Research Article

Pancancer Analyses Reveal Genomics and Clinical Characteristics of the SETDB1 in Human Tumors

Xin Lin [iD],[1,2] **Min Xiao** [iD],[1] **Zhitao Chen** [iD],[1] **Chenchen Ding** [iD],[1] **Ting Zhang** [iD],[1] **and Qiyong Li** [iD][1]

[1]*Department of Hepatobiliary Surgery, Shulan (Hangzhou) Hospital Affiliated to Zhejiang Shuren University Shulan International Medical College, Hangzhou, China*
[2]*School of Medicine, Zhejiang University, Hangzhou, China*

Correspondence should be addressed to Qiyong Li; liqiyong@zju.edu.cn

Background. Malignant tumor is one of the most common diseases that seriously affect human health. The prior literature has reported the biological function and potential therapeutic targets of SET domain bifurcated histone lysine methyltransferase 1 (SETDB1) as an oncogene. However, SETDB1 has rarely been analyzed from a pan-cancer perspective. *Methods.* Bioinformatics analysis tools and databases, including GeneCards, National Center for Biotechnology Information (NCBI), UniProt, Illustrator for Biological Sequences (IBS), Human Protein Atlas (HPA), GEPIA, TIMER2, Sangerbox 3.0, UALCAN, Kaplan-Meier (K-M) plotter, cBioPortal, Catalogue Of Somatic Mutations In Cancer (COSMIC), PhosphoSitePlus, TISIDB, STRING, and GeneMANIA, were utilized to clarify the biological functions and clinical significance of SETDB1 from a pan-cancer perspective. *Results.* In this study, the pan-cancer analysis demonstrated that SETDB1 showed significantly differential expression in most tumor tissues and paracancerous tissues, and SETDB1 expression was associated with clinicopathological features and clinical prognosis. We also found that SETDB1 mutations occurred in most tumors and were related to tumorigenesis. In addition, DNA methylation of SETDB1 primarily occurred at the cg10444928 site and was associated with prognosis in several human tumors. The predicted phosphorylation site of SETDB1 was Ser1006. We found that SETDB1 was significantly related to the specific tumor-infiltrating immune cell populations and expression of clinically targetable immune checkpoints and may be a promising immunotherapy target. The Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analyses also indicated that SETDB1 may function as crucial regulator in carcinogenesis of human cancers. *Conclusions.* SETDB1 is an important oncogene involved in tumorigenesis and tumor progression through different biological mechanisms. Furthermore, SETDB1 may be a potential therapeutic target for cancer treatment.

1. Introduction

Malignant tumor is becoming a major disease endangering human health [1]. At present, there are no curative strategies for malignant tumors. Antitumor therapies, including radical surgical resection, radiofrequency ablation, transplantation, chemotherapy, immunotherapy, and targeted therapy, have been developed [2]. However, the overall survival (OS) for patients with cancer, especially pancreatic adenocarcinoma (PAAD), lung adenocarcinoma (LUAD), and breast invasive carcinoma (BRCA), remains low because of the complexity and heterogeneity of tumorigenesis [3–5].

Germline mutations caused by abnormal activation and expression of oncogenes have also been confirmed as major inducements of tumorigenesis [6]. The investigation of epigenetic changes, expression levels, potential molecular basis, and clinical significance of oncogene can help understand the mechanisms of tumorigenesis and improve the treatment of various cancers.

SET domain bifurcated histone lysine methyltransferase 1 (SETDB1) protein, also known as ERG-associated protein with SET domain (ESET), KG1T, KMT1E, TDRD21, and H3-K9-HMTase4, is a member of the SET family involved in chromatin gene silencing, chromatin remodeling,

transcriptional suppression, and histone methylation in cells [7, 8]. The SET family also shows epigenetic regulation, participating in gene expression and function changes without altering the DNA sequence. SET family is a significant regulator of tumorigenesis and is important for tumor-targeted therapy [9, 10]. SETDB1 was first reported in 1999, and increasing studies have found that SETDB1 is significantly related to human tumorigenesis and immune cell functions [11–13]. SETDB1 is also a well-known histone H3 lysine 9 (H3K9) methyltransferase that associates with methylation in various euchromatic regions, which causes gene silencing [12]. Therefore, it is important to conduct a comprehensive genomic analysis of SETDB1 and explore its relation with clinical outcome and potential target of oncotherapy in human malignant tumors.

In this study, for the first time, we conducted a structure/function pan-cancer analysis of SETDB1 based on several online databases to explore its oncogenic role and clinical significance in various cancers.

2. Material and Methods

2.1. Omics Analysis of SETDB1. Firstly, we acquired the chromosome localization, coding sequence (CDS), and exon counts of SETDB1 based on the GeneCards database (https://www.genecards.org/). Subsequently, the biological information of the SETDB1 gene and its encoded six protein isoforms was obtained from the "gene" and "protein" module of the National Center for Biotechnology Information (NCBI) (https://www.ncbi.nlm.nih.gov/), with its 3D (three-dimensional) protein structure explored in the UniProt database (https://www.uniprot.org/). In addition, the CDS in nucleotide sequence and conserved domains in amino acid sequence were visualized using Illustrator for Biological Sequences (IBS, version 1.0) [14] (http://ibs .biocuckoo.org/). The position of conserved domains of histone-lysine N-methyltransferase SETDB1 isoform 1 protein was obtained from the "HomoloGene" of NCBI. Conserved amino acid sequences encoded by SETDB1 and phylogenetic tree of SETDB1 family were explored by Constraint-based Multiple Alignment Tool (https://www .ncbi.nlm.nih.gov/tools/cobalt/) in the NCBI. Finally, the distribution of SETDB1 protein was obtained from Human Protein Atlas (HPA) (https://www.proteinatlas.org/) database.

2.2. Gene Expression Analysis. The expression levels of mRNA and encoded protein of SETDB1 in normal tissues were obtained from several online databases, including GEPIA [15] (http://gepia.cancer-pku.cn/), HPA, and the University of California, Santa Cruz (UCSC) Xena browser (https://xenabrowser.net). The datasets were derived from The Cancer Genome Atlas (TCGA) (https://www.cancer .gov/) and Genotype-Tissue Expression Project (GTEx) (https://www.genome.gov/). Additionally, the immunohistochemical (IHC) staining and hematoxylin-eosin (H&E) staining displayed the top five significantly expressed tissues of SETDB1. In this study, we also compared the mRNA expression of the SETDB1 gene in different cell lines and single cell specificity. TIMER2 (http://timer.cistrome.org/),

Sangerbox 3.0 (http://vip.sangerbox.com/), UALCAN (http://ualcan.path.uab.edu/), and GEPIA2 [16] (http://gepia2.cancer-pku.cn) databases were also applied to explore and verify the SETDB1 differential expression levels in tumor tissues and paracancerous tissues.

2.3. Clinicopathological Features Analysis. The "Pathological Stage Plot" module of GEPIA2 was applied to assess the relationship between the SETDB1 gene expression and cancer stage based on TCGA. $P < 0.05$ was set as the significance threshold. We also used the UALCAN database to explore the relationship between the SETDB1 mRNA expression level and clinicopathological stage, including the within-stage correlation. Sangerbox 3.0 was applied to confirm the connection between the SETDB1 mRNA transcription level and other clinicopathological features, including TNM classification and clinicopathological grade.

2.4. Survival Analysis. GEPIA2 is also an available comprehensive prognosis analysis database, and its target gene was used as input for survival analysis in various human cancers. In this study, we first used the "Survival Map" module of GEPIA2 to explore the OS and disease-free survival (DFS) significance map data of SETDB1 among all tumors from TCGA datasets. The patients with cancer were divided into the high- and low-expression subgroups according to the median expression levels of SETDB1. Subsequently, the "survival analysis" module of GEPIA2 was used to draw the significance of the Kaplan-Meier (K-M) curves for patients with cancer, and $P < 0.05$ was set as the significance threshold. Furthermore, the Cox analyses based on Sangerbox 3.0 database were performed for disease-specific survival (DSS) and progression-free interval (PFI) of SETDB1 across various cancers samples, and the results were displayed by a forest plot. Then, we explored the prognostic value of SETDB1 in ovarian cancer, liver hepatocellular carcinoma (LIHC), LUAD, and BRCA based on the K-M plotter [17] (http://kmplot.com/analysis/). The survival datasets were derived from TCGA and Gene Expression Omnibus (GEO) (https://www.ncbi.nlm.nih.gov/geo/).

2.5. Genetic Alteration Analysis. A total of 32 studies containing 10967 samples were selected from the "TCGA Pan-Cancer Atlas Studies" module of cBioPortal (https://www .cbioportal.org) online database. The genetic alteration levels of SETDB1 were further explored using 110443 samples with mutation data. The totality genetic alteration samples of SETDB1 across TCGA tumors were generated from the "OncoPrint" module of cBioPortal. According to the "Cancer Types Summary" module of cBioPortal, the alteration frequency, number of genetic mutations, type of SETDB1 mutations, and copy number variation (CNV) in each tumor type were analyzed. The mutated site information of SETDB1 was shown in the amino acid sequence containing conserved domain sites and the 3D structure by the "Mutations" module. The Sangerbox 3.0 online serve was used to comprehensively analyze the mutational landscape of SETDB1. The GSCA database (http://bioinfo.life.hust .edu.cn/GSCA/#/) was used to analyze the CNV percentage

in each cancer and the relationship between SETDB1 expression and CNV. The Catalogue Of Somatic Mutations In Cancer (COSMIC) (https://cancer.sanger.ac.uk/cosmic) is also a comprehensive alteration analysis database for exploring the mutation of SETDB1 in human cancers.

2.6. Methylation and Protein Phosphorylation Analysis. We first assessed the differential expression of SETDB1 promoter methylation in tumor tissues and normal tissues with the UALCAN online database. Furthermore, the relationship between SETDB1 expression and RNA modification-related genes was explored using the Sangerbox 3.0 online service. The MethSurv [18] (https://biit.cs.ut.ee/methsurv/) online database was used to obtain the relative expression level of single CpGs of SETDB1 methylation and their prognostic value. MethSurv database is specifically designed to compare the relative expression level of a single CpG and perform multivariable survival analysis using DNA methylation data. The prognostic value of single CpG of SETDB1 in 25 cancers was also assessed using the "all cancers" and "single CpG" modules of the MethSurv database.

Subsequently, the PhosphoSitePlus (version 6.6.0.2, https://www.phosphosite.org/) was used to investigate the protein phosphorylation sites of SETDB1 in amino acid sequence. The UALCAN online database was also used to compare SETDB1 phosphorylation levels between tumor tissues and paracancerous tissues. The protein phosphorylation data were sourced from the Clinical Proteomic Tumor Analysis Consortium (CPTAC) (https://proteomics.cancer.gov/programs/cptac), and BRCA, glioblastoma multiforme (GBM), PAAD, head and neck squamous cell carcinoma (HNSC), and LUAD.

2.7. Immune and Molecular Subtype Analysis, Immune Infiltration Analysis, and Immune Checkpoint Inhibitor-Related Gene Analysis of SETDB1. We then logged into the TISIDB [19] (http://cis.hku.hk/TISIDB/index.php) database with submitting "SETDB1" to assess the association of SETDB1 expression status with immune cell and molecular subtypes in various human cancers. The "Immune-Gene" module of the TIMER2, which is specifically designed to analyze the immune infiltration across all TCGA cancers, was applied to explore the association between the SETDB1 expression and tumor-related immune cell infiltration levels. The SETDB1 expression was related to the abundance of tumor-infiltrating cells, including cancer-associated fibroblasts (CAFs), CD8$^+$ T cells, CD4$^+$ T cells, regulatory T cells (Tregs), and B cells. The scatterplots were used to present the correlation between SETDB1 mRNA expression and the abundance of infiltrating CAFs. The TISIDB database was also used to analyze the association of SETDB1 expression with immune checkpoint inhibitor-related genes, including immunoinhibitor and immunostimulator. Part of the results with significant correlation was presented as scatterplots. To identify potential groups that may benefit from immunotherapy, we used the radar chart to display the association between microsatellite instability (MSI) and tumor mutational burden (TMB) in various cancers based on the Sangerbox 3.0.

