# Identification of miR-200c and miR141-Mediated lncRNA-mRNA Crosstalks in Muscle-Invasive Bladder Cancer Subtypes

Guojun Liu[1†], Zihao Chen[2†], Irina G. Danilova[1], Mikhail A. Bolkov[3], Irina A. Tuzankina[3] and Guoqing Liu[4*]

[1] Institute of Natural Sciences and Mathematics, Ural Federal University, Yekaterinburg, Russia, [2] Department of Urology, Nanfang Hospital, Southern Medical University, Guangzhou, China, [3] Institute of Immunology and Physiology, Ural Branch of the Russian Academy of Sciences, Yekaterinburg, Russia, [4] School of Life Sciences and Technology, Inner Mongolia University of Science and Technology, Baotou, China

Basal and luminal subtypes of muscle-invasive bladder cancer (MIBC) have distinct molecular profiles and heterogeneous clinical behaviors. The interactions between mRNAs and lncRNAs, which might be regulated by miRNAs, have crucial roles in many cancers. However, the miRNA-dependent crosstalk between lncRNA and mRNA in specific MIBC subtypes still remains unclear. In this study, we first classified MIBC into two conservative subtypes using miRNA, mRNA and lncRNA expression data derived from The Cancer Genome Atlas. Then we investigated subtype-related biological pathways and evaluated the subtype classification performance using Decision Trees, Random Forest and eXtreme Gradient Boosting (XGBoost). At last, we explored potential miRNA-mediated lncRNA-mRNA crosstalks based on co-expression analysis. Our results show that: (1) the luminal subtype is primarily characterized by upregulation of metabolism-related pathways while the basal subtype is predominantly characterized by upregulation of epithelial-mesenchymal transition, metastasis, and immune system process-related pathways; (2) the XGBoost prediction model is consistently robust for classification of the molecular subtypes of MIBC across four datasets (The area under the ROC curve > 0.9); (3) the expression levels of the molecules in the miR-200c and miR141-mediated lncRNA-mRNA crosstalks differ considerably between the two subtypes and have close relationships with the prognosis of MIBC. The miR-200c and miR-141-dependent mRNA-lncRNA crosstalks might be of great significance in tumorigenesis and tumor progression and may serve as the novel prognostic predictors and classification markers of MIBC subtypes.

**Keywords: muscle-invasive bladder cancer, subtypes, miR200c, miR-141, random forest, XGBoost**

**Abbreviations:** ASW, Average of Silhouette Width; CC, cellular component; CC, consensus clustering; CDF, cumulative density function; CoC, cluster of cluster; CPCC, cophenetic correlation coefficient; DEFGs, differentially expressed feature genes; DEGs, differentially expressed genes; DTs, decision trees; EMT, epithelial-mesenchymal transition; FDR, False Discovery Rate; GO-BP, Gene Ontology Biological Process; GSEA, gene set enrichment analysis; KEGG, Kyoto Encyclopedia of Genes and Genomes; K–M curve, Kaplan–Meier curve; MAD, median absolute deviation; MF, molecular function; MIBC, muscle-invasive bladder cancer; NES, normalized enrichment scores; PAC, proportion of ambiguous clustering; RF, random forest; ROC, receiver operating characteristics curve; XGBoost, eXtreme Gradient Boosting.

# INTRODUCTION

Urothelial bladder cancer (UBC) is one of the most common malignant tumors of urinary system. UBC can generally be classified into non-muscle-invasive bladder cancer (NMIBC) and muscle-invasive bladder cancer (MIBC), according to whether the cancer cells are restricted locally in the lamina propria or invade the muscularis propria (Kamat et al., 2016). A great number of studies have reported that according to shared RNA expression patterns or specific genomic alterations MIBC can be further classified into two major subtypes, namely basal and luminal (Sjodahl et al., 2012; Iyer et al., 2013; Choi et al., 2014a,b; Damrauer et al., 2014; Network, 2014; Robertson et al., 2017), which are strikingly similar to the molecular subtypes first described in breast cancer (Perou et al., 2000; Prat et al., 2010). The basal subtype has drawn much attention because it is associated with a more aggressive phenotype and has a higher risk of distant metastasis than luminal subtype (Choi et al., 2014a; Robertson et al., 2017). One reason for the difference is that the two subtypes develop from etiologically different pathways. Pathways that are involved in EMT and immune-associated pathways are upregulated in the basal subtype (Choi et al., 2014a). The molecular biomarkers and pathways involved in MIBC subtypes are the key to understanding its subtype heterogeneity and identifying subtype-specific biomarkers that can be used to better manage MIBC patients.

MicroRNAs (miRNAs) represent one of the most exciting areas of modern medical and biological sciences as they can modulate an immense and complex regulatory network of gene expression in a broad spectrum of developmental and cellular processes, such as cell proliferation, metabolism, apoptosis, and viral infection (Johnston and Hobert, 2003; Hatfield et al., 2005; Zhao et al., 2005; Chen et al., 2006; Oliveira-Carvalho et al., 2012; Huang M. et al., 2016). miRNAs not only have a well-established inhibitory effect on gene expression but also promote gene expression in some cases (Sayed and Abdellatif, 2011; Song et al., 2014), and long non-coding RNAs (lncRNAs) exhibit facilitative or suppressive effects on the gene regulatory network during tumor development (Gontan et al., 2012; Sun et al., 2013). Furthermore, aberration or perturbation in miRNA-mediated mRNA and lncRNA expression levels has a significant correlation with serious clinical consequences, including diseases of diverse origins and malignancy (Salmena et al., 2011; Valinezhad Orang et al., 2014; Tay et al., 2014; Yuan et al., 2014; Zeng et al., 2016; Hu et al., 2017).

Regarding molecular drivers of cancer development, oncogenic mutations and downstream signaling pathways in the pre-cancerous or cancerous cell have been thought to play a crucial role in the cancer formation and progression. In addition, recent studies have shown that metabolic reprogramming plays much more important roles than previously thought in cancer development (Cairns et al., 2011). It is possible that a great number of genomic mutations detected in cancer provide a selective advantage for the cancer cell in the stressful tumor microenvironment by reprogramming cell metabolic processes (Zhang et al., 2015). No matter what is the primary cause of cancer development, it is clear that both the oncogenic signaling and reprogrammed metabolisms involve numerous genes, working in a concerted manner in a complex network. Gene regulatory network-based view can, therefore, provide a deeper insight into the cancer development.

The aim of this study is to identify subtype-specific dysregulated miRNA-mediated mRNA-lncRNA interactions and discover new critical subtype-related genes in MIBC.

# MATERIALS AND METHODS

## Data Acquisition and Pre-processing

The MIBC RNA-Seq (FPKM) and clinical data were obtained from The Cancer Genome Atlas (TCGA) public data portal[1], and miRNA-Seq (RPM) data was downloaded from the Broad GDAC Firehose[2]. The gene expression datasets of 403 tumor samples and 19 adjacent normal tissue samples contain 19181 mRNAs, 14376 lncRNAs, and 2588 mature miRNAs. The microarray datasets (GSE32894, GSE13507, and GSE31684) derived from Gene Expression Omnibus (GEO) were used to evaluate the performance of classifiers and verify the prognostic use of marker genes[3].

## Clustering Analysis and Gene Set Enrichment Analysis

Consensus clustering (Monti et al., 2003) is a method that provides quantitative evidence for determining the number and membership of possible clusters within a dataset, such as RNA-Seq and microarray. For CC analysis, the RPKM gene expression data was pre-processed to detect the most highly expressed and variable genes across samples. We removed 25% genes that have the low arithmetic mean of the given gene across samples. Then the MAD was used to select the most highly expressed and variable 3,000 mRNAs, 300 miRNAs, and 3,000 lncRNAs. CC available in the R package "ConsensusClusterPlus" was performed on 3,000 mRNAs, 300 miRNAs, and 3,000 lncRNAs with 403 tumor samples, using the following key parameters: reps = 50, innerLinkage = complete, clusterAlg = hc, k = 6, and distance = pearson (Wilkerson and Hayes, 2010).

Cluster of cluster analysis is a method of integrating the primary clustering results into final cluster assignments. Each sample is represented as a binary vector, whose length is $\sum_{i=1}^{t} K_i$ (where $t$ is the number of datasets and $K_i$ is the number of clusters for dataset ($i$), to implement subsequent clustering analysis. We first conducted the CoC analysis on the clustering results of mRNA, miRNA, and lncRNA dataset to obtain a binary dataset. The CC was once more performed on the binary dataset for generating final clusters. Number of final clusters (K) was estimated by commonly used methods including ASW, CPCC, Relative Change in Area under Cumulative density function [$^{\Delta}(K)$], and PAC (Şenbabaoglu et al., 2014).

In order to explore subtype-associated biological processes, GSEA (Subramanian et al., 2005) was conducted using three

---

[1]https://cancergenome.nih.gov
[2]https://gdac.broadinstitute.org
[3]https://www.ncbi.nlm.nih.gov

gene set datasets (GO-BP, KEGG, and Hallmark gene sets]. The following parameters were taken for GSEA: Number of permutations = *1000*, Permutation type = *gene_set*, Enrichment statistic = *weighted*, Metric for ranking genes = *Signal2Noise*.

## Differentially Expressed Genes and Machine Learning

"Ballgown" (R package) was used to identify DEGs between tumor and normal samples (Frazee et al., 2015). *F*-test was used in "Ballgown", and DEGs here were defined as those with FDR adjusted *p*-value < 0.05 (Benjamini–Hochberg method) and $|\log_2 \text{fold change}| > 0.57$).

