Research article

Elevated expression of the MYB proto-oncogene like 2 (MYBL2)-encoding gene as a prognostic and predictive biomarker in human cancers

Zekun Xin[1,†], Yang Li[2,†], Lingyin Meng[3,†], Lijun Dong[1], Jing Ren[2,*] and Jianlong Men[2,*]

[1] Urology Surgery, Hebei Petro China Central Hospital, Langfang 065000, China
[2] Precision Medicine Center, Tianjin Medical University General Hospital, Tianjin 300000, China
[3] Central Laboratory, Linyi People's Hospital, Linyi 276000, China
[a] Zekun Xin, Yang Li and Lingyin Meng contributed equally to this work.

* Correspondence: Email: renjing_tmu@163.com, menjianlong@163.com.

† These authors contributed equally to this work.

Abstract: Recently, MYBL2 is frequently found to be overexpressed and associated with poor patient outcome in breast cancer, colorectal cancer, bladder carcinoma, hepatocellular carcinoma, neuroblastoma and acute myeloid leukemia. In view of the fact that there is an association between MYBL2 expression and the clinicopathological features of human cancers, most studies reported so far are limited in their sample size, tissue type and discrete outcomes. Furthermore, we need to verify which additional cancer entities are also affected by MYBL2 deregulation and which patients could specifically benefit from using MYBL2 as a biomarker or therapeutic target. We characterized the up-regulated expression level of MYBL2 in a large variety of human cancer via TCGA and oncomine database. Subsequently, we verified the elevated MYBL2 expression effect on clinical outcome using various databases. Then, we investigate the potential pathway in which MYBL2 may participate in and find 4 TFs (transcript factors) that may regulate MYBL2 expression using bioinformatic methods. At last, we confirmed elevated MYBL2 expression can be useful as a biomarker and potential therapeutic target of poor patient prognosis in a large variety of human cancers. Additionally, we find E2F1, E2F2, E2F7 and ZNF659 could interact with MYBL2 promotor directly or indirectly, indicating the four TFs may be the upstream regulator of MYBL2. TP53 mutation or TP53 signaling altered may lead to elevated MYBL2 expression. Our findings indicate that elevated MYBL2 expression represents a prognostic biomarker for a large number of cancers. What's more, patients with both P53 mutation and elevated MTBL2 expression showed a worse survival in PRAD and BRCA.

1. Introduction

MYBL2 (alias B-myb), a highly conserved member of the Myb transcription factor family, is ubiquitously expressed in proliferating cells, particularly in embryonic stem cells and adult haematopoietic precursors [1−3]. Previous studies have shown that the DREAM complex (DP, RB-like, E2F, and MuvB), which is critical for coordinating cell cycle-dependent gene expression, represses most cell cycle genes expression during quiescence [4]. Recently, it has been shown that the MuvB core dissociates from P107/P130, sequentially recruits MYBL2 and FoxM1 to coordinate the expression of the late cell cycle G2/M genes. The expression of MYBL2 involved in cell cycle regulation and performed essential functions in proliferating cells, is hardly detectable at G0 phase and induced at the G1/S transition of the cell cycle [5,6]. Taking those considerations together, these data indicate that MYBL2 is involved in cell proliferation and carcinogenesis.

So far, MYBL2 is frequently found to be overexpressed and associated with poor patient outcome in breast cancer, colorectal cancer, bladder carcinoma, hepatocellular carcinoma, neuroblastoma and acute myeloid leukemia [7−12]. In view of the fact that there is an association between MYBL2 expression and the clinicopathological features of human cancers, most studies reported so far are limited in their sample size, tissue type and discrete outcomes. Furthermore, which additional cancer entities are also affected by MYBL2 deregulation and which patients could specifically benefit from using MYBL2 as a biomarker or therapeutic target? In the present study, we have investigated whether elevated expression of MYBL2 could be used as a prognostic and predictive factor in a variety of human cancers. To this end, we used publicly available clinical information to generate a comprehensive molecular profile of MYBL2 in human cancers at the genomic and transcriptomic levels. Our findings indicate that elevated MYBL2 expression represents a prognostic biomarker for a large number of cancers. What's more, patients with both P53 mutation and elevated MYBL2 expression showed a worse survival in PRAD and BRCA.

2. Material and methods

2.1. Catalogue of somatic mutations in cancer (COSMIC) database analysis

The COSMIC database is an online resource for exploring the impact of somatic mutations [13], gene fusions, genomic rearrangements, and copy number variations in human cancers. Based on this authoritative database, we performed a summary of alterations affecting MYBL2. All data were extracted on November 19, 2021 (COSMIC v95 version).

2.2. Oncomine database analysis

Oncomine is a cancer transcriptomic database and online discovery platform with genome-wide expression analyses of 715 datasets including 86,733 samples (Oncomine v4.5 version) [14]. Gene expression analysis for a single gene or a set of genes can be conducted across various types of cancer and include comparisons relative to normal tissues, other cancer subtypes and various clinicopathological features.

2.3. Xena database analysis

The gene expression of MYBL2 in human cancer was downloaded from TOIL (scalable and efficient workflow engine) in Xena public data hubs (https://xena.ucsc.edu/public-hubs/) [15]. We used the RSEM tpm normalized expression to perform further analysis. The clinical traits and survival data was also downloaded.

2.4. cBio cancer genomics portal (cBioPortal) analysis

The cBioPortal is an open-access resource for interactive exploration of multidimensional cancer genomics data sets. Among these datasets [16,17], we used TCGA provisional data about prostate, breast and lung carcinoma, including clinical traits and TP53 signalling pathway alteration. We also downloaded the genes correlated with MYBL2 in prostate, breast and lung cancer for further analysis. The annotation of the TCGA cancer type is summarized in Table S1.

2.5. PrognoScan database analysis

PrognoScan is a database for meta-analysis of the prognostic value of genes [18]. We used the database to assess the biological relationship between MYBL2 gene expression and prognosis. The meta-analysis was also downloaded.

2.6. Kaplan-Meier plotter analysis

The Kaplan-Meier plotter is a database that can be used to assess the effect of 54,675 genes on survival using 18,647 cancer samples (breast, ovarian, lung and gastric cancer) [19]. Data were obtained before November 19, 2021. Patients with higher and lower expression divided by median of MYBL2 (Probe ID: 201710_at, Jetset best probe) were analyzed the 5-years survival using the log-rank test. The hazard ratio with 95% confidence intervals and log-rank p values were noted. In this database, there is only 201710_at linked to MYBL2. The database has not affy U95 platform associated probes. Indeed, there are a lot of RNA-seq data for AML/COAD, but we have not found RNA-seq data for both normal samples and tumor samples.

2.7. GO and KEGG analysis

GOstats is a bioconductor package written in R [20], that allows users to test GO terms for over or under-representation using either a classical hypergeometric test or a conditional hypergeometric that uses the relationships among GO terms to decorrelate the results. We overlapped the genes correlated with MYBL2 among prostate, breast and lung cancer from cBioportal, and the overlapped genes conducted GO and KEGG analysis used the GOstats packages. The cutoff p-value below 0.05 was selected.

2.8. Candidate transcription factors analysis

We first get 154 genes that are co-expressed with MYBL2. Then we obtain the transcription factor list from HumanTFDB [21], and take the intersection with the co-expressed genes. After that, we get the position frequency matrix of the relevant transcription factor through the R package Motifdb [22]. Lastly, we use the matchPWM function from the R Biostrings package [23] to get the relevant binding site on the promoter region of MYBL2. The more detail method can refer to gene regulation workflow (https://www.bioconductor.org/help/workflows/generegulation/).

