Int J Clin Exp Med 2019;12(7):8438-8449
www.ijcem.com /ISSN:1940-5901/IJCEM0091545

Original Article

Identification and characterization of divergent regulated genes correlates with reduced survival times in cancer genome atlas (TCGA) datasets

Juchuanli Tu[1,2,3], Xiaolu Li[3], Jianjun Wang[3]

[1]College of Life Sciences, Beijing Normal University, Beijing, China; [2]National Institute of Biological Sciences, Beijing, China; [3]State Key Laboratory of Microbial Resources, Institute of Microbiology, Chinese Academy of Sciences, Beijing, China

Received January 20, 2019; Accepted May 7, 2019; Epub July 15, 2019; Published July 30, 2019

Abstract: Abnormally-expressed genes play a vital role in tumorigenesis. Dissecting their roles in tumorigenesis, the current study identified differentially-expressed genes in 13 Cancer Genome Atlas (TCGA) datasets. It was found that 33.5% of differentially-expressed genes were both upregulated and downregulated in different TCGA datasets. Furthermore, analysis showed that 82.7% divergent regulated genes correlated with differential DNA methylation in at least one TCGA dataset, indicating that DNA methylation may contribute to regulation of these genes. Of these divergent regulated genes, 86.7% of the genes correlated with prognosis, influencing clinical outcomes based on survival analysis in at least one TCGA dataset. In contrast, 71.8% of the genes that did not show this divergent regulation correlated with prognosis in at least one TCGA dataset. Upregulation and downregulation of 336 divergent regulated genes correlated with reduced survival times in different TCGA datasets. Furthermore, upregulation and downregulation of 22 divergent regulated genes, classified as oncogene (OG) or tumor suppressor gene (TSG) genes, correlated with reduced survival times in different TCGA datasets. The current study sheds new light on the complexities of gene regulation in cancer, suggesting novel targets for mechanistic and therapeutic studies.

Keywords: Differentially expressed genes, divergent regulation, DNA methylation, survival analysis

Introduction

Previous studies have shown that abnormally-expressed genes play a crucial role in tumorigenesis [1, 2]. Expression profiles of genes have drawn much attention from cancer investigators. Indeed, numerous studies have identified genes expressed in a tumor-specific manner in various types of cancer [3-5]. Although much effort has been dedicated to dissecting the complexity of gene function and regulation in tumorigenesis, it remains poorly understood.

The current study identified differentially-expressed genes in 13 TCGA datasets. Surprisingly, it was found that 33.5% of the differentially-expressed genes were both upregulated and downregulated in at least one TCGA dataset. Results suggest that DNA methylation may contribute to regulation of these genes. Of these genes, 336 genes were upregulated in some TCGA datasets, correlating with reduced survival times. In contrast, the same genes were downregulated in other TCGA datasets. Low expression of these genes also correlated with reduced survival times. Moreover, 24 genes belonging to oncogene (OG) and tumor suppressor gene (TSG) families were both upregulated and downregulated in different TCGA datasets. Upregulation and downregulation of these 24 OG and TSG genes correlated with reduced survival times in different TCGA datasets.

Present results indicate that these genes may be under divergent regulation and correlate with reduced survival times in different cancers. Identification and characterization of these genes sheds new light on the complexities of gene function and regulation in cancer, suggesting novel targets for mechanistic and therapeutic studies.

Materials and methods

TCGA datasets and expression analysis

All TCGA datasets were downloaded from the TCGA website (https://portal.gdc.cancer.gov/). Gene annotation was also retrieved from the TCGA website (https://gdc.cancer.gov/about-data/data-harmonization-and-generation/gdc-reference-files). Raw reads counts were extracted from files with the suffix "htseq.counts", evaluating gene expression levels.

This study only incorporated cancer datasets with ≥ 200 cancer samples and ≥ 15 normal samples, ensuring accurate identification of differentially-expressed genes. This stud analyzed a total of 13 cancer datasets. All subsequent analyses were performed based on these 13 cancer datasets, ensuring the accuracy of analysis, except for methylation data analysis.

Identification of differentially-expressed genes was carried out using the "edgeR" package (v3.24.0) in R [6]. Genes in each TCGA dataset were kept when their CPM was ≥ 0.5 in at least the size of normal samples. This study then identified differentially-expressed genes by comparing expression profiles of specific genes between groups of cancer and normal samples. Genes with FDR ≤ 0.05 and expression fold changes ≥ 1.5 (up-regulated genes) or ≤ 0.67 (down-regulated genes) were defined as differentially-expressed genes [7, 8].

Correlation analysis of methylation levels and target genes

DNA methylation data from Human Methylation 450 BeadChip platform was downloaded from the TCGA website (https://portal.gdc.cancer.gov/). Beta-values were extracted to evaluate DNA methylation levels of each probe. Annotations of probes to specific genes were extracted from the retrieved files. This study excluded the STAD dataset for the methylation data available for only two normal samples.

Moreover, the "champ.DMP" function in the "ChAMP" package in R was used to identify differential methylation probes [9]. Probes with adjusted *p*-values ≤ 0.05 are defined as differential methylation probes.

A gene was defined as being correlated to methylation when it was upregulated in cancer, with at least one hypo-methylated probe annotated to the gene, or when the gene was down-regulated in cancer, with at least one hyper-methylated probe annotated.

List of oncogenes (OG) and tumor suppressor genes (TSG)

The list of oncogenes (OG) was downloaded from the ONGene database (http://ongene.bioinfo-minzhao.org/index.html) [10], while the list of tumor suppressor genes (TSG) was downloaded from the TSGene database (https://bioinfo.uth.edu/TSGene/index.html) [11]. Overlapping between OG/TSG lists and the specific gene list was determined based on gene names.

Gene ontology (GO) analysis

Gene ontology (GO) analysis was performed with the DAVID GO bioinformatics platform (V6.8) (https://david.ncifcrf.gov/). Ensembl gene IDs for each gene were used as the input. Ensembl gene IDs for each gene can be found in the gene annotation file (https://gdc.cancer.gov/about-data/data-ha rmonization-and-generation/gdc-reference-files).

Survival analysis of TCGA datasets

Survival analysis of TCGA datasets was conducted, as previously reported [8, 12]. Briefly, survival information was retrieved from "clinical follow-up" and "clinical patient" data in each cancer dataset, keeping the most recent follow-up information for each patient. Only differentially-expressed genes (defined above) were included in the analysis of each cancer dataset.

Cox's proportional hazards regression model was employed, with the "coxph" function from the "survival" library in R. The equation for the model was: "coxph (Surv (time, censor) ~ exprs)", where time is survival time (for dead patients) or last follow-up time (for live patients). Censor is dead or alive for each cancer sample and exprs is the gene expression value measured as log-transformed TMM normalized-counts. *P*-values and prognosis-relevant coefficients were obtained from Cox's proportional

hazards regression model. Genes with *P*-values ≤ 0.05 are defined as prognostic genes.