Three tree-based machine learning methods, namely DTs, RF, and eXtreme Gradient Boosting (XGBoost or XG), were performed on 3000 mRNAs, 300 miRNAs, and 3000 lncRNAs for MIBC subtype classification. The area under the ROC curve (AUC) was used to estimate the performance of the classification methods. For each classification method, MIBC samples were randomly divided into training (60%) and testing (40%). We performed RF with different parameter values of *ntree* and *mtry*, and used 10-fold cross-validation to acquire the mean accuracy. XGBoost was implemented with the following parameters: gamma = *1*, min_child_weight = *1*, max_depth = *14*, nrounds = *2000*. In order to optimize the parameter *iter* (number of iterations) of XGBoost, we obtained 10-fold cross-validation performance for each *iter* and selected the *iter* value that generated the best performance. For DTs, the following parameters were taken: minCases = 20 and CF = 0.25. Moreover, the well-performed classifiers in this study were trained on the TCGA-derived RNA expression data and were tested on the GSE32894 to further evaluate their performance. All machine learning methods were implemented using R packages including "C5.0", "randomForest", and "XGBoost" packages (Liaw and Wiener, 2002; Chen and Guestrin, 2016; Kuhn et al., 2018).

The overlap between the feature genes obtained by the well-performed classifiers and DEGs was referred to as DEFGs. GO enrichment analysis available in the R package "clusterProfiler" was performed on DEFGs to identify their enriched GO terms (Yu et al., 2012). A multiple-test correction was done using the method proposed by Benjamini and Hochberg, in which an adjusted *p*-value < 0.05 was considered to represent statistical significance.

## Construction of a Subtype-Related mRNA-miRNA-lncRNA Network

Pairwise Pearson's correlation analysis was carried out on the DEFGs. The lncRNA-miRNA pairs, miRNA-mRNA pairs, and lncRNA-mRNA pairs with |r| > = 0.4 and *p*-value < 0.05 were considered to be co-expressed gene pairs. If both elements in a co-expressed lncRNA-mRNA pair are simultaneously co-expressed with a miRNA, it is defined as a miRNA-dependent lncRNA-mRNA co-expressed interaction. A miRNA-dependent lncRNA-mRNA network was established using Cytoscape software (version 3.5.1). miRWalk2.0 (Dweep et al., 2011) is an integration of six widely used databases (miRWalk, miRanda, miRDB, miRNAMap, RNA22, and Targetscan) and supplies the biggest available collection of predicted and experimentally

verified miRNA-target interactions. Our inferred co-expressed interactions including mRNA-miRNA and lncRNA-miRNA interactions were compared to those derived from miRWalk2.0. An mRNA is considered to be a true target of miRNA if their interaction occurs in at least four databases, and an lncRNA is considered to be a true target of miRNA if their interaction is supported in at least one database among miRWalk, miRanda, and Targetscan.
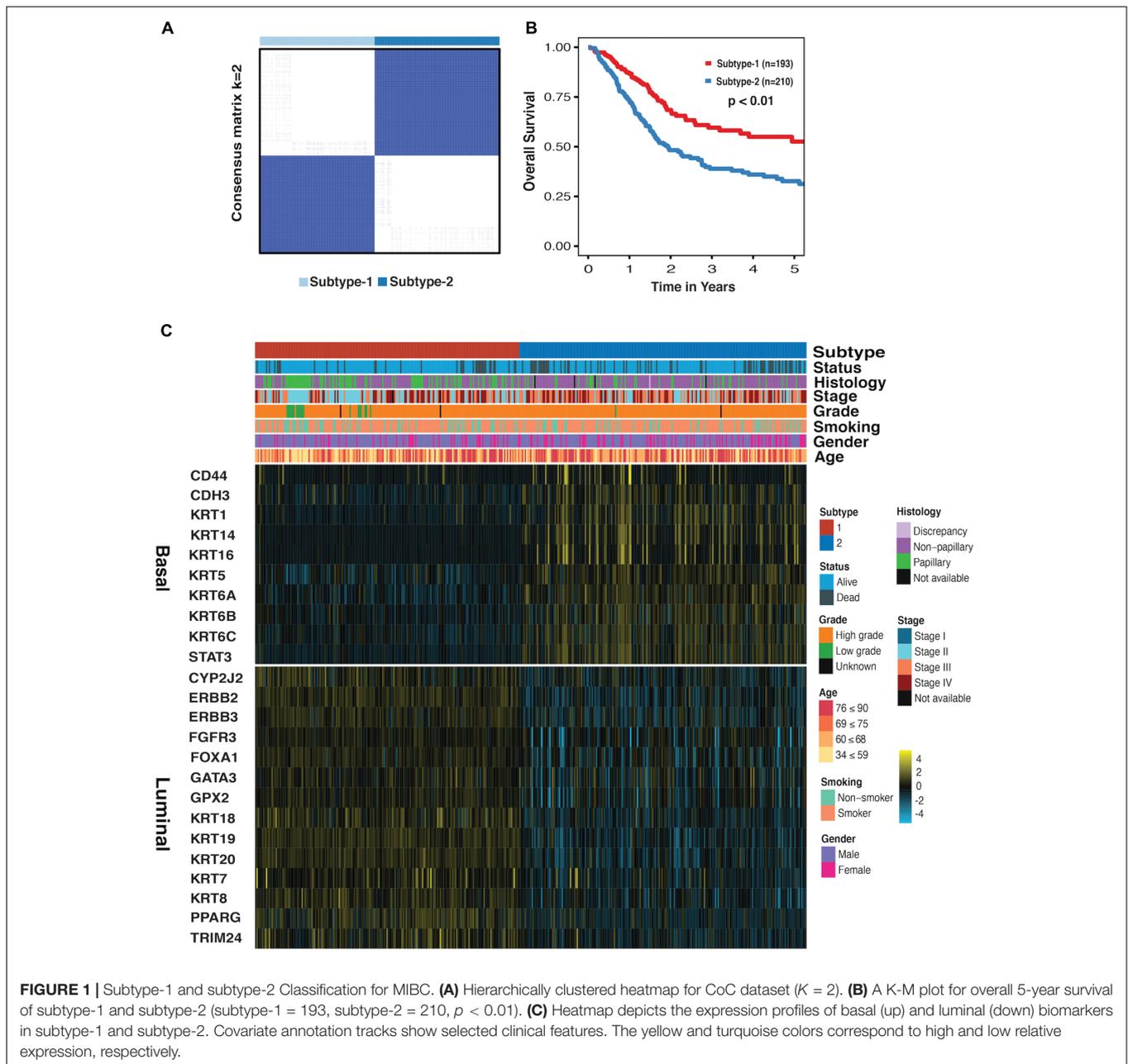
## Survival Analysis

We further assessed whether the genes in the inferred interactions are correlated with the overall survival of MIBC patients. Based on the mean expression level of the genes, patient samples were divided into high and low expression groups. We performed survival analysis available in R package "survival" (Therneau, 2015) using the Kaplan–Meier curve (K–M curve) method. A log-rank test was used to compare survival times between two groups, and *p* < 0.05 was considered to represent the statistical significance.

# RESULTS

## Clustering Analysis and GSEA

We first performed the CC on mRNA, miRNA, and lncRNA expression datasets to obtain the clustering results. By applying the CoC analysis to the clustering outcomes of CC, a binary dataset was obtained, which was referred to as CoC dataset. The CC was once again performed on the CoC dataset to generate the different Ks, and the ASW, CPCC, ΔK, and PAC were used to evaluate the optimal K (**Supplementary Figure S1**). As a result, for the CoC dataset, ASW evaluation suggests the optimal K of 6 and CPCC, ΔK, and PAC indicate the optimal K of 2. Given that K = 2 is the consistent optimal value, we chose K = 2 as a solution, dividing MIBC samples into two subtypes, namely subtype-1 and subtype-2. The hierarchically clustered heatmap of K = 2 for CoC dataset was shown in **Figure 1A**. Survival curves regarding two subtypes were plotted using the K-M method. Our results have shown that 5-year overall survival rate with regard to subtype-1 is 55% and 30% for subtype-2, indicating that they differ considerably in clinical prognosis (**Figure 1B**, *p* < 0.01). The heatmap depicting basal biomarkers, luminal biomarkers, and clinical indicators for the two subtypes was shown in the **Figure 1C**. The subtype-1 is characterized by the high expression of luminal markers such as CYP2J2, ERBB2, and KRT18, while the subtype 2 is characterized by high expression of basal markers such as CD44, CDH3, and KRT1. The Pearson's chi-squared test is utilized to compare clinical indicators between the two subtypes. The histology, stage, grade, and status are significantly different between the two subtypes, and gender almost differs between the two subtypes (**Supplementary Table S1**). The subtype-1 and subtype-2 resemble the luminal and basal subtype, respectively, in terms of K–M curves, biomarkers, and clinical indicators, therefore, which were redefined as luminal and basal subtypes (Choi et al., 2014a).

Gene set enrichment analysis was done for the basal and luminal subtypes, and the results were shown in **Tables 1**, **2**.

**FIGURE 1 |** Subtype-1 and subtype-2 Classification for MIBC. **(A)** Hierarchically clustered heatmap for CoC dataset ($K = 2$). **(B)** A K-M plot for overall 5-year survival of subtype-1 and subtype-2 (subtype-1 = 193, subtype-2 = 210, $p < 0.01$). **(C)** Heatmap depicts the expression profiles of basal (up) and luminal (down) biomarkers in subtype-1 and subtype-2. Covariate annotation tracks show selected clinical features. The yellow and turquoise colors correspond to high and low relative expression, respectively.

