup> & Afshin Beheshti<sup>7,8</sup>

Bladder urothelial carcinoma (BLC) is one of the most common cancers in men, and its heterogeneity challenges the treatment to cure this disease. Recently, microRNAs (miRNAs) gained promising attention as biomarkers due to their potential roles in cancer biology. Identifying survival-associated miRNAs may help identify targets for therapeutic interventions in BLC. This work aims to identify a miRNA signature that could estimate the survival in patients with BLC. We developed a survival estimation method called BLC-SVR based on support vector regression incorporated with an optimal feature selection algorithm to select a robust set of miRNAs as a signature to estimate the survival in patients with BLC. BLC-SVR identified a miRNA signature consisting of 29 miRNAs and obtained a mean squared correlation coefficient and mean absolute error of  $0.79 \pm 0.02$  and  $0.52 \pm 0.32$  year between actual and estimated survival times, respectively. The prediction performance of BLC-SVR had a better estimation capability than other standard regression methods. In the identified miRNA signature, 14 miRNAs, hsa-miR-432-5p, hsa-let-7e-3p, hsa-miR-652-3p, hsa-miR-629-5p, and hsa-miR-203a-3p, hsa-miR-129-5p, hsa-miR-769-3p, hsa-miR-570-3p, hsa-miR-320c, hsa-miR-642a-5p, hsa-miR-496, hsa-miR-5480-3p, hsa-miR-221-5p, and hsa-miR-7-1-3p, were found to be good biomarkers for BLC diagnosis; and the six miRNAs, hsa-miR-652-5p, hsa-miR-193b-5p, hsa-miR-129-5p, hsa-miR-143-5p, hsa-miR-496, and hsa-miR-7-1-3p, were found to be good biomarkers of prognosis. Further bioinformatics analysis of this miRNA signature demonstrated its importance in various biological pathways and gene ontology annotation. The identified miRNA signature would further help in understanding of BLC diagnosis and prognosis in the development of novel miRNA-target based therapeutics in BLC.

Bladder urothelial carcinoma (BLC) is one of the major causes of cancer mortalities with nearly 17,200 deaths and 83,730 estimated new cases in 2021 in United States alone; and 549,000 estimated new cases and 200,000 deaths globally<sup>1,2</sup>. According to estimates, 440,864 cases in men and 132,414 cases in women have been reported in 2020<sup>1</sup>. A male predominance is observed in all BLC cases which was ranked as the 6th most common cancer and 9th leading causes of cancer among men globally<sup>1</sup>. The risk factors of BLC include occupational exposure to carcinogenic substances and cigarette smoking which is considered as the major risk factor in both genders and accounts for 47% of all these cases<sup>3,4</sup>. BLC presents in two different forms, non-muscle-invasive tumors (NMIBC) and muscle-invasive tumors (MIBC). The NMIBC is benign with a higher incidence rate whereas MIBC is aggressive, could metastasize but lower incidence<sup>5</sup>. The standard treatment includes the combination of cytology and cystoscopy for the prognosis and diagnosis of BLC. There are some outstanding issues remaining in treatment conditions such as poor sensitivity of cytology in tumor detection, and invasiveness of cystoscopy<sup>6</sup>. Additionally, difference exists in response to similar treatment among patients because tumor heterogeneity makes it challenging to cure cancer and therapeutic modalities greatly affects the quality of life in elderly patients<sup>7</sup>. The five-year survival rate for patients with MIBC is 45%, and lymph node metastasis causes poor survival of 5% regardless

<sup>1</sup>Center for Precision Medicine Research, Marshfield Clinic Research Institute, Marshfield, WI 54449, USA. <sup>2</sup>Hinda and Arthur Marcus Institute for Aging Research at Hebrew Senior Life, Boston, MA, USA. <sup>3</sup>Department of Medicine, Beth Israel Deaconess Medical Center and Harvard Medical School, Boston, MA, USA. <sup>4</sup>Institute of Bioinformatics and Systems Biology, National Yang Ming Chiao Tung University, Hsinchu, Taiwan. <sup>5</sup>College of Health Sciences, Kaohsiung Medical University, Kaohsiung, Taiwan. <sup>6</sup>Biomedical Engineering, National Yang Ming Chiao Tung University, Hsinchu, Taiwan. <sup>7</sup>KBR, Space Biosciences Division, NASA Ames Research Center, Moffett Field, CA 94035, USA. <sup>8</sup>Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA. ✉email: sathipathi.srinivasulu@marshfieldclinic.org



**Figure 1.** System flowchart of BLC-SVR. miRNA expression profiles of BLC patients with survival information were used as input data of the BLC-SVR method, and the outputs were the miRNA signature with the predicted survival time.

of type of the treatment<sup>8,9</sup>. Although considerable advancements in adjuvant chemotherapy and surgery, BLC continues to be a common cancer. Therefore, identifying the survival related variants that could contribute to development of novel therapeutic strategies is necessary to improve the survival in patients with BLC.

MicroRNAs (miRNAs) are small, endogenous, non-coding RNAs involved in translation repression by post-transcriptional regulation of gene expression<sup>10</sup>. MiRNAs have been implicated in cancer progression and they can either promote or suppress tumor progression and metastasis<sup>11–13</sup>. A growing body of evidences has shown the association of miRNAs with cancer progression, diagnosis, and prognosis<sup>14,15</sup>, especially that of BLC<sup>16,17</sup>. For instance, significant changes in miRNA expression were observed in clinical samples, and three miRNAs, miR-129, miR-133b, and miR-518c, were discovered as potential prognostic predictors associated with BLC progression<sup>18</sup>. Li et al. identified urothelial carcinoma associated miRNAs that were significantly expressed in urine and plasma samples of patients with chronic kidney disease<sup>19</sup>. Ichimi et al. identified seven miRNAs namely, miR-145, miR-30a-3p, miR-133a, miR-133b, miR-195, miR-125b, and miR-199a, which were significantly downregulated in BLC when compared to the normal samples<sup>20</sup>. Lin and colleagues used hybridization-based miRNA array on BLC and found that miR-143 functioning as a tumor suppressor, and 14 down-regulated miRNAs that were significantly expressed between tumor and normal samples<sup>21</sup>. Studies have also noted the aberrant expression of miRNAs in tumor compared to normal bladder tissues<sup>22–25</sup>. However, there are limited studies on estimating survival using machine learning techniques to explore the roles of miRNAs in terms of survival association in BLC.

Previously, we identified miRNA signatures for predicting the cancer stages in breast and hepatocellular carcinoma<sup>26,27</sup>, and estimating the survival in glioblastoma, lung adenocarcinoma, and ovarian cancers<sup>28–31</sup>. In this study, we developed a survival time estimator called BLC-SVR to estimate the survival in BLC patients using miRNA expression profiles. BLC-SVR was developed based on support vector regression (SVR) incorporating with an optimal feature selection algorithm IBCGA<sup>32</sup> to identify the survival associated miRNA signature and estimate the survival in patients with BLC. BLC-SVR achieved a promising accuracy on estimating the survival time of patients with BLC. Furthermore, bioinformatics analysis on the identified miRNAs to explore their diagnostic and prognostic abilities in BLC. The overview of the BLC-SVR is shown in Fig. 1.

## Results

**Identification of miRNA signature for estimating survival time.** We retrieved 106 miRNA expression profiles of patients with BLC from The Cancer Genome Atlas (TCGA) database. Each miRNA profile consisted of 485 miRNAs which were the variables for survival estimation. BLC-SVR identified a set of miRNAs as a signature for estimating the survival time in patients with BLC. A robust miRNA signature was selected by performing 50 independent runs of BLC-SVR. The appearance score (ASC) for each miRNA signature of the prediction model was measured and scored according to their frequency among independent runs. The miRNA signature with a highest ASC accommodates the more frequent miRNAs among the independent runs of BLC-SVR. The average and highest ASCs obtained from 50 independent runs were  $13.52 \pm 1.60$ , and 17.27,

Method	R <sup>2</sup>	MAE (years)	Features selected
Ridge regression	0.42	0.81	485
Lasso	0.50	0.74	26
Elastic net	0.52	0.73	33
BLC-SVR-ASC	0.81	0.51	29
BLC-SVR-Best	0.83	0.51	32
BLC-SVR-Mean	0.79 ± 0.02	0.63 ± 0.32	32.64 ± 4.22

**Table 1.** The comparison of prediction performance. R<sup>2</sup>, squared correlation coefficient; MAE, mean absolute error.

respectively. The robust signature with the highest ASC consisted of 29 miRNAs and obtained a squared correlation coefficient (R<sup>2</sup>) and mean absolute error (MAE) of 0.81 and 0.51 year, between actual and estimated survival times, respectively. The diagnostic and prognostic prediction ability of the identified miRNA signature is discussed in the following sections. The ASCs for all the independent runs of BLC-SVR are depicted in Supplementary Fig. S1.