Results

Identification of differentially-expressed genes in TCGA datasets

Abnormal expression of genes is vital in tumorigenesis [1, 2]. Hence, the current study set out to identify differentially-expressed genes in 13 TCGA datasets. Detailed information concerning the 13 analyzed TCGA datasets is shown in **Figure 1A**.

The "edgeR" package in R was used to identify differentially-expressed genes [6]. The method used involved first removing the genes that were lowly-expressed. After filtration, about 50% of the genes in each TCGA dataset were defined as lowly-expressed. They were removed from subsequent analysis (**Figure 1B** and Supplemental Table 1). Next, this study identified differentially-expressed genes by comparing expression profiles between groups of cancer and normal samples in 13 TCGA datasets. In total, ~15% and ~10% genes were identified as upregulated and downregulated isoforms in each TCGA dataset (**Figure 1C** and Supplemental Table 2).

Gene enrichment analysis was also conducted. Results showed that upregulated genes in at least one TCGA dataset were enriched in various functions, such as calcium ion binding, cell adhesion, and inflammatory response (**Figure 1D** and Supplemental Table 3). In contrast, downregulated isoforms in at least one TCGA dataset were enriched in several functions, including calcium ion binding, signal transduction, and angiogenesis (**Figure 1E** and Supplemental Table 3).

Differentially-expressed isoforms (33.5%) were both downregulated and upregulated in different TCGA datasets

Some GO terms, such as cell adhesion and calcium ion binding, were both enriched in gene ontology analysis when using upregulated and downregulated genes as input (**Figure 1D** and **1E**). It was speculated that the overlapping of enriched functions may reflect the overlapping of genes between upregulated and downregulated genes in different TCGA datasets. Overlapping of genes was checked with upregulation and downregulation in at least one TCGA dataset. Of the up/downregulated genes, ~35% of upregulated and ~60% of downregulated genes in one TCGA dataset were downregulated or upregulated in at least one other TCGA dataset (**Figure 2A** and Supplemental Table 4). For example, 37% of upregulated genes in the BLCA dataset were downregulated in at least one of the other 12 TCGA datasets. Conversely, 51% of downregulated genes in the BLCA dataset were upregulated in at least one of the other 12 TCGA datasets (**Figure 2A**). Overall, 33.5% of differentially-expressed genes across all 13 analyzed TCGA datasets were both upregulated and downregulated in different TCGA datasets.

Of these genes, 9,651 genes were upregulated in more than one TCGA dataset, while downregulated in more than one TCGA dataset (Group 1). Only 1,102 isoforms were upregulated in only one TCGA dataset, while downregulated in only one other TCGA dataset (Group 2). Overall, 90% of the genes were both upregulated and downregulated in multiple TCGA datasets (**Figure 2B**).

As shown in **Figure 2C**, this study listed the top 15 genes that were both upregulated and downregulated in multiple TCGA datasets, sorted by the number of upregulated and downregulated TCGA datasets. This table shows the genes which were upregulated in multiple TCGA datasets and also downregulated in multiple TCGA datasets. Taking the top gene, CA2 (ENSG00000104267.8) as an example, the distribution of normalized counts clearly shows that expression levels of this gene in cancer samples were significantly higher than those in normal samples in BLCA, BRCA, HNSC, THCA, and UCEC datasets. They were significantly lower than those in the normal samples in the COAD, KIRC, KIRP, LIHC, LUAD, LUSC, PRAD, and STAD datasets (**Figure 2D**). Genes that were both upregulated and downregulated in different TCGA datasets are defined as genes with divergent regulation.

Next, gene ontology (GO) analysis was conducted for genes which were both upregulated and downregulated in different TCGA datasets. Results showed enrichment in functions, including cell adhesion and calcium ion binding. These were also enriched, according to gene

A

Cancer type		Number of RNA-seq libraries analyzed			Number of methylation libraries analyzed			Number of patients with survival data		
Abbreviation	Full name	Cancer	Healthy	Combined	Cancer	Healthy	Combined	Alive	Dead	Combined
BLCA	Bladder Urothelial Carcinoma	414	19	433	418	21	439	228	180	408
BRCA	Breast invasive carcinoma	1102	113	1215	793	97	890	939	152	1091
COAD	Colon adenocarcinoma	478	41	519	313	38	351	352	102	454
HNSC	Head and Neck squamous cell carcinoma	500	44	544	528	50	578	283	218	501
KIRC	Kidney renal clear cell carcinoma	538	72	610	324	160	484	357	173	530
KIRP	Kidney renal papillary cell carcinoma	288	32	320	275	45	320	245	44	289
LIHC	Liver hepatocellular carcinoma	371	50	421	377	50	427	241	130	371
LUAD	Lung adenocarcinoma	533	59	592	473	32	505	328	187	515
LUSC	Lung squamous cell carcinoma	502	49	551	370	42	412	284	218	502
PRAD	Prostate adenocarcinoma	498	52	550	502	50	552	485	10	495
STAD	Stomach adenocarcinoma	375	32	407	395	2	397	230	150	380
THCA	Thyroid carcinoma	502	58	560	507	56	563	486	16	502
UCEC	Uterine Corpus Endometrial Carcinoma	551	35	586	438	46	484	452	91	543