2.9. Statistical analysis

For Oncomine analysis, we used the parameter of FC > 1.5, pvalue < 0.01 and gene rank (all) for two class differential expression analyses (e.g., cancer tissues versus normal tissues). For the RNA-seq analysis in TCGA and GTEx samples, the Mann-Whitney U test was used for two class differential expression analyses. With the Kaplan-Meier plotter and TCGA survival data, the log-rank test and uni-variable COX analysis was used to analyze the data. With the PrognoScan analysis, the data was filtered by pvalue below 0.05. For the correlated genes with MYBL2, the data was filter by absolute Spearman correlation upper 0.5 and the pvalue below 0.05. In the GO and KEGG analysis, the pvalue below 0.05 was considered significant. For the relationship between MYBL2 and TP53 signialling pathway, the permutation test and Mann-Whitney U test was used.

3. Result

3.1. MYBL2 DNA mutations

Table 1. The number and percentage of the genetic alteration.

Genetic alteration	Number	Percentage (%)
Substitution-coding silent	73	0.20
Substitution-missense	178	0.49
Substitution-nonsense	8	0.02
Deletion-frameshift	1	0.00
Insertion-frameshift	3	0.01
Gain	871	2.38
Loss	169	0.46
Unknown	11	0.03

The MYBL2 gene in cancers of patients was assessed for mutations using the COSMIC database (COSMIC v95 version) [13], a comprehensive resource for exploring somatic mutations in human cancer. Before November 19, 2021, the MYBL2 gene was tested in 36,635 patient specimens in 38 different types of cancer. There were no mutations in 18 of these cancer types and 263 point mutations were found (overall frequency = 0.72%) in the 20 remaining types of cancer; 1040 copy number variations (CNV) were found (overall frequency = 2.84%) in 22 of the cancer types (Table 1). Out of the 263 point mutations in the MYBL2 gene, 8 were nonsense and 178 were missense; 76 mutations were coding silent (synonymous mutation) and the other 4 were frame shift (Table 1). The frequency of point mutations was not high in most of the cancer types, with the highest mutation frequency being 3.8%, found in cancer of small intestine (Table S2). As shown in that table, the

highest CNVs were found in upper pancreas cancer, breast cancer and large intestine cancer, account for 1.5% of all samples, CNVs of other cancer account were much little. Taken together, the results indicate that there were no major alterations in the sequence or copy number of the MYBL2 gene that could account for the development of the malignancies.
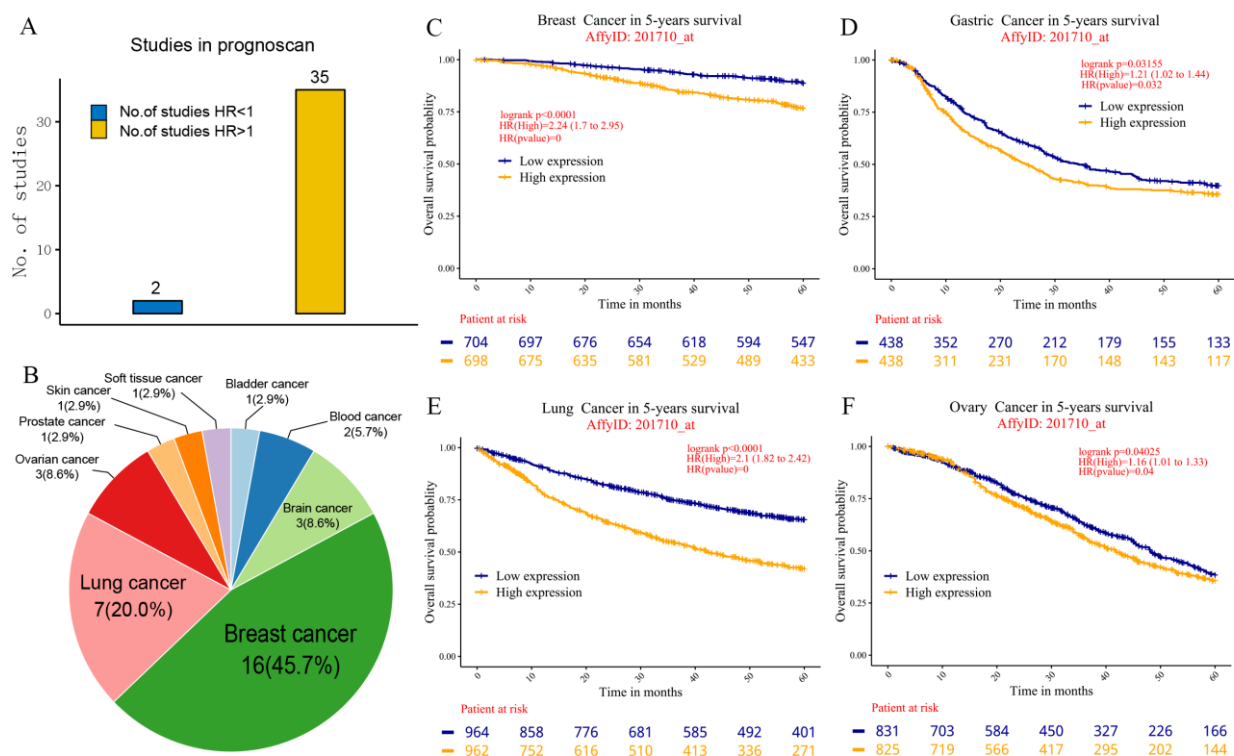
3.2. Elevated MYBL2 expression in human cancers
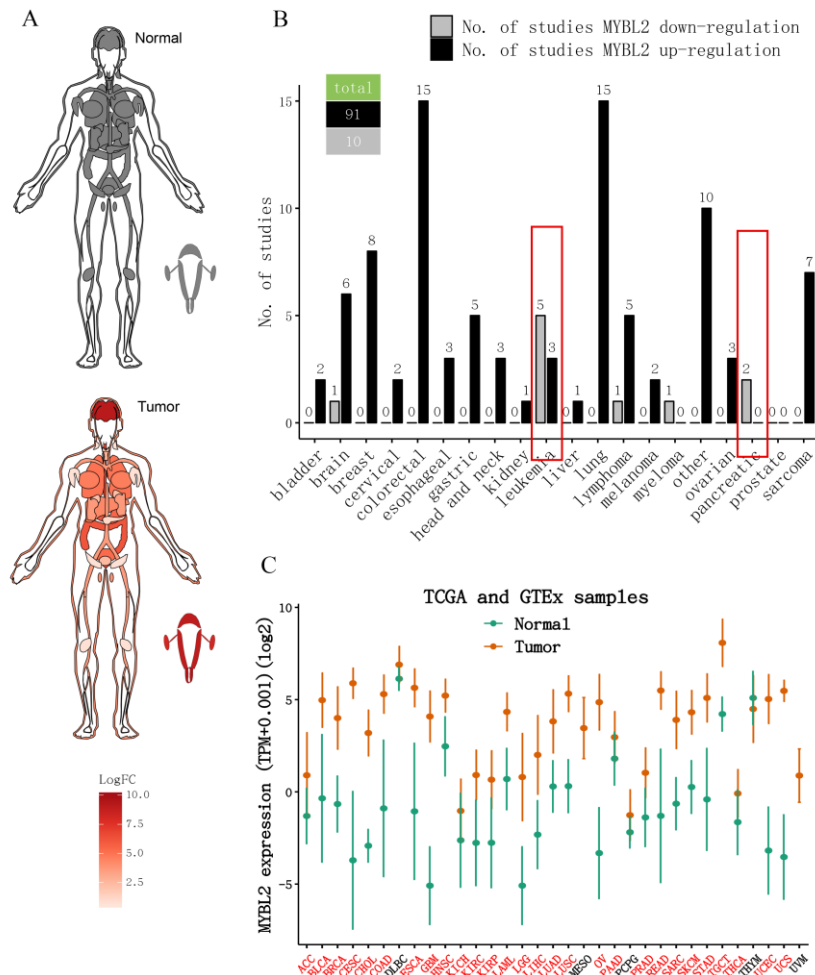
Figure 1. MYBL2 expression is significantly un-regulated in cancer tissues and normal tissues as indicated by Oncomine and TCGA dataset analysis. (A) The number of studies that significantly correlated with MYBL2 up-regulation and down-regulation in cancer tissues versus normal tissues, is shown at different cancer types by Oncomine analysis (fold change > 2, $p < 0.01$). Black columns stand for MYBL2 up-regulated studies (91 studies), white columns stand for MYBL2 down-regulated studies (5 studies). (B-C) The gene expression profile of 33 different cancer tissues and paired normal tissues is shown by TCGA and GTEx dataset analysis. Brown error-bars stand for the cancer tissues, green error-bars stand for the normal tissues. The red x-axis tick titles stand for the cancer types of significantly up-regulated MYBL2 expression. The black x-axis tick titles stand for no significant change of MYBL2 in cancer tissues versus normal tissues. The abbreviations of x-axis tick title refer to Table S1.