**Prediction performance comparison.** We compared the prediction performance of BLC-SVR with standard machine learning methods, including ridge regression, least absolute shrinkage selection operator (Lasso), and elastic net. Ridge regression achieved a R<sup>2</sup> and MAE of 0.42 and 0.81 year, between actual and estimated survival times, respectively; Lasso obtained a R<sup>2</sup> and MAE of 0.50 and 0.74 year, between actual and estimated survival times, respectively; and elastic net obtained a R<sup>2</sup> and MAE of 0.52 and 0.73 year, between actual and estimated survival times, respectively. BLC-SVR obtained a mean performance of 50 runs (BLC-SVR-Mean) with R<sup>2</sup> and MAE of 0.79 ± 0.02 and 0.52 ± 0.32 year, between actual and estimated survival times, respectively. Whereas the best performance (BLC-SVR-Best) has the largest R<sup>2</sup> and MAE of 0.83 and 0.516 year using 32 miRNAs, respectively. The prediction comparison results showed the better estimation capability of BLC-SVR than the popular regression methods. The comparison of prediction performance of BLC-SVR with some regression methods is shown in Table 1. The correlation plots of BLC-SVR, ridge regression, Lasso, and elastic net are shown in Supplementary Fig. S2.

Next, we validated the estimation ability of BLC-SVR using an independent test cohort from the TCGA database. The independent test cohort consisting of 123 patients with BLC along with their follow-up times up to one year with an average 5.54 months. BLC-SVR estimated mean survival time of these patients was 15.50 months. There were 93 patients whose predicted survival time was longer than the actual follow-up time. The estimation performance of BLC-SVR achieved 75.60% accuracy on estimating the patients' survival time. However, for the remaining 30 patients, estimated survival time was slightly shorter than the follow-up time. The prediction performance of BLC-SVR on 123 patients is shown in Supplementary Fig. S3.

Next, we prioritized the miRNAs of the signature using main effect difference (MED) analysis based on their contribution to the estimation of survival time as described in the study<sup>33</sup>. The top 10 ranked miRNAs of the signature, including hsa-miR-432-5p, hsa-let-7e-3p, hsa-miR-146b-5p, hsa-miR-505-3p, hsa-miR-652-3p, hsa-miR-629-5p, hsa-miR-193b-5p, hsa-miR-203a-3p, hsa-miR-542-5p, and hsa-miR-128-3p, were analyzed further. The miRNAs signature and their corresponding MED scores and ranks are listed in Table 2.

**The roles of top 10 ranked miRNAs in cancer.** The literature validation on top ranked miRNAs revealed that these miRNAs possess different functions and active involvement in BLC progression (Table 3). For instance, the up-regulated hsa-miR-432-5p targets RNA-binding motif protein 5 and regulate apoptosis in bladder cancer cells<sup>34</sup>. Hsa-let-7 family is known to be differentially expressed in various cancers including BLC<sup>35,36</sup>. Hsa-miR-146b expression was upregulated in bladder cancer tissues when compared to the normal tissues<sup>37</sup>. A real-time quantitative polymerase chain reaction (RT-qPCR) study on BLC cell lines reported that hsa-miR-652-3p expression levels were upregulated and knockdown of this miRNA significantly affected cell proliferation, migration, and invasion in BLC<sup>38</sup>. A PCR based miRNA screening study revealed that hsa-miR-193 down-regulated the expression of oncogenes, Cyclin D1 and *EST1*, and inhibited cell migration in human urothelial cells<sup>39</sup>. The hsa-miR-203a has been identified as a tumor suppressor and its overexpression inhibits cell proliferation, invasion and migration in BLC<sup>40</sup>. Hsa-miR-542 expression was downregulated and negatively correlated with the expression of *surviving* protein resulting in the inhibition of proliferation in BLC cells<sup>41</sup>.

The roles of three of the top 10 ranked miRNAs (hsa-miR-505-3p, hsa-miR-629-5p, and hsa-miR-128-3p) have not been previously reported in BLC. However, these miRNAs are implicated in other major cancers. For instance, hsa-miR-505-3p acts as a tumor suppressor in pancreatic cancer and hepatocellular carcinoma<sup>42,43</sup>. Hsa-miR-629-5p promotes tumor progression by targeting *AKAP13* in prostate cancer<sup>44</sup>, and hsa-miR-128-3p acts as tumor suppressor in breast cancer by regulating the *LIMK1/CFL1* signaling pathway<sup>45</sup>. Hence, their involvement in other major cancers suggests that their expression is biologically consistent and important in BLC. A summary of miRNAs and their regulation in BLC is shown in Table 3.

**Diagnostic ability of the miRNAs.** To determine the diagnostic ability of the identified miRNA signature, receiver operating curve (ROC) analysis was performed using BLC tumor and normal samples. The ROC

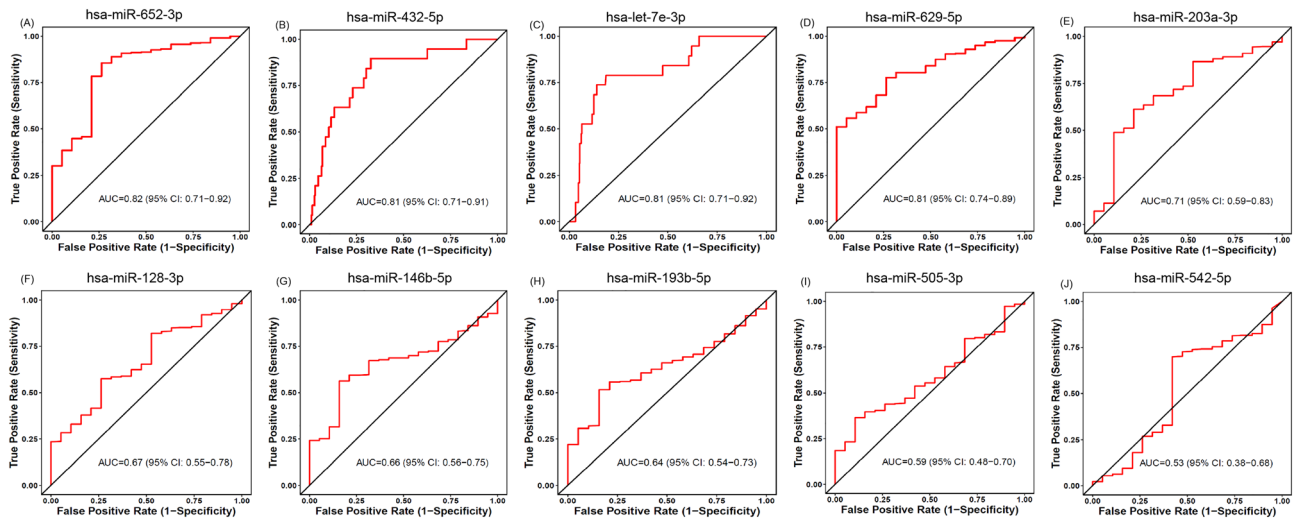
RANK	miRNA	MIMAT_ID	MED	AUC
1	hsa-miR-432-5p	MIMAT0002814	1.53	0.81
2	hsa-let-7e-3p	MIMAT0004485	1.50	0.81
3	hsa-miR-146b-5p	MIMAT0002809	1.47	0.66
4	hsa-miR-505-3p	MIMAT0002876	1.34	0.59
5	hsa-miR-652-3p	MIMAT0003322	1.12	0.82
6	hsa-miR-629-5p	MIMAT0004810	0.96	0.81
7	hsa-miR-193b-5p	MIMAT0004767	0.95	0.64
8	hsa-miR-203a-3p	MIMAT0000264	0.80	0.71
9	hsa-miR-542-5p	MIMAT0003340	0.79	0.53
10	hsa-miR-128-3p	MIMAT0000424	0.79	0.67
11	hsa-miR-129-5p	MIMAT0000242	0.66	0.85
12	hsa-miR-769-3p	MIMAT0003887	0.64	0.79
13	hsa-miR-224-3p	MIMAT0009198	0.57	0.59
14	hsa-miR-570-3p	MIMAT0003235	0.56	0.77
15	hsa-miR-1254	MIMAT0005905	0.54	–
16	hsa-miR-143-5p	MIMAT0004599	0.53	0.67
17	hsa-miR-320c	MIMAT0005793	0.43	0.7
18	hsa-miR-642a-5p	MIMAT0003312	0.38	0.74
19	hsa-miR-496	MIMAT0002818	0.35	0.7
20	hsa-miR-421	MIMAT0003339	0.35	0.5
21	hsa-miR-2116-3p	MIMAT0011161	0.31	0.59
22	hsa-miR-361-5p	MIMAT0000703	0.29	0.66
23	hsa-miR-548o-3p	MIMAT0005919	0.26	0.73
24	hsa-miR-26a-1-3p	MIMAT0004499	0.16	0.64
25	hsa-miR-339-3p	MIMAT0004702	0.08	0.49
26	hsa-miR-23a-5p	MIMAT0004496	0.07	0.63
27	hsa-miR-508-3p	MIMAT0002880	0.06	0.58
28	hsa-miR-221-5p	MIMAT0004568	0.02	0.72
29	hsa-miR-7-1-3p	MIMAT0004553	0.012	0.71