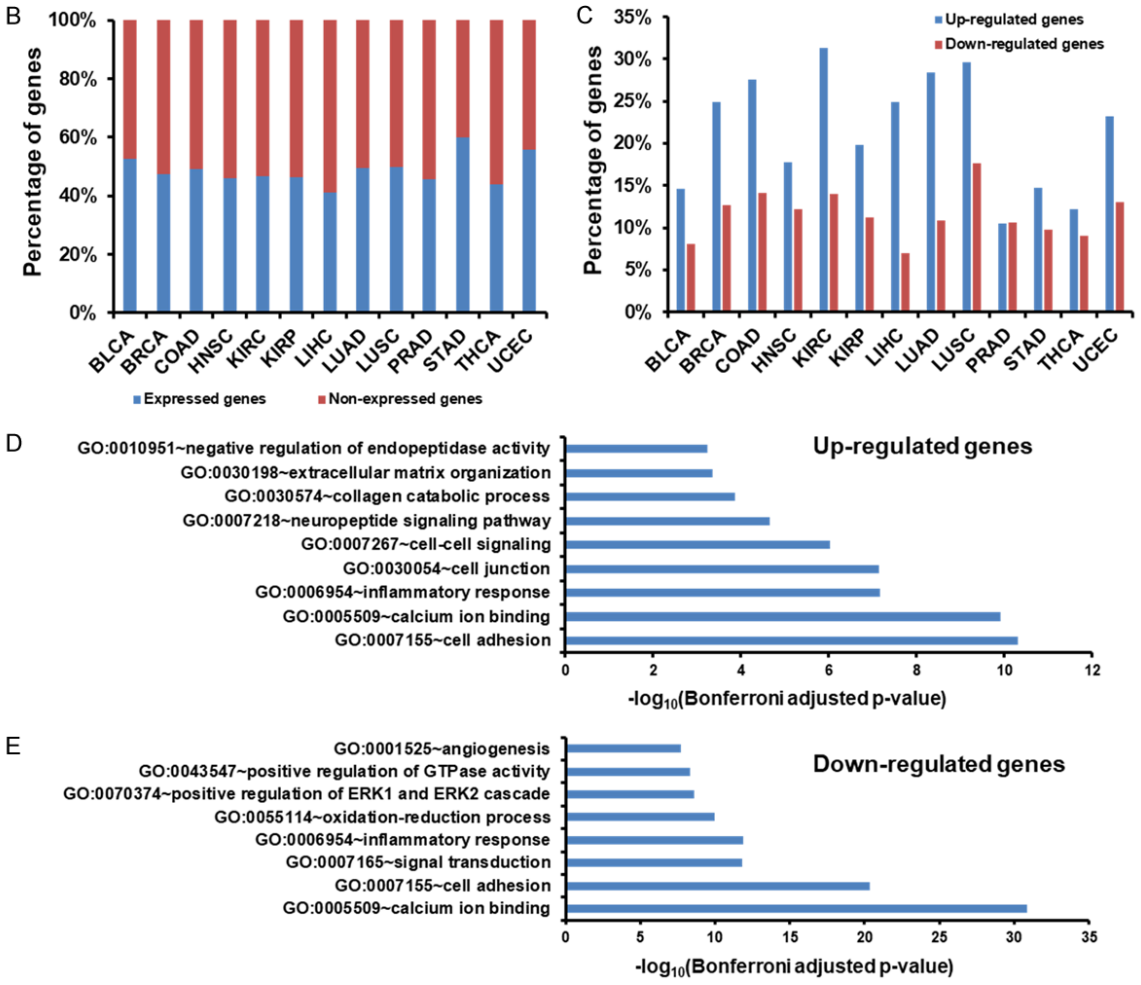
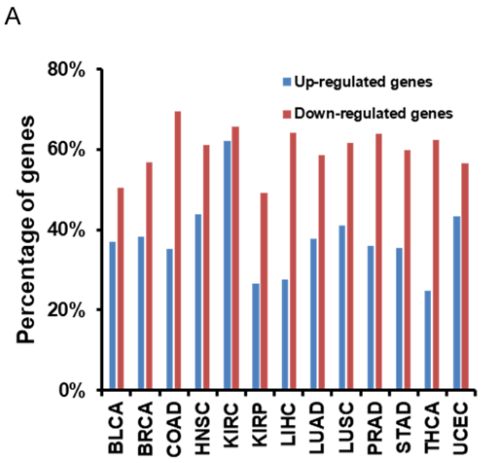
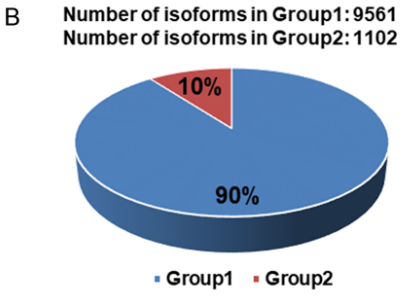


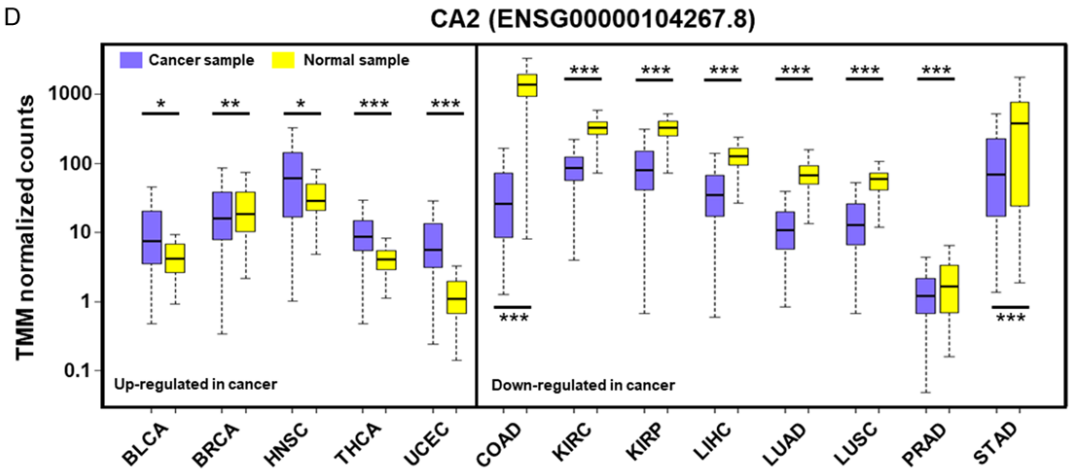
Figure 1. Characterization of differentially-expressed genes in the TCGA datasets. (A) Summary of the number of cancer and normal samples in each of the 13 analyzed TCGA datasets. (B) Bar graph showing the percentage of expressed genes in each TCGA dataset. Genes were defined as expressed isoforms when their CPM was ≥ 0.5 in at least the size of normal samples. (C) Bar graph showing the percentage of upregulated and downregulated isoforms in each TCGA dataset. (D and E) Gene ontology (GO) enrichment analysis showing the function of upregulated (D) and downregulated (E) genes. X-axis: -$\log_{10}$ transformed Bonferroni corrected *p*-value.