Upregulated pathways in luminal subtype are mainly involved in metabolism (e.g., oxidative phosphorylation, cytochrome P450, and fatty acid metabolism) (**Table 1**). Whereas, upregulated pathways in the basal subtype are principally related to immune system process (e.g., extracellular structure organization, allograft rejection, mTORC1 signaling, and TNF-a signaling via NF-kB), metastasis, and EMT (**Table 2**).

## Differentially Expressed Genes and Machine Learning

The DEGs that could distinguish tumor from normal samples were analyzed and visualized as volcano plots (**Supplementary Figures S2A–C**). In total, 208 miRNAs (148 upregulated and 60 downregulated), 2488 lncRNAs (1402 upregulated and 1086 down-regulated), and 4167 mRNAs (2314 upregulated and 1853 downregulated) are differentially expressed.

We applied DTs, RF, and XGBoost for the basal and luminal subtype classification based on mRNA, miRNA, and lncRNA expression dataset, and AUC was used to evaluate their performance. As shown in **Figure 2A**, XGBoost outperforms RF and DTs, having AUC values of 98.6, 94.5, and 98.7%, respectively, in mRNA, miRNA and lncRNA-based classification. Details regarding 10-fold cross-validation procedure can be found in **Supplementary Figure S3**. DTs was excluded in the following comparison, as it is significantly inferior to RF and XG on average. By using the CC method, the GSE32894 dataset

**TABLE 1** | Top-ranked terms of GO-BP, KEGG and Hallmark gene sets for the luminal subtype.

| Gene set name | Size | NES | FDR q-value |
|---|---|---|---|
| **GO-BP** | | | |
| GO monocarboxylic acid catabolic process | 95 | 2.5151 | 0 |
| GO oxidative phosphorylation | 73 | 2.4333 | 0 |
| GO fatty acid catabolic process | 80 | 2.4284 | 0 |
| GO fatty acid beta oxidation | 51 | 2.3211 | 0 |
| GO electron transport chain | 78 | 2.3116 | 0 |
| GO organic acid catabolic process | 92 | 2.2781 | 2.26E-04 |
| GO mitochondrial respiratory chain complex assembly | 42 | 2.1788 | 0.0014 |
| GO lipid oxidation | 63 | 2.1586 | 0.0015 |
| GO mitochondrial respiratory chain complex i biogenesis | 199 | 2.1449 | 0.0018 |
| GO establishment of protein localization to endoplasmic reticulum | 70 | 2.1285 | 0.0022 |
| **KEGG** | | | |
| KEGG ribosome | 87 | 2.3289 | 0 |
| KEGG alpha linolenic acid metabolism | 19 | 2.0770 | 5.29E-04 |
| KEGG metabolism of xenobiotics by cytochrome p450 | 68 | 2.0424 | 5.36E-04 |
| KEGG valine leucine and isoleucine degradation | 44 | 1.9767 | 0.00137 |
| KEGG drug metabolism cytochrome p450 | 70 | 1.9727 | 0.00120 |
| KEGG oxidative phosphorylation | 116 | 1.9372 | 0.00209 |
| KEGG peroxisome | 78 | 1.9320 | 0.00214 |
| KEGG fatty acid metabolism | 42 | 1.9184 | 0.00229 |
| KEGG retinol metabolism | 63 | 1.8697 | 0.00391 |
| KEGG linoleic acid metabolism | 29 | 1.8393 | 0.00482 |
| **Hallmark gene sets** | | | |
| Hallmark oxidative phosphorylation | 198 | 1.5145 | 0.05830 |
| Hallmark bile acid metabolism | 112 | 1.4110 | 0.07668 |
| Hallmark peroxisome | 103 | 1.4095 | 0.05174 |
| Hallmark adipogenesis | 191 | 1.3794 | 0.05125 |
| Hallmark fatty acid metabolism | 156 | 1.2596 | 0.11892 |

*NES, normalized enrichment score; GO-BP, Gene Ontology Biological Process; KEGG, Kyoto Encyclopedia of Genes and Genomes. Size is the number of genes in the gene set. A positive NES means that genes over-represented in the gene set are upregulated in luminal subtypes.*

**TABLE 2** | Top-ranked categories of GO-BP, KEGG and Hallmark gene sets for the basal subtype.

| Gene set name | Size | NES | FDR q-value |
|---|---|---|---|
| **GO-BP** | | | |
| GO extracellular structure organization | 297 | −2.8256 | 0 |
| GO antigen processing and presentation of exogenous peptide antigen via mhc class i | 65 | −2.7258 | 0 |
| GO antigen processing and presentation | 206 | −2.6334 | 0 |
| GO antigen processing and presentation of peptide antigen | 170 | −2.6246 | 0 |
| GO antigen processing and presentation of peptide antigen via mhc class i | 90 | −2.6134 | 0 |
| GO chondroitin sulfate biosynthetic process | 25 | −2.6008 | 0 |
| GO collagen fibril organization | 36 | −2.5958 | 0 |
| GO regulation of innate immune response | 349 | −2.5825 | 0 |
| GO positive regulation of defense response | 360 | −2.5802 | 0 |
| GO cytokine mediated signaling pathway | 440 | −2.5675 | 0 |
| **KEGG** | | | |
| KEGG focal adhesion | 197 | −2.6862 | 0 |
| KEGG cytokine cytokine receptor interaction | 257 | −2.5127 | 0 |
| KEGG ecm receptor interaction | 84 | −2.512 | 0 |
| KEGG proteasome | 43 | −2.4802 | 0 |
| KEGG leishmania infection | 69 | −2.4718 | 0 |
| KEGG viral myocarditis | 68 | −2.4178 | 0 |
| KEGG hematopoietic cell lineage | 85 | −2.4134 | 0 |
| KEGG regulation of actin cytoskeleton | 211 | −2.3911 | 0 |
| KEGG allograft rejection | 35 | −2.3902 | 0 |
| KEGG autoimmune thyroid disease | 50 | −2.3778 | 0 |
| **Hallmark gene sets** | | | |
| Hallmark epithelial-mesenchymal transition | 197 | −3.2473 | 0 |
| Hallmark inflammatory response | 197 | −3.0190 | 0 |
| Hallmark interferon gamma response | 197 | −2.9964 | 0 |
| Hallmark interferon alpha response | 94 | −2.9491 | 0 |
| Hallmark allograft rejection | 199 | −2.9010 | 0 |
| Hallmark G2M checkpoint | 194 | −2.6389 | 0 |
| Hallmark E2F targets | 196 | −2.6177 | 0 |
| Hallmark TNF-a signaling via NF-kB | 198 | −2.5512 | 0 |
| Hallmark complement | 195 | −2.5475 | 0 |
| Hallmark mTORC1 signaling | 198 | −2.441 | 0 |