**Table 2.** ROC analysis of top ranked miRNAs between tumor and normal samples. MED, main effect difference; AUC, area under the receiver operating curve.

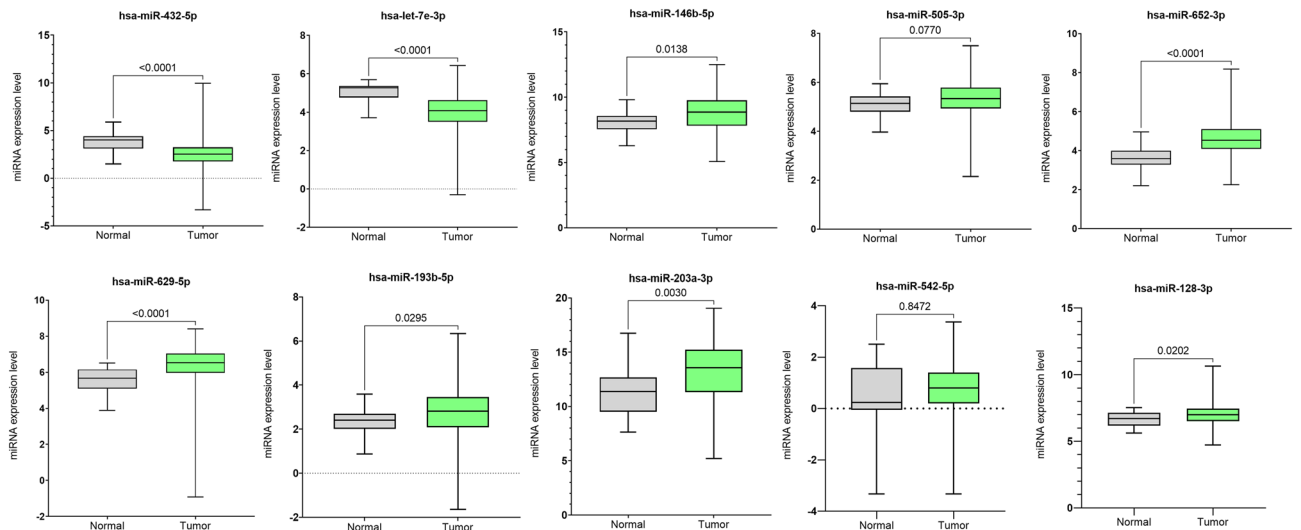
miRNA	Regulation	Functional impact on tumors	Target gene for bladder cancer	References
hsa-miR-432-5p	Up	Tumor progression	RBM5	34
hsa-let-7e-3p	Up	Tumor progression	LIN28, HMGA2, MYC, BCL2L	35,72
hsa-miR-146b-5p	Up	Tumor progression	ETS2	37
hsa-miR-505-3p	N/A	N/A	N/A	N/A
hsa-miR-652-3p	Up	Tumor progression	KCNN3	38
hsa-miR-629-5p	N/A	N/A	N/A	N/A
hsa-miR-193b-5p	down	Tumor suppression	CCND1 and ETS2	39
hsa-miR-203a-3p	Down	Tumor suppression	SIX4	40
hsa-miR-542-5p	Down	Tumor suppression	BIRC5	41
hsa-miR-128-3p	N/A	N/A	N/A	N/A

**Table 3.** The summary of the roles of miRNAs in bladder cancer. N/A, not available.

analysis showed that 14 miRNAs of the signature have the good diagnostic ability ( $AUC \geq 0.7$ ) while distinguishing the tumor and normal samples, shown in Table 2. The top 10 ranked miRNAs obtained an average area under the ROC curve (AUC) of  $0.70 \pm 0.10$  and five of these miRNAs showed good diagnostic ability. The five miRNAs, hsa-miR-432-5p, hsa-let-7e-3p, hsa-miR-652-3p, hsa-miR-629-5p, and hsa-miR-203a-3p obtained AUCs of 0.81, 0.81, 0.82, 0.81, and 0.71, respectively. The ROC curves for the top 10 ranked miRNAs are shown in Fig. 2. Further, we combined the top 10 ranked miRNAs to predict the diagnosis of BLC using Random forest classifier<sup>46</sup>. We used a dataset consisting of 418 tumor samples and 18 normal samples retrieved from the TCGA. The prediction model was selected after 100 iterations of 10-CV. Random forest obtained a 10-CV accuracy, sensitivity, specificity, and AUC of 96.02%, 0.96, 0.85, and 0.95, respectively, while distinguishing tumor and



**Figure 2.** ROC curve analysis of the top 10 ranked miRNAs. (A) hsa-miR-652-3p, (B) hsa-miR-432-5p, (C) hsa-let-7e-3p, (D) hsa-miR-629-5p, (E) hsa-miR-203a-3p, (F) hsa-miR-128-3p, (G) hsa-miR-146b-5p, (H) hsa-miR-193b-5p, (I) hsa-miR-505-3p, and (J) hsa-miR-542-5p.