C

ID	Up-regulation		Down-regulation		
	N of datasets	Datasets	N of datasets	Datasets	Gene
ENSG00000104267.8	5	BLCA,BRCA,HNSC,THCA,UCEC	8	COAD,KIRC,KIRP,LIHC,LUAD,LUSC,PRAD,STAD	CA2
ENSG00000118160.12	8	BRCA,HNSC,KIRC,KIRP,LIHC,LUAD,LUSC,THCA	5	BLCA,COAD,PRAD,STAD,UCEC	SLC8A2
ENSG00000146197.8	7	BRCA,COAD,HNSC,LIHC,LUAD,LUSC,UCEC	5	BLCA,KIRC,KIRP,PRAD,THCA	SCUBE3
ENSG00000141744.3	6	BLCA,BRCA,COAD,LIHC,THCA,UCEC	6	HNSC,KIRC,KIRP,LUAD,LUSC,PRAD	PNMT
ENSG00000167080.7	7	BLCA,BRCA,HNSC,LIHC,LUAD,LUSC,UCEC	5	COAD,KIRC,KIRP,STAD,THCA	B4GALNT2
ENSG00000162779.19	6	BRCA,HNSC,LIHC,LUAD,LUSC,UCEC	6	COAD,KIRC,KIRP,PRAD,STAD,THCA	AXDND1
ENSG00000147614.3	6	BLCA,BRCA,HNSC,LIHC,STAD,UCEC	6	COAD,KIRC,KIRP,LUAD,LUSC,PRAD	ATP6V0D2
ENSG00000140519.11	7	BRCA,COAD,LIHC,LUAD,LUSC,THCA,UCEC	5	BLCA,HNSC,KIRP,PRAD,STAD	RHCG
ENSG00000171401.13	6	COAD,KIRP,LUAD,LUSC,THCA,UCEC	5	BLCA,HNSC,KIRC,PRAD,STAD	KRT13
ENSG00000248290.1	5	BRCA,KIRC,KIRP,LIHC,THCA	6	BLCA,COAD,HNSC,LUSC,STAD,UCEC	TNXA
ENSG00000176490.4	5	BRCA,LIHC,LUAD,LUSC,THCA	6	BLCA,COAD,HNSC,KIRC,STAD,UCEC	DIRAS1
ENSG00000151632.15	6	HNSC,KIRP,LIHC,LUAD,LUSC,UCEC	5	BRCA,COAD,KIRC,PRAD,STAD	AKR1C2
ENSG00000235770.4	6	BRCA,COAD,KIRC,KIRP,LIHC,THCA	5	BLCA,LUAD,LUSC,PRAD,UCEC	LINC00607
ENSG00000132874.12	5	LIHC,LUAD,LUSC,THCA,UCEC	5	BRCA,COAD,HNSC,KIRC,KIRP	SLC14A2
ENSG00000180251.4	5	BRCA,LUAD,LUSC,THCA,UCEC	5	HNSC,KIRC,KIRP,PRAD,STAD	SLC9A4



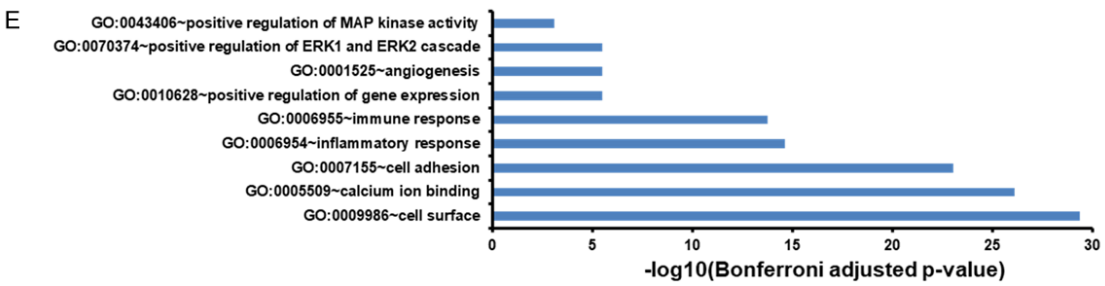


Figure 2. Characterization of genes that were both upregulated and downregulated in different TCGA datasets. A. Bar graph showing the percentage of genes which were both upregulated and downregulated in different TCGA datasets. The percentage was calculated as the ratio between the numbers of upregulated or downregulated genes with divergent regulation in one TCGA dataset versus the total number of upregulated or downregulated genes in the same dataset. B. Pie plot showing the number of genes which were both upregulated and downregulated in different numbers of TCGA datasets. Group 1: Number of genes that were both upregulated and downregulated in more than

one TCGA datasets; Group 2: Number of isoforms both upregulated in only one TCGA dataset and downregulated in other one TCGA dataset. C. Summary list of the top 15 genes that were both upregulated and downregulated in different TCGA datasets, sorted by the number of TCGA datasets showing up/downregulation. D. Boxplots showing the distribution of the TMM normalized counts for the indicated genes in indicated TCGA datasets. E. Gene ontology (GO) enrichment analysis showing the function of genes that were both upregulated and downregulated in different TCGA datasets. X-axis: -$\log_{10}$ transformed Bonferroni's corrected *p*-value.

ontology analysis, using upregulated and downregulated genes as input (**Figure 2E** and Supplemental Table 3).

DNA methylation contributed to regulation of divergent regulated genes

DNA methylation has been increasingly recognized as an important process underlying tumorigenesis [13]. Recent studies have identified DNA methylation in normal tissues and cancer [13, 14]. The current study explored the relationship between DNA methylation and regulation of divergent regulated genes. It was hypothesized that DNA methylation may contribute to regulation of divergent regulated genes in different TCGA datasets.

Pan-cancer analysis was conducted, indicating that methylation correlated to divergent regulated genes. Methylation probes were identified as hypo-methylated probes when the beta-value of the probe in the cancer group was significantly lower than that in the normal group. Hyper-methylated probes were identified when the beta-value in the cancer group was significantly higher than that in the normal group. A divergent regulated gene was defined when correlating to methylation when it was upregulated in cancer, while at least one hypo-methylated probe was annotated to this gene, or downregulated in cancer, while at least one hyper-methylated probe was annotated to this gene.

Moreover, ~800 and ~400 divergent regulated genes correlated with hypo- and hyper-methylation (**Figure 3A** and Supplemental Table 5). Comparing the percentage of DNA methylation associated genes showing vs. not showing divergent regulation in different TCGA datasets (gene set 1 and gene set 2), it was found that 82.7% of the genes showing both upregulation and downregulation in different TCGA datasets correlated with DNA methylation in at least one TCGA dataset. In contrast, only 61.3% of the genes that did not show this divergent regulation correlated with DNA methylation in at least one TCGA dataset (**Figure 3B**). The difference was significant (*p*-value < 2.2e-16). Hence, DNA methylation was prone to be correlated with expression of divergent regulated genes.

Taking gene DLL4 as an example, expression levels were significantly upregulated in KIRC, LIHC, and THCA. Levels were downregulated in BRCA, LUAD, LUSC, and UCEC (**Figure 3C**). The methylation probe cg14209037 was annotated to DLL4. The beta-value of this probe was significantly lower in cancer in KIRC, LIHC, and THCA, while higher in BRCA, LUAD, LUSC, and UCEC (**Figure 3D**). The gene expression profile was negatively correlated with methylation levels.

Genes under divergent regulation more likely to be prognostic genes

To explore the association of genes which were both upregulated and downregulated in different TCGA datasets with clinical outcomes, expression levels were correlated with patient survival. The number of differentially-expressed and prognostic genes varied among different TCGA datasets. In the KIRC dataset, 50% of upregulated genes were identified as prognostic. In contrast, less than 10% of upregulated or downregulated genes in COAD or PRAD datasets were identified as prognostic (**Figure 4A** and Supplemental Table 6).