*All abbreviations are the same as in **Table 1**. A negative NES value indicates that genes over-represented in the gene set are upregulated in the basal subtype.*
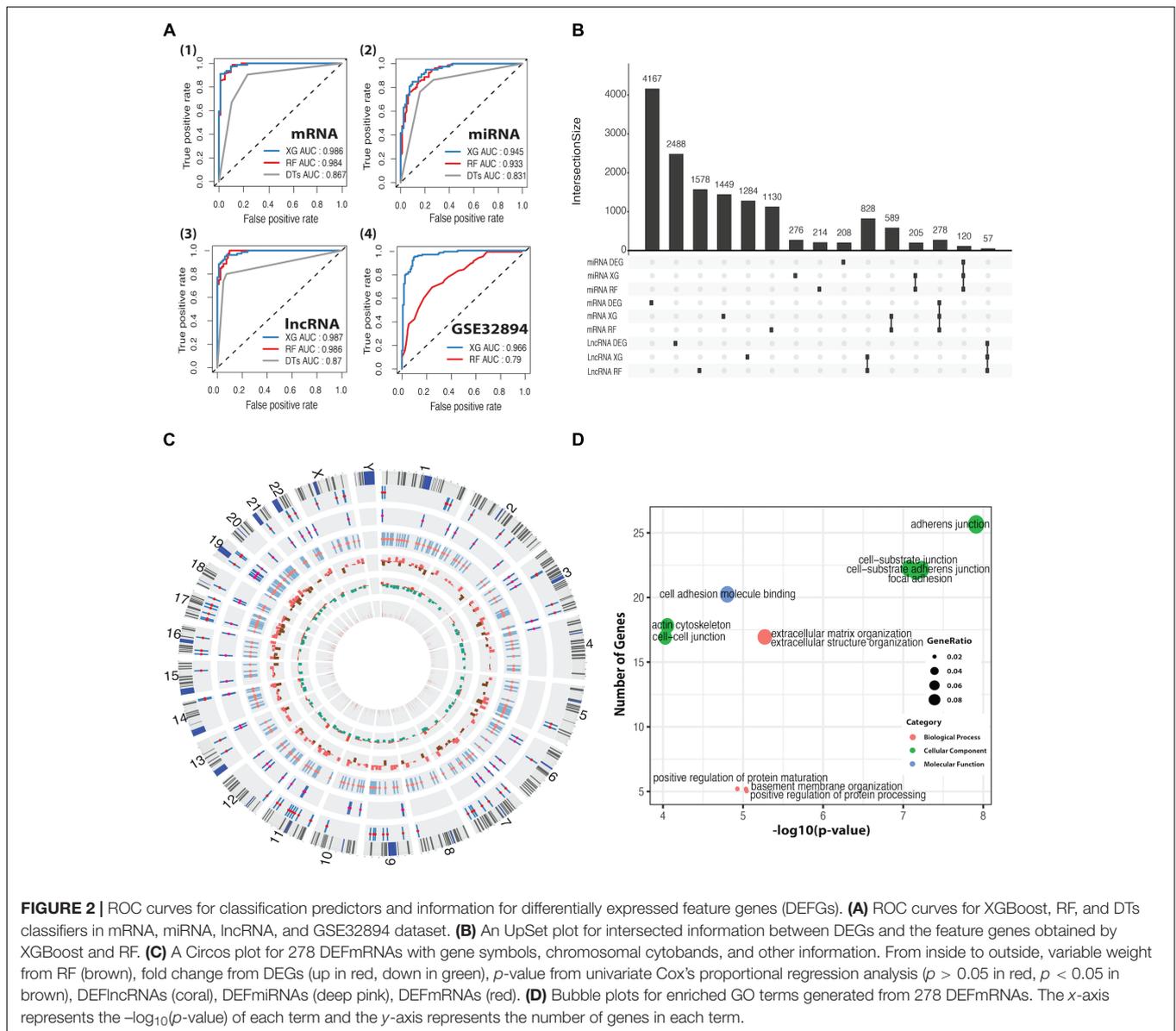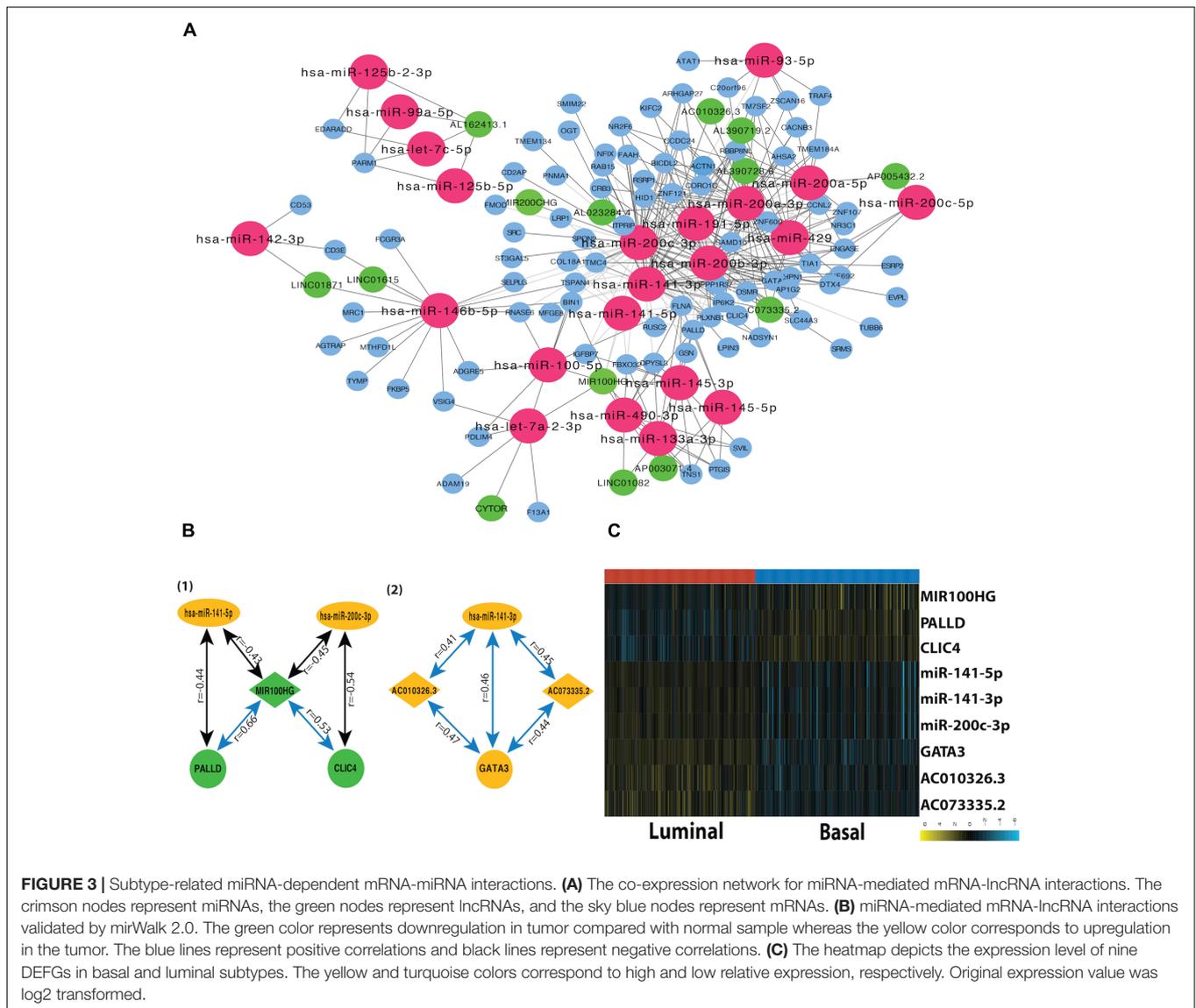
containing 28 biomarkers and 190 samples was grouped into two subtypes prepared for the classification task. The heatmap plots and the K–M curves for the two subtypes were shown in **Supplementary Figure S4**. We trained the well-performed classifiers (RF and XG) on mRNA dataset that was derived from TCGA and tested them on GSE32894 dataset. The results demonstrated that XGBoost has a better performance than RF (**Figure 2A4**).

The intersection between DEGs and feature genes obtained by RF and XG was defined as DEFGs, which includes 57 lncRNAs, 120 miRNAs, and 278 mRNAs. The Upset plot and heatmap

**FIGURE 2 |** ROC curves for classification predictors and information for differentially expressed feature genes (DEFGs). **(A)** ROC curves for XGBoost, RF, and DTs classifiers in mRNA, miRNA, lncRNA, and GSE32894 dataset. **(B)** An UpSet plot for intersected information between DEGs and the feature genes obtained by XGBoost and RF. **(C)** A Circos plot for 278 DEFmRNAs with gene symbols, chromosomal cytobands, and other information. From inside to outside, variable weight from RF (brown), fold change from DEGs (up in red, down in green), *p*-value from univariate Cox's proportional regression analysis (*p* > 0.05 in red, *p* < 0.05 in brown), DEFlncRNAs (coral), DEFmiRNAs (deep pink), DEFmRNAs (red). **(D)** Bubble plots for enriched GO terms generated from 278 DEFmRNAs. The *x*-axis represents the −log₁₀(*p*-value) of each term and the *y*-axis represents the number of genes in each term.

plots for DEFGs were shown in **Figure 2B** and **Supplementary Figures S2D–F**. The genetic and clinical information of DEFGs was visualized in **Figure 2C**. GO enrichment analysis indicated that differentially expressed feature mRNAs are enriched with adherens junction, cell-substrate junction, cell-cell junction, cell-substrate adherens junction, and focal adhesion (**Figure 2D**). These GO terms have been found to play roles in tumorigenesis and tumor progression by regulating T-cell signaling, innate immunity, TGF-β signaling, and Wnt signaling through post-translational modification (Kikuchi et al., 2006; Lönn, 2010; Liu et al., 2016; Cho et al., 2018; Kuwabara et al., 2018).

## Construction of Subtype-Related mRNA-miRNA-lncRNA Network

A miRNA-dependent mRNA-lncRNA co-expression network was constructed, which consists of 90 mRNAs, 22 miRNAs,

and 14 lncRNAs (**Figure 3A**). The miRNA-dependent mRNA-lncRNA crosstalks verified in miRWalk database contain four miRNA-mediated mRNA-lncRNA interactions (**Figure 3B**). To be specific, two co-expressed lncRNA-mRNA pairs, AC010326.3-GATA3 and AC073335.2-GATA3, are positively regulated by miR-141-3p; The lncRNA-mRNA pairs, such as MIR100HG–CLIC4 and MIR100HG–PALLD, are negatively regulated by miR-200c-3p and miR-141-5p, respectively. All the nine genes in the network differ in their expression between the two subtypes (**Figure 3C**). For instance, as compared to the luminal subtype, the basal subtype is characterized by a lower expression level of six genes (miR-200c-3p, miR-141-3p, miR-141-5p, GATA3, AC010326.3, and AC073335.2) and a higher expression level of the other three genes (MIR100HG, PALLD, and CLIC4), suggesting that all the nine genes can be used as potential markers for the two MIBC subtypes. In addition, GO analysis showed that

**FIGURE 3 |** Subtype-related miRNA-dependent mRNA-miRNA interactions. **(A)** The co-expression network for miRNA-mediated mRNA-lncRNA interactions. The crimson nodes represent miRNAs, the green nodes represent lncRNAs, and the sky blue nodes represent mRNAs. **(B)** miRNA-mediated mRNA-lncRNA interactions validated by mirWalk 2.0. The green color represents downregulation in tumor compared with normal sample whereas the yellow color corresponds to upregulation in the tumor. The blue lines represent positive correlations and black lines represent negative correlations. **(C)** The heatmap depicts the expression level of nine DEFGs in basal and luminal subtypes. The yellow and turquoise colors correspond to high and low relative expression, respectively. Original expression value was log2 transformed.