Using Oncomine analysis, 8 we investigated whether expression of the MYBL2 gene was

altered in human cancers. It was found that MYBL2 was elevated in 91 studies and down-regulated in 10 studies (Figure 1A; fold change > 2, p < 0.01). Furthermore, 5 of the down-regulated studies clustered in leukemia and pancreatic cancer. The most common cancer types showing elevated MYBL2 expression in these studies were cancers of the lung, colorectal, breast, sarcoma and brain cancer. For further validate whether MYBL2 expression was elevated in other cancer, we used the TCGA tumor and GTEx normal sample to perform the differential expression analysis between normal and tumor tissue using Mann-Whitney U test (Figure 1B,C and Table S3) [15]. The study involved 15408 samples, with logFC range from 0.76 to 9.6, and the pvalue range from 1.80E-02 to 1.74E-158. Together, the results indicate that elevated MYBL2 expression is a diagnosis marker between normal and tumor tissue for most of human cancers.

## 3.3. Elevated MYBL2 expression effect on clinical outcome

Figure 2. Elevated MYBL2 expression is significantly correlated with poor patient outcome as indicated by PrognoScan and Kaplan-Meier plotter analysis. (A-B) As indicated by PrognoScan analysis, the most studies with elevated MYBL2 show poor prognosis, the threshold was set as cox p-value < 0.05. The pie plot shows studies distribution of poor prognosis (HR > 1). HR stands for hazard ratio. (C-F) As shown by Kaplan-Meier plotter analysis, the breast, lung, ovarian and gastric cancer patients with higher expression of MYBL2 show poor overall survival.
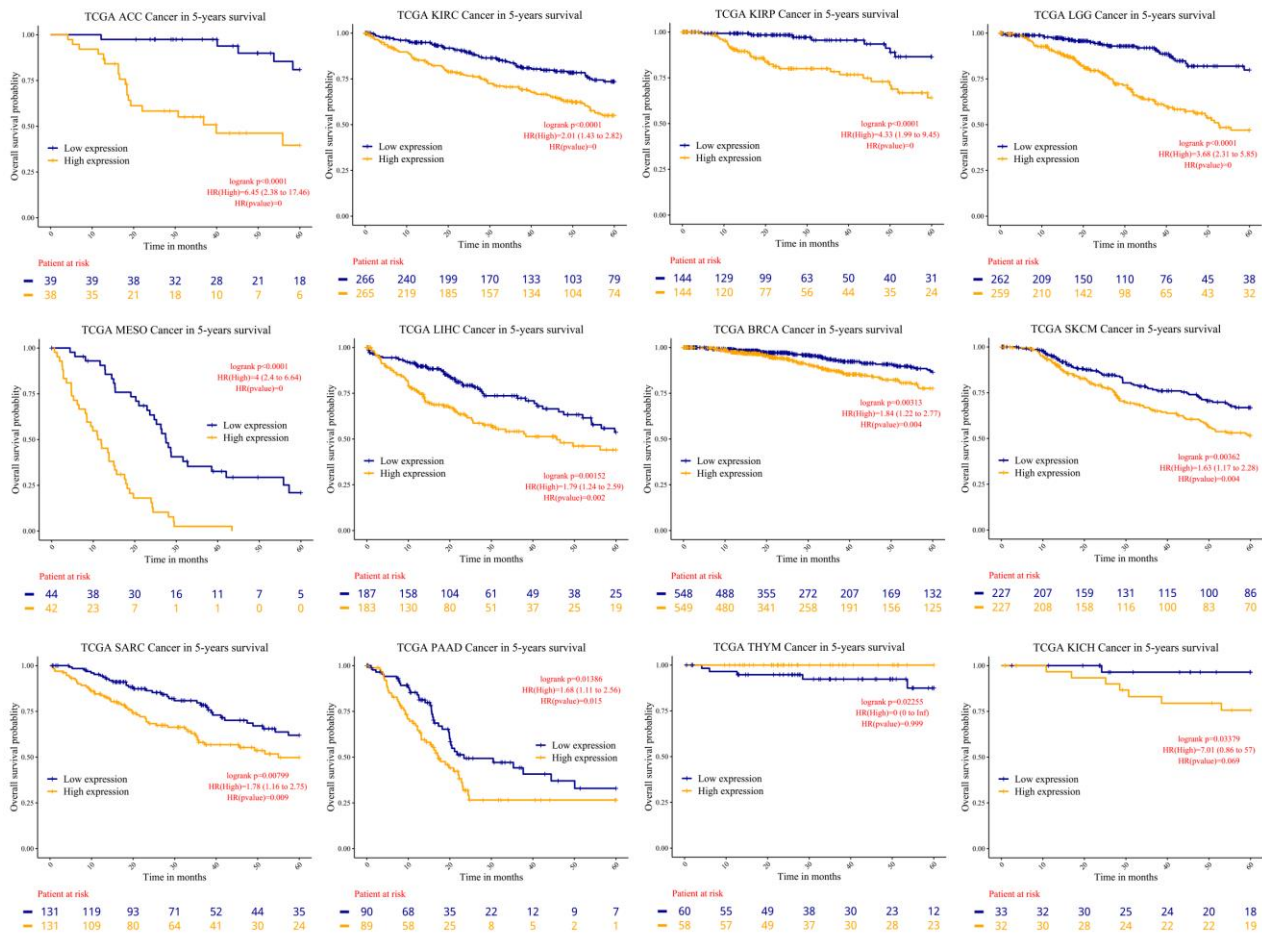
Figure 3. Elevated MYBL2 expression significantly correlated with lower 5-years overall survival of ACC, KIRC, KIRP, LGG, MESO, LIHC, BRCA, SKCM, SARC, PAAD, THYM, KICH.

Next, we wondered whether elevated MYBL2 expression had an effect on clinical outcome. Using the Prognoscan database [18], we noticed that elevated MYBL2 expression was associated with higher hazard ratio in 35 studies, which indicated patients with higher MYBL2 expression had poor prognosis (Figure 2A,B and Table S4). While in two studies elevated MYBL2 expression was associated with lower hazard ratio, considering the duplicated probe (216421_at, 201710_at) and the cox-pvalue, the probe (216421_at) made these studies less definitive. Then, we focused on the relationship between elevated MYBL2 expression and overall survival in 5-years using a Kaplan-Meier plotter [19]. Elevated MYBL2 expression significantly correlated with lower 5-years overall survival of breast, lung, gastric and ovarian cancer (Figure 2C-F and Table S5). We also validated the relationship between elevated MYBL2 expression and clinical outcome using 5-years OS (overall survival), 5-years DSS (disease-specific survival) and 5-years PFI (progressive-free interval) of TCGA data [15]. We found elevated MYBL2 expression significantly correlated with lower 5-years OS of ACC, KIRC, KIRP, LGG, MESO, LIHC, BRCA, SKCM, SARC, PAAD, THYM, KICH (Figure 3 and Table S6). Furthermore, elevated MYBL2 expression was significantly correlated with lower 5-years DSS of ACC, KIRC, KIRP, LGG, MESO, BRCA, LIHC, SKCM, KICH, PAAD, LUAD, UVM and SARC (Figure 4 and Table S7). Lastly, we found elevated MYBL2 expression