Of these prognostic genes, around 20% under divergent regulation were also predicted to correlate with prognosis in each TCGA dataset (**Figure 4B**). This study then compared the percentage of prognostic genes showing vs. not showing divergent regulation in different TCGA datasets (gene set 1 and gene set 2). It was found that 86.7% of the genes showing both upregulation and downregulation in different TCGA datasets correlated with prognosis in at least one TCGA dataset. In contrast, only 71.8% of genes that did not show this divergent regulation correlated with prognosis in at least one TCGA dataset (**Figure 4C**). The difference between the percentages of prognostic isoforms in

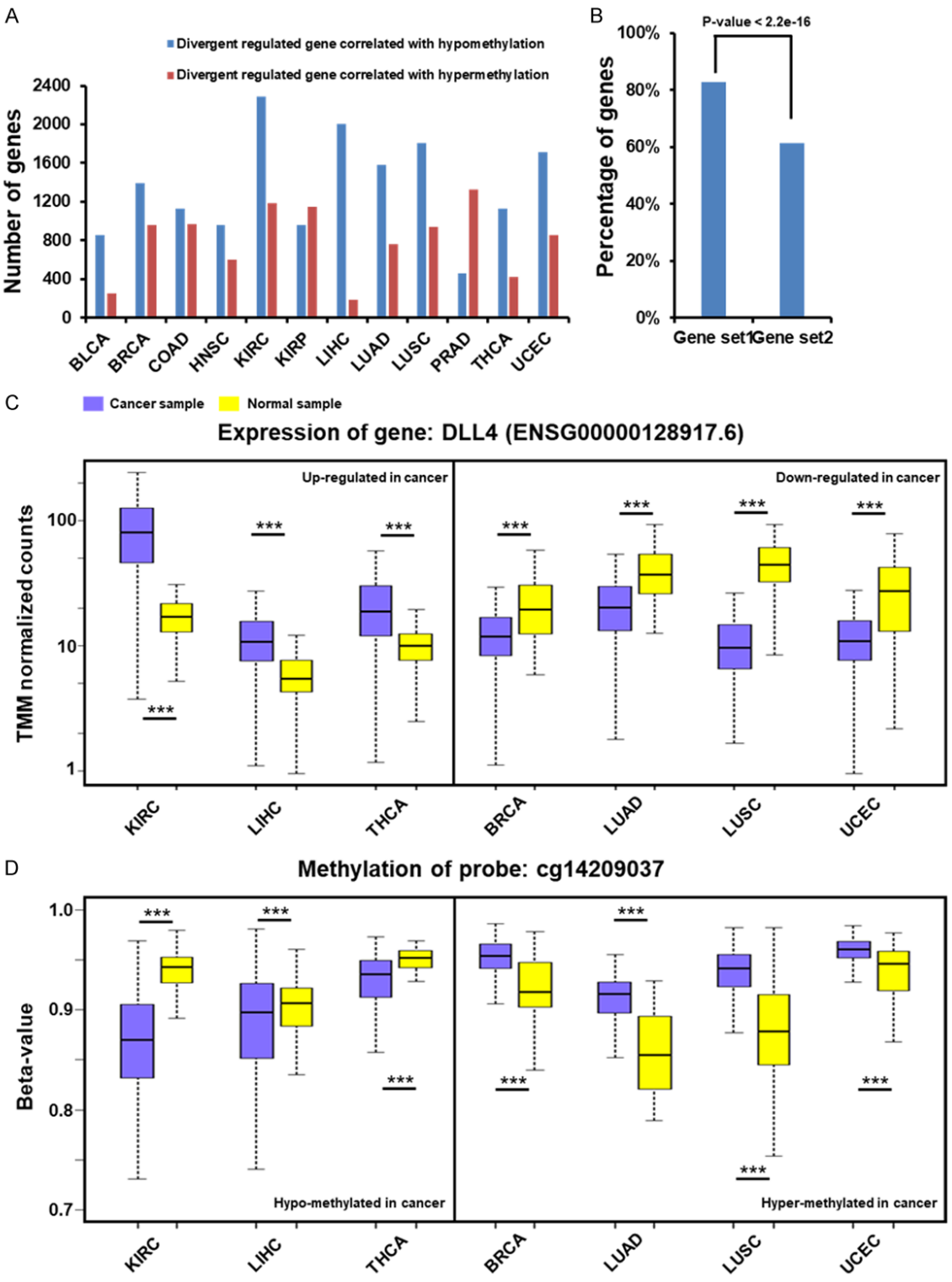


Figure 3. Correlation between the expression profile of divergent regulated genes and DNA methylation. A. Bar graph showing the number of divergent regulated genes correlated with hypo- or hyper-methylated probes. B. Bar graph comparing the percentage of genes correlated with DNA methylation among those which were both upregulated and downregulated in different TCGA datasets versus the remaining up/downregulated genes. *P*-values were calculated by the Chi-square test in R. C. Boxplots showing the distribution of the TMM normalized counts for the indicated gene in indicated TCGA datasets. D. Boxplots showing the distribution of the beta-value for the indicated probe in indicated TCGA datasets.