2.8. Function and Pathway Analysis. The STRING (https://string-db.org/) database was used to explore the targeting gene-binding proteins by searching protein name "SETDB1" and organism type "Homo sapiens." By setting parameters of STRING, the experimentally determined SETDB1-binding proteins were obtained. Using GeneMANIA (http://genemania.org/) online database, we predicted the possible function of SETDB1-related genes according to their association with genes with assigned biological functions. To clarify the functions of the target genes, the Gene Ontology (GO) functional enrichment analysis and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis of SETDB1-related genes were performed by R (version R 3.6.3, https://www.r-project.org/). Furthermore, the HALLMARK terms were analyzed by Sangerbox 3.0. In order to construct the mRNA-miRNA-lncRNA network, we first predicted the miRNA targeting SETDB1 based on TargetScanHuman (Release 7.2 March 2018, http://www.targetscan.org/), mirDIP (http://ophid.utoronto.ca/mirDIP/), and miRWalk (http://mirwalk.umm.uni-heidelberg.de/). Then, the complementary sequences of SETDB1 and miRNA targeting SETDB1 were displayed using the TargetScan-Human database. Finally, the lncRNA targeting miRNAs were predicted by the LncBase Predicted v2 module of DIANA tools (http://carolina.imis.athena-innovation.gr/diana_tools/web/index.php?r=site/index).

2.9. Drug Sensitivity and Resistance Analysis. The drug sensitivity analysis and resistance analysis were performed based on the Genomics of Drug Sensitivity in Cancer (GDSC) (https://www.cancerrxgene.org/) database, and the volcano plot was displayed. The IC50 values of Ara-G and Bleomycin (50 μM) for SETDB1 mutation were analyzed. GSCALite (http://bioinfo.life.hust.edu.cn/web/GSCALite/) is a comprehensive web-based analysis platform for gene set cancer analysis and drug sensitivity analysis. DrugBank (https://www.drugbank.com/) database was used to explore the chemical formula and structural formula for Ara-G and Bleomycin (50 μM).

2.10. Validating Expression of SETDB1 by IHC. Six pairs of paraffin-embedded digestive system tumors, including LIHC, CHOL, COAD, ESCA, PAAD, and STAD, and corresponding adjacent tissues were collected in the Shulan (Hangzhou) Hospital. Collected tissues were embedded in paraffin and sliced into 4 μm sections, then baked in an oven at 65°C for 2 hours, and hydrated. These tissues were incubated with 1 : 25 dilution of anti-SETDB1 monoclonal antibody (catalog number: KHC0067). After incubation with the anti-rabbit secondary antibody (ORIGENE) at room temperature for 1 h, diaminobenzidine (DAB) was used to reveal the color of antibody staining. Finally, the stained sections were observed under the microscope.

3. Results

3.1. Multiomics Analysis of SETDB1. The details of the pan-cancer analysis are summarized and presented in Figure 1. This study was aimed at investigating the oncogenic role of
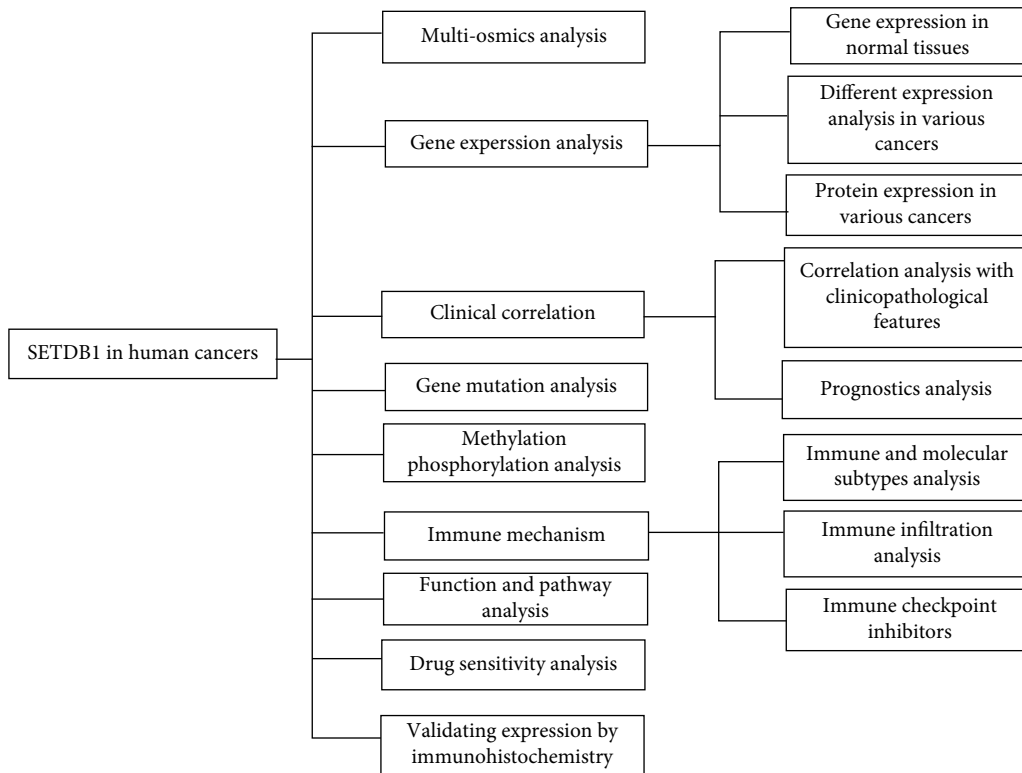
FIGURE 1: The workflow of the study.

SETDB1 in human cancers. SETDB1 (Gene ID: 9869) is located at 1q21.3 and contains 23 exons (Figure 2(a)). The CDS of SETDB1 in nucleotide sequence was displayed (Figure 2(a)). The SETDB1 encoded six protein isoforms, including histone-lysine N-methyltransferase SETDB1 isoforms 1-6, which were mainly distributed in the nucleoplasm (Figure S1A). The mRNA, protein reference sequences (Refseq), and the conserved domains of SETDB1 were summarized (Table 1). Histone-lysine N-methyltransferase SETDB1 isoform 1 was the dominant isoform and the main undertaker of histone-lysine N-methyltransferase SETDB1 functions. To better understand the biological function and structural information of histone-lysine N-methyltransferase SETDB1 protein isoform 1, structure-function analysis was conducted, and the protein domains, region, and nucleotide compositional bias were displayed. As shown in Figure 2(b), the protein structure of SETDB1 consists of six domains, five regions, one coiled coil, and ten nucleotide compositional biases. For multiple species, the SETDB1 contains six domains, including two Tudors (cl02573), MBD (cl00110), pre-SET (cl02622), SET (cl02566), post-SET, and SEEEED (cl19208) (Figure 2(d)). The pre-SET, SET, and post-SET domains are required for methyltransferase activity. Additionally, the protein 3D structure was also displayed (Figure 2(c)). According to the NCBI online database, the SETDB1 protein was conserved in different species, such as chimpanzee, rhesus monkey, dog, cow, mouse, rat, chicken, zebrafish, and frog (Figure S1B). The phylogenetic tree of SETDB1 protein was produced using the fast minimum evolution, and it

presented the evolutionary relationship among different species (Figure S1C). We also found that SETDB1 protein was mainly localized in the nucleoplasm of A-431 (human epithelial carcinoma cell line), U-2 OS (human osteosarcoma cells), and U-251 MG (human brain glioblastoma astrocytoma cancer cells) and vesicles of U-2 OS cell lines (Figure S1D).

3.2. Gene Expression Analysis of SETDB1. We first confirmed that SETDB1 was widely expressed in human normal and tumor tissues (Figure S2A and Table S1). Then, the SETDB1 mRNA expression levels were compared in nontumor tissues based on the HPA and GTEx database (Figure 3(a) and Figure S2B). The bar charts showed that SETDB1 had the highest expression level in the testis, followed by the thymus, tonsil, spleen, and lymph node, indicating that the SETDB1 was mainly expressed in the bone marrow and lymphoid tissues. The SETDB1 expression level was high in most normal tissues, indicating the low tissue specificity of the SETDB1 mRNA expression. Additionally, the IHC and H&E staining results of the top five normal tissues in terms of SETDB1 expression level were displayed based on HPA online database (Figures 3(b) and 3(c)). No data related to the IHC and H&E staining assessment of SETDB1 in thymus tissues were obtained. The expression levels of SETDB1 in different cell lines were assessed. The results from the HPA database showed that SETDB1 was significantly enriched in U-698, followed by the BEWO, THP-1, NTERA-2, and SH-SY5Y (Figure 3(d)). In addition, SETDB1 single cell

UCSC genome browser on human

Chr 1

chrl:150,926,263-150,964,744(GRCh38/hg38)

SETDB1

1

38475

5′UTR

3′UTR

Coding sequency

(a)

Coiled coil   Tudor   MBD   Pre–SET   SET   Post–SET

1

1291

S-adenosyl-L-methonine binding

Disordered

Basic and acidic residues   Pro residues   Acidic residues

Polar residues

(b)

Front   Top   Side

(c)

SETDB1, *H.sapiens*
SET domain, bifurcated 1
SETDB1, *P.troglodytes*
SET domain, bifurcated 1
SETDB1, *M.mulatta*
SET domain, bifurcated 1
SETDB1, *C.lupus*
SET domain, bifurcated 1
SETDB1, *B.taurus*
SET domain, bifurcated 1

NP_001138887.1
1291 aa
XP_524864.3
1291 aa
XP_001106711.2
1290 aa
XP_540304.0
1305 aa
NP_001178317.1
1290 aa

Setdb1, *M.musculus*
SET domain, bifurcated 1
Setdb1, *R.norvegicus*
SET domain, bifurcated 1
LOC101751150, *G.gallus*
Histone-lysine N-methyltransferase
Setdb1, *X.tropicalls*
SET domain, bifurcated 1
Setdb1b, *D.rerio*
SET domain, bifurcated 1b

NP_061365.3
1308 aa
NP_001258104.1
1302 aa
XP_004948398.1
917 aa
NP_001238764.1
1284 aa
NP_001070745.2
1214 aa