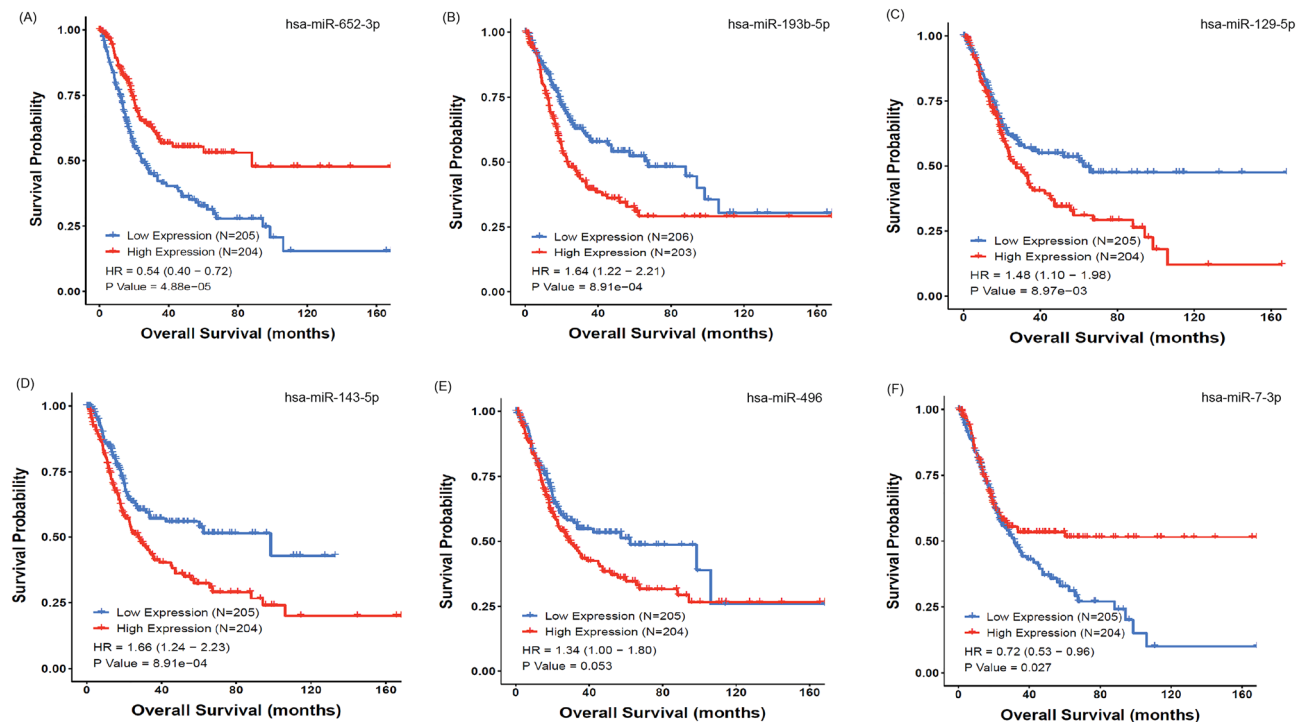
**Figure 3.** Expression differences of the top 10 ranked miRNAs between tumor and normal samples using box plots.

normal samples, shown in Supplementary Fig. S4. The combination of top 10 ranked miRNAs showed better diagnostic ability.

Additionally, expression differences of the top 10 ranked miRNAs between normal and tumor samples were analyzed using box-plot analysis. The analysis showed that eight miRNAs, hsa-miR-432-5p, hsa-let-7e-3p, hsa-miR-146b-5p, hsa-miR-652-3p, hsa-miR-629-5p, hsa-miR-193b-5p, hsa-miR-203a-3p, and hsa-miR-128-3p were significantly expressed ( $p < 0.05$ ) between normal and tumor samples. The box plot analysis of the top 10 ranked miRNAs is shown in Fig. 3.

**Prognostic ability of the miRNAs.** The prognostic performance of miRNA signature was analyzed by Kaplan-Meier (KM) survival curves using CancerMIRNome<sup>47</sup>. Six miRNAs of the signature showed significant prognosis capability in overall survival analysis. These six miRNAs, hsa-miR-652-5p, hsa-miR-193b-5p, hsa-miR-129-5p, hsa-miR-143-5p, hsa-miR-496, and hsa-miR-7-1-3p, obtained p-values of  $4.88e-05$ ,  $8.91e-04$ ,  $8.97e-03$ ,  $8.91e-04$ , 0.05, and 0.027, respectively, between high and low expression groups. The KM survival curves for the six miRNAs are shown in Fig. 4.

**Biological significance of the miRNA signature.** To determine the biological relevance of the miRNA signature that could aid in understanding the functional information and involvement in disease-associated pathways, Kyoto Encyclopedia of Genes and Genomes (KEGG) and Gene Ontology (GO) analysis were



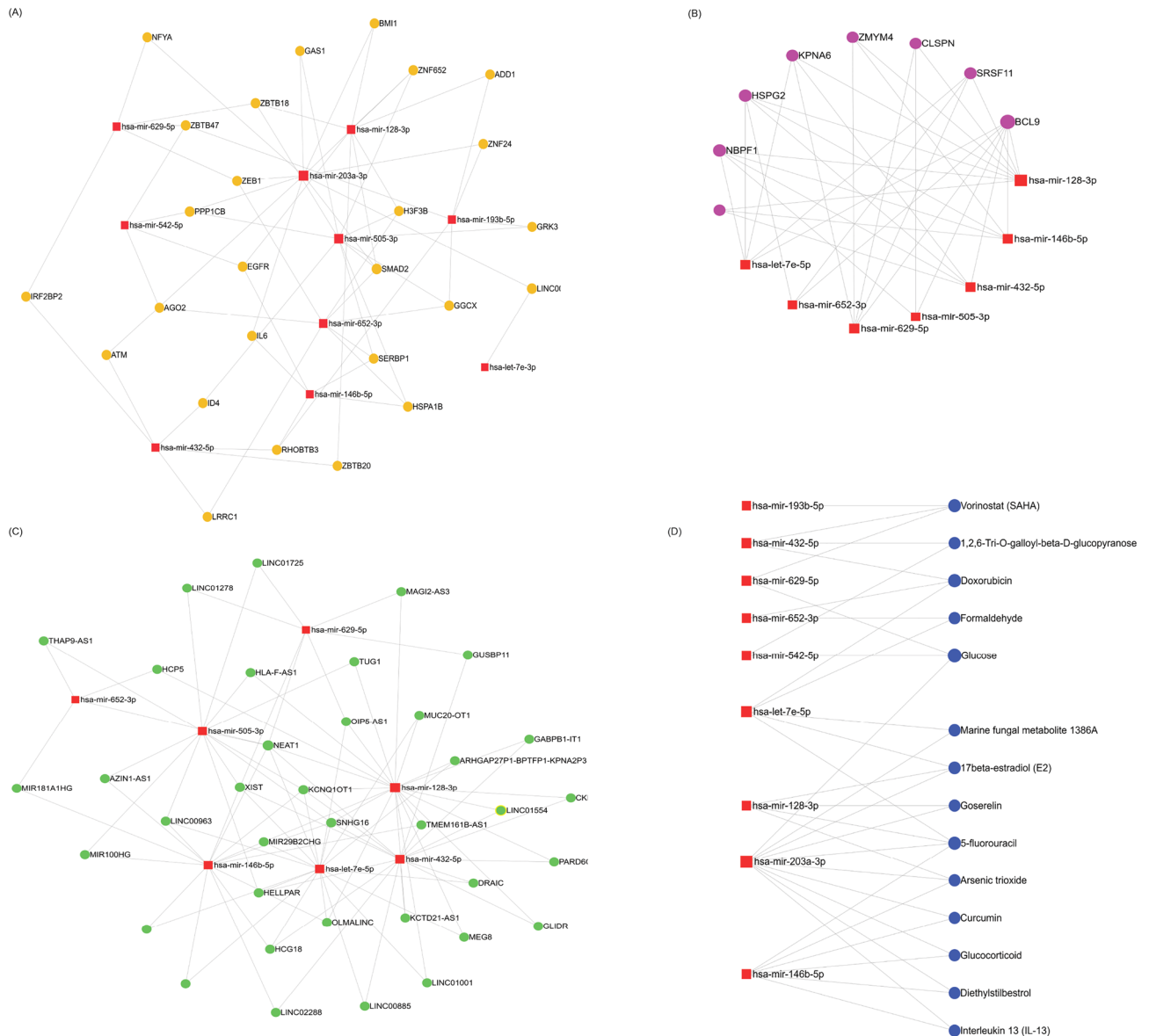
**Figure 4.** Kaplan-Meier survival analysis. The survival probability of patients with high expression of miRNAs (A–D) is poorer than that of patients with low expression. The high expression of hsa-miR-652-5p and hsa-miR-7-3p (E, F) has positive effect on survival probability.

employed. The miRNA signature is significantly involved in various biological pathways such as prion disease, fatty acid biosynthesis, fatty acid metabolism, ECM-receptor interaction, hippo signaling pathway, adherence junction, steroid biosynthesis, lysine degradation, TGF-beta signaling pathway, and proteoglycans in cancer. The number of targeted genes involved in KEGG pathways of the miRNA signature is shown in supplementary Table S1. The miRNA signature enriched in KEGG pathways is shown in Supplementary Fig. S5.