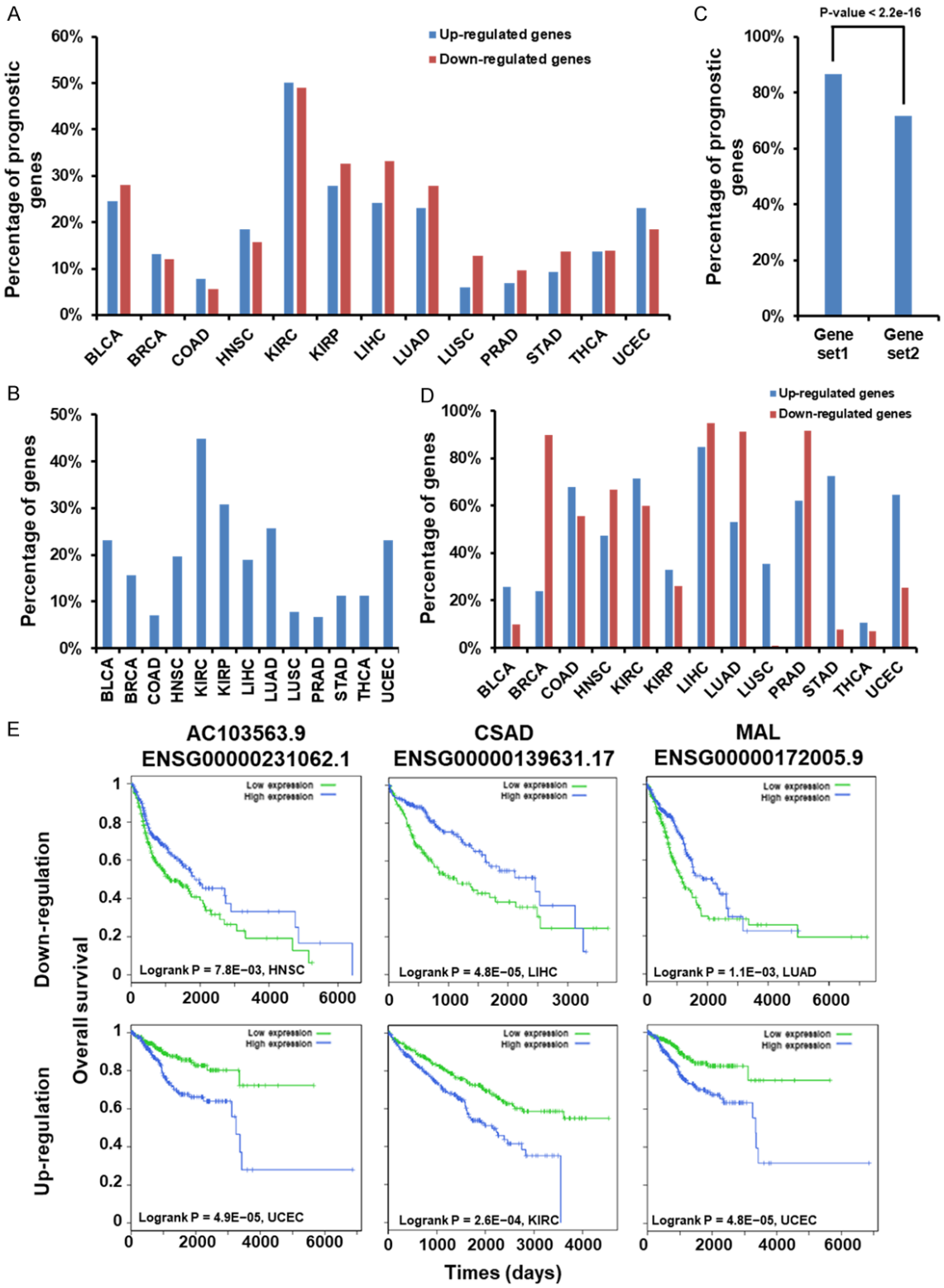


Figure 4. Identification of prognostic genes among those that were both upregulated and downregulated in different TCGA datasets. A. Bar graph showing the percentage of upregulated or downregulated and prognostic genes among upregulated or downregulated genes in each TCGA dataset. B. Bar graph showing the percentage of upregulated or downregulated prognostic genes with divergent regulation among upregulated or downregulated genes with divergent regulation in each TCGA dataset. C. Bar graph comparing the percentage of prognostic genes among those

which were both upregulated and downregulated in different TCGA datasets versus the remaining up/downregulated genes. *P*-values were calculated by the Chi-square test in R. D. Bar graph showing the percentage of upregulated prognostic genes with divergent regulation and prognosis-relevant coefficient > 0 among upregulated prognostic genes with divergent regulation in each TCGA dataset and downregulated prognostic genes with divergent regulation and prognosis-relevant coefficient < 0 among downregulated prognostic genes with divergent regulation in each TCGA dataset. E. Survival plots corresponding to indicated genes with dual function. Isoforms showing genes with divergent regulation. Upregulation and downregulation of these genes in different TCGA datasets reduced survival times. Groups of high or low expression were defined as > or ≤ the median value of TMM normalized counts.

these two gene sets was significant (*p*-value < 2.2e-16). Hence, genes that were both upregulated and downregulated in different TCGA datasets were more likely to be prognostic genes.

Identification of divergent regulated genes correlated with different prognosis in TCGA datasets

A straightforward and simple explanation for the divergent regulation of genes is that the transcription of certain genes is activated in some types of cancer, while other types of cancer suppress the transcription of the same genes. Activation and suppression may promote tumorigenesis and decrease survival times in different cancers.

Cox's model for survival analysis provides a coefficient for each gene. This is related to its contribution to the hazard ratio. A positive coefficient indicates that this isoform increases hazard ratios, while a negative coefficient indicates that expression of this isoform increases survival times [12]. The question was considered whether a gene which was upregulated and which increased the hazard ratio and decreased the survival time in one TCGA dataset could be downregulated in a different TCGA dataset, while also increasing the hazard ratio and decreasing the survival time in that second TCGA dataset.

Results showed that upregulation or downregulation of a certain gene showing divergent regulation in each TCGA dataset could decrease survival times (**Figure 4D**). In total, the current study found that both upregulation and downregulation of 336 genes with divergent regulation decreased survival times in corresponding TCGA datasets (Supplemental Table 6).

Taking three genes (AC103563.9, CSAD, and MAL) as examples, analysis showed that both upregulation and downregulation of these three genes in the indicated TCGA datasets reduced survival times (Supplemental Figure 1 and **Figure 4E**). Gene CSAD was upregulated in the KIRC dataset and downregulated in the LIHC dataset (Supplemental Figure 1). Upregulation of CSAD in the KIRC dataset decreased survival times (log-rank P = 2.6E-04), while downregulation of this same gene in the LIHC dataset also decreased survival times (log-rank P = 4.8E-05) (**Figure 4E**).

Current results suggest that 336 genes with divergent regulation correlated with different prognosis in different TCGA dataset. Upregulation and downregulation of these 336 isoforms associated with decreased survival times in different TCGA datasets.

Identification of divergent regulated genes correlated with different prognosis classified as OG and TSG

Oncogenes (OG) and tumor suppressor genes (TSG) are two classes of genes that strongly promote or suppress tumorigenesis, respectively. Cancer cells may induce expression of OG and suppress expression of TSG, promoting tumorigenesis [15, 16]. Surprising, 100 genes with divergent regulation were classified as OG or TSG and upregulated or downregulated in each TCGA dataset (**Figure 5A**). Furthermore, it was found that a certain number of prognostic genes with divergent regulation were classified as OG or TSG (**Figure 5B**). A total of 22 genes with dual function were classified as OG or TCG (Supplemental Table 7). Results indicated that both upregulation and downregulation of these OGs or TSGs decreased survival times in different TCGA datasets. Thus, these OGs and TSGs may also play a dual role in tumorigenesis.

NR1I2, which belongs to OG, was upregulated in KIRC. However, in the LIHC dataset, it was downregulated (**Figure 5D**). High expression of NR1I2 in KIRC was significantly associated with reduced survival times (log-rank P = 1.6E-02). In contrast, low expression of NR1I2 in LIHC was significantly associated with reduced survival times (log-rank P = 1.4E-03) (**Figure 5C**).