was significantly correlated with lower 5-years PFI of ACC, KIRC, KIRP, PRAD, MESO, LGG, LIHC, UVM, KICH, THCA, PAAD, BRCA, SARC, PCPG and TGCT (Figure 5 and Table S8). Noticeably, while the log-rank p value of some cancer types was no significant, elevated MYBL2 expression may still have effect on 5-years OS, 5-years DSS, 5-years PFI (Figures 3−5 and Tables S6−S8). Together, the results indicate that elevated MYBL2 expression is prognostic of poor clinical outcome for a large variety of human cancers. In Figures 1−5, high MYBL2 expression means that the gene expression of MYBL2 is high, which is significantly correlated with poor patient outcome in numerous cancer entities. Low MYBL2 expression denotes that the gene expression amount of MYBL2 is low. Generally, MYBL2 expression is higher than other normal tissues in almost all cancers. Therefore, MYBL2 may become a potential diagnostic and prognostic marker in the cancer.
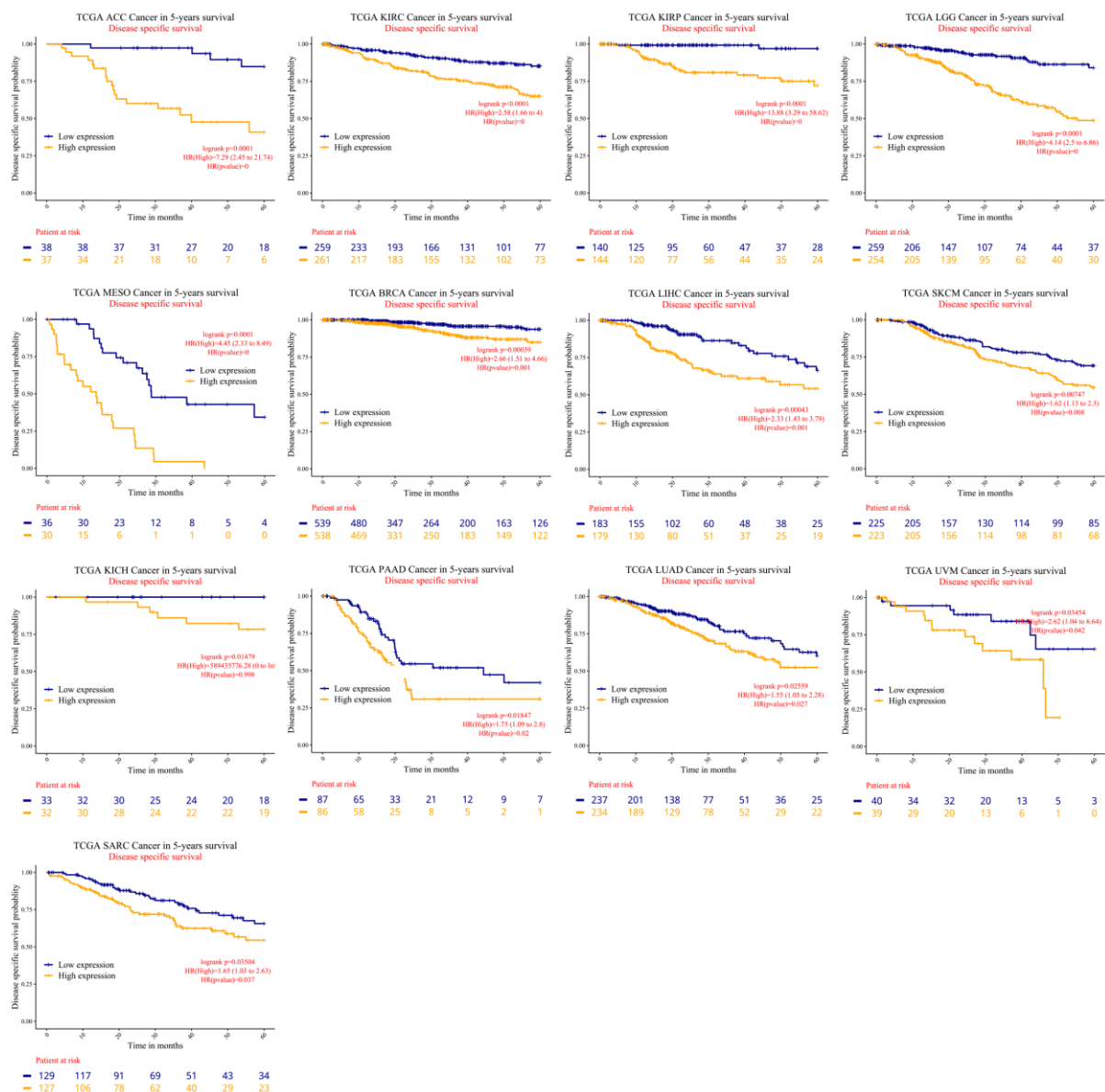
Figure 4. Elevated MYBL2 expression significantly correlated with lower 5-years disease-specific survival of ACC, KIRC, KIRP, LGG, MESO, BRCA, LIHC, SKCM, KICH, PAAD, LUAD, UVM and SARC.
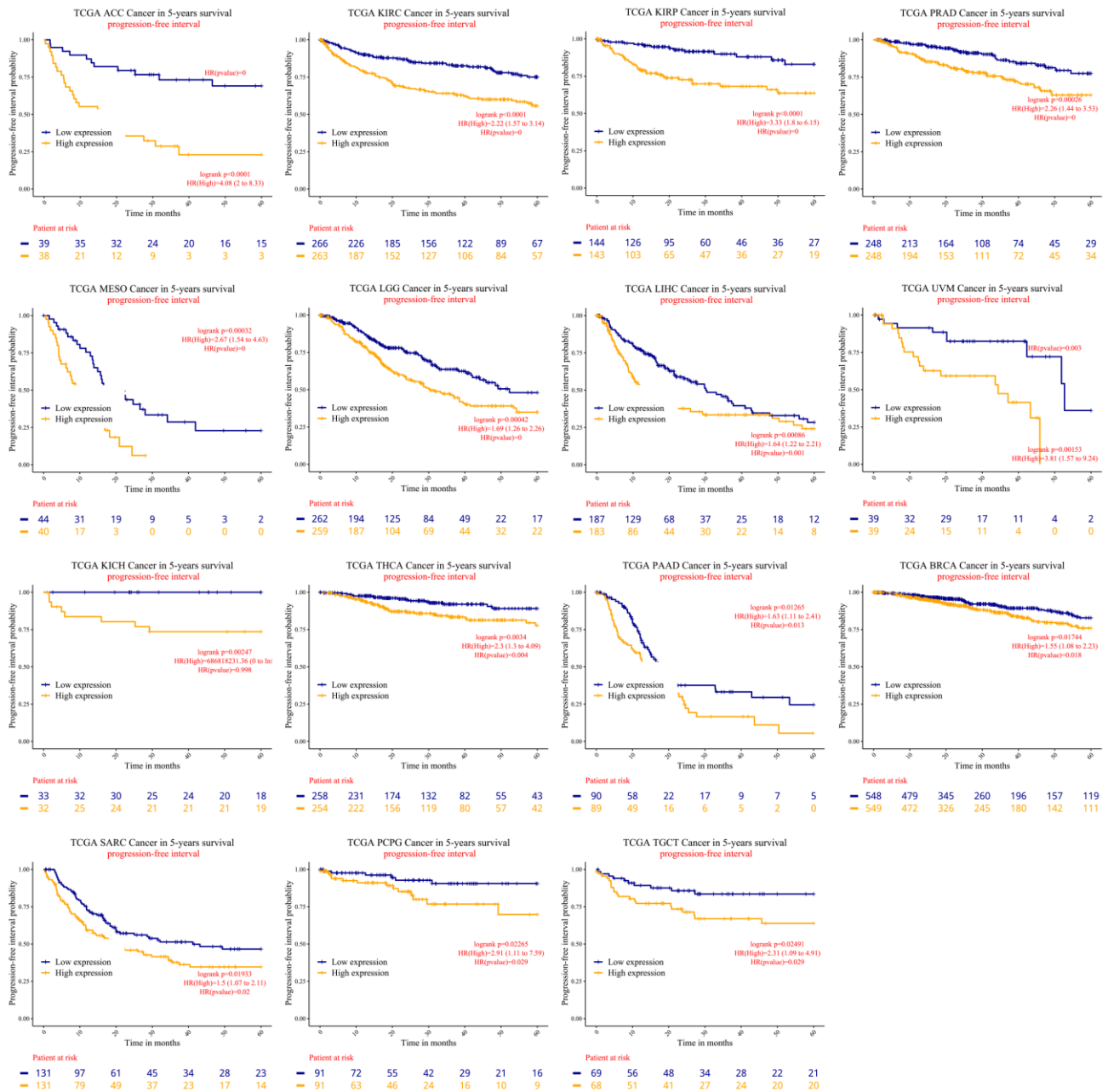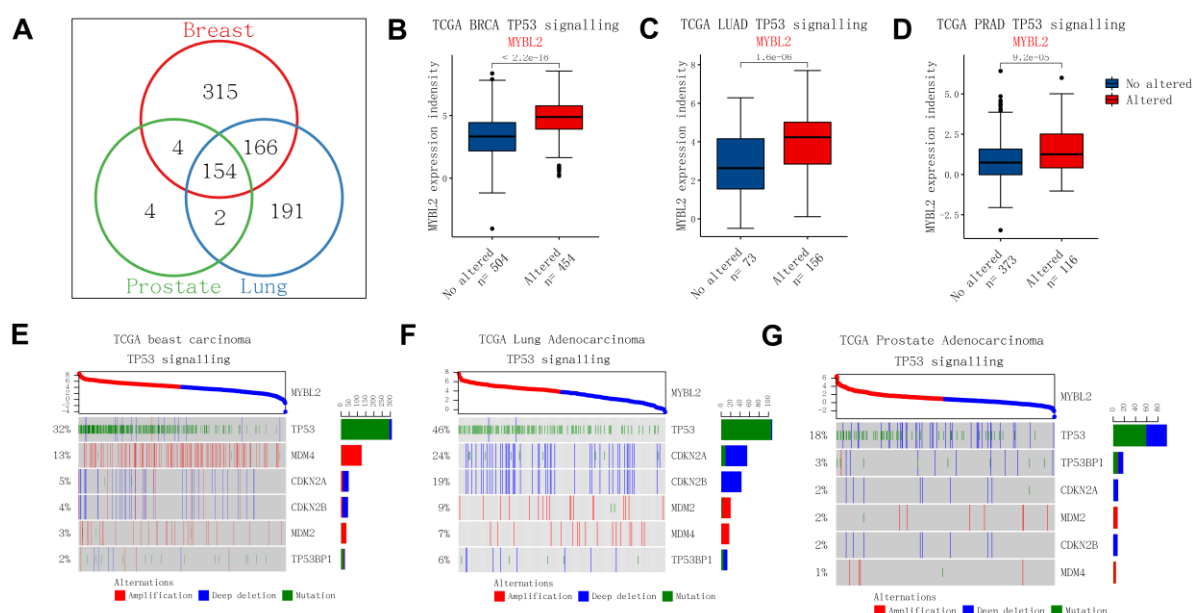
Figure 5. Elevated MYBL2 expression was significantly correlated with lower 5-years progressive-free interval of ACC, KIRC, KIRP, PRAD, MESO, LGG, LIHC, UVM, KICH, THCA, PAAD, BRCA, SARC, PCPG and TGCT.