Next, the biological significance of the miRNA signature was analyzed in different stages of BLC using KEGG pathway analysis. The differentially expressed miRNAs of the signature were identified between stage II, III, and IV. There were nine miRNAs, including hsa-miR-496, hsa-miR-146b-5p, hsa-miR-652-3p, hsa-miR-26a-1-3p, hsa-miR-193b-5p, hsa-miR-642a-5p, hsa-miR-432-5p, hsa-miR-143-5p, and hsa-miR-505-3p which were differentially expressed between stage II&III, and two miRNAs, hsa-miR-143-5p and hsa-let-7e-3p, between stage III&IV. The significant biological pathways in stage II&III were thyroid hormone synthesis, oxytocin signaling pathway, ErbB signaling pathway, long-term depression, and hippo signaling pathway, to name a few. The significant pathways in stage III&IV were biosynthesis of unsaturated fatty acids, ErbB signaling pathway, GABAergic synapse, morphine addiction, and estrogen signaling pathway. There were some common pathways, including ErbB signaling pathway, thyroid hormone signaling, morphine addiction, non-small cell lung cancer, and estrogen signaling pathway in stage II&III and stage III&IV. However, some targeted pathways were different across cancer stages of BLC. The complete list of significant pathways across cancer stages are listed in Supplementary Table S2. The bubble plots showing the KEGG pathways in BLC stages are shown in Supplementary Figs. S6&S7.

Next, GO annotations of the miRNA signature was employed in three categories, including biological process, molecular functions, and cellular components. The GO analysis showed that the miRNA signature involved several biological processes that the top five significant biological processes were DNA metabolic process, cellular protein metabolic process, membrane organization, RNA metabolic process, and nucleobase-containing compound catabolic process (Supplementary Table S3). The top five molecular functions were ion binding, nucleic acid binding transcription factor activity, protein binding transcription factor activity, enzyme binding, and enzyme regulatory activity. The top five cellular components were organelle, cytosol, nucleoplasm, protein complex, and focal adhesion. The details of GO annotations for the miRNA signature are listed in Supplementary Tables S3–S5. The GO enrichment analysis of the miRNA signature is depicted in Supplementary Figs. S8–S10.

**Gene interaction network.** The complex networks in which miRNAs engaged with other functional molecules can influence cell biological responses and human diseases<sup>48</sup>. Hence, a miRNA network analysis was employed for the top 10 ranked miRNAs with genes, long non-coding RNAs (lncRNAs), circular RNAs (ciRNAs), and small molecules to explore the miRNA-target interactions using miRNet 2.0: a miRNA-centric network visual analytics platform<sup>49</sup>. The miRNA-gene target interaction network was built with experimentally validated gene target networks using the miRTarBase V8.0<sup>50</sup>. There were 1594 gene interactions with top 10 ranked miRNAs in a miRNA-gene target network. The miRNA-gene target network is shown in Fig. 5A.



**Figure 5.** MiRNA-interaction network. (A) miRNA-gene interaction network, (B) miRNA-ciRNA interaction network, (C) miRNA-lncRNA interaction network, and (D) miRNA-small molecule interaction network.

In the miRNA-lncRNA interaction network, seven of the top 10 ranked miRNAs targeted 197 lncRNAs formed with 255 edges. There were 4508 circular RNAs (ciRNAs) formed with 7751 edges in the miRNA-ciRNA interaction network. In the miRNA-small molecule interaction network, there were 43 compounds interacting with top 10 ranked miRNAs and forming 65 edges. The miRNA interaction networks for genes, lncRNA, ciRNA, and small molecules are shown in Fig. 5A–D.

## Discussion

The critical role of miRNAs in cancer biology has opened up a new direction for oncology research. Numerous evidences have demonstrated the development of miRNA-based cancer progression, diagnosis, and therapeutics<sup>51–53</sup>. Bladder cancer is one of the common cancers and a heterogeneous disease with prognostic and therapeutic challenges. Identifying the survival related variants could help understand the cancer survival at various stages and may contribute to the therapeutic improvements in BLC. Due to cost and time consumption in experimental methods to predict the targets and identify biomarkers, computational methods are often used in miRNA biology and cancer prognosis predictions. Advances in machine learning methods have significant importance in developing fast and accurate models to aid in cancer prognosis, diagnosis, and medical decision-making<sup>54</sup>. Recent developments on miRNA-disease associations revealed the importance of computational models in understanding the disease associated variants<sup>55</sup>. However, there are limited studies on identifying miRNA signatures to estimate the survival time in patients with BLC using machine learning techniques.

The machine learning methods often suffer from higher dimensionality issues<sup>56</sup>, especially in biomedical data. The used feature selection algorithms could work well in coping with the curse of dimensionality issue resulting

from genomic data and select a robust signature for cancer prognosis<sup>31,57</sup>. To address the dimensionality issue, we used an optimal feature selection algorithm IBCGA to identify a small set of miRNAs from a large number of candidate miRNAs that are associated with survival time in patients with BLC. In our previous studies, optimized survival estimation methods were developed to estimate the survival time in patients with glioblastoma, lung adenocarcinoma, and ovarian cancers<sup>28–30</sup>. In this study, we developed an optimized SVR-based method BLC-SVR to identify a miRNA signature associated with survival and estimate the survival time in patients with BLC. The identified miRNA signature consisted of 29 miRNAs as a signature and obtained a  $R^2$  and MAE of 0.81 and 0.51 year, between actual and predicted survival times, respectively. The estimation capability of BLC-SVR was compared with some standard regression methods and results showed its promising estimation performance. Further, the identified miRNAs of the signature were ranked based on their contribution to the estimation performance. The literature survey on top 10 ranked miRNAs demonstrated that seven of top 10 ranked miRNAs are actively involved in BLC progression except the three miRNAs hsa-miR-505-3p, hsa-miR-629-5p, and hsa-miR-128-3p. These three miRNAs are important contributors to the estimated performance. Therefore, hsa-miR-505-3p, hsa-miR-629-5p, and hsa-miR-128-3p may be novel targets for BLC and further studies are needed to validate their roles in BLC.

In addition, the diagnostic ability prediction results showed that five of the top 10 ranked miRNAs, hsa-miR-432-5p, hsa-let-7e-3p, hsa-miR-652-3p, hsa-miR-629-5p, and hsa-miR-203a-3p, obtained an AUC greater than 0.70 while distinguishing the tumor and normal samples, proving their discrimination ability. The differential expression analysis on top 10 ranked miRNAs showed that eight of the top 10 ranked miRNAs were significantly expressed between tumor and normal samples. Next, KM survival analysis of the miRNA signature revealed that six miRNAs, hsa-miR-652-5p, hsa-miR-193b-5p, hsa-miR-129-5p, hsa-miR-143-5p, hsa-miR-496, and hsa-miR-7-1-3p, were good prognostic predictors of overall survival in patients with BLC.

The functional analysis of miRNAs revealed the involvement of miRNAs in physiological process that are essential for disease mechanism. Biological relevance of the identified miRNA signature concluded that the miRNA signature was involved in several biological pathways, including biological processes, molecular functions, and cellular components. The top-3 KEGG pathways were prion diseases, fatty acid biosynthesis, and fatty acid metabolism. In human prion diseases, point mutations in the prion protein gene (*PRNP*), which encodes *PrP*, induce familial forms of human prion diseases<sup>58</sup>. Somatic missense mutation in the prion protein gene (*PRNP*) have been identified in patients with BLC<sup>59</sup>. Urinary retention as an early symptom was observed in patients with prion disease<sup>60</sup>. The prion proteins were detected in urine and involved in disease infection<sup>61</sup>. Fatty acid synthesis and fatty acid metabolism pathways are associated with various cancers including BLC<sup>62</sup>. A previous study showed that the change in the fatty acid composition may be an indicator of altered lipid metabolism occurring in vivo during human bladder tumorigenesis. The bladder cancer tissue showed a significant reduction in total n-6 polyunsaturated fatty acid (−15.1%;  $P < 0.001$ )<sup>63</sup>.

The pathway analysis demonstrated that the group of miRNAs target specific pathways and target genes that might contribute to the cancer progression. To investigate the numbers of genes, lncRNAs, ciRNAs, and small molecules targeted by the top ranked miRNAs, miRNA-gene target interaction networks were constructed. A miRNA network showed some key molecules that were connected to top ranked miRNAs which might act as underlying drivers of survival in BLC. In conclusion, the identified miRNA signature would guide the understanding of the survival associated miRNAs and help develop miRNA target-based therapeutic strategies in BLC.