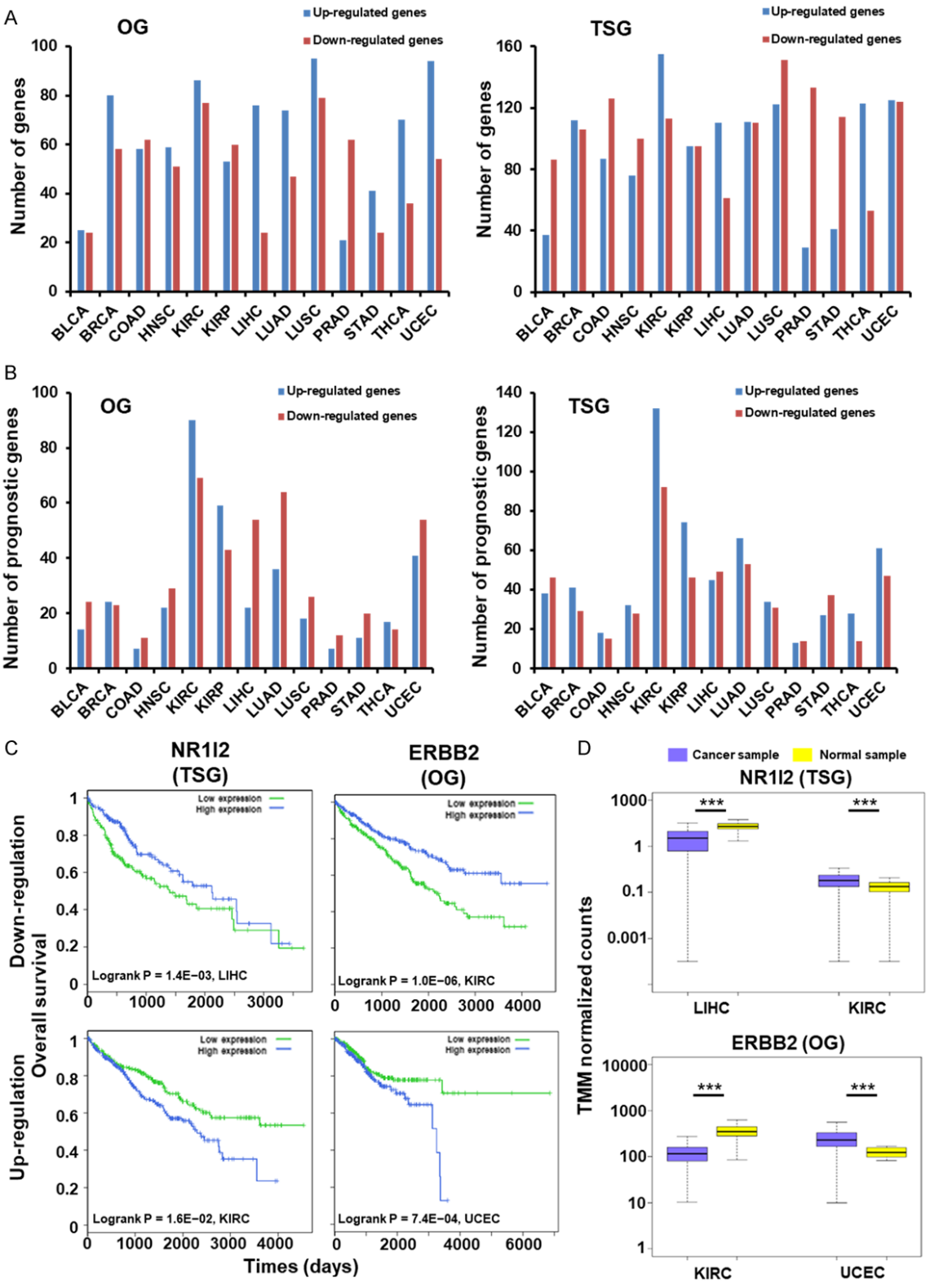


Figure 5. Characterization of genes classified as OG and TSG. A. Bar graph showing the number of genes with divergent regulation that were classified as OG (left panel) or TSG (right panel) in each TCGA dataset. B. Bar graph showing the number of prognostic genes with divergent regulation that were classified as OG (left panel) or TSG (right panel) in each TCGA dataset. C. Survival plots corresponding to indicated genes that were classified as OG (left panel) or TSG (right panel). D. Boxplots showing the distribution of the TMM normalized counts for indicated genes in the indicated TCGA datasets.

ERBB2, which belongs to TSG, was upregulated in UCEC. In contrast, it was downregulated in KIRC (**Figure 5D**). High expression of ERBB2 in UCEC significantly reduced survival times (log-rank P = 7.4E-04). In contrast, low expression of ERBB2 in KIRC also significantly reduced survival times (log-rank P = 1.0E-06) (**Figure 5C**).

Discussion

Genes which show both upregulation and downregulation and are associated with reduced survival times may reflect the complexity of gene regulation in tumorigenesis. Indeed, previous studies have identified several genes which play a dual role in tumorigenesis [17-20]. Some of these studies have identified several targets with opposing functions [18, 19]. The current study identified hundreds of genes with divergent regulation that correlated with reduced survival times, constituting novel targets for further studies. Understanding the dual function of these genes could expand current knowledge about cancer. It also serves as a warning that upregulation or downregulation of a specific gene which promotes or suppresses tumorigenesis in one type of cancer does not mean that this same pattern of expression will also promote or suppress tumorigenesis in a different type of cancer. The actual function of these genes may be very complex.

Indeed, the current study found that both upregulation and downregulation of these divergent regulated genes correlated with reduced survival times in different TCGA datasets. The percentage of prognostic genes among these divergent regulated genes was significantly higher than that in control genes. Results showed that these divergent regulated genes strongly influenced clinical outcomes. The current study identified 336 genes that decreased survival times when upregulated or downregulated in corresponding TCGA datasets, providing novel targets for future studies.

Regulation mechanisms of these divergent regulated genes could be an interesting topic for future studies. The current study provided a link between DNA methylation and regulation of these divergent regulated genes. An important way of regulation in cancer, DNA methylation changes were prone to be correlated with and contribute to expression of divergent regulated genes.

Additionally, it was found that certain genes that were classified as OG or TSG were also upregulated or downregulated. They were associated with reduced survival times in different TCGA datasets. Inhibition of the activity of OG by chemical compounds is one of the ways to cure cancer [21]. However, the current study indicated that low expression of some OGs correlated with reduced survival times in some types of cancer. Thus, inhibiting the activity of these OGs may not be a suitable cancer therapy approach in these cases. It is also notable that overexpression of some TSGs could paradoxically promote tumorigenesis in certain types of cancer.

Acknowledgements

We acknowledge the contributions of the TCGA Research Network. Without their efforts, this type of analysis would not be possible.

Disclosure of conflict of interest

None.

Address correspondence to: Jianjun Wang, State Key Laboratory of Microbial Resources, Institute of Microbiology, Chinese Academy of Sciences, Beijing, China, NO.1 Beichen West Road, Chaoyang District, Beijing 100101, China. Fax: +86-10-64807429; Tel: +86-10-64807417; E-mail: wangjj@im.ac.cn

References

[1] Lee TI and Young RA. Transcriptional regulation and its misregulation in disease. Cell 2013; 152: 1237-1251.