3.4. Genes co-expressed with MYBL2 and potential MYBL2 regulatory mechanisms

According to the newest incidence of cancer [24], we selected breast, prostate and lung cancer to perform further investigating. To identify the potential function of elevated MYBL2 expression, we searched for genes co-expressed with MYBL2 in breast, prostate and lung cancer (TCGA provisional) from cBioportal database [16,17]. We identified 441, 380 and 164 genes with a co-expression score ≥ 0.5 or ≤ -0.5 and got 154 gene common in the three data-sets (Figure 6A). Using

GOstats R package to perform GO and KEGG analysis, the GO terms primarily related to cell process and cell cycle (Table 2) [20] and the KEGG terms primarily related to the cell cycle, p53 signaling pathway and some types of cancer pathway (Table 3). Furthermore, we investigated the potential relationship between the elevated MYBL2 expression and p53 signaling pathway. The mutation and copy number variant data of TP53 signaling pathway were downloaded from cBioportal database. We plotted the oncoprint according the rank of MYBL2 expression and observed that most of the TP53 mutant patients are accompanied by elevated MYBL2 expression (Figure 6E-G). There are 29.22% of the PRAD samples in MYBL2 high expression group had TP53 signaling pathway altered, compared with 18.29% of the PRAD samples in low group (Figure 6E). The similar results of BRCA and LUAD were 77.45% in high group and 28.6% in low group, 79.82% in high group and 56.52% in low group (Figure 6F-G and Table S9). Simultaneously, the box-plot showed that the TP53 altered group had higher MYBL2 expression (Figure 6B-D). We concluded that TP53 mutation or TP53 signaling altered may lead to elevated expression of MYBL2. Lastly, we divided the samples into four groups based on the presence or absence of TP53 mutations and the median expression of MYBL2. We found low and no-altered group had the better prognosis compared with high and altered group in PRAD and BRCA (Figure 7).

Figure 6. Elevated MYBL2 expression is significantly correlated with p53 signaling alter as indicated by cBioPortal analysis. (A) The venn plot shows the common genes correlated with MYBL2 expression in breast, lung and prostate cancer. The 154 genes can be enriched to p53 pathway by KEGG analysis. (B-D) The group of p53 pathway alter has higher MYBL2 expression compared with its counterpart in breast, lung and prostate cancer. (E-G) The oncoprint plot according the rank of MYBL2 expression show that most of the TP53 mutant patients are accompanied by elevated MYBL2 expression.
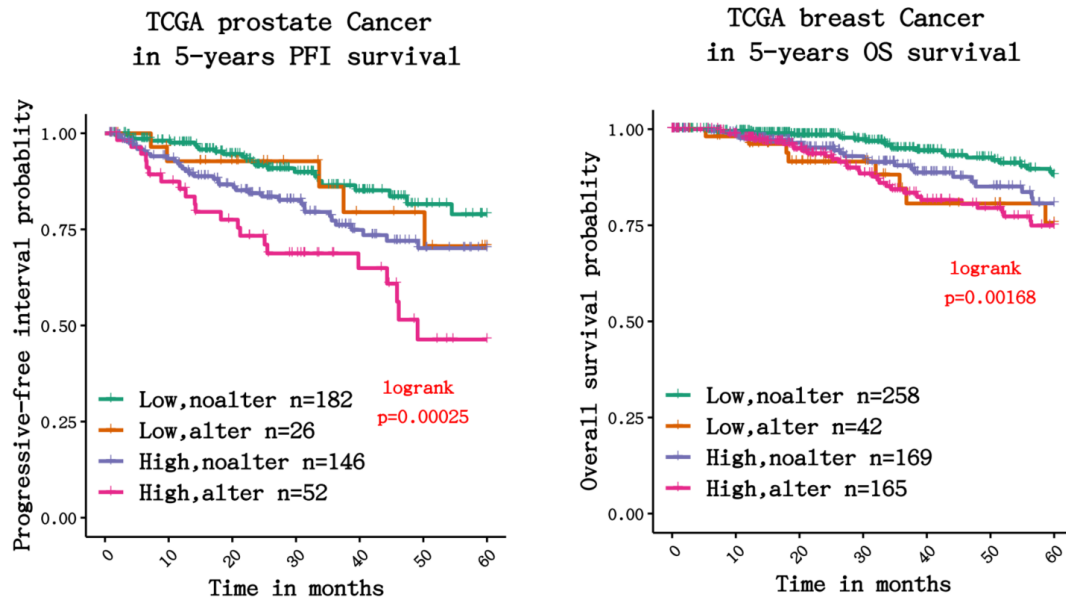
Figure 7. Low and no-altered group had a better prognosis compared with high and altered group in PRAD and BRCA.
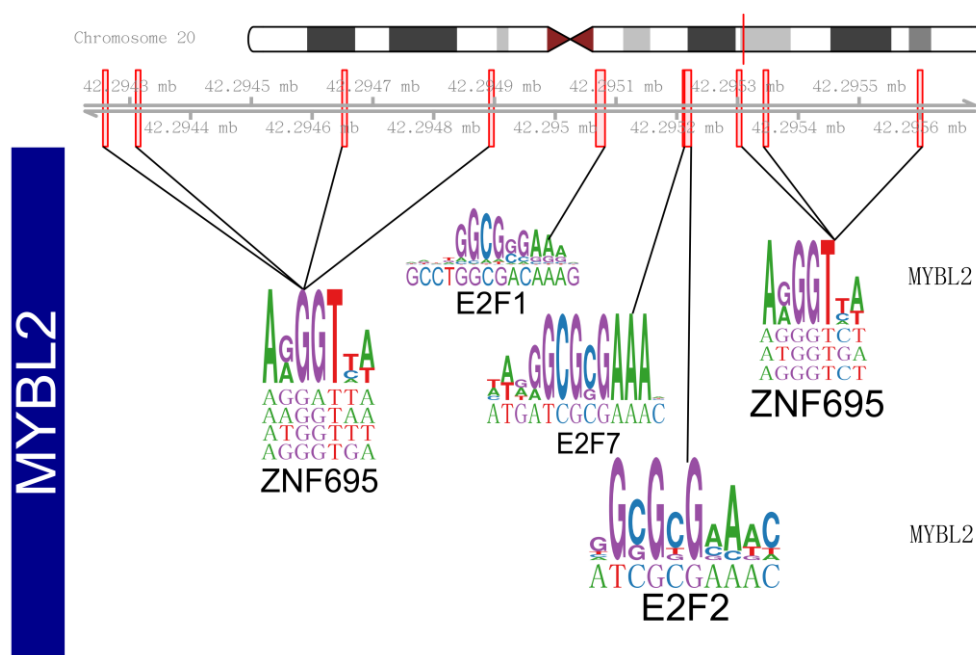
Figure 8. The TFs (transcription factors) from the 154 co-expressed genes may bind to the 1500bp upstream of MYBL2. The red columns stand for the binding sites. The upper of TF gene stand for the motif. The lower of TF gene stand for the matched sequence of 1500bp upstream of MYBL2.