## Material and methods

**Dataset.** *The clinical characteristics.* The miRNA expression profiles of 409 patients along with their survival times were retrieved from the TCGA database. All the data extraction methods were carried out in accordance with the TCGA guidelines and regulations. The patient selection criteria included patients with survival times and miRNA expression profiles. The miRNA expression was considered if the expression levels of mature miRNAs were presented in more than 70% of the samples. After the filtration process, there were 106 patients with miRNA expression profiles where each miRNA profile consisted of 485 miRNAs in the final dataset.

The clinical characteristics of the 106 patients with BLC are presented in Supplementary Fig. S11. The majority of BLC patients were male, and 69% were male and 31% female. The average age at diagnosis was  $70.46 \pm 9.43$  and average height of the patients was  $173.4 \pm 11.14$  cms. The total numbers of patients in stages 2, 3, and 4 were 14, 37, and 55, respectively. The range of survival times of patients were between 0.63 and 94.26 months.

**BLC-SVR method.** BLC-SVR was designed to identify a set of miRNAs as a signature that could estimate the survival time in patients with BLC. BLC-SVR method was developed based on SVR incorporated with the optimal feature selection algorithm IBCGA. Two main parts of BLC-SVR are feature selection and survival estimation. BLC-SVR adopted the optimization technique from our previous study<sup>29</sup>.

**Feature selection algorithm IBCGA.** BLC-SVR utilized the optimal feature selection algorithm IBCGA to select a minimum number of features from a large number of candidate features (miRNAs) while maximizing the prediction performance<sup>32</sup>. The IBCGA uses an intelligent evolutionary algorithm to solve the large parameter combinatorial optimization problems<sup>64</sup>. Here, we used genetic algorithm (GA) terms GA-chromosomes and GA-genes for the feature representation. The chromosome of IBCGA comprises 485 GA-genes and three 4-bit GA-genes to encode parameters  $C$ ,  $\gamma$ , and  $\nu$  for  $\nu$ -SVR. The encoded GA-chromosomes were designed as described in previous studies<sup>29–31</sup>. The best prediction model of BLC-SVR was generated from the 50 independent runs of IBCGA. The main steps in IBCGA are described as follows: where the detailed description can be refer to the work<sup>32</sup>:



- Step 1: (Initialization) Randomly generate an initial population of individuals.
- Step 2: (Evaluation) Evaluate the fitness value of all individuals using the fitness function, which is to maximize the prediction accuracy ( $R^2$ ) in terms of 10-CV.
- Step 3: (Selection) Use a conventional method of tournament selection that selects a winner from two randomly selected individuals to generate a mating pool.
- Step 4: (Crossover) Select two parents from the mating pool to perform an orthogonal array crossover operation.
- Step 5: (Mutation) Apply a conventional mutation operator to the randomly selected individuals in the new population.
- Step 6: (Termination test) If the stopping condition for obtaining the solution is satisfied, output the best individual as the solution. Otherwise, go to Step 2.
- Step 7: (Inheritance) If  $r$  is less than a predefined number of features, randomly changes one bit in the binary GA-genes for each individual from 0 to 1, increase the number  $r$  by one and go to Step 2. Otherwise, stop the algorithm.

The applications of support vector machines (SVM) have diverse importance in biomedical sciences and precision medicine due to their capability in solving complications in predictions<sup>65</sup>. The SVM has two modules, support vector classifier (SVC) and support vector regression (SVR)<sup>66</sup>. The SVMs are used in various cancer diagnosis and prognosis predictions. The optimized SVMs were used to predict the cancer stage in breast cancer and hepatocellular carcinoma<sup>26,27</sup>, and estimation of survival in patients with lung adenocarcinoma, glioblastoma, neuroblastoma, and ovarian cancers<sup>28–31</sup>. The LibSVM package<sup>67</sup> was used to implement the BLC-SVR. The optimization technique of SVR can be written as follows:

$$\min \left\{ \frac{1}{2} w^T (\phi(x_i) + b) + C(v\varepsilon + \frac{1}{m} \sum_{i=1}^m (\xi_i + \xi_i^*)) \right\} \quad (1)$$

where  $0 \leq v \leq 1$ ,  $\xi_i \geq 0$ ,  $\xi_i^* \geq 0$ ,  $(x_1, y_1) \dots (x_m, y_m)$  are the input data points,  $C$  is the regularization parameter,  $\varepsilon$  is an insensitive loss function, and  $b$  is a constant.

**Performance measures.** We used squared correlation coefficient ( $R^2$ ) and mean absolute error (MAE) as the estimation measures to evaluate the prediction performance of BLC-SVR.

$$R = \frac{\sum_{i=1}^N (y_i - \bar{y})(z_i - \bar{z})}{\left[ \sum_{i=1}^N (y_i - \bar{y})^2 \right] \left[ \sum_{i=1}^N (z_i - \bar{z})^2 \right]} \quad (2)$$

where  $y_i$  and  $z_i$  are the actual and predicted survival times of the  $i$ th miRNA, respectively,  $\bar{y}$  and  $\bar{z}$  are the corresponding means, and  $N$  is the total number of BLC patients in the validation set. The mean absolute error (MAE) is also used for the evaluation of prediction performance, defined as follows:

$$MAE = \frac{1}{N} \sum_{i=1}^N |z_i - y_i| \quad (3)$$

**Appearance score ASC.** The robust signatures among the 50 independent runs of BLC-SVR has the highest ASC obtained using the following procedure<sup>29</sup>.

- Step 1: Perform  $N_s$  independent runs of BLC-SVR for obtaining  $N_s$  miRNA signatures. There are  $m_i$  features in the  $i$ th signatures,  $i = 1, \dots, N_s$  (in this study  $N_s = 50$ ).
- Step 2: The ASC of a miRNA signature is calculated as follows:

- (1) Calculate the appearance frequency  $f(miR)$  for each feature  $miR$  that appears in the  $N_s$  signatures.
- (2) Calculate the score  $F_i$ ,  $i = 1, \dots, N_s$ . Where  $miR_{it}$  is the  $t$ th feature in the  $i$ th signature:

$$F_i = \sum_{t=1}^{m_i} f(miR_{it}) / m_i \quad (4)$$

- (3) Obtain the  $i$ -th feature set with the highest appearance score  $F_i$  as the robust signature.

**Ridge regression, Lasso and elastic net.** The estimation performance of BLC-SVR was compared with some standard regression methods, ridge regression, Lasso, and elastic net. Ridge regression is a penalized regression approach where the Euclidean norm was used as the penalty<sup>68</sup>. Lasso uses L1 regularization to identify features and regression coefficients by regularizing the coefficients to zero that lead to minimize the prediction error<sup>69</sup>. Elastic net is a combination of Lasso and ridge regression<sup>70</sup>. The minimum  $\lambda$  was chosen after 100 iterations of 10-CV for ridge, Lasso and elastic net. The prediction performance was evaluated in terms of the correlation coefficient and mean absolute error.

**KEGG pathway and GO annotation analysis.** The DIANA-micro-T-CDS algorithm provided the predicted miRNA targets for the pathway analysis<sup>71</sup>. The p-value threshold was set to 0.05, and Fishers's exact test (hypergeometric distribution) was used for the enrichment analysis.

**miRNA-interaction network.** The miRNA-target interaction networks were built using miRNet 2.0: a miRNA-centric network visual analytics platform<sup>49</sup>. For better visualization of target genes, we reduced the less important edges based on the shortest path measures, where the number of edges within the network can be reduced significantly by keeping the shortest path between hub-nodes. We used short distance and minimum layout filters for lncRNA and ciRNA networks, respectively.

## Data availability

All the data used in this analysis can be found on the TCGA data portal [<https://portal.gdc.cancer.gov/>].