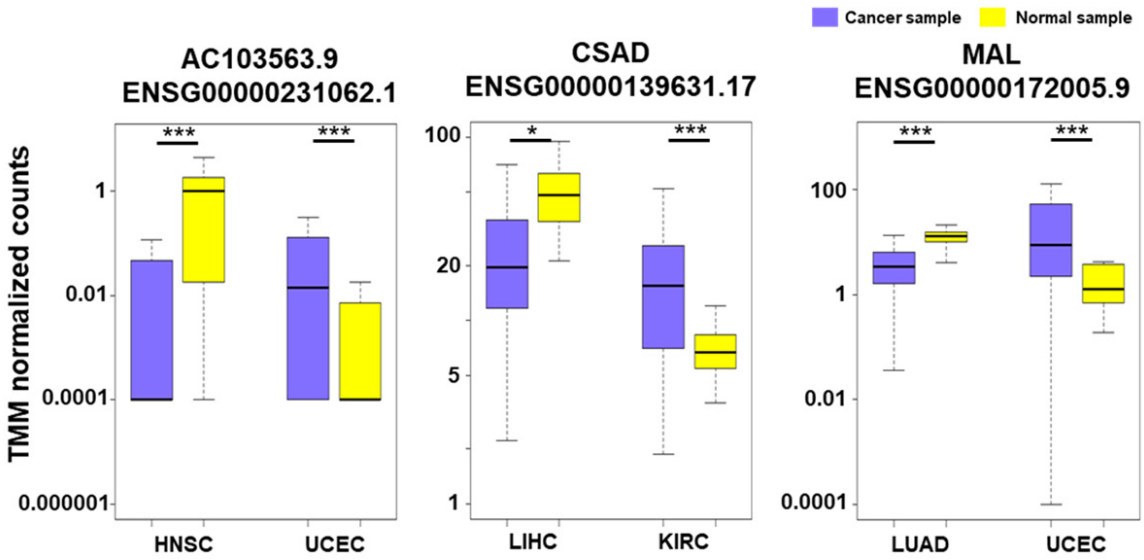
[2] Hanahan D and Weinberg RA. Hallmarks of cancer: the next generation. Cell 2011; 144: 646-674.

[3] Zhao W, Hoadley KA, Parker JS and Perou CM. Identification of mRNA isoform switching in breast cancer. BMC Genomics 2016; 17: 181.

[4] Stricker TP, Brown CD, Bandlamudi C, McNerney M, Kittler R, Montoya V, Peterson A, Grossman R and White KP. Robust stratification of breast cancer subtypes using differential patterns of transcript isoform expression. PLoS Genet 2017; 13: e1006589.

[5] Bjorklund SS, Panda A, Kumar S, Seiler M, Robinson D, Gheeya J, Yao M, Alnaes GIG, Toppmeyer D, Riis M, Naume B, Borresen-Dale AL, Kristensen VN, Ganesan S and Bhanot G. Widespread alternative exon usage in clinically distinct subtypes of invasive ductal carcinoma. Sci Rep 2017; 7: 5568.

[6] Robinson MD, McCarthy DJ and Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 2010; 26: 139-140.
[7] Xiao B, Chen L, Ke Y, Hang J, Cao L, Zhang R, Zhang W, Liao Y, Gao Y, Chen J, Li L, Hao W, Sun Z and Li L. Identification of methylation sites and signature genes with prognostic value for luminal breast cancer. BMC Cancer 2018; 18: 405.
[8] Trincado JL, Sebestyen E, Pages A and Eyras E. The prognostic potential of alternative transcript isoforms across human tumors. Genome Med 2016; 8: 85.
[9] Morris TJ, Butcher LM, Feber A, Teschendorff AE, Chakravarthy AR, Wojdacz TK and Beck S. ChAMP: 450k chip analysis methylation pipeline. Bioinformatics 2014; 30: 428-430.
[10] Liu Y, Sun J and Zhao M. ONGene: a literature-based database for human oncogenes. J Genet Genomics 2017; 44: 119-121.
[11] Zhao M, Sun J and Zhao Z. TSGene: a web resource for tumor suppressor genes. Nucleic Acids Res 2013; 41: D970-976.
[12] Anaya J, Reon B, Chen WM, Bekiranov S and Dutta A. A pan-cancer analysis of prognostic genes. PeerJ 2015; 3: e1499.
[13] Fernandez AF, Assenov Y, Martin-Subero JI, Balint B, Siebert R, Taniguchi H, Yamamoto H, Hidalgo M, Tan AC, Galm O, Ferrer I, Sanchez-Cespedes M, Villanueva A, Carmona J, Sanchez-Mut JV, Berdasco M, Moreno V, Capella G, Monk D, Ballestar E, Ropero S, Martinez R, Sanchez-Carbayo M, Prosper F, Agirre X, Fraga MF, Grana O, Perez-Jurado L, Mora J, Puig S, Prat J, Badimon L, Puca AA, Meltzer SJ, Lengauer T, Bridgewater J, Bock C and Esteller M. A DNA methylation fingerprint of 1628 human samples. Genome Res 2012; 22: 407-419.
[14] Berman BP, Weisenberger DJ, Aman JF, Hinoue T, Ramjan Z, Liu Y, Noushmehr H, Lange CP, van Dijk CM, Tollenaar RA, Van Den Berg D and Laird PW. Regions of focal DNA hypermethylation and long-range hypomethylation in colorectal cancer coincide with nuclear lamina-associated domains. Nat Genet 2011; 44: 40-46.
[15] Croce CM. Oncogenes and cancer. N Engl J Med 2008; 358: 502-511.
[16] Sun W and Yang J. Functional mechanisms for human tumor suppressors. J Cancer 2010; 1: 136-140.
[17] Gao Y, Zhao Y, Zhang J, Lu Y, Liu X, Geng P, Huang B, Zhang Y and Lu J. The dual function of PRMT1 in modulating epithelial-mesenchymal transition and cellular senescence in breast cancer cells through regulation of ZEB1. Sci Rep 2016; 6: 19874.
[18] Lebrun JJ. The dual role of TGFbeta in human cancer: from tumor suppression to cancer metastasis. ISRN Mol Biol 2012; 2012: 381428.
[19] Menegon S, Columbano A and Giordano S. The dual roles of NRF2 in cancer. Trends Mol Med 2016; 22: 578-593.
[20] Li J, Zhu Y, Hazeldine ST, Li C and Oupicky D. Dual-function CXCR4 antagonist polyplexes to deliver gene therapy and inhibit cancer cell invasion. Angew Chem Int Ed Engl 2012; 51: 8740-8743.
[21] Hertz DL and Rae J. Pharmacogenetics of cancer drugs. Annu Rev Med 2015; 66: 65-81.



Supplemental Figure 1. Boxplots showing the distribution of TMM normalized counts for indicated isoforms in the indicated TCGA datasets.