TUDOR (c102573)
MBD (c100110)
Pre-SET (pfam05033)
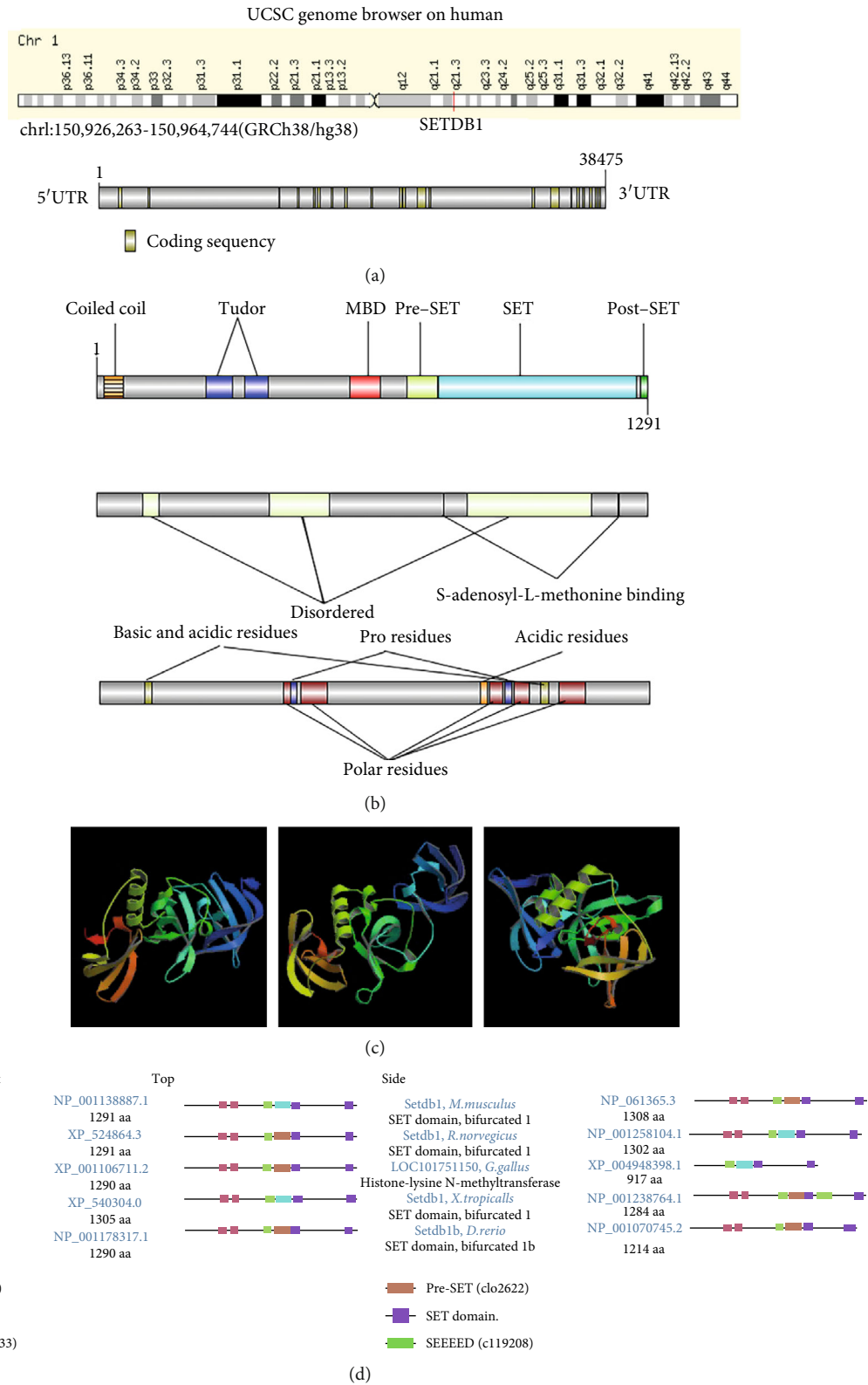
Pre-SET (clo2622)
SET domain.
SEEEED (c119208)

(d)

FIGURE 2: Gene structure, protein structure, and conserved domain of SETDB1. (a) Chromosome localization and gene coding sequence (CDS) of SETDB1 in human. (b) The six domains, five regions, one coiled coil, and ten nucleotide compositional biases of protein structure of SETDB1. (c) The protein structure of SETDB1 gene. (d) Conservation of SETDB1 protein among different species.

TABLE 1: The mRNA and protein information of SETDB1.

| mRNA Accession number	Protein Accession number	Protein isoform	Conserved domains		
			Superfamily	Location	Conserved domain
NM_001145415.2	NP_001138887.1	Histone-lysine N-methyltransferase SETDB1 isoform 1	smart00391	597 → 672	MBD
NM_001393960.1	NP_001380889.1		cd10517	674 → 875	SET_SETDB1
			cl40432	1202 → 1291	SET
			pfam18300	193 → 250	DUF5604
			cd20382	260 → 341	Tudor_SETDB1_rpt1
			cd21181	348 → 401	Tudor_SETDB1_rpt2
			smart00391	597 → 672	MBD
NM_012432.4	NP_036564.3	Histone-lysine N-methyltransferase SETDB1 isoform 2	smart00468	679 → 786	PreSET
			smart00333	348 → 400	TUDOR
			smart00317	1206 → 1267	SET
NM_001243491.2	NP_001230420.1	Histone-lysine N-methyltransferase SETDB1 isoform 3	pfam18300	193 → 250	DUF5604
NM_001393964.1	NP_001380893.1		cd20382	260 → 341	Tudor_SETDB1_rpt1
NM_001393965.1	NP_001380894.1		cl02573	348 → 380	Tudor_SF
NM_001393966.1	NP_001380895.1		smart00391	598 → 673	MBD
NM_001366417.1	NP_001353346.1	Histone-lysine N-methyltransferase SETDB1 isoform 4	cd10517	675 → 876	SET_SETDB1
NM_001366418.1	NP_001353347.1		cl40432	1203 → 1292	SET
NM_001393958.1	NP_001380887.1		pfam18300	193 → 250	DUF5604
NM_001393959.1	NP_001380888.1		cd20382	260 → 341	Tudor_SETDB1_rpt1
			cd21181	348 → 401	Tudor_SETDB1_rpt2
			smart00391	598 → 673	MBD
NM_001393961.1	NP_001380890.1	Histone-lysine N-methyltransferase SETDB1 isoform 5	smart00333	348 → 400	TUDOR
			cd10517	675 → 876	SET_SETDB1
			cl40432	1203 → 1291	SET
NM_001393967.1 NM_001393968.1	NP_001380896.1 NP_001380897.1	Histone-lysine N-methyltransferase SETDB1 isoform 6	pfam18300	193 → 225	DUF5604

MBD: methyl-CpG binding domain; SET_SETDB1: SET domain (including pre-SET and post-SET domains) found in SET domain bifurcated 1 (SETDB1) and similar proteins; SET: SET (Su(var)3-9, Enhancer-of-zeste, Trithorax) domain superfamily; DUF5604: domain of unknown function (DUF5604); Tudor_SETDB1_rpt1: first Tudor domain found in SET domain bifurcated 1 (SETDB1) and similar proteins; Tudor_SETDB1_rpt2: second Tudor domain found in SET domain bifurcated 1 (SETDB1) and similar proteins; TUDOR: Tudor domain; Tudor_SF: Tudor domain superfamily.

(a)

IHC staining

Testis tissues   Tonsil tissues   Spleen tissues   Lymph nodes tissues
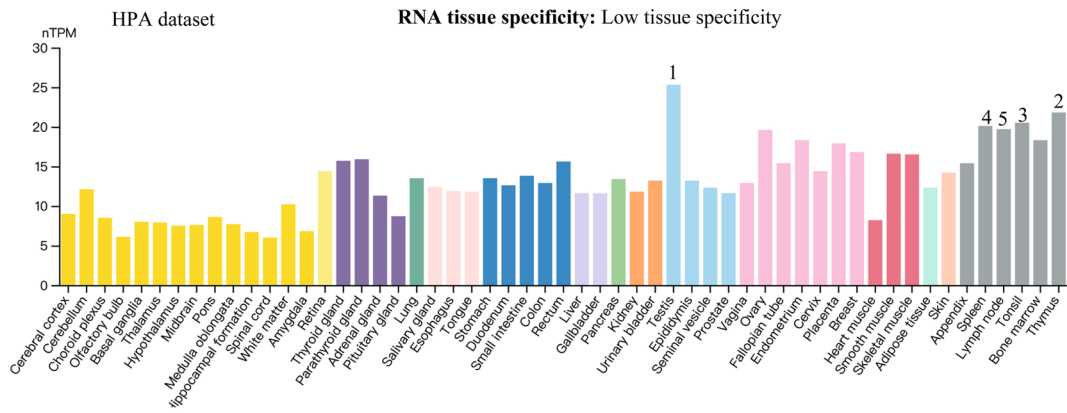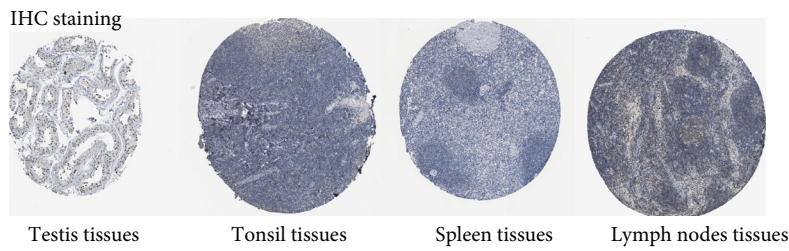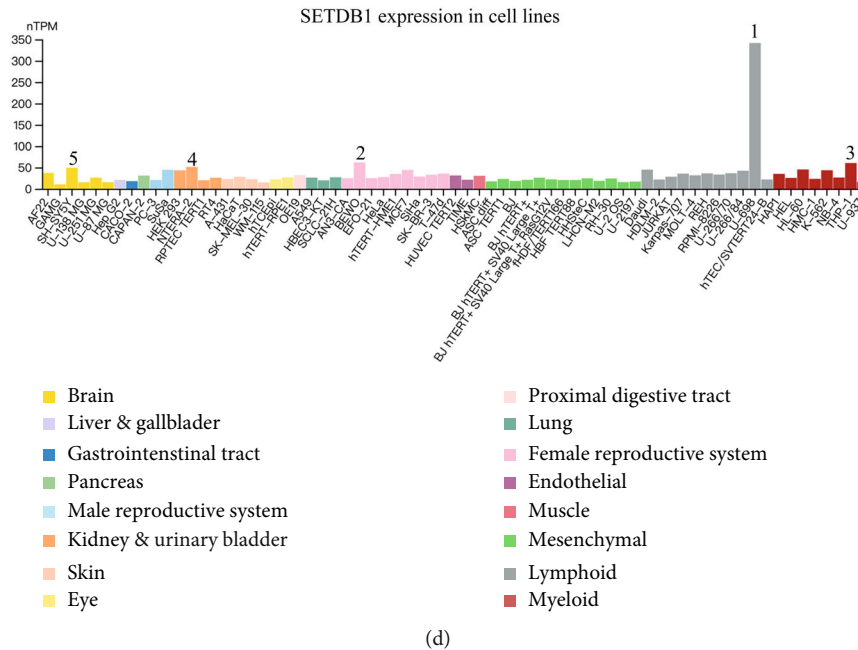
(b)
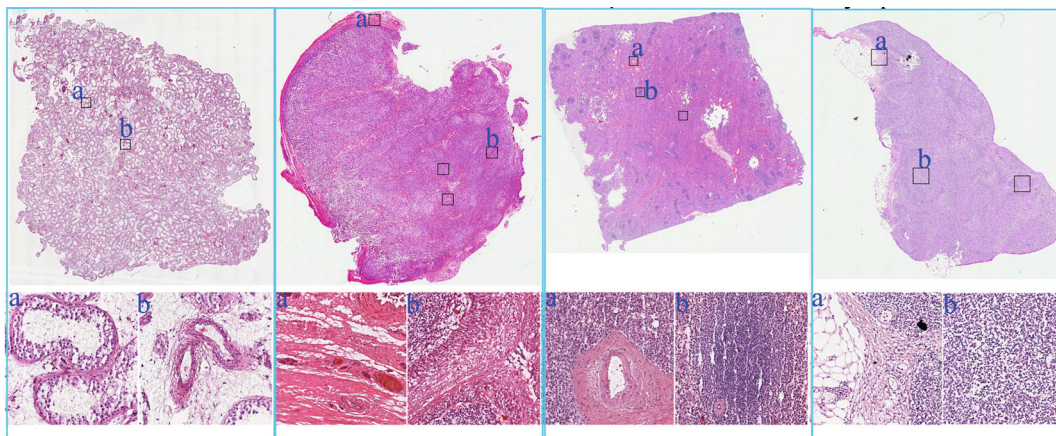
(c)

FIGURE 3: Continued.

FIGURE 3: The expression levels of SETDB1 in normal tissues and cell lines. (a) The mRNA expression levels of SETDB1 in normal tissues (data from HPA dataset). (b) Immunohistochemistry results of SETDB1 assessed in testis tissues, tonsil tissues, spleen tissues, and lymph node tissues. (c) Hematoxylin-eosin results of SETDB1 assessed in testis tissues, tonsil tissues, spleen tissues, and lymph node tissues. (d) The mRNA expression levels of SETDB1 in cell lines.

specificity is displayed in Figure S2C. The SETDB1 expression level was significantly higher in late spermatids, early spermatids, spermatocytes, oligodendrocytes, and microglial cells. Finally, the SETDB1 expression patterns in testis tissues were assessed using published RNA-sequencing data (Figure S2D).