To explore the upstream regulator of MYBL2, we selected the correlated TF (transcription factors) from the 154 co-expressed genes, and used their motif sequence to match the 1500bp sequence of MYBL2 at the upstream of TSS (transcription start site). We observed that E2F1, E2F2, E2F7 and

ZNF659 could interact with MYBL2 promotor directly or indirectly, indicating the four TFs may be the upstream regulator of MYBL2 (Figure 8 and Table S10). In the cancer cells from the public database Cristrome [25], we find the chip-seq data of E2F1 and E2F7. Then, we show the actual motif occupancy of E2F1 and E2F7 on the promoter region of MYBL2. The result is summarized in Figure S1. As observed from Figure S1, it can be found that E2F1 and E2F7 have binding sites in the MYBL2 promoter region, which further verified our analysis that E2F1 and E2F7 can regulate MYBL2.

Table 2. GO terms.

Term	Count	Pvalue
cell cycle	117	1.83E-92
mitotic cell cycle	91	2.03E-80
mitotic cell cycle process	87	2.06E-80
cell cycle process	98	9.16E-78
cell division	67	1.96E-61
nuclear division	58	1.35E-57
chromosome segregation	53	4.57E-57
organelle fission	59	1.43E-56
mitotic nuclear division	49	1.90E-53
sister chromatid segregation	43	3.39E-52
nuclear chromosome segregation	46	3.38E-50
mitotic cell cycle phase transition	56	2.98E-49
cell cycle phase transition	56	3.05E-47

Table 3. KEGG terms.

Term	Count	Pvalue
Cell cycle	25	5.10E-31
Oocyte meiosis	14	5.36E-14
Progesterone-mediated oocyte maturation	9	1.44E-08
DNA replication	6	2.70E-07
p53 signaling pathway	6	1.26E-05
Homologous recombination	4	5.90E-05
Base excision repair	3	0.002085686
Mismatch repair	2	0.013694815
Pancreatic cancer	3	0.017106232
Small cell lung cancer	3	0.02846077
Prostate cancer	3	0.032025374
Pyrimidine metabolism	3	0.041914498
Bladder cancer	2	0.042434636

3.5. Elevated MYBL2 expression as a potential predictive biomarker in prostate, breast and lung cancer

To investigate whether elevated MYBL2 expression in prostate, lung and breast cancer could be used to predict the prognosis, we used Gleason score of prostate cancer, tipple negative status of breast cancer, and adenocarcinoma or not of lung cancer. The result showed that unregulated MYBL2 expression may be used as a predictive biomarker for poor clinical trait of prostate and breast cancer, and for distinguished the adenocarcinoma and squamous carcinoma of lung (Figure 9A-C). Furthermore, we performed multivariate analysis. The analysis result is summarized in Figure S2A,C. From those figures, we found MYBL2 can be a significant prognostic marker in BRCA (HR = 1.2, p < 0.01) and PRAD (HR = 1.2, p < 0.05). We also compare the prognostic value of MYBL2 expression with Gleason score in prostate cancer and hormonal receptor expression in breast cancer, which is summarized in Figure S2B, D. From those figures, we can see that in BRCA, the TNBC group and the Her2 group with high MYBL2 expression have a worse prognosis; while in PRAD, the high Gleason group with high MYBL2 expression have a worse prognosis.
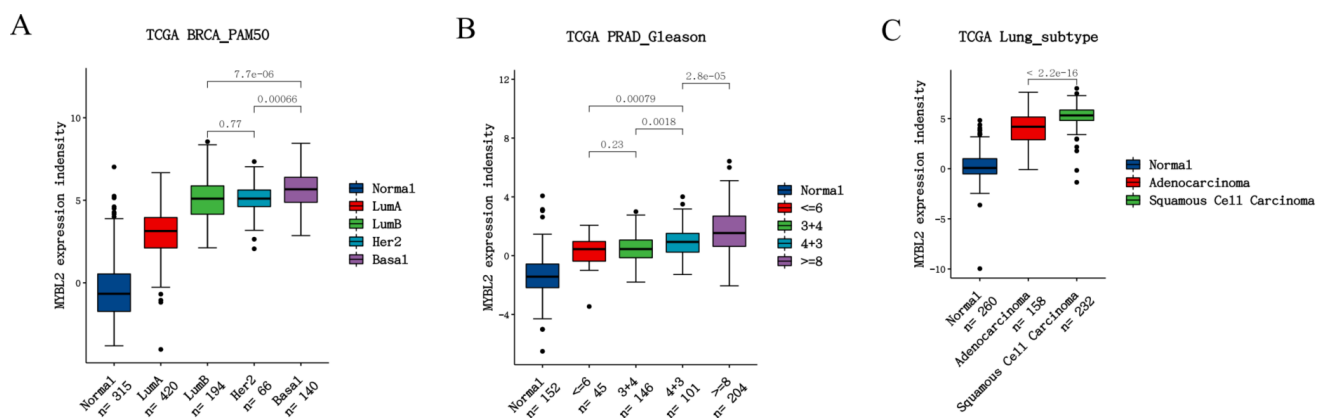
Figure 9. Elevated MYBL2 expression is significantly correlated with clinical traits. (A) Elevated MYBL2 expression is significantly correlated with PAM50 subtypes of breast cancer. (B) Elevated MYBL2 expression is significantly correlated with Gleason score of prostate cancer. (C) Elevated MYBL2 expression may be a bio-marker to distinguish the lung adenocarcinoma and squamous cell carcinoma.