the mRNAs in the network (CLIC4, PALLD, and GATA3) are related to cytoskeleton.
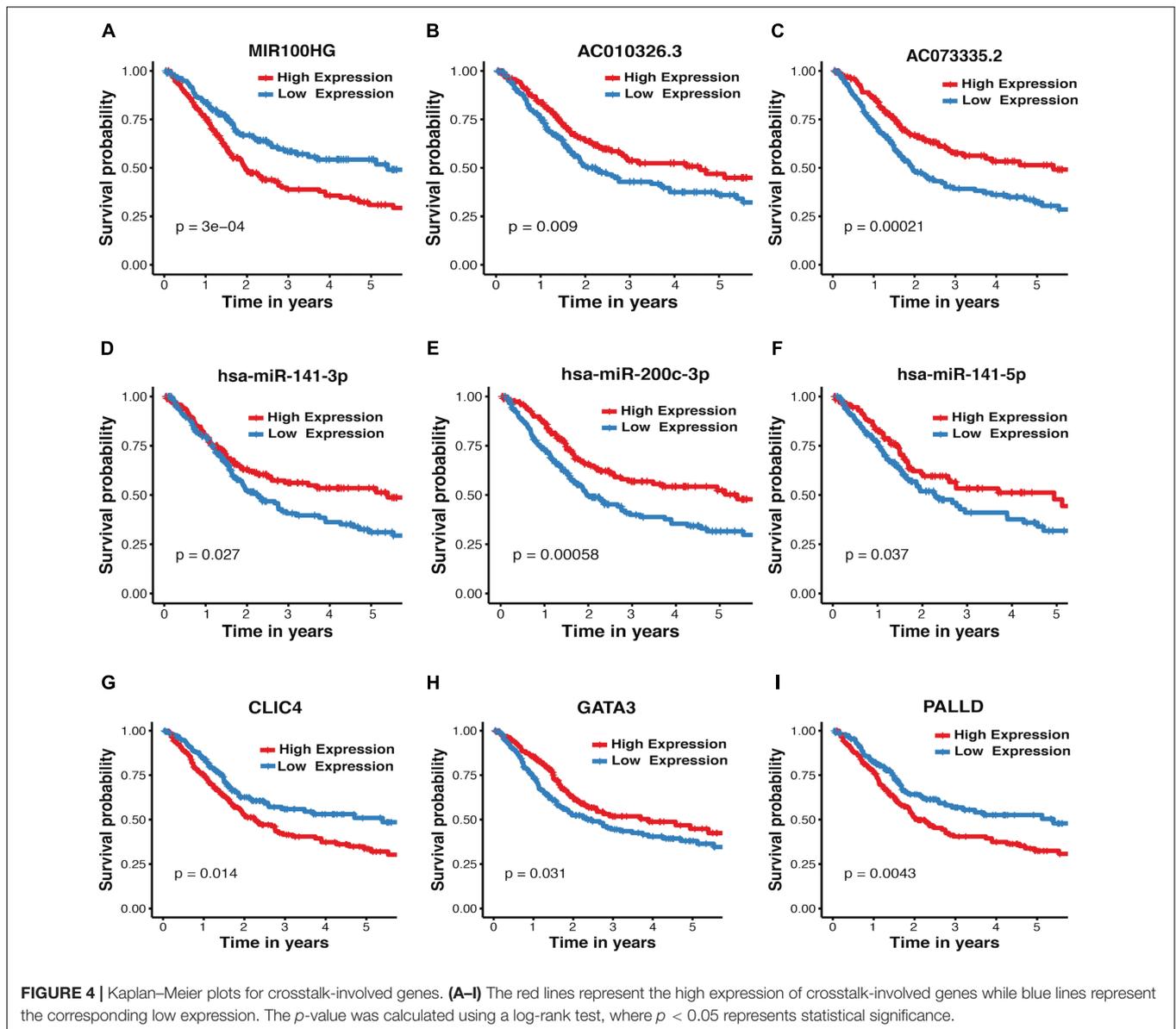
## Survival Analysis of Crosstalk-Involved Genes

The association between expression levels of crosstalk-involved genes and MIBC prognosis was analyzed by K–M method. Strikingly, the results revealed that all of them are closely related to prognosis of MIBC. Specifically, the higher expression level of miR-141-5p, miR-141-3p, AC010326.3, AC073335.2, miR-200c-3p, and GATA3 predicts better prognosis, indicating that they may function as tumor suppressors (**Figures 4B–F,H**); In contrast, the higher expression level of MIR100HG, PALLD, and CLIC4 is associated with worse prognosis, suggesting that they may play an oncogenic role (**Figures 4A,G,I**). In addition, the association between MIBC prognosis and expression levels of crosstalk-related mRNAs (CLIC4, PALLD, and GATA3) was

validated in two independent microarray datasets (GSE13507 and GSE31684), suggesting again their prognosis value in MIBC (**Supplementary Figure S5**).

## DISCUSSION

In this study, we have investigated miRNA-dependent mRNA-lncRNA interactions in MIBC basal and luminal subtypes using bioinformatics approaches. On the basis of MIBC mRNA, miRNA, and lncRNA expression datasets obtained from TCGA, 403 MIBC samples were reliably classified into two intrinsic molecular types, which resemble basal and luminal subtypes identified previously (Choi et al., 2014a). A number of subtype-related pathways were identified through GSEA. Moreover, we conducted and compared subtype classification performance among tree-based machine learning algorithms, and found XGBoost outperforms other classifiers. Additionally,

**FIGURE 4 |** Kaplan–Meier plots for crosstalk-involved genes. **(A–I)** The red lines represent the high expression of crosstalk-involved genes while blue lines represent the corresponding low expression. The *p*-value was calculated using a log-rank test, where $p < 0.05$ represents statistical significance.

we implemented a gene co-expression analysis on DEFGs and successfully identified subtype-specific mRNA-lncRNA crosstalks, which differ considerably between basal and luminal subtypes and have close relationships with the prognosis of MIBC.

Subtype-related pathways presented in this study (**Tables 1, 2**) are largely consistent with the previously identified (Choi et al., 2014a; Hurst and Knowles, 2014; McConkey et al., 2016; Ochoa et al., 2016; Hau et al., 2017; Seiler et al., 2017; Baker et al., 2018). In general, pathways that are involved in the EMT, metastasis, and immune system process, are upregulated in the basal subtype, whereas, metabolic-related pathways are upregulated in the luminal subtype. Th pathways enriched in basal and luminal subtypes provide a biological explanation for their distinctively different clinical and pathological behaviors. However, the mechanisms by which some other pathways

shown in our results, like valine leucine, isoleucine degradation, autoimmune thyroid disease, hematopoietic cell lineage and viral myocarditis, play a role in MIBC subtypes deserve further investigation.

Many machine learning methods have been broadly applied in many areas of biology such as gene family classification, hepatotoxicity prediction, RNA methylation prediction, cancer prediction and classification (Zou et al., 2014; Kourou et al., 2015; Liao et al., 2017, 2018; Su et al., 2018; Wei et al., 2018a,b). As suggested in previous studies, RF is a powerful classifier for classifying gene expression data (Wu et al., 2003; Lee et al., 2005; Ishwaran et al., 2010). And XGBoost keeps winning in "every" Kaggle competition and has become a really popular tool among data scientists (Ren et al., 2017; Torlay et al., 2017; Zhang and Zhan, 2017). Recently, XGBoost has been successfully applied to many classification problems, such as pan-cancer classification

(Li et al., 2017) and prediction of RNA-protein interactions (Jain et al., 2018). However, no comparison between RF and XGBoost in terms of cancer classification has been made in the past. In this study, we compared the performance of DTs, RF, XGBoost in classifying basal and luminal subtypes. Our results clearly demonstrated the advantage of XGBoost in gene expression data-based cancer classification (**Figure 2A**).

Previous studies investigated MIBC-associated miRNAs and their target genes without considering the genetic heterogeneity of MIBC subtypes (Martens-Uzunova et al., 2014; Huang T. et al., 2016; Xue et al., 2016; Zhong et al., 2016). It is therefore important to elucidate the subtype-related molecular pathways and identify novel biomarkers for MIBC subtypes. In this study, we systematically explored MIBC subtype-related gene co-expression networks. A total of three mRNAs (GATA3, CLIC4, and PALLD), three miRNAs (miR-200c-3p, miR-141-3p, and miR-141-5p), and three lncRNAs (AC010326.3, AC073335.2, and MIR100HG) were found in miRNA-mediated mRNA-lncRNA crosstalks, which differ considerably in their expression between basal and luminal subtypes (**Figure 3**), and their expression level is significantly associated with the prognosis of MIBC (**Figure 4**). It was previously observed that miR-141-5p, miR-141-3p, miR-200c-3p, and GATA3 are the most important markers of luminal subtype, which is consistent with our results (Robertson et al., 2017). Besides, previous studies found that the down-regulation of miR-200c and miR-141 is associated with elevated ZEB1 (Wiklund et al., 2011; Shan et al., 2013; Mahdavinezhad et al., 2015), and the down-regulation of miR-200c is also coupled with the down-regulation of BMI-1 and E2F3 (Liu et al., 2014), which play an important role in the invasion, migration, and EMT of bladder cancer.