In order to further explore the expression levels of SETDB1 in different tumor tissues and paracancerous tissues, we further performed the differential expression analysis using several online databases. In the TIMER database, the SETDB1 expression level was elevated in bladder urothelial carcinoma (BLCA), BRCA, CHOL, colon adenocarcinoma (COAD), esophageal carcinoma (ESCA), GBM, HNSC, KIRC, LIHC, LUAD, lung squamous cell carcinoma (LUSC), rectum adenocarcinoma (READ), stomach adenocarcinoma (STAD), thyroid carcinoma (THCA), and uterine corpus endometrial carcinoma (UCEC) (Figure 4(a)). However, compared with the SETDB1 expression level in paracancerous tissue, that in KICH (kidney chromophobe) was lower (Figure 4(a)). The mRNA expression levels of SETDB1 in HNSC-HPV$^+$ tumor and SKCM-Metastasis tissues were higher than those in HNSC-HPV$^-$ tumor and SKCM primary tumor tissues (Figure 4(a)). These data are in agreement with the expression levels of SETDB1 in tumor tissues and paracancerous tissues in Sangerbox and UALCAN databases (Figure S3). The expression of SETDB1 in SKCM was lower in patients with primary tumors than that in patients with metastasis tumors, indicating that a high expression level of SETDB1 in SKCM may imply metastasis (Figure 4(a)). In the GEPIA2 online database, combined with the TCGA and GTEx datasets, the

differential expression analysis showed that SETDB1 was highly expressed in the thymoma (THYM), lower grade glioma (LGG), acute myeloid leukemia (LAML), GBM, lymphoid neoplasm diffuse large B-cell lymphoma (DLBC), and CHOL, while it was lowly expressed in THCA, prostate adenocarcinoma (PRAD), KICH, and adrenocortical carcinoma (ACC) and was not expressed in other cancers (Figure 4(b)).

For acquiring more comprehensive expression information, HPA and UALCAN online databases were combined to assess the protein expression of SETDB1 in various cancers and normal tissues. In HPA online dataset, the protein of SETDB1 was observed in 45 normal tissue samples. Among them, 15 samples showed a high expression score, nine samples exhibited a medium expression score, 11 samples exhibited a low expression score, and 10 samples had no expression score (Figure 5(a)). In tumor samples, most cancers were weakly stained or negative. Moderate to strong nuclear and cytoplasmic positivity was observed in several gliomas, lymphomas, melanomas, colorectal cancer, endometrial cancer, and testicular cancer (Figure 5(b)). Furthermore, the classic IHC staining was performed. The results showed that SETDB1 had high expression levels in brain glioma and Hodgkin's lymphoma, had medium levels in THCA and BLCA, and had a low level in endometrium adenocarcinoma (Figure 5(c)). In the "CPTAC" module of UALCAN, we accessed the differences in protein expression of SETDB1 between various cancers and paracancerous tissues. Compared with SETDB1 protein expression levels in the adjacent normal tissues, those in BRCA ($p = 2.85E - 12$), OV ($p = 1.05E - 05$), KIRC ($p = 7.92E - 45$), UCEC
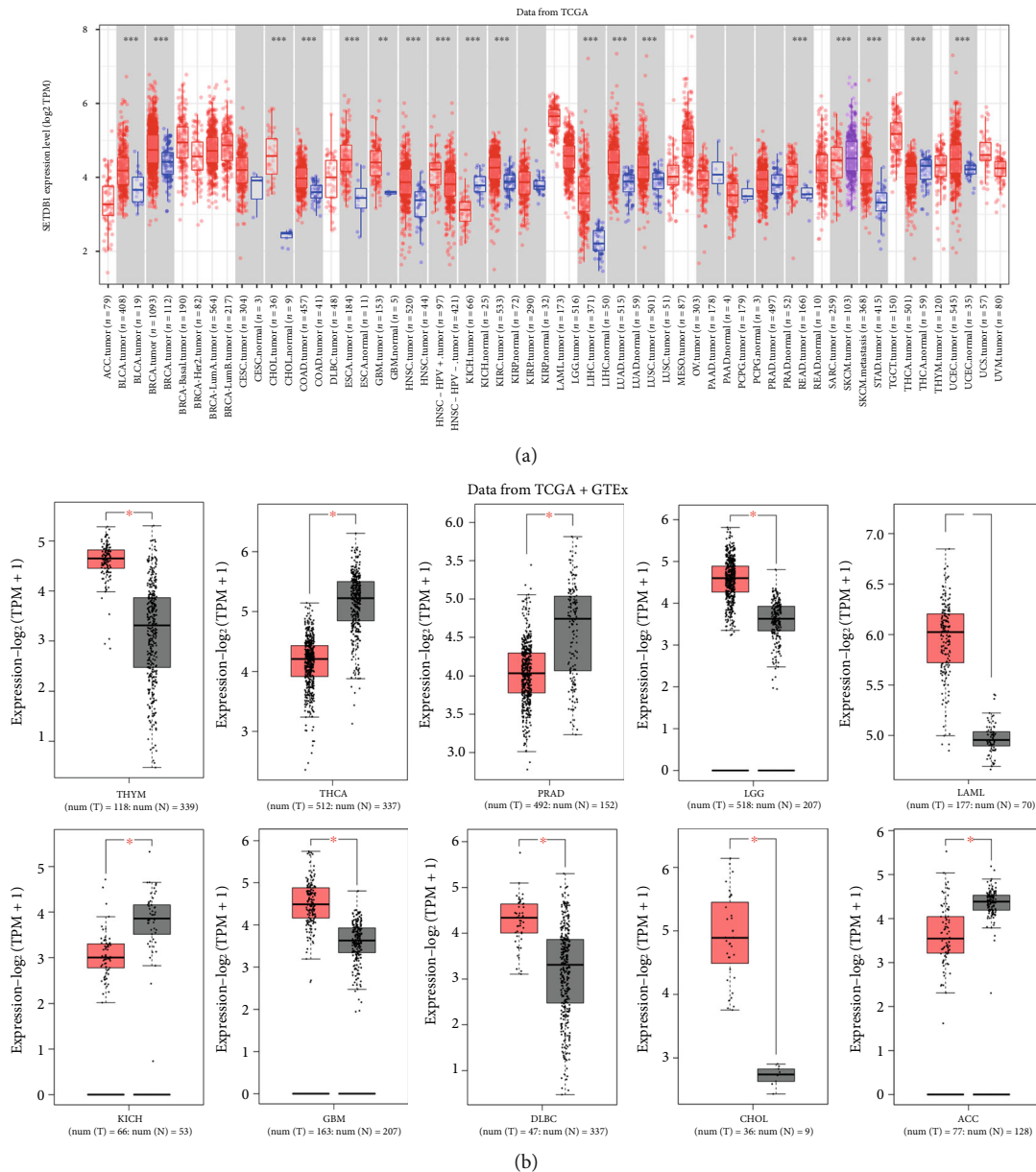
(a)

(b)

FIGURE 4: The mRNA expression levels of SETDB1 between cancer and normal tissues. (a) SETDB1 mRNA expression levels in different tumor types and corresponding normal tissue from TCGA datasets. (b) The differences in expression levels of SETDB1 in different tumors and normal tissues from TCGA and GTEx datasets.

($p = 1.91E - 12$), LUAD ($p = 2.91E - 31$), HNSC ($p = 2.21E - 08$), PAAD ($p = 3.71E - 02$), GBM ($p = 2.65E - 19$), and LIHC ($p = 1.56E - 41$) were significantly higher (Figure 5(d)).
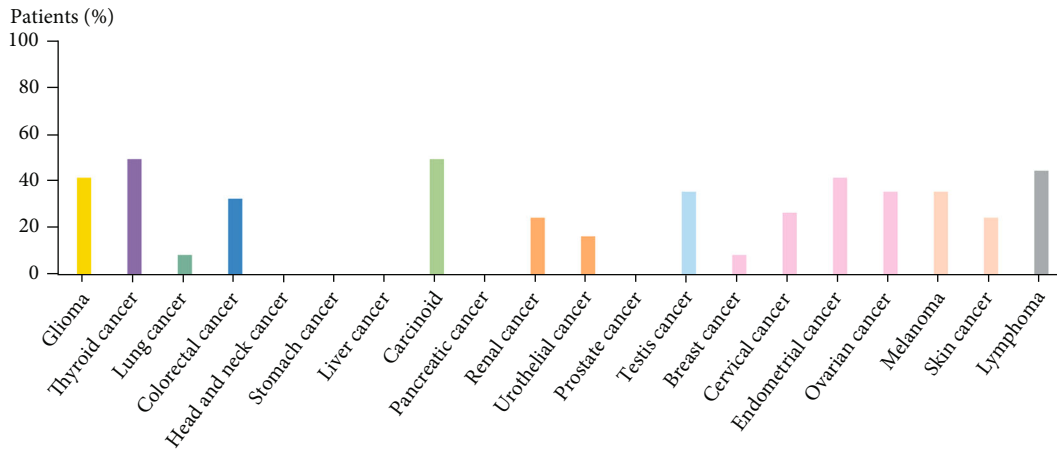
3.3. Association of SETDB1 Expression with Clinicopathological Features.
GEPIA2 and UALCAN online databases were also applied to assess the association between SETDB1 expression and the clinicopathological stages of various cancers. The results showed that SETDB1 expression was significantly related to clinicopathological stages of tumors, including PAAD ($p = 0.0444$), LIHC ($p = 0.0132$), KICH ($p = 0.00939$), and testicular germ cell tumors (TGCT) ($p = 0.0468$) (Figure 6(a)). Additionally, in the UALCAN online databases, we found that 13 tumors at dif-

ferent clinicopathological stages showed higher SETDB1 expression levels than corresponding normal tissues, and these tumors are BLCA, BRCA, CHOL, COAD, ESCA, HNSC, kidney renal clear cell carcinoma (KIRC), LIHC, LUAD, LUSC, READ, STAD, and UCEC (Figures 6(b)–6(n)). These findings suggested that SETDB1 may be an oncogenic gene in these human malignant tumors. In contrast, the SETDB1 expression level was higher in normal tissues than that in three tumor tissues, namely, KICH, PAAD, and THCA, at clinicopathological stages (Figures 6(o)–6(q)). The results indicate that SETDB1 may be a protective factor in these tumors. Furthermore, for uveal melanoma (UVM) and TGCT patients, the SETDB1 expression levels in cancers at advanced clinicopathological stages were significantly
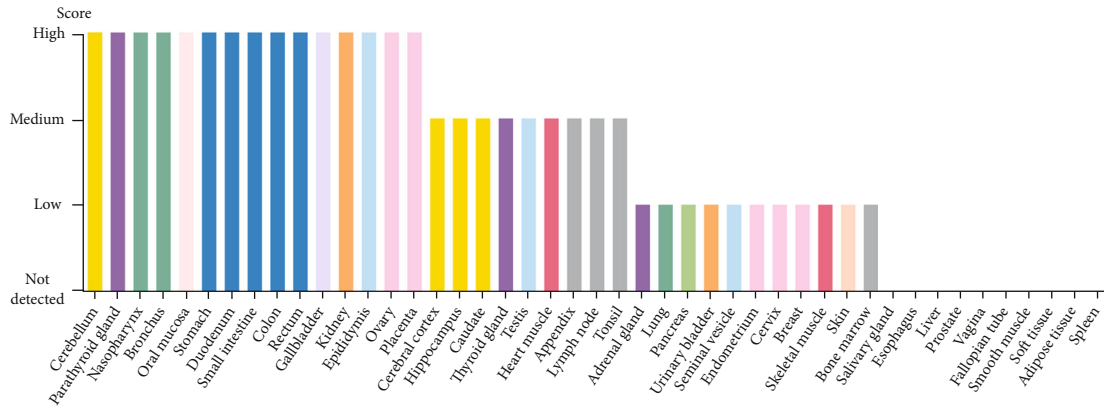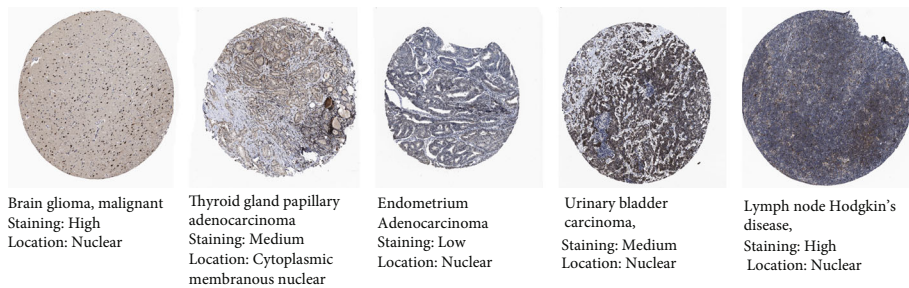
(a)

Most cancers were weakly stained or negative. Moderate to strong nuclear and/or cytoplasmic positivity was observed in several gliomas, lymphomas, melanomas, colorectal, endometrial and testicular cancers.