4. Discussion

While elevated levels of MYBL2 have been reported in many cancer [26−29], the present study, based on TCGA and Oncomine analysis, shows that elevated MYBL2 expression can be found in almost 29 cancer types. The study involved 15408 samples, with logFC range from 0.76 to 9.6 and the pvalue range from 1.80E-02 to 1.74E-158. While the pvalue was no significant in Thymoma and Pheochromocytoma, the relatively low number of patients in normal tissue made this finding less definitive. Based on prognoscan and Kaplan-Meier plotter database [18,19], we found the unregulated MYBL2 expression is correlated poor prognosis (HR > 1) in bladder, blood, brain, breast, lung, prostate, skin, ovarian, gastric and soft tissue cancer. We further used TCGA data to analyze the OS, DSS and PFI in 33 cancer types and observed the same result. Taken together, the data showed that the unregulated MYBL2 expression not only could be used as a prognostic biomarker of poor survival outcome in breast cancer, colorectal cancer, bladder carcinoma, hepatocellular carcinoma,

neuroblastoma and acute myeloid leukemia [30], but also in adrenocortical, kidney, lung, skin, pancreatic, thyroid, brain, mesothelioma, uveal melanoma, prostate, testicular, sarcoma, uveal melanoma and paraganglioma.

Based on the newest incidence of cancer [24], we selected breast, prostate and lung cancer to further investigate the mechanism (Figure S3). We used the 154 genes, which were co-expressed with MYBL2, to perform GO and KEGG analysis using the bioconductor package GOstats [20]. We found the major KEGG terms were p53 signaling, cell cycle and other cancer pathway, and the major GO terms were cell process and cell cycle. The GO and KEGG analysis indicated the elevated expression of MYBL2 may associate with p53 signaling and cell cycle related genes. This is in accordance with the results from Fischer et al. showing that the p53-p21-DREAM-CDE/CHR pathway represses MYBL2 expression [31]. We further plotted the oncoprint according the rank of MYBL2 expression and observed that most of the TP53 mutant patients are accompanied by elevated MYBL2 expression. Lastly, we divided the samples into four groups based on the presence or absence of TP53 mutations and the median expression of MYBL2. We found in low and no-altered group had the best prognosis compared with the worst prognosis of high and altered group in PRAD and BRCA. These data further elucidated the MYBL2 related mechanism and identified patients could specifically benefit from using MYBL2 as a biomarker.

In order to find the potential upstream of MYBL2, we selected the 6 TF co-expressed with MYBL2 from the 154 genes and used the bioconductor packages (https://www.bioconductor.org/help/workflows/ generegulation) to find the potential binding site at the 1500bp sequence of MYBL2 at the upstream of TSS [32]. As a result, only E2F1, E2F2, E2F7 and ZNF695 can find the matched sequence. In accordance with this, previous studies have shown that E2F1 and E2F2 could transactivate MYBL2 and thus regulated cell cycle [4]. All these data suggested that E2F7 and ZNF695 may be alternative upstream regulator of MYBL2.

In addition, to investigate whether the elevated MYBL2 expression in prostate, lung cancer and breast cancer could be used as predictive marker, we used Gleason score of prostate cancer, tipple negative status of breast cancer and adenocarcinoma or not of lung cancer. And the data showed that the elevated MYBL2 expression was correlated with high gleason score of prostate cancer, correlated with tipple negative status of breast cancer and correlated with squamous lung cancer.

In conclusion, the present study indicates that elevated expression of MYBL2 can be used as a biomarker of poor patient prognosis in a large variety of human cancers. Patients with low MYLB2 expression and no-TP53 altered had the best prognosis compared to the worst prognosis of patients with high MYLB2 expression and TP53 altered in PRAD and BRCA. In our future work, we will collect clinical samples from different cancers, such as prostate, breast, and lung cancer. It can be noted that artificial intelligence techniques can also be employed in the further analysis for those cancers [33−36]. Then, we employ the FFPE tissues from those cancers on the protein level to explore the effect of the protein expression by IHC in the clinic. Moreover, we will further verify the regulatory relationship between p35, E2F2, ZNF659 and MYBL2 in cancer cells using the clinical samples experimentally. The detailed mechanism that mediates the poor prognosis and the study about the downstream targets of MYBL2 can also be treated as one of the research directions.

Abbreviations

TCGA	The Cancer Genome Atlas
COSMIC	Catalogue of somatic mutations in cancer
cBioPortal	cBio cancer genomics portal
OS	Overall survival
DSS	Disease-specific survival
PFI	Progressive-free interval
ACC	Adrenocortical carcinoma
BLCA	Bladder Urothelial Carcinoma
BRCA	Breast invasive carcinoma
CESC	Cervical squamous cell carcinoma and endocervical adenocarcinoma
CHOL	Cholangio carcinoma
COAD	Colon adenocarcinoma
DLBC	Lymphoid Neoplasm Diffuse Large B-cell Lymphoma
ESCA	Esophageal carcinoma
GBM	Glioblastoma multiforme
HNSC	Head and Neck squamous cell carcinoma
KICH	Kidney Chromophobe
KIRC	Kidney renal clear cell carcinoma
KIRP	Kidney renal papillary cell carcinoma
LAML	Acute Myeloid Leukemia
LGG	Brain Lower Grade Glioma
LIHC	Liver hepatocellular carcinoma
LUAD	Lung adenocarcinoma
LUSC	Lung squamous cell carcinoma
MESO	Mesothelioma
OV	Ovarian serous cystadenocarcinoma
PAAD	Pancreatic adenocarcinoma
PCPG	Pheochromocytoma and Paraganglioma
PRAD	Prostate adenocarcinoma
READ	Rectum adenocarcinoma
SARC	Sarcoma
SKCM	Skin Cutaneous Melanoma
STAD	Stomach adenocarcinoma
TGCT	Testicular Germ Cell Tumors
THCA	Thyroid carcinoma
THYM	Thymoma
UCEC	Uterine Corpus Endometrial Carcinoma

UCS   Uterine Carcinosarcoma
UVM   Uveal Melanoma

Conflict of interest

The authors declare no competing interests.

References

1. A. Sala, R. Watson, B-Myb protein in cellular proliferation, transcription control, and cancer: Latest developments, *J. Cell. Physiol.*, **3** (1999), 245−250. doi: 10.1002/(SICI)1097-4652(199906)179:3<245:AID-JCP1>3.0.CO;2-H.

2. M. Bessa, M. Joaquin, F. Tavner, M. K. Saville, R. J. Watson, Regulation of the cell cycle by B-Myb, *Blood Cells Mol. Dis.*, **2** (2001), 416−421. doi: 10.1006/bcmd.2001.0399.

3. K. V. Tarasov, Y. S. Tarasova, W. L. Tam, D. R. Riordon, S. T. Elliott, G. Kania, et al., B-MYB is essential for normal cell cycle progression and chromosomal stability of embryonic stem cells, *PloS one*, **6** (2008), e2478. doi: 10.1371/journal.pone.0002478.

4. S. Sadasivam, J. A. DeCaprio, The DREAM complex: master coordinator of cell cycle-dependent gene expression, *Nat. Rev. Cancer*, **8** (2013), 585−595. doi: 10.1038/nrc3556.

5. M. Joaquin, R. J. Watson, Cell cycle regulation by the B-Myb transcription factor, *Cell. Mol. Life Sci.*, **11** (2003), 2389-2401. doi: 10.1007/s00018-003-3037-4.

6. S. Sadasivam, S. Duan, J. A. DeCaprio, The MuvB complex sequentially recruits B-Myb and FoxM1 to promote mitotic gene expression, *Genes Dev.*, **5** (2012), 474−489. doi: 10.110r1/gad.181933.111.

7. R. Bayley, C. Ward, P. Garcia, MYBL2 amplification in breast cancer: Molecular mechanisms and therapeutic potential, *Biochim. Biophys. Acta Rev. Cancer*, **2020** (2020), 188407. doi: 10.1016/j.bbcan.2020.188407.