Received: 10 September 2021; Accepted: 2 March 2022

Published online: 09 March 2022

## References

- Sung, H. *et al.* Global Cancer Statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA A Cancer J. Clin.* **71**, 209–249. <https://doi.org/10.3322/caac.21660> (2021).
- Siegel, R. L., Miller, K. D. & Jemal, A. Cancer statistics, 2020. *CA Cancer J. Clin.* **70**, 7–30 (2020).
- Thun, M., Linet, M. S., Cerhan, J. R., Haiman, C. A. & Schottenfeld, D. *Cancer Epidemiology and Prevention*. (Oxford University Press, 2017).
- Freedman, N. D., Silverman, D. T., Hollenbeck, A. R., Schatzkin, A. & Abnet, C. C. Association between smoking and risk of bladder cancer among men and women. *JAMA* **306**, 737–745 (2011).
- Cookson, M. S. *et al.* The treated natural history of high risk superficial bladder cancer: 15-year outcome. *J. Urol.* **158**, 62–67 (1997).
- Kaufman, D. S. Challenges in the treatment of bladder cancer. *Ann. Oncol.* **17**(Suppl 5), v106–112. <https://doi.org/10.1093/annonc/mdj963> (2006).
- Shariat, S. F., Milowsky, M. & Droller, M. J. Bladder cancer in the elderly. *Urol. Oncol. Semin. Original Investig.* **27**, 653–667. <https://doi.org/10.1016/j.urolonc.2009.07.020> (2009).
- von der Maase, H. *et al.* Long-term survival results of a randomized trial comparing gemcitabine plus cisplatin, with methotrexate, vinblastine, doxorubicin, plus cisplatin in patients with bladder cancer. *J. Clin. Oncol.* **23**, 4602–4608 (2005).
- Stein, J. P. *et al.* Radical cystectomy in the treatment of invasive bladder cancer: long-term results in 1054 patients. *J. Clin. Oncol.* **19**, 666–675 (2001).
- Ambros, V. The functions of animal microRNAs. *Nature* **431**, 350–355. <https://doi.org/10.1038/nature02871> (2004).
- Ventura, A. & Jacks, T. MicroRNAs and cancer: short RNAs go a long way. *Cell* **136**, 586–591 (2009).
- Baranwal, S. & Alahari, S. K. miRNA control of tumor cell invasion and metastasis. *Int. J. Cancer* **126**, 1283–1290 (2010).
- Pencheva, N. & Tavazoie, S. F. Control of metastatic progression by microRNA regulatory networks. *Nat. Cell Biol.* **15**, 546–554 (2013).
- Bartel, D. P. MicroRNAs: Genomics, biogenesis, mechanism, and function. *Cell* **116**, 281–297 (2004).
- Michael, I. P., Saghafinia, S. & Hanahan, D. A set of microRNAs coordinately controls tumorigenesis, invasion, and metastasis. *Proc. Natl. Acad. Sci.* **116**, 24184. <https://doi.org/10.1073/pnas.1913307116> (2019).
- Mahdavinzhad, A. *et al.* Evaluation of miR-141, miR-200c, miR-30b expression and clinicopathological features of bladder cancer. *Int. J. Mol. Cell. Med.* **4**, 32 (2015).
- Adam, L. *et al.* miR-200 expression regulates epithelial-to-mesenchymal transition in bladder cancer cells and reverses resistance to epidermal growth factor receptor therapy. *Clin. Cancer Res.* **15**, 5060–5072 (2009).
- Dyrskjot, L. *et al.* Genomic profiling of MicroRNAs in bladder cancer: miR-129 is associated with poor outcome and promotes cell death in vitro. *Can. Res.* **69**, 4851. <https://doi.org/10.1158/0008-5472.CAN-08-4043> (2009).
- Li, A.-L. *et al.* The microRNA prediction models as ancillary diagnosis biomarkers for urothelial carcinoma in patients with chronic kidney disease. *Front. Med.* **2021**, 1758 (2021).
- Ichimi, T. *et al.* Identification of novel microRNA targets based on microRNA signatures in bladder cancer. *Int. J. Cancer* **125**, 345–352. <https://doi.org/10.1002/ijc.24390> (2009).
- Lin, T. *et al.* MicroRNA-143 as a tumor suppressor for bladder cancer. *J. Urol.* **181**, 1372–1380 (2009).
- Chen, Y.-H. *et al.* 2 edn 219–227 (Elsevier).
- Han, Y. *et al.* MicroRNA expression signatures of bladder cancer revealed by deep sequencing. *PLoS ONE* **6**, e18286 (2011).
- Yoshino, H. *et al.* The tumour-suppressive function of miR-1 and miR-133a targeting TAGLN2 in bladder cancer. *Br. J. Cancer* **104**, 808–818 (2011).
- Wang, G. *et al.* Up-regulation of microRNA in bladder tumor tissue is not common. *Int. Urol. Nephrol.* **42**, 95–102 (2010).
- Yerukala Sathipati, S. & Ho, S. Y. Identifying a miRNA signature for predicting the stage of breast cancer. *Sci. Rep.* **8**, 16138. <https://doi.org/10.1038/s41598-018-34604-3> (2018).
- Yerukala Sathipati, S. & Ho, S. Y. Novel miRNA signature for predicting the stage of hepatocellular carcinoma. *Sci. Rep.* **10**, 14452. <https://doi.org/10.1038/s41598-020-71324-z> (2020).
- Yerukala Sathipati, S., Huang, H. L. & Ho, S. Y. Estimating survival time of patients with glioblastoma multiforme and characterization of the identified microRNA signatures. *BMC Genomics* **17**, 1022. <https://doi.org/10.1186/s12864-016-3321-y> (2016).
- Yerukala Sathipati, S. & Ho, S. Y. Identifying the miRNA signature associated with survival time in patients with lung adenocarcinoma using miRNA expression profiles. *Sci. Rep.* **7**, 7507. <https://doi.org/10.1038/s41598-017-07739-y> (2017).
- Sathipati, S. Y. & Ho, S. Y. Identification of the miRNA signature associated with survival in patients with ovarian cancer. *Aging (Albany NY)* **13**, 12660–12690. <https://doi.org/10.18632/aging.202940> (2021).
- Yerukala Sathipati, S., Sahu, D., Huang, H. C., Lin, Y. & Ho, S. Y. Identification and characterization of the lncRNA signature associated with overall survival in patients with neuroblastoma. *Sci. Rep.* **9**, 5125. <https://doi.org/10.1038/s41598-019-41553-y> (2019).
- Ho, S. Y., Chen, J. H. & Huang, M. H. Inheritable genetic algorithm for biobjective 0/1 combinatorial optimization problems and its applications. *IEEE Trans. Syst. Man Cybern. B Cybern.* **34**, 609–620. <https://doi.org/10.1109/tsmcb.2003.817090> (2004).
- Tung, C.-W. & Ho, S.-Y. Computational identification of ubiquitylation sites from protein sequences. *BMC Bioinformatics* **9**, 1–15 (2008).
- Zhang, Y.-P. *et al.* Down-regulated RBM5 inhibits bladder cancer cell apoptosis by initiating an miR-432-5p/ $\beta$ -catenin feedback loop. *FASEB J.* **33**, 10973–10985. <https://doi.org/10.1096/fj.201900537R> (2019).