(b)

Brain glioma, malignant
Staining: High
Location: Nuclear

Thyroid gland papillary
adenocarcinoma
Staining: Medium
Location: Cytoplasmic
membranous nuclear

Endometrium
Adenocarcinoma
Staining: Low
Location: Nuclear

Urinary bladder
carcinoma,
Staining: Medium
Location: Nuclear

Lymph node Hodgkin's
disease,
Staining: High
 Location: Nuclear
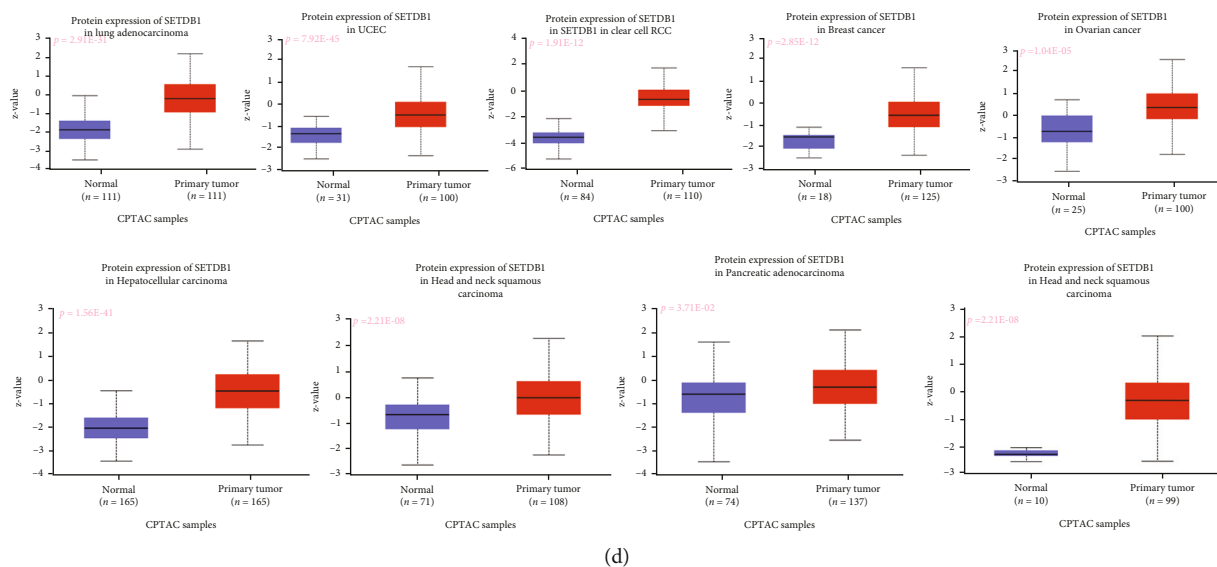
(c)

Figure 5: Continued.

(d)

Figure 5: The protein expression levels SETDB1 in normal and cancer tissues. (a) SETDB1 protein expression data in 44 normal tissues. (b) The percentage of cancer patients (maximum 12 patients) with high and medium protein expression level (HPA018142). (c) Immunohistochemistry results of SETDB1 protein assessed in brain glioma, papillary adenocarcinoma, endometrium adenocarcinoma, urinary bladder carcinoma, and lymph node Hodgkin's disease. (d) The differences in protein expression levels of SETDB1 in different tumors and normal tissues from TCGA.
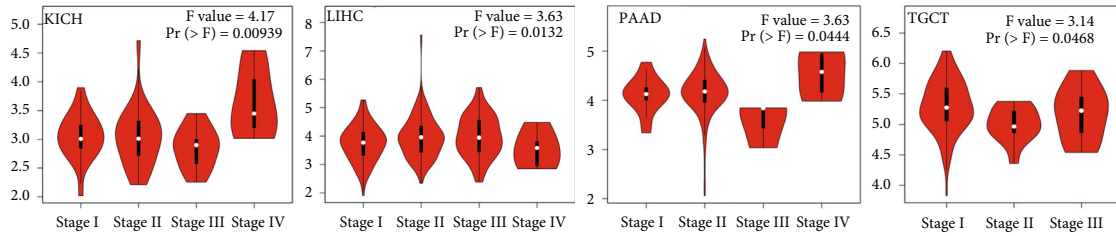
lower than those in tumors at early stages, implying that decreased SETDB1 expression may indicate the tumor progression in these patients (Figures 6(r) and 6(s)). These results demonstrated that abnormal expression of SETDB1 may be associated with the initiation and progression of various human cancers.

We further investigated the association of SETDB1 expression with other clinicopathological features, such as TNM classification and clinicopathological grade based on Sangerbox 3.0. SETDB1 expression in GBM, BRCA, ESCA, SARC (Sarcoma), and PRAD was significantly associated with tumor T classification (Figure 6(t)). SETDB1 expression is significantly associated with N classification in LGG, LIHC, and READ and is related to M classification in LUAD, PAAD, and UVM (Figures 6(u) and 6(v)). Additionally, the clinicopathological grade of LUSC, UCS (uterine carcinosarcoma), and SARC is associated with SETDB1 expression (Figure 6(w)).
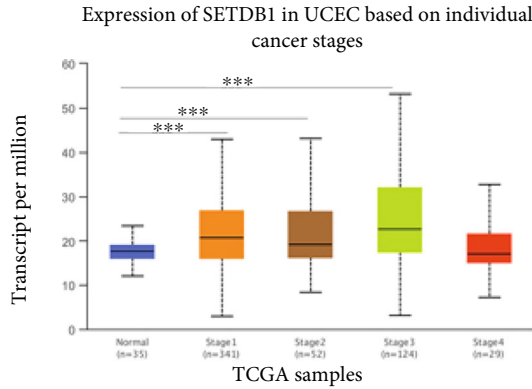
3.4. Survival Analysis of SETDB1. To explore the prognostic value of SETDB1 in various human cancers, we classified the cancer samples into high- and low-expression subgroups according to the median expression value of SETDB1. First, the GEPIA2 database was used to perform the OS and DFS analyses in pan-cancer cohorts. In terms of OS, higher expression of SETDB1 was associated with poorer clinical outcomes in ACC ($p = 0.0055$) and LIHC ($p = 0.0290$) (Figure 7(a)). The results also revealed a correlation between high SETDB1 expression levels and poor DFS in ACC ($p = 3.9e − 05$), READ ($p = 0.0180$), and PRAD ($p = 0.0150$) (Figure 7(a)). It is indicated that abnormally expressed SETDB1 may be a prognostic indicator in these tumors. For acquiring a more comprehensive prognostic value of SETDB1, the Cox analyses based on Sangerbox database

were performed for DFI and DSS of various cancer samples. The results revealed that SETDB1 expression influenced DFI in patients with ACC ($p = 4.1e − 7$, HR = 3.27, 95% CI 2.07-5.16), PRAD ($p = 2.1e − 3 2$, HR = 2.14, 95% CI 1.32-3.48), KIPAN ($p = 3.5e − 3$, HR = 1.36, 95% CI 1.11-1.67), KICH ($p = 0.01$, HR = 4.38, 95% CI 1.43-13.43), LIHC ($p = 0.02$, HR = 1.24, 95% CI 1.03-1.49), and cervical squamous cell carcinoma and endocervical adenocarcinoma (CESC) ($p = 0.04$, HR = 1.66, 95% CI 1.03-2.66), and READ ($p = 0.04$, HR = 3.90, 95% CI 1.06-14.30) (Figure 7(b)). However, abnormally expressed SETDB1 was associated with DSS for patients with ACC ($p = 5.1e − 5$, HR = 2.91, 95% CI 1.75-4.85), KICH ($p = 2.1e − 4$, HR = 34.02, 95% CI 4.43-260.98), pheochromocytoma and paraganglioma (PCPG) ($p = 1.4e − 3$, HR = 61.75, 95% CI 3.38-1127.99), THCA ($p = 7.6e − 3$, HR = 26.07, 95% CI 2.86-237.32), KIPAN ($p = 0.01$, HR = 1.38, 95% CI 1.07-1.79), LUSC ($p = 0.02$, HR = 1.45, 95% CI 1.07-1.97), and LIHC ($p = 0.04$, HR = 1.33, 95% CI 1.01-1.75) (Figure 7(c)). We focused on the association between SETDB1 expression and the breast cancer, ovarian cancer, lung cancer, and gastric cancer prognosis. In order to evaluate the prognostic abilities of SETDB1 in these cancers, independent clinical factors were selected as the subgroups (Table S2-S6).
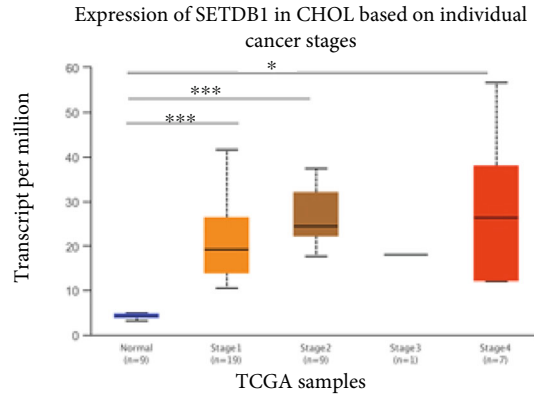
3.5. Genetic Alteration Analysis of SETDB1. Malignant tumor is caused by genetic alterations, and mutated genes offer potential molecular therapeutic targets [20, 21]. Given that SETDB1 genetic alterations were associated with molecular therapeutic targets for various human cancers, we investigated the genetic alteration levels of SETDB1 in various human cancers based on TCGA datasets. The results showed that SETDB1 altered 630 cases (6%) out of 10439 cases (data from PanCancer Atlas and TCGA) (Figure 8A). We also
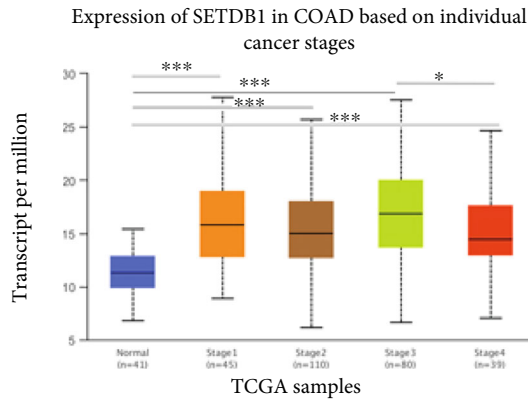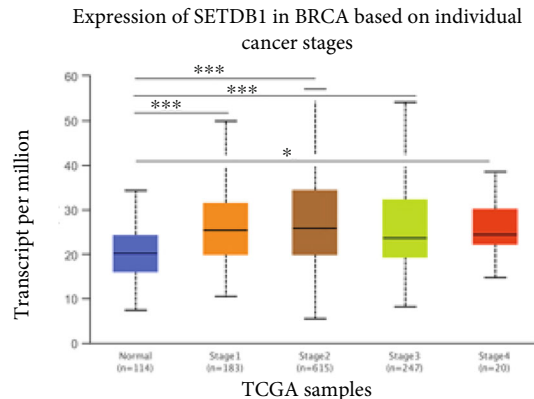
(a)

Expression of SETDB1 in UCEC based on individual cancer stages

(b)

Expression of SETDB1 in CHOL based on individual cancer stages

(c)

Expression of SETDB1 in COAD based on individual cancer stages

(d)

Expression of SETDB1 in BRCA based on individual cancer stages

(e)

Expression of SETDB1 in STAD based on individual cancer stages

(f)

Expression of SETDB1 in READ based on individual cancer stages

(g)

FIGURE 6: Continued.