It has been shown that some other genes in the crosstalk are also closely related to cancer. For example, AC073335.2, a highly expressed lncRNA in human glioblastoma, is involved in tumorigenesis via acting as a competing endogenous RNA of miR-940 (Shi et al., 2017). MIR100HG was previously reported to act as a regulator of hematopoiesis and oncogenes in many cancers (Emmrich et al., 2014; Nair, 2016; Shang et al., 2016; Wieczorek and Reszka, 2018; Zhang et al., 2018). In agreement with our findings, MIR100HG was reported to be down-regulated in MIBC and may serve as a significant biomarker for MIBC (Wang et al., 2016). As reported previously, GATA3 is a prognostic marker and inhibits cell migration and invasion in MIBC (Miyamoto et al., 2012; Choi et al., 2014a,b). And, GATA3 is differentially expressed between basal and luminal subtypes and can be used as a luminal-infiltrated marker (Robertson et al., 2017). CLIC4 has a complicated role in cancer. For instance, it functions as a tumor suppressor in lung adenocarcinomas (Okudela et al., 2014). And it promotes the metastasis and development of colorectal cancer (Deng et al., 2014; Peretti et al., 2015). Previous studies have established that the expression of CLIC4 in MIBC has a subtype-dependent pattern (Robertson et al., 2017). And the overexpression of CLIC4 in stroma increases cell migration and invasion and promotes epithelial to mesenchymal transition in multiple human cancers (Shukla et al., 2014). PALLD SNPs were reported to be a significant predictor of prostate cancer-specific mortality (Bao et al., 2011). Our findings are largely consistent with previously reported

results, suggesting crosstalk-implicated genes might be of great significance in MIBC pathogenesis and post-transcriptional gene regulation.

The combination of bioinformatics and several machine learning approaches in this study have achieved reliable results regarding the MIBC subtype classification, subtype-associated pathways, and the network-associated markers for MIBC subtypes. The subtype-related genes can not only be used for subtype classification but also serve as a good predictor of cancer prognosis. It is worth noting that we can enhance our study in the following aspects in the future: (1) the crosstalks discovered through computational analyses need to be verified by biological experiments. (2) DEFGs were defined as the overlap between DEGs and feature genes that were determined by XGBoost based on the ranking approximates of Information Gain. This procedure may result in the missing of some highly correlated genes that are also biologically important.

## CONCLUSION

By conducting bioinformatics analyses, we identified two subtypes of MIBC and lncRNA-mRNA crosstalks mediated by miR-200c and miR-141, which are found to be significantly associated with prognosis, formation, and metastasis of bladder cancer. Our results should be informative for molecular subtype classification, prognosis and molecule-targeted therapy of bladder cancer.

## AUTHOR CONTRIBUTIONS

GjL and ZC performed the computations. MB and IT contributed to data preparation and analysis. GjL and GqL wrote the manuscript. GqL and ID conceived and designed the study.

## FUNDING

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2018.00422/full#supplementary-material

**FIGURE S1 |** The graphs show the evaluation output of ACW, CPCC, ΔK, and PAC. CoC datasets represented by green line were used as the criteria to infer optimal K. **(A)** ASW allows us to inference the optimal K by high ASW. **(B)** The optimal K according to CPCC is that the magnitude of CPCC should be very close to one. **(C)** The optimal K according to ΔK is the K value before the 'elbow' or the K where D(K) reaches its maximum. **(D)** PAC allows us to inference the optimal K by the lowest PAC.

FIGURE S2 | Volcano plots for DEGs and heatmap plots for DEFGs. **(A–C)** Volcano plots for differentially expressed 4167 mRNAs, 208 miRNAs, and 2488 lncRNAs between tumor and normal samples (adjusted $p$-value $< 0.05$ and $|\log_2$fold change$| > 0.57$). **(D–F)** Heatmap plots for 278 DEFmRNAs, 120 DEFmiRNAs, and 57 DEFlncRNAs. Basal, luminal, and normal samples are represented by the red, blue, and yellow bar, respectively.

FIGURE S3 | Parameter selection and Performance of RF and XG in mRNA, miRNA and lncRNA dataset. **(A)** The $x$-axis represents the number of mtry set for RF classifier (1, 5, 10, 15, 20, 25). The $y$-axis represents the corresponding AUC. **(B)** The $x$-axis represents the number of ntree set for RF (20, 400, 600, 800). The $y$-axis represents corresponding obb error rates. The colors correspond to mtry numbers. **(C)** The $x$-axis represents the number of fold set for RF. The $y$-axis represents corresponding accuracy. The red color shows mean accuracy. **(D)** The

$x$-axis represents the number of *iter* set for XG (1, 400, 800, 1200, 1600, 2000, 2400) and the y-axis represents the corresponding accuracy.

FIGURE S4 | Heatmap and K–M plots for basal and luminal subtypes of GSE32894. **(A)** Heatmap depicts the expression profiles of basal (up) and luminal (down) biomarkers in GSE32894. The yellow and turquoise color corresponds to high and low relative expression, respectively. **B**. A K-M plot for the overall 5-year survival of basal and luminal subtypes (basal = 52, luminal = 62, $p < 0.01$).

FIGURE S5 | Kaplan-Meier plots for CLIC4, PALLD, and GATA3 in GSE13507 and GSE31684. **(A–C)** K–M survival curves showing overall survival according to high expression and low expression of CLIC4, PALLD, and GATA3 in GSE13507. **(D–F)** K–M survival curves showing overall survival according to high expression and low expression of CLIC4, PALLD, GATA3, and MIR100HG in GSE31684.

# REFERENCES

Baker, S. C., Arlt, V. M., Indra, R., Joel, M., Stiborová, M., Eardley, I., et al. (2018). Differentiation-associated urothelial cytochrome P450 oxidoreductase predicates the xenobiotic-metabolizing activity of "luminal" muscle-invasive bladder cancers. *Mol. Carcinog.* 57, 606–618. doi: 10.1002/mc.22784

Bao, B.-Y., Pao, J.-B., Huang, C.-N., Pu, Y.-S., Chang, T.-Y., Lan, Y.-H., et al. (2011). Polymorphisms inside microRNAs and microRNA target sites predict clinical outcomes in prostate cancer patients receiving androgen-deprivation therapy. *Clin. Cancer Res.* 17, 928–936. doi: 10.1158/1078-0432.CCR-10-2648

Cairns, R. A., Harris, I. S., and Mak, T. W. (2011). Regulation of cancer cell metabolism. *Nat. Rev. Cancer* 11, 85–95. doi: 10.1038/nrc2981

Chen, J. F., Mandel, E. M., Thomson, J. M., Wu, Q., Callis, T. E., Hammond, S. M., et al. (2006). The role of microRNA-1 and microRNA-133 in skeletal muscle proliferation and differentiation. *Nat. Genet.* 38, 228–233. doi: 10.1038/ng1725

Chen, T., and Guestrin, C. (2016). "Xgboost: a scalable tree boosting system," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Francisco, CA, 785–794. doi: 10.1145/2939672.2939785

Cho, Y., Kang, H. G., Kim, S.-J., Lee, S., Jee, S., Ahn, S. G., et al. (2018). Post-translational modification of OCT4 in breast cancer tumorigenesis. *Cell Death Differ.* doi: 10.1038/s41418-018-0079-6 [Epub ahead of print].

Choi, W., Czerniak, B., Ochoa, A., Su, X., Siefker-Radtke, A., Dinney, C., et al. (2014a). Intrinsic basal and luminal subtypes of muscle-invasive bladder cancer. *Nat. Rev. Urol.* 11, 400–410. doi: 10.1038/nrurol.2014.129

Choi, W., Porten, S., Kim, S., Willis, D., Plimack, E. R., Hoffman-Censits, J., et al. (2014b). Identification of distinct basal and luminal subtypes of muscle-invasive bladder cancer with different sensitivities to frontline chemotherapy. *Cancer Cell* 25, 152–165. doi: 10.1016/j.ccr.2014.01.009

Damrauer, J. S., Hoadley, K. A., Chism, D. D., Fan, C., Tiganelli, C. J., Wobker, S. E., et al. (2014). Intrinsic subtypes of high-grade bladder cancer reflect the hallmarks of breast cancer biology. *Proc. Natl. Acad. Sci. U.S.A.* 111, 3110–3115. doi: 10.1073/pnas.1318376111

Deng, Y.-J., Tang, N., Liu, C., Zhang, J.-Y., An, S.-L., Peng, Y.-L., et al. (2014). CLIC4, ERp29, and Smac/DIABLO derived from metastatic cancer stem-like cells stratify prognostic risks of colorectal cancer. *Clin. Cancer Res.* 20, 3809–3817. doi: 10.1158/1078-0432.CCR-13-1887

Dweep, H., Sticht, C., Pandey, P., and Gretz, N. (2011). miRWalk – Database: Prediction of possible miRNA binding sites by "walking" the genes of three genomes. *J. Biomed. Inform.* 44, 839–847. doi: 10.1016/j.jbi.2011.05.002

Emmrich, S., Streltsov, A., Schmidt, F., Thangapandi, V., Reinhardt, D., and Klusmann, J.-H. (2014). LincRNAs MONC and MIR100HG act as oncogenes in acute megakaryoblastic leukemia. *Mol. Cancer* 13:171. doi: 10.1186/1476-4598-13-171

Frazee, A. C., Pertea, G., Jaffe, A. E., Langmead, B., Salzberg, S. L., and Leek, J. T. (2015). Ballgown bridges the gap between transcriptome assembly and expression analysis. *Nat. Biotechnol.* 33, 243–246. doi: 10.1038/nbt.3172

Gontan, C., Achame, E. M., Demmers, J., Barakat, T. S., Rentmeester, E., van IJcken, W., et al. (2012). RNF12 initiates X-chromosome inactivation by targeting REX1 for degradation. *Nature* 485, 386–390. doi: 10.1038/nature11070

Hatfield, S. D., Shcherbata, H. R., Fischer, K. A., Nakahara, K., Carthew, R. W., and Ruohola-Baker, H. (2005). Stem cell division is regulated by the microRNA pathway. *Nature* 435, 974–978. doi: 10.1038/nature03816