35. Spagnuolo, M. *et al.* Urinary expression of let-7c cluster as non-invasive tool to assess the risk of disease progression in patients with high grade non-muscle invasive bladder Cancer: A pilot study. *J. Exp. Clin. Cancer Res. CR* **39**, 68–68. <https://doi.org/10.1186/s13046-020-01550-w> (2020).
36. Yin, X.-H. *et al.* Development of a 21-miRNA signature associated with the prognosis of patients with bladder cancer. *Front. Oncol.* **9**, 729 (2019).
37. Zhu, J. *et al.* MicroRNA-146b overexpression promotes human bladder cancer invasion via enhancing ETS2-mediated mmp2 mRNA transcription. *Mol. Ther. Nucleic Acids* **16**, 531–542 (2019).
38. Zhu, Q. L., Zhan, D. M., Chong, Y. K., Ding, L. & Yang, Y. G. MiR-652-3p promotes bladder cancer migration and invasion by targeting KCNN3. *Eur. Rev. Med. Pharmacol. Sci.* **23**, 8806–8812. [https://doi.org/10.26355/eurrev\\_201910\\_19275](https://doi.org/10.26355/eurrev_201910_19275) (2019).
39. Lin, S.-R. *et al.* MiR-193b mediates CEBPD-induced cisplatin sensitization through targeting ETS1 and cyclin D1 in human urothelial carcinoma cells. *J. Cell. Biochem.* **118**, 1563–1573. <https://doi.org/10.1002/jcb.25818> (2017).
40. Na, X. Y., Shang, X. S., Zhao, Y., Ren, P. P. & Hu, X. Q. MiR-203a functions as a tumor suppressor in bladder cancer by targeting SIX4. *Neoplasma* **66**, 211–221. [https://doi.org/10.4149/neo\\_2018\\_180512N312](https://doi.org/10.4149/neo_2018_180512N312) (2019).
41. Zhang, J. *et al.* MicroRNA-542-3p suppresses cellular proliferation of bladder cancer cells through post-transcriptionally regulating survivin. *Gene* **579**, 146–152. <https://doi.org/10.1016/j.gene.2015.12.048> (2016).
42. Wei, G., Lu, T., Shen, J. & Wang, J. LncRNA ZEB1-AS1 promotes pancreatic cancer progression by regulating miR-505-3p/TRIB2 axis. *Biochem. Biophys. Res. Commun.* **528**, 644–649. <https://doi.org/10.1016/j.bbrc.2020.05.105> (2020).
43. Ren, L., Yao, Y., Wang, Y. & Wang, S. MiR-505 suppressed the growth of hepatocellular carcinoma cells via targeting IGF-1R. *Biosci. Rep.* <https://doi.org/10.1042/BSR20182442> (2019).
44. Liu, Y. *et al.* MiR-629-5p promotes prostate cancer development and metastasis by targeting AKAP13. *Front. Oncol.* <https://doi.org/10.3389/fonc.2021.754353> (2021).
45. Zhao, J., Li, D. & Fang, L. MiR-128-3p suppresses breast cancer cellular progression via targeting LIMK1. *Biomed. Pharmacother.* **115**, 108947. <https://doi.org/10.1016/j.biopha.2019.108947> (2019).
46. Hall, M. *et al.* The WEKA data mining software: An update. *ACM SIGKDD Explor. Newsl.* **11**, 10–18 (2009).
47. Li, R. *et al.* CancerMIRNome: an interactive analysis and visualization database for miRNome profiles of human cancer. *bioRxiv.* <https://doi.org/10.1101/2020.10.04.325670> (2021).
48. Anastasiadou, E., Jacob, L. S. & Slack, F. J. Non-coding RNA networks in cancer. *Nat. Rev. Cancer* **18**, 5–18 (2018).
49. Chang, L., Zhou, G., Soufan, O. & Xia, J. miRNet 20: Network-based visual analytics for miRNA functional analysis and systems biology. *Nucleic Acids Res.* **48**, W244–W251. <https://doi.org/10.1093/nar/gkaa467> (2020).
50. Huang, H. Y. *et al.* miRTarBase 2020: Updates to the experimentally validated microRNA-target interaction database. *Nucleic Acids Res.* **48**, D148–d154. <https://doi.org/10.1093/nar/gkz896> (2020).
51. Jung, E.-J. *et al.* Plasma microRNA 210 levels correlate with sensitivity to trastuzumab and tumor presence in breast cancer patients. *Cancer* **118**, 2603–2614. <https://doi.org/10.1002/ncr.26565> (2012).
52. Yoshino, H. *et al.* Aberrant expression of microRNAs in bladder cancer. *Nat. Rev. Urol.* **10**, 396–404 (2013).
53. Khan, M. T. *et al.* A miRNA signature predicts benefit from addition of hypoxia-modifying therapy to radiation treatment in invasive bladder cancer. *Br. J. Cancer* **125**, 1–9 (2021).
54. Kourou, K., Exarchos, T. P., Exarchos, K. P., Karamouzis, M. V. & Fotiadis, D. I. Machine learning applications in cancer prognosis and prediction. *Comput. Struct. Biotechnol. J.* **13**, 8–17. <https://doi.org/10.1016/j.csbj.2014.11.005> (2015).
55. Chen, X., Sun, L. G. & Zhao, Y. NCMCMDA: miRNA-disease association prediction through neighborhood constraint matrix completion. *Brief Bioinform.* **22**, 485–496. <https://doi.org/10.1093/bib/bbz159> (2021).
56. Bolón-Canedo, V., Sánchez-Maróño, N. & Alonso-Betanzos, A. Recent advances and emerging challenges of feature selection in the context of big data. *Knowl. Based Syst.* **86**, 33–45 (2015).
57. Abeel, T., Helleputte, T., Van de Peer, Y., Dupont, P. & Saeys, Y. Robust biomarker identification for cancer diagnosis with ensemble feature selection methods. *Bioinformatics* **26**, 392–398. <https://doi.org/10.1093/bioinformatics/btp630> (2010).
58. Jeong, B. H. & Kim, Y. S. Genetic studies in human prion diseases. *J. Korean Med. Sci.* **29**, 623–632. <https://doi.org/10.3346/jkms.2014.29.5.623> (2014).
59. Kim, Y.-C., Won, S.-Y. & Jeong, B.-H. Identification of prion disease-related somatic mutations in the prion protein gene (PRNP) in cancer patients. *Cells* **9**, 1480. <https://doi.org/10.3390/cells9061480> (2020).
60. Mead, S. *et al.* A novel prion disease associated with diarrhea and autonomic neuropathy. *N. Engl. J. Med.* **369**, 1904–1914. <https://doi.org/10.1056/NEJMoa1214747> (2013).
61. Gonzalez-Romero, D., Barria, M. A., Leon, P., Morales, R. & Soto, C. Detection of infectious prions in urine. *FEBS Lett.* **582**, 3161–3166. <https://doi.org/10.1016/j.febslet.2008.08.003> (2008).
62. Koundouros, N. & Pouligiannis, G. Reprogramming of fatty acid metabolism in cancer. *Br. J. Cancer* **122**, 4–22. <https://doi.org/10.1038/s41416-019-0650-z> (2020).
63. Lee, H. Y. *et al.* Sulfatase-1 overexpression indicates poor prognosis in urothelial carcinoma of the urinary bladder and upper tract. *Oncotarget* **8**, 47216–47229. <https://doi.org/10.18632/oncotarget.17590> (2017).
64. Ho, S.-Y., Shu, L.-S. & Chen, J.-H. Intelligent evolutionary algorithms for large parameter optimization problems. *IEEE Trans. Evol. Comput.* **8**, 522–541 (2004).
65. Huang, S. *et al.* Applications of support vector machine (SVM) learning in cancer genomics. *Cancer Genomics Proteomics* **15**, 41–51. <https://doi.org/10.21873/cgp.20063> (2018).
66. Noble, W. S. What is a support vector machine?. *Nat. Biotechnol.* **24**, 1565–1567 (2006).
67. Chang, C.-C. & Lin, C.-J. LIBSVM: a library for support vector machines. *ACM Trans. Intell. Syst. Technol. (TIIST)* **2**, 1–27 (2011).
68. Hoerl, A. E. & Kennard, R. W. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics* **12**, 55–67 (1970).
69. Tibshirani, R. Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. Ser. B (Methodol.)* **58**, 267–288 (1996).
70. Zou, H. & Hastie, T. Regularization and variable selection via the elastic net. *J. R. Stat. Soc. Ser. B (Stat. Methodol.)* **67**, 301–320 (2005).
71. Vlachos, I. S. *et al.* DIANA-miRPath v3.0: Deciphering microRNA function with experimental support. *Nucleic Acids Res.* **43**, W460–W466 (2015).
72. Vinall, R. L. *et al.* Decreased expression of let-7c is associated with non-response of muscle-invasive bladder cancer patients to neoadjuvant chemotherapy. *Genes Cancer* **7**, 86–97. <https://doi.org/10.18632/genesandcancer.103> (2016).

## Acknowledgements

The authors acknowledge David Puthoff, PhD, from the Marshfield Clinic Research Institute for manuscript editing assistance.

## Author contributions

S.Y.S. designed the system and carried out the detail study. S.Y.S., MT, S.K.S., S.Y.H., Y.L. and A.B. participated in data analysis, manuscript preparation and discussed the results. All authors have read and approved the final manuscript.

## Funding

This work was supported in part by the Marshfield Clinic Research Institute, Marshfield, WI. The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-08082-7>.

**Correspondence** and requests for materials should be addressed to S.Y.S.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022, corrected publication 2022