Expression of SETDB1 in ESCA based on individual cancer stages

(h)

Expression of SETDB1 in LUSC based on individual cancer stages

(i)

Expression of SETDB1 in LUAD based on individual cancer stages

(j)

Expression of SETDB1 in LIHC based on individual cancer stages

(k)

Expression of SETDB1 in KIRC based on individual cancer stages

(l)

Expression of SETDB1 in BLCA based on individual cancer stages

(m)

FIGURE 6: Continued.

Expression of SETDB1 in HNSC based on individual
cancer stages

(n)

Expression of SETDB1 in PAAD based on individual
cancer stages

(o)

Expression of SETDB1 in THCA based on individual
cancer stages

(p)

Expression of SETDB1 in KICH based on individual
cancer stages

(q)

Expression of SETDB1 in UVM based on individual
cancer stages

(r)

Expression of SETDB1 in TCCT based on individual
cancer stages

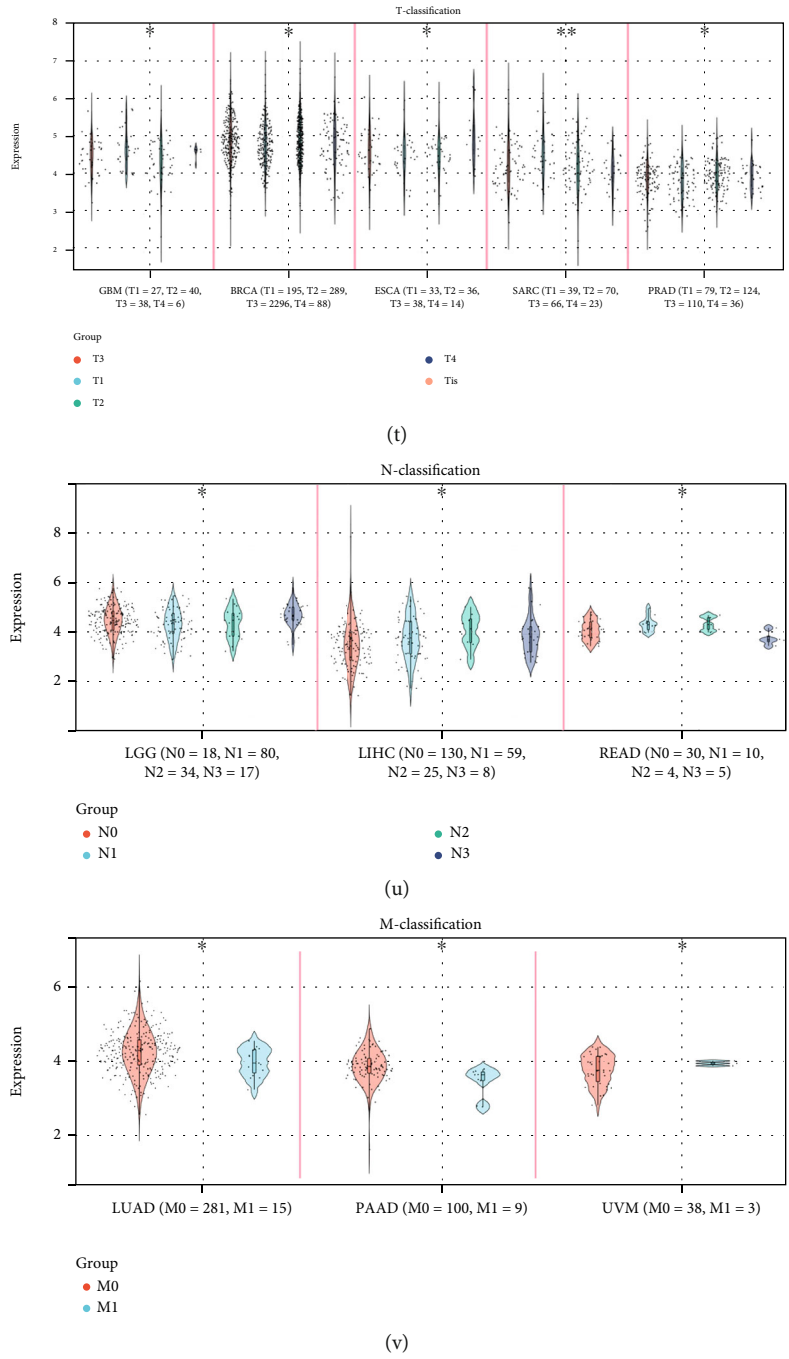
(s)

Figure 6: Continued.

(t)
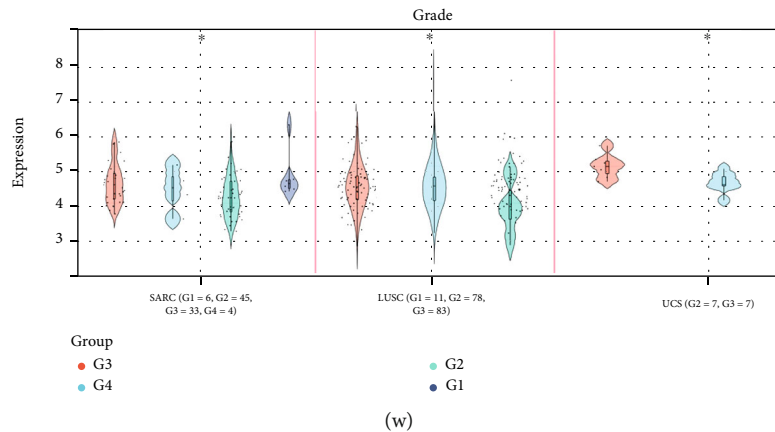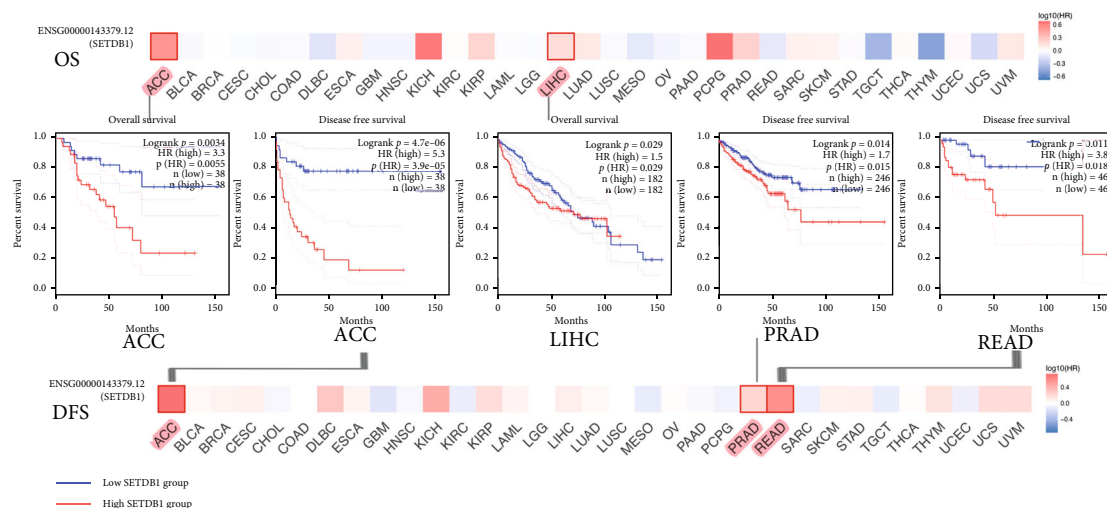

(u)

(v)

Figure 6: Continued.

FIGURE 6: Association analysis of SETDB1 expression with clinicopathological features. (a) The association between SETDB1 expression and pathological clinical stage. (b–s) Correlation between SETDB1 expression and normal as well as pathological clinical stage. (t–w) Correlation between SETDB1 expression and TNM classification (t–v) and clinicopathological grade (w).

found that missense mutation was the main type of SETDB1 mutation, followed by the truncating mutation and splice mutation (Figure 8(a) and Figure S4A). Furthermore, the primary SNV class was C > T (29.81%), followed by G > A (21.12%), A > G (11.47%), and G > T (10.56%) (Figure S4B). By analyzing genetic alterations and expression, we found that SETDB1 genetic alterations induced a switch in the mRNA expression levels of SETDB1 in human tumors. However, few differences were observed in the genetic alteration by deep deletion (Figure S4C). Furthermore, the genetic alteration type in cholangiocarcinoma (CHOL) (8.99% of 523 cases), PCPG (3.37% of 178 cases), DLBC (7.5% of 440 cases), and THYM (0.81% of 123 cases) (Figure 8(b)) was amplification. The genetic alteration type in kidney renal papillary cell carcinoma (KIRP) (1.81% of 276 cases), LAML (0.5% of 200 cases), and THCA (0.2% of 490 cases) (Figure 8(b)) was missense mutation. Additionally, the SETDB1 mutation frequency in patients with mixed endometrial carcinomas was the highest (15.09% of 517 cases), including 8.51% (44 cases) mutation and 6.58% (34 cases) amplification (Figure 8(b)). The structural variant and deep deletion were rare in human cancers and were only identified in five cancers among cancers included in this research, namely, LIHC, BRCA, and SKCM (structural variant) and ESCA and SARC (deep deletion) (Figure 7(b)). We used the "mutation" module of the cBioPortal database to investigate the type and site of SETDB1 mutation (NM_001145415/ENST00000271640) in each sample. The R1256W/L/Q mutation and translation from R (Arginine) to W (Tryptophan) or L (Leucine) or Q (Glutamine) were observed in the SET conserved domain and occurred in one case of GBM (R1256W), one case of skin cutaneous melanoma (SKCM) (R1256L), one case of STAD (R1256W), and two cases of colorectal adenocarcinoma (R1256W and R1256Q). However, the function of R1256W/L/Q mutation remained unknown (Figure 8(d)). The mutation spectrum of SETDB1 was explored by Sangerbox 3.0 (Figure S4D). Finally, the 3D structure of SETDB1 protein and the mutation of the