Hau, A. M., Nakasaki, M., Nakashima, K., Krish, G., and Hansel, D. E. (2017). Differential mTOR pathway profiles in bladder cancer cell line subtypes to predict sensitivity to mTOR inhibition. *Urol. Oncol.* 35, 593–599. doi: 10.1016/j.urolonc.2017.03.025

Hu, H., Wilson, K. D., Zhong, S., and He, C. (2017). LincRNAs: systemic computational identification and functional exploration. *Curr. Bioinform.* 12, 34–42. doi: 10.2174/1574893611666160923159

Huang, M., Zhong, Z., Lv, M., Shu, J., Tian, Q., and Chen, J. (2016). Comprehensive analysis of differentially expressed profiles of lncRNAs and circRNAs with associated co-expression and ceRNA networks in bladder carcinoma. *Oncotarget* 7, 47186–47200. doi: 10.18632/oncotarget.9706

Huang, T., Li, B.-Q., and Cai, Y.-D. (2016). The integrative network of gene expression, MicroRNA, methylation and copy number variation in colon and rectal cancer. *Curr. Bioinform.* 11, 59–65. doi: 10.2174/1574893611666151119215823

Hurst, C. D., and Knowles, M. A. (2014). Molecular subtyping of invasive bladder cancer: time to divide and rule? *Cancer Cell* 25, 135–136. doi: 10.1016/j.ccr.2014.01.026

Ishwaran, H., Kogalur, U. B., Gorodeski, E. Z., Minn, A. J., and Lauer, M. S. (2010). High-dimensional variable selection for survival data. *J. Am. Stat. Assoc.* 105, 205–217. doi: 10.1198/jasa.2009.tm08622

Iyer, G., Al-Ahmadie, H., Schultz, N., Hanrahan, A. J., Ostrovnaya, I., Balar, A. V., et al. (2013). Prevalence and co-occurrence of actionable genomic alterations in high-grade bladder cancer. *J. Clin. Oncol.* 31, 3133–3140. doi: 10.1200/JCO.2012.46.5740

Jain, D. S., Gupte, S. R., and Aduri, R. (2018). A data driven model for predicting RNA-protein interactions based on gradient boosting machine. *Sci. Rep.* 8:9552. doi: 10.1038/s41598-018-27814-2

Johnston, R. J. Jr., and Hobert, O. (2003). A microRNA controlling left/right neuronal asymmetry in *Caenorhabditis elegans*. *Nature* 426, 845–849. doi: 10.1038/nature02255

Kamat, A. M., Hahn, N. M., Efstathiou, J. A., Lerner, S. P., Malmström, P.-U., Choi, W., et al. (2016). Bladder cancer. *Lancet* 388, 2796–2810. doi: 10.1016/S0140-6736(16)30512-8

Kikuchi, A., Kishida, S., and Yamamoto, H. (2006). Regulation of Wnt signaling by protein-protein interaction and post-translational modifications. *Exp. Mol. Med.* 38, 1–10. doi: 10.1038/emm.2006.1

Kourou, K., Exarchos, T. P., Exarchos, K. P., Karamouzis, M. V., and Fotiadis, D. I. (2015). Machine learning applications in cancer prognosis and prediction. *Comput. Struct. Biotechnol. J.* 13, 8–17. doi: 10.1016/j.csbj.2014.11.005

Kuhn, M., Weston, S., Culp, M., Coulter, N., and Quinlan, R. (2018). *Package "C50"*.

Kuwabara, T., Matsui, Y., Ishikawa, F., and Kondo, M. (2018). Regulation of T-cell signaling by post-translational modifications in autoimmune disease. *Int. J. Mol. Sci.* 19:819. doi: 10.3390/ijms19030819

Lee, J. W., Lee, J. B., Park, M., and Song, S. H. (2005). An extensive comparison of recent classification tools applied to microarray data. *Comput. Stat. Data Anal.* 48, 869–885. doi: 10.1016/j.csda.2004.03.017

Li, Y., Kang, K., Krahn, J. M., Croutwater, N., Lee, K., Umbach, D. M., et al. (2017). A comprehensive genomic pan-cancer classification using The Cancer Genome

Atlas gene expression data. *BMC Genomics* 18:508. doi: 10.1186/s12864-017-3906-0

Liao, Z., Li, D., Wang, X., Li, L., and Zou, Q. (2018). Cancer diagnosis through IsomiR expression with machine learning method. *Curr. Bioinform.* 13, 57–63. doi: 10.2174/1574893611666160609081155

Liao, Z., Wan, S., He, Y., and Zou, Q. (2017). Classification of small GTPases with hybrid protein features and advanced machine learning techniques. *Curr. Bioinform.* 13, 492–500. doi: 10.2174/1574893612666171121162552

Liaw, A., and Wiener, M. (2002). Classification and regression by randomForest. *R News* 2, 18–22.

Liu, J., Qian, C., and Cao, X. (2016). Post-translational modification control of innate immunity. *Immunity* 45, 15–30. doi: 10.1016/J.IMMUNI.2016.06.020

Liu, L., Qiu, M., Tan, G., Liang, Z., Qin, Y., Chen, L., et al. (2014). miR-200c Inhibits invasion, migration and proliferation of bladder cancer cells through down-regulation of BMI-1 and E2F3. *J. Transl. Med.* 12:305. doi: 10.1186/s12967-014-0305-z

Lönn, P. (2010). *Regulation of TGF-β Signaling by Post-Translational Modifications.* Doctoral dissertation, Acta Universitatis Upsaliensis, Uppsala.

Mahdavinezhad, A., Mousavi-Bahar, S. H., Poorolajal, J., Yadegarazari, R., Jafari, M., Shabab, N., et al. (2015). Evaluation of miR-141, miR-200c, miR-30b expression and clinicopathological features of bladder cancer. *Int. J. Mol. Cell. Med.* 4, 32–39.

Martens-Uzunova, E. S., Böttcher, R., Croce, C. M., Jenster, G., Visakorpi, T., and Calin, G. A. (2014). Long noncoding RNA in prostate, bladder, and kidney cancer. *Eur. Urol.* 65, 1140–1151. doi: 10.1016/j.eururo.2013.12.003

McConkey, D. J., Choi, W., Ochoa, A., and Dinney, C. P. (2016). Intrinsic subtypes and bladder cancer metastasis. *Asian J. Urol.* 3, 260–267. doi: 10.1016/j.ajur.2016.09.009

Miyamoto, H., Izumi, K., Yao, J. L., Li, Y., Yang, Q., McMahon, L. A., et al. (2012). GATA binding protein 3 is down-regulated in bladder cancer yet strong expression is an independent predictor of poor prognosis in invasive tumor. *Hum. Pathol.* 43, 2033–2040. doi: 10.1016/j.humpath.2012.02.011

Monti, S., Tamayo, P., Mesirov, J., and Golub, T. (2003). Consensus clustering: a resampling-based method for class discovery and visualization of gene expression microarray data. *Mach. Learn.* 52, 91–118. doi: 10.1023/A:1023949509487

Nair, S. (2016). Current insights into the molecular systems pharmacology of lncRNA-miRNA regulatory interactions and implications in cancer translational medicine. *AIMS Mol. Sci.* 3, 104–124. doi: 10.3934/molsci.2016.2.104

Network, T. C. G. A. R. (2014). Comprehensive molecular characterization of urothelial bladder carcinoma. *Nature* 507, 315–322. doi: 10.1038/nature12965

Ochoa, A. E., Choi, W., Su, X., Siefker-Radtke, A., Czerniak, B., Dinney, C., et al. (2016). Specific micro-RNA expression patterns distinguish the basal and luminal subtypes of muscle-invasive bladder cancer. *Oncotarget* 7, 80164–80174. doi: 10.18632/oncotarget.13284

Okudela, K., Katayama, A., Woo, T., Mitsui, H., Suzuki, T., Tateishi, Y., et al. (2014). Proteome analysis for downstream targets of oncogenic KRAS - the potential participation of CLIC4 in carcinogenesis in the lung. *PLoS One* 9:e87193. doi: 10.1371/journal.pone.0087193

Oliveira-Carvalho, V., Carvalho, V. O., Silva, M. M., Guimarães, G. V., and Bocchi, E. A. (2012). MicroRNAs: a new paradigm in the treatment and diagnosis of heart failure? *Arq. Bras. Cardiol.* 98, 362–370. doi: 10.1590/S0066-782X2012000400011

Peretti, M., Angelini, M., Savalli, N., Florio, T., Yuspa, S. H., and Mazzanti, M. (2015). Chloride channels in cancer: focus on chloride intracellular channel 1 and 4 (CLIC1 AND CLIC4) proteins in tumor development and as novel therapeutic targets. *Biochim. Biophys. Acta* 1848, 2523–2531. doi: 10.1016/J.BBAMEM.2014.12.012

Perou, C. M., Sørlie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., Rees, C. A., et al. (2000). Molecular portraits of human breast tumours. *Nature* 406, 747–752. doi: 10.1038/35021093

Prat, A., Parker, J. S., Karginova, O., Fan, C., Livasy, C., Herschkowitz, J. I., et al. (2010). Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer. *Breast Cancer Res.* 12:R68. doi: 10.1186/bcr2635

Ren, X., Guo, H., Li, S., Wang, S., and Li, J. (2017). "A novel image classification method with CNN-XGBoost model," in *Proceedings of the International Workshop on Digital Watermarking*, Magdeburg, 378–390. doi: 10.1007/978-3-319-64185-0_28