8. F. Ren, L. Wang, X. Shen, X. Xiao, Z. Liu, P. Wei, et al., MYBL2 is an independent prognostic marker that has tumor-promoting functions in colorectal cancer, *Am. J. Cancer Res.*, **4** (2015), 1542.

9. M. Zhang, H. Li, D. Zou, J. Gao, Ruguo key genes and tumor driving factors identification of bladder cancer based on the RNA-seq profile, *Onco Targets Ther.*, **9** (2016), 2717. doi: 10.2147/ott.s92529.

10. Z. Guan, W. Cheng, D. Huang, A. Wei, High MYBL2 expression and transcription regulatory activity is associated with poor overall survival in patients with hepatocellular carcinoma, *Curr. Res. Transl. Med.*, **1** (2018), 27−32. doi: 10.1016/j.retram.2017.11.002.

11. G. Raschellà, V. Cesi, R. Amendola, A. Negroni, B. Tanno, P. Altavista, et al., Expression of B-myb in neuroblastoma tumors is a poor prognostic factor independent from MYCN amplification, *Cancer Res.*, **14** (1999), 3365−3368.

12. O. Fuster, M. Llop, S. Dolz, P. Garcíab, E. Suchc, M. Ibáñez, et al., Adverse prognostic value of MYBL2 overexpression and association with microRNA-30 family in acute myeloid leukemia patients, *Leuk. Res.*, **12** (2013), 1690−1696. doi: 10.1016/j.leukres.2013.09.015.

13. S. A. Forbes, D. Beare, P. Gunasekaran, K. Leung, N. Bindal, H. Boutselakis, et al., COSMIC: exploring the world's knowledge of somatic mutations in human cancer, *Nucleic Acids Res.*, D**1**

(2015), D805−D811. doi: 10.1093/nar/gku1075.

14. D. R. Rhodes, S. Kalyana-Sundaram, V. Mahavisno, R. Varambally, J. Yu, B. B. Briggs, et al., Oncomine 3.0: genes, pathways, and networks in a collection of 18,000 cancer gene expression profiles, *Neoplasia*, **2** (2007), 166−180. doi: 10.1593/neo.07112.

15. J. Vivian, A. A. Rao, F. A. Nothaft, C. Ketchum, J. Armstrong, A. Novak, et al., Toil enables reproducible, open source, big biomedical data analyses, *Nat. Biotechnol.*, **4** (2017), 314−316. doi: 10.1038/nbt.3772.

16. J. Gao, B. A. Aksoy, U. G. Dogrusoz, Dresdner, B. Gross, S. O. Sumer, et al., Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal, *Sci. Signal.*, **269** (2013), pl1−pl1. doi: 10.1126/scisignal.2004088.

17. E. Cerami, J. Gao, U. Dogrusoz, B. E. Gross, S. O. Sumer, B. A. Aksoy, et al., The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data, *Cancer Discov*, **2** (2012). doi: 10.1158/2159-8290.CD-12-0095.

18. H. Mizuno, K. Kitada, K. Nakai, A. Sarai, PrognoScan: a new database for meta-analysis of the prognostic value of genes, *BMC Med. Genomics*, **1** (2009), 1−11. doi: 10.1186/1755-8794-2-18.

19. B. Györffy, A. Lanczky, A. C. Eklund, C. Denkert, J. Budczies, Q. Li, et al., An online survival analysis tool to rapidly assess the effect of 22,277 genes on breast cancer prognosis using microarray data of 1,809 patients, *Breast Cancer Res. Treat.*, **3** (2010), 725−731. doi: 10.1007/s10549-009-0674-9.

20. S, R. Falcon, R. Gentleman, Using GOstats to test gene lists for GO term association, *Bioinformatics*, **2** (2007), 257−258. doi: 10.1093/bioinformatics/btl567.

21. H. Hu, Y. R. Miao, L. H. Jia, Q. Y. Yu, Q. Zhang, A. Y. Guo, et al., AnimalTFDB 3.0: a comprehensive resource for annotation and prediction of animal transcription factors, *Nucleic Acids Res.*, **D1** (2019), D33−D38. doi: 10.1093/nar/gky822.

22. P. Shannon, M. Richards, *An annotated collection of protein-DNA binding sequence motifs*, 2021. Available from: https://xueshu.baidu.com/usercenter/paper/show?paperid=73fac988c60c44137 363f1c554acbdea.

23. H. Pagès, P. Aboyoun, R. Gentleman, S. DebRoy, Biostrings: Efficient manipulation of biological strings, *Bioconductor*, **2021** (2021). doi: 10.18129/B9.bioc.Biostrings.

24. F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, A. Jema, Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries, *CA Cancer J. Clin.*, **6** (2018), 394−424. doi: 10.3322/caac.21492.

25. R. Zheng, C. Wan, S. Mei, Q. Qin, Q. Wu, H. Sun, et al., Cistrome Data Browser: expanded datasets and new tools for gene regulatory analysis, *Nucleic Acids Res.*, **D1** (2019), D729−D735. doi: 10.1093/nar/gky1094.

26. H. Nord, U. Segersten, J. Sandgren, K. Wester, C. Busch, U. Menzel, et al., Focal amplifications are associated with high grade and recurrences in stage Ta bladder carcinoma, *Int. J. Cancer*, **6** (2010), 1390−1402. doi: 10.1002/ijc.24954.

27. K. Inoue, E. A. Fry, Novel molecular markers for breast cancer, *Biomark. Cancer*, **8** (2016), S38394. doi: 10.4137/BIC.S38394.

28. F. Ren, L. Wang, X. Shen, X. Xiao, Z. Liu, P. Wei, et al., MYBL2 is an independent prognostic marker that has tumor-promoting functions in colorectal cancer, *Am. J. Cancer Res.*, **4** (2015), 1542.

29. H. D. Qin, X. Y. Liao, Y. B. Chen, S. Yi. Huang, W. Q. Xue, F. F. Li, et al., Genomic characterization of esophageal squamous cell carcinoma reveals critical genes underlying tumorigenesis and poor prognosis, *Am. J. Hum. Genet.*, **4** (2016), 709−727. doi: 10.1016/j.ajhg.2016.02.021.

30. J. Musa, M. M. Aynaud, O. Mirabeau, O. Delattre, T. G. Grünewald, MYBL2 (B-Myb): a central regulator of cell proliferation, cell survival and differentiation involved in tumorigenesis, *Cell Death Dis.*, **6** (2017), e2895−e2895. doi: 10.1038/cddis.2017.244.

31. M. Fischer, M. Quaas, L. Steiner, The p53-p21-DREAM-CDE/CHR pathway regulates G2/M cell cycle genes, *Nucleic Acids Res.*, **1** (2016), 164−174. doi: 10.1093/nar/gkv927.

32. Bioconductor package maintainer, Finding candidate binding sites for known transcription factors via sequence matching, *Bioconductor*, **2018** (2018). doi: 10.18129/B9.bioc.generegulation.

33. X. Zhao, X. Li, Z. Ma, M. H. Yin, Identify DNA-binding proteins with optimal Chou's amino acid composition, *Protein Pept. Lett.*, **4** (2012), 398−405. doi: 10.2174/092986612799789404.

34. Y. Wang, Z. Ma, K. C. Wong, X. Li, Nature-inspired multiobjective patient stratification from cancer gene expression data, *Inf. Sci.*, **526** (2020), 245−262. doi: 10.1016/j.ins.2020.03.095.

35. Y. Wang, B. Liu, Z. Ma, K. C. Wong, X. Li, Nature-inspired multiobjective cancer subtype diagnosis, *IEEE J. Transl. Eng. Health Med.*, **7** (2019), 1−12. doi: 10.1109/jtehm.2019.2891746.

36. Y. Wang, Z. Ma, K. C. Wong, X. Li, Evolving multiobjective cancer subtype diagnosis from cancer gene expression data, *IEEE/ACM Trans. Comput. Biol. Bioinform.*, **6** (2020), 2431−2444. doi: 10.1109/tcbb.2020.2974953.