sequence were displayed (Figure 8(c)). However, the R1256W/L/Q mutation was not displayed in the 3D structure of the SETDB1 protein. The CNV pie chart also showed that the heterozygous amplification of CNV was distributed in most cancers, whereas the heterozygous deletion was predominantly distributed in the KICH (Figure S4E). A significant positive correlation was observed between SETDB1 expression and CNV in various cancers (Figure S4F).
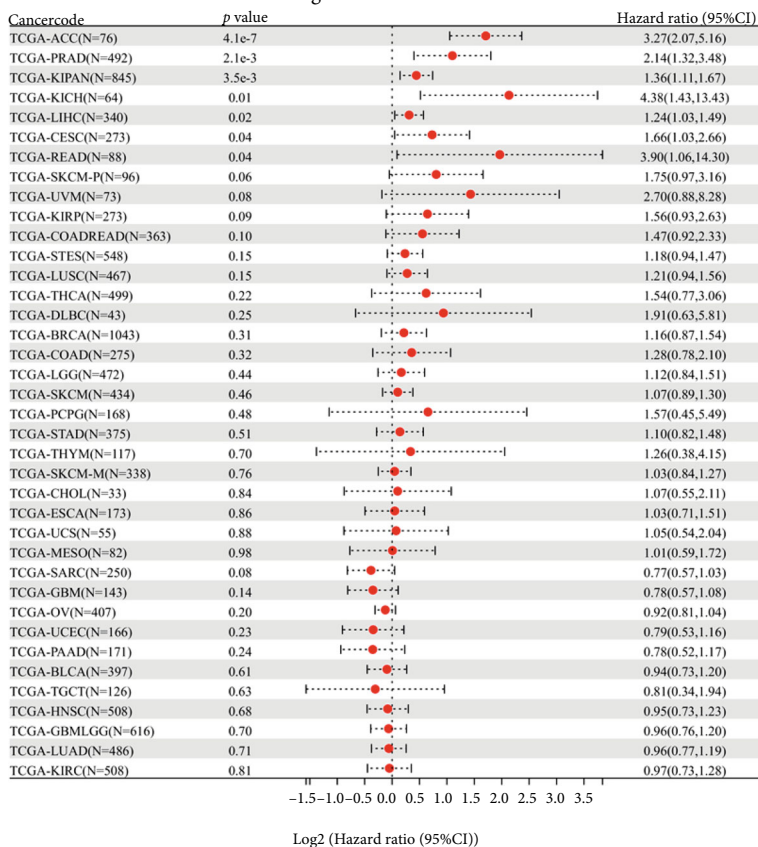
3.6. Methylation Analysis of SETDB1. Growing evidence showed that aberrant methylation was associated with oncogenesis and may have a significant clinical value [22]. Therefore, we assessed the DNA methylation levels of SETDB1 and its prognosis value in various human cancers. Firstly, we compared the levels of SETDB1 promoter methylation in tumors and paracancerous tissues based on the UALCAN database. The results showed that the promoter methylation levels of SETDB1 in BLCA, BRCA, COAD, ESCA, HNSC, LIHC, LUAD, LUSC, PRAD, READ, TGCT, and UCEC were significantly reduced compared with those in paracancerous tissues (Figures 9(a)–9(l)). Correlation analysis showed that SETDB1 expression was significantly positively correlated with RNA modification-related genes (Figure S5). In the MethSurv online database, we evaluated the DNA methylation level and prognostic value of SETDB1 in various human cancers, and the relative methylation level was displayed in Figure S6. It can be seen that cg10444928 site of SETDB1 in 25 human tumors showed the highest DNA methylation level. To analyze the association of the cg10444928 site of SETDB1 with prognosis across various human cancers, we explored the prognosis value of single CpG (cg10444928) of SETDB1 based on the "single CpG" module of MethSurv database. The results showed that cg10444928 of SETDB1 was significantly associated with the prognosis of UCS ($p = 1.50E − 02$, HR = 2.514), UVM ($p = 6.30E − 04$, HR = 0.196), mesothelioma (MESO) ($p = 8.40E − 03$, HR = 1.893), LGG ($p = 7.00E − 03$, HR = 1.628), KIRP ($p = 2.10E − 02$, HR = 2.673), KIRC ($p = 2.30E − 02$, HR = 0.638), and CESC ($p = 0.03$, HR =
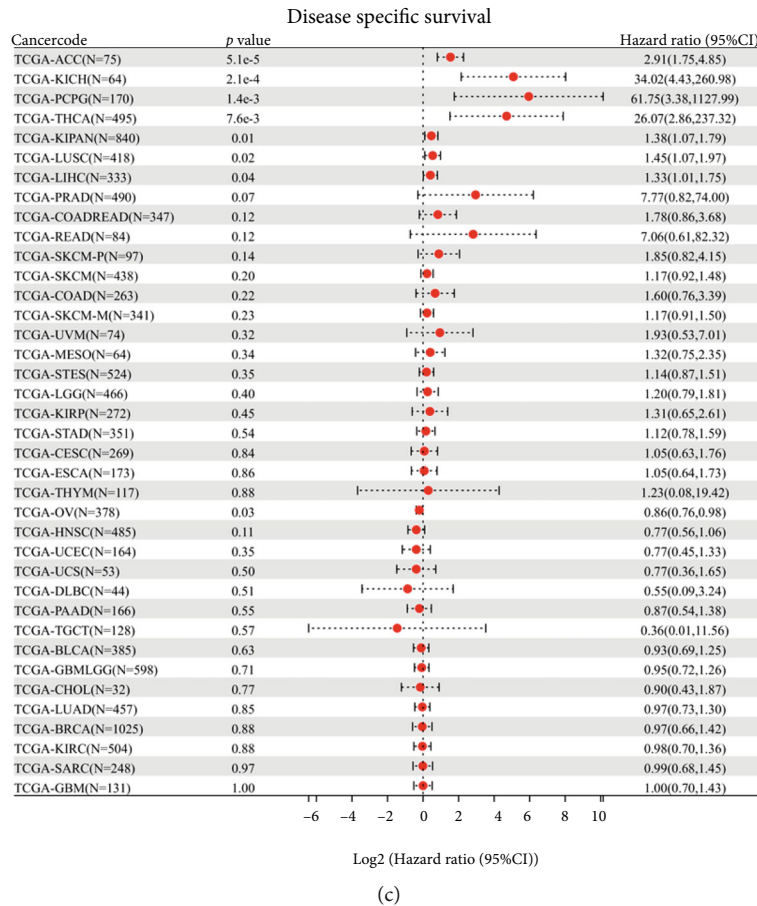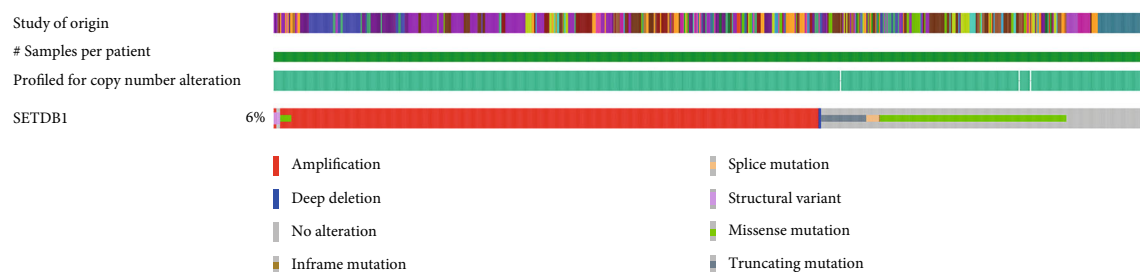
(a)

Progression free survival

Cancercode	p value		Hazard ratio (95%CI)
TCGA-ACC(N=76)	4.1e-7		3.27(2.07,5.16)
TCGA-PRAD(N=492)	2.1e-3		2.14(1.32,3.48)
TCGA-KIPAN(N=845)	3.5e-3		1.36(1.11,1.67)
TCGA-KICH(N=64)	0.01		4.38(1.43,13.43)
TCGA-LIHC(N=340)	0.02		1.24(1.03,1.49)
TCGA-CESC(N=273)	0.04		1.66(1.03,2.66)
TCGA-READ(N=88)	0.04		3.90(1.06,14.30)
TCGA-SKCM-P(N=96)	0.06		1.75(0.97,3.16)
TCGA-UVM(N=73)	0.08		2.70(0.88,8.28)
TCGA-KIRP(N=273)	0.09		1.56(0.93,2.63)
TCGA-COADREAD(N=363)	0.10		1.47(0.92,2.33)
TCGA-STES(N=548)	0.15		1.18(0.94,1.47)
TCGA-LUSC(N=467)	0.15		1.21(0.94,1.56)
TCGA-THCA(N=499)	0.22		1.54(0.77,3.06)
TCGA-DLBC(N=43)	0.25		1.91(0.63,5.81)
TCGA-BRCA(N=1043)	0.31		1.16(0.87,1.54)
TCGA-COAD(N=275)	0.32		1.28(0.78,2.10)
TCGA-LGG(N=472)	0.44		1.12(0.84,1.51)
TCGA-SKCM(N=434)	0.46		1.07(0.89,1.30)
TCGA-PCPG(N=168)	0.48		1.57(0.45,5.49)
TCGA-STAD(N=375)	0.51		1.10(0.82,1.48)
TCGA-THYM(N=117)	0.70		1.26(0.38,4.15)
TCGA-SKCM-M(N=338)	0.76		1.03(0.84,1.27)
TCGA-CHOL(N=33)	0.84		1.07(0.55,2.11)
TCGA-ESCA(N=173)	0.86		1.03(0.71,1.51)
TCGA-UCS(N=55)	0.88		1.05(0.54,2.04)
TCGA-MESO(N=82)	0.98		1.01(0.59,1.72)
TCGA-SARC(N=250)	0.08		0.77(0.57,1.03)
TCGA-GBM(N=143)	0.14		0.78(0.57,1.08)
TCGA-OV(N=407)	0.20		0.92(0.81,1.04)
TCGA-UCEC(N=166)	0.23		0.79(0.53,1.16)
TCGA-PAAD(N=171)	0.24		0.78(0.52,1.17)
TCGA-BLCA(N=397)	0.61		0.94(0.73,1.20)
TCGA-TGCT(N=126)	0.63		0.81(0.34,1.94)
TCGA-HNSC(N=508)	0.68		0.95(0.73,1.23)
TCGA-GBMLGG(N=616)	0.70		0.96(0.76,1.20)
TCGA-LUAD(N=486)	0.71		0.96(0.77,1.19)
TCGA-KIRC(N=508)	0.81		0.97(0.73,1.28)

−1.5 −1.0 −0.5 0.0 0.5 1.0 1.5 2.0 2.5 3.0 3.5

Log2 (Hazard ratio (95%CI))

(b)

Figure 7: Continued.

Disease specific survival

Cancercode	p value		Hazard ratio (95%CI)
TCGA-ACC(N=75)	5.1e-5		2.91(1.75,4.85)
TCGA-KICH(N=64)	2.1e-4		34.02(4.43,260.98)
TCGA-PCPG(N=170)	1.4e-3		61.75(3.38,1127.99)
TCGA-THCA(N=495)	7.6e-3		26.07(2.86,237.32)
TCGA-KIPAN(N=840)	0.01		1.38(1.07,1.79)
TCGA-LUSC(N=418)	0.02		1.45(1.07,1.97)
TCGA-LIHC(N=333)	0.04		1.33(1.01,1.75)
TCGA-PRAD(N=490)	0.07		7.77(0.82,74.00)
TCGA-COADREAD(N=347)	0.12		1.78(0.86,3.68)
TCGA-READ(N=84)	0.12		7.06(0.61,82.32)
TCGA-SKCM-P(N=97)	0.14		1.85(0.82,4.15)
TCGA-SKCM(N=438)	0.20		1.17(0.92,1.48)
TCGA-COAD(N=263)	0.22		1.60(0.76,3.39)
TCGA-SKCM-M(N=341)	0.23		1.17(0.91,1.50)
TCGA-UVM(N=74)	0.32		1.93(0.53,7.01)
TCGA-MESO(N=64)	0.34		1.32(0.75,2.35)
TCGA-STES(N=524)	0.35		1.14(0.87,1.51)
TCGA-LGG(N=466)	0.40		1.20(0.79,1.81)
TCGA-KIRP(N=272)	0.45		1.31(0.65,2.61)
TCGA-STAD(N=351)	0.54		1.12(0.78,1.59)
TCGA-CESC(N=269)	0.84		1.05(0.63,1.76)
TCGA-ESCA(N=173)	0.86		1.05(0.64,1.73)
TCGA-THYM(N=117)	0.88		1.23(0.08,19.42)
TCGA-OV(N=378)	0.03		0.86(0.76,0.98)
TCGA-HNSC(N=485)	0.11		0.77(0.56,1.06)
TCGA-UCEC(N=164)	0.35		0.77(0.45,1.33)
TCGA-UCS(N=53)	0.50		0.77(0.36,1.65)
TCGA-DLBC(N=44)	0.51		0.55(0.09,3.24)
TCGA-PAAD(N=166)	0.55		0.87(0.54,1.38)
TCGA-TGCT(N=128)	0.57		0.36(0.01,11.56)
TCGA-BLCA(N=385)	0.63		0.93(0.69,1.25)
TCGA-GBMLGG(N=598)	0.71		0.95(0.72,1.26)
TCGA-CHOL(N=32)	0.77		0.90(0.43,1.87)
TCGA-LUAD(N=457)	0.85		0.97(0.73,1.30)
TCGA-BRCA(N=1025)	0.88		0.97(0.66,1.42)
TCGA-KIRC(N=504)	0.88		0.98(0.70,1.36)
TCGA-SARC(N=248)	0.97		0.99(0.68,1.45)
TCGA-GBM(N=131)	1.00		1.00(0.70,1.43)

−6   −4   −2   0   2   4   6   8   10

Log2 (Hazard ratio (95%CI))

(c)

FIGURE 7: Correlation between SETDB1 expression and survival prognosis of cancers in TCGA datasets. (a) Overall survival. (b) Disease-free survival. (c) Progression-free survival. (d) Disease-specific survival.