Robertson, A. G., Kim, J., Al-Ahmadie, H., Bellmunt, J., Guo, G., Cherniack, A. D., et al. (2017). Comprehensive molecular characterization of muscle-invasive bladder cancer. *Cell* 171, 540–556.e25. doi: 10.1016/j.cell.2017.09.007

Salmena, L., Poliseno, L., Tay, Y., Kats, L., and Pandolfi, P. P. (2011). A ceRNA hypothesis: the Rosetta stone of a hidden RNA language? *Cell* 146, 353–358. doi: 10.1016/J.CELL.2011.07.014

Sayed, D., and Abdellatif, M. (2011). MicroRNAs in development and disease. *Physiol. Rev.* 91, 827–887. doi: 10.1152/physrev.00006.2010

Seiler, R., Ashab, H. A. D., Erho, N., van Rhijn, B. W. G., Winters, B., Douglas, J., et al. (2017). Impact of molecular subtypes in muscle-invasive bladder cancer on predicting response and survival after neoadjuvant chemotherapy. *Eur. Urol.* 72, 544–554. doi: 10.1016/J.EURURO.2017.03.030

Sjodahl, G., Lauss, M., Lovgren, K., Chebil, G., Gudjonsson, S., Veerla, S., et al. (2012). A molecular taxonomy for urothelial carcinoma. *Clin. Cancer Res.* 18, 3377–3386. doi: 10.1158/1078-0432.CCR-12-0077-T

Shan, Y., Zhang, L., Bao, Y., Li, B., He, C., Gao, M., et al. (2013). Epithelial-mesenchymal transition, a novel target of sulforaphane via COX-2/MMP2, 9/Snail, ZEB1 and miR-200c/ZEB1 pathways in human bladder cancer cells. *J. Nutr. Biochem.* 24, 1062–1069. doi: 10.1016/J.JNUTBIO.2012.08.004

Shang, C., Zhu, W., Liu, T., Wang, W., Huang, G., Huang, J., et al. (2016). Characterization of long non-coding RNA expression profiles in lymph node metastasis of early-stage cervical cancer. *Oncol. Rep.* 35, 3185–3197. doi: 10.3892/or.2016.4715

Shi, J., Wang, Y.-J., Sun, C.-R., Qin, B., Zhang, Y., and Chen, G. (2017). Long noncoding RNA lncHERG promotes cell proliferation, migration and invasion in glioblastoma. *Oncotarget* 8, 108031–108041. doi: 10.18632/oncotarget.22446

Shukla, A., Edwards, R., Yang, Y., Hahn, A., Folkers, K., Ding, J., et al. (2014). CLIC4 regulates TGF-β-dependent myofibroblast differentiation to produce a cancer stroma. *Oncogene* 33, 842–850. doi: 10.1038/onc.2013.18

Song, R., Walentek, P., Sponer, N., Klimke, A., Lee, J. S., Dixon, G., et al. (2014). miR-34/449 miRNAs are required for motile ciliogenesis by repressing cp110. *Nature* 510, 115–120. doi: 10.1038/nature13413

Su, R., Wu, H., Xu, B., Liu, X., and Wei, L. (2018). Developing a multi-dose computational model for drug-induced hepatotoxicity prediction based on toxicogenomics data. *IEEE/ACM Trans. Comput. Biol. Bioinform.* doi: 10.1109/TCBB.2018.2858756 [Epub ahead of print].

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., et al. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U.S.A.* 102, 15545–15550. doi: 10.1073/pnas.0506580102

Sun, Q., Csorba, T., Skourti-Stathaki, K., Proudfoot, N. J., and Dean, C. (2013). R-loop stabilization represses antisense transcription at the *Arabidopsis FLC* locus. *Science* 340, 619–621. doi: 10.1126/science.1234848

Tay, Y., Rinn, J., and Pandolfi, P. P. (2014). The multilayered complexity of ceRNA crosstalk and competition. *Nature* 505, 344–352. doi: 10.1038/nature12986

Therneau, T. (2015). *A Package for Survival Analysis in S. version 2.38*

Torlay, L., Perrone-Bertolotti, M., Thomas, E., and Baciu, M. (2017). Machine learning–XGBoost analysis of language networks to classify patients with epilepsy. *Brain Inform.* 4, 159–169. doi: 10.1007/s40708-017-0065-7

Valinezhad Orang, A., Safaralizadeh, R., and Kazemzadeh-Bavili, M. (2014). Mechanisms of miRNA-mediated gene regulation from common downregulation to mRNA-specific upregulation. *Int. J. Genomics* 2014: 970607. doi: 10.1155/2014/970607

Wang, H., Niu, L., Jiang, S., Zhai, J., Wang, P., Kong, F., et al. (2016). Comprehensive analysis of aberrantly expressed profiles of lncRNAs and miRNAs with associated ceRNA network in muscle-invasive bladder cancer. *Oncotarget* 7, 86174–86185. doi: 10.18632/oncotarget.13363

Wei, L., Chen, H., and Su, R. (2018a). M6APred-EL: a sequence-based predictor for identifying N6-methyladenosine sites using ensemble learning. *Mol. Ther. Nucleic Acids* 12, 635–644. doi: 10.1016/j.omtn.2018.07.004

Wei, L., Zhou, C., Chen, H., Song, J., and Su, R. (2018b). ACPred-FL: a sequence-based predictor using effective feature representation to improve the

prediction of anti-cancer peptides. *Bioinformatics* doi: 10.1093/bioinformatics/bty451 [Epub ahead of print].

Wieczorek, E., and Reszka, E. (2018). mRNA, microRNA and lncRNA as novel bladder tumor markers. *Clin. Chim. Acta.* 477, 141–153. doi: 10.1016/j.cca.2017.12.009

Wiklund, E. D., Bramsen, J. B., Hulf, T., Dyrskjøt, L., Ramanathan, R., Hansen, T. B., et al. (2011). Coordinated epigenetic repression of the miR-200 family and miR-205 in invasive bladder cancer. *Int. J. Cancer* 128, 1327–1334. doi: 10.1002/ijc.25461

Wilkerson, M. D., and Hayes, D. N. (2010). ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. *Bioinformatics* 26, 1572–1573. doi: 10.1093/bioinformatics/btq170

Wu, B., Abbott, T., Fishman, D., McMurray, W., Mor, G., Stone, K., et al. (2003). Comparison of statistical methods for classification of ovarian cancer using mass spectrometry data. *Bioinformatics* 19, 1636–1643. doi: 10.1016/j.csda.2004.03.017

Xue, M., Pang, H., Li, X., Li, H., Pan, J., and Chen, W. (2016). Long non-coding RNA urothelial cancer-associated 1 promotes bladder cancer cell migration and invasion by way of the hsa-miR-145-ZEB1/2-FSCN1 pathway. *Cancer Sci.* 107, 18–27. doi: 10.1111/cas.12844

Yu, G., Wang, L.-G., Han, Y., and He, Q.-Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* 16, 284–287. doi: 10.1089/omi.2011.0118

Yuan, J., Yang, F., Wang, F., Ma, J., Guo, Y., Tao, Q., et al. (2014). A long noncoding RNA activated by TGF-β promotes the invasion-metastasis cascade in hepatocellular carcinoma. *Cancer Cell* 25, 666–681. doi: 10.1016/J.CCR.2014.03.010

Zeng, X., Zhang, X., and Zou, Q. (2016). Integrative approaches for predicting microRNA function and prioritizing disease-related microRNA using biological interaction networks. *Brief. Bioinform.* 17, 193–203. doi: 10.1093/bib/bbv033

Zhang, C., Cao, S., Toole, B. P., and Xu, Y. (2015). Cancer may be a pathway to cell survival under persistent hypoxia and elevated ROS: a model for solid-cancer initiation and early development. *Int. J. Cancer* 136, 2001–2011. doi: 10.1002/ijc.28975

Zhang, J., Yuan, Y., Wei, Z., Ren, J., Hou, X., Yang, D., et al. (2018). Crosstalk between prognostic long noncoding RNAs and messenger RNAs as transcriptional hallmarks in gastric cancer. *Epigenomics* 10, 433–443. doi: 10.2217/epi-2017-0136

Zhang, L., and Zhan, C. (2017). "Machine learning in rock facies classification: an application of XGBoost," in *Proceedings of the International Geophysical Conference, (Society of Exploration Geophysicists and Chinese Petroleum Society) 17-20 April 2017*, Qingdao, 1371–1374. doi: 10.1190/IGC2017-351

Zhao, Y., Samal, E., and Srivastava, D. (2005). Serum response factor regulates a muscle-specific microRNA that targets Hand2 during cardiogenesis. *Nature* 436, 214–220. doi: 10.1038/nature03817

Zhong, Z., Lv, M., and Chen, J. (2016). Screening differential circular RNA expression profiles reveals the regulatory role of circTCF25-miR-103a-3p/miR-107-CDK6 pathway in bladder carcinoma. *Sci. Rep.* 6:30919. doi: 10.1038/srep30919

Zou, Q., Mao, Y., Hu, L., Wu, Y., and Ji, Z. (2014). miRClassify: an advanced web server for miRNA family classification and annotation. *Comput. Biol. Med.* 45, 157–160. doi: 10.1016/j.compbiomed.2013.12.007

Şenbabaoglu, Y., Michailidis, G., and Li, J. Z. (2014). Critical limitations of consensus clustering in class discovery. *Sci. Rep.* 4:6207. doi: 10.1038/srep06207