0.526) (Figures 9(m)–9(s)). The prognostic value of other single CpGs of SETDB1 in 25 cancers was also assessed using the "all cancers" module of the MethSurv database (Table S7).

3.7. Protein Phosphorylation Analysis of SETDB1. Protein phosphorylation may be a promoter or a suppressor of oncogenesis. Therefore, exploring protein phosphorylation is beneficial to developing a novel antitumor agent in human tumors [23]. We first explored the protein phosphorylation site of SETDB1 based on the PhosphoSitePlus database. As shown in Figure 9(t), the most predominant protein phosphorylation locus for the SETDB1 is Ser1006 (flanking sequence: RNYGYNPsPVkPEGL) located in the SET conserved domain. Subsequently, we assessed the differences in phosphorylation levels at the single phosphorylation site of SETDB1 between tumor tissues and paracancerous tissues using the CPTAC dataset. The Ser1006 locus of SETDB1 possessed a higher phosphorylation level in BRCA ($p = 1.55E − 08$), GBM ($p = 1.27E − 02$), PAAD ($p = 1.00E − 11$), HNSC ($p = 1.16E − 33$), LUAD ($p = 1.91E − 32$), and KIRC ($p = 8.36E − 17$) (Figures 9(u)–9(z)). These results implied that protein phosphorylation of SETDB1 at Ser1006 locus may play an important role in the development and progres-

sion of those tumors. Our previous findings suggested that SETDB1 protein was mainly located in the nucleoplasm. However, whether the protein phosphorylation of SETDB1 at Ser1006 locus affects its location or its function remains unknown and requires more investigations.
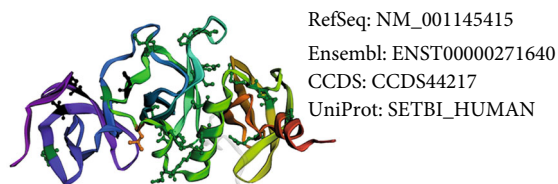
3.8. Immune and Molecular Subtype Analysis of SETDB1. We assessed the relationship between SETDB1 expression status and immune activity as well as molecular subtypes in human cancers based on the TISIDB database. According to immune activity, the tumor tissues were divided into C1 (wound healing), C2 (IFN-gamma dominant), C3 (inflammatory), C4 (lymphocyte depleted), C5 (immunologically quiet), and C6 (TGF-b dominant). In order to verify the dynamic relationship between SETDB1 expression status and immune activity, we assessed the immune activity levels of the six subtypes in various cancers. The results showed that SETDB1 expression was significantly associated with immune subtypes in BLCA, COAD, KICH, KIRC, LIHC, LUAD, LUSC, ovarian serous cystadenocarcinoma (OV), SARC, STAD, and TGCT (Figure 10(a)). Similarly, the SETDB1 expression was associated with molecular subtypes in ACC, BRCA, COAD, GBM, HNSC, KIRP, LGG, LUSC OV, PCPG, PRAD, SKCM, STAD, and UCEC (Figure 10(b)).
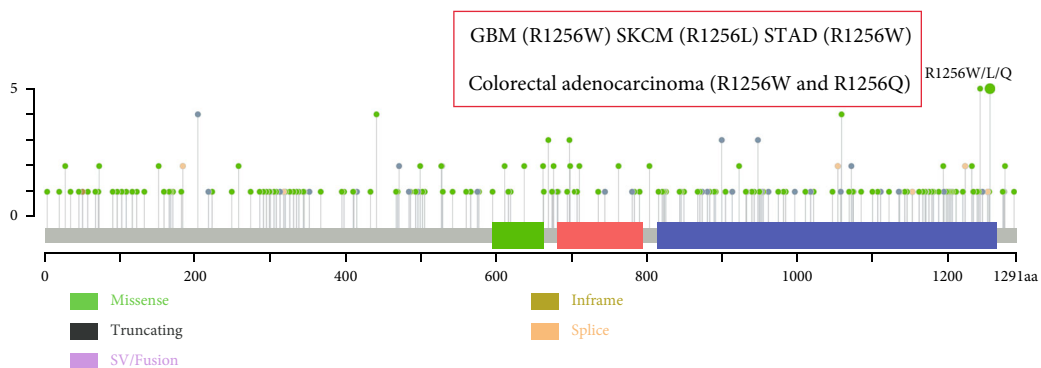
(a)

(b)

RefSeq: NM_001145415
Ensembl: ENST00000271640
CCDS: CCDS44217
UniProt: SETBI_HUMAN

(c)

GBM (R1256W) SKCM (R1256L) STAD (R1256W)

Colorectal adenocarcinoma (R1256W and R1256Q)
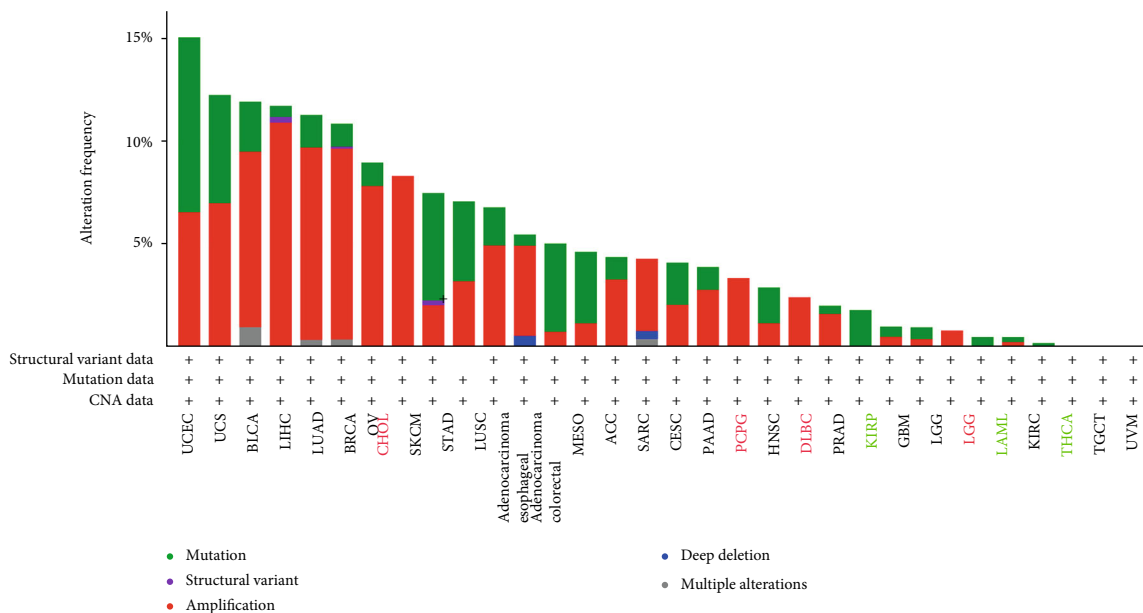
(d)

FIGURE 8: Mutation feature of SETDB1 in different tumors of TCGA. (a) Summary of alterations in SETDB1 expression in different tumors. (b) The alteration frequency with mutation type. (c) Some SETDB1 mutations are shown on the 3D structure of protein. (d) The mutation site of SETDB1 in amino acid sequence.
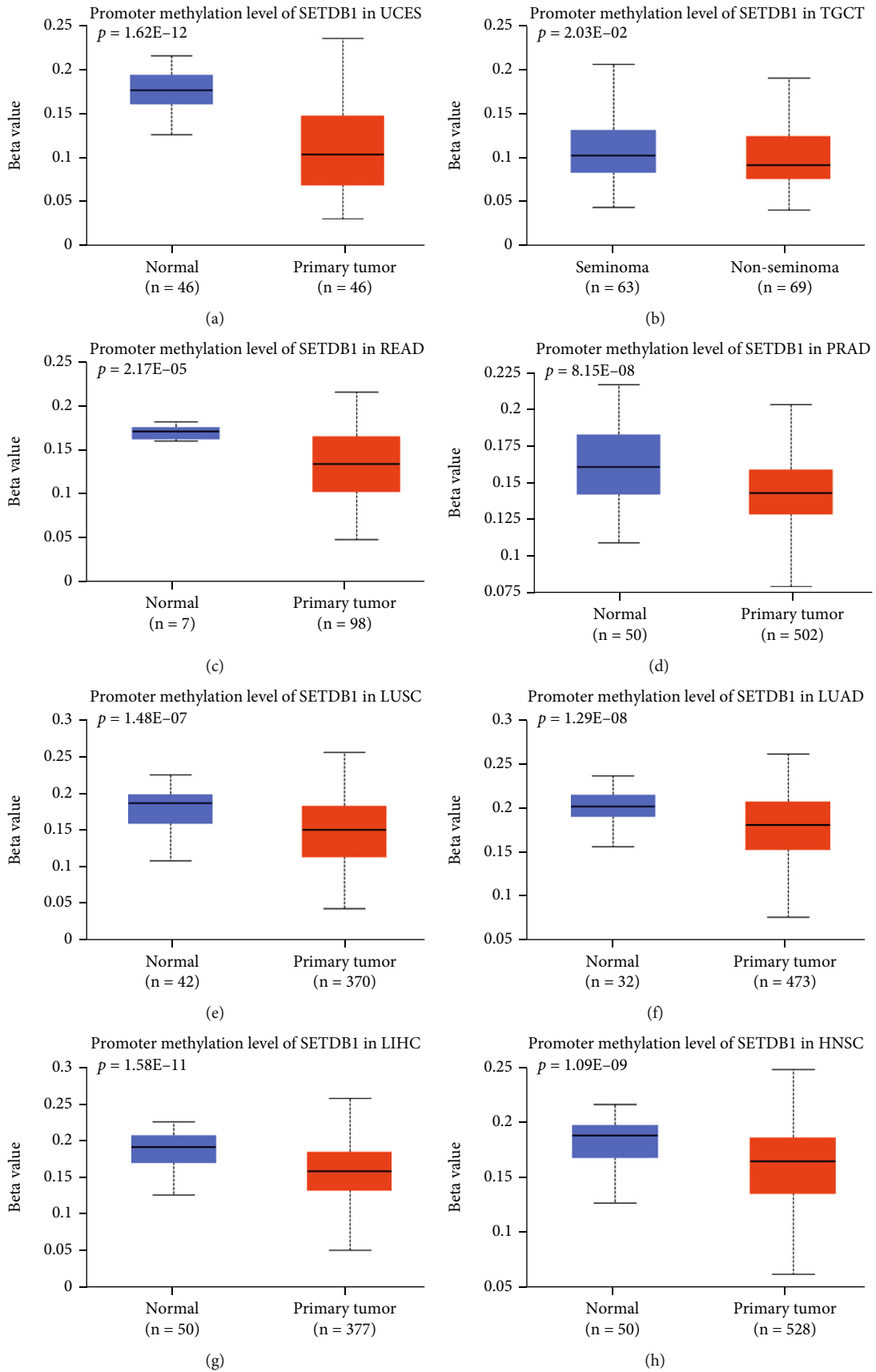
FIGURE 9: Continued.

Promoter methylation level of SETDB1 in ESCA

$p = 3.24E-02$

Promoter methylation level of SETDB1 in COAD

$p = 2.47E-06$

Promoter methylation level of SETDB1 in BRCA

$p = 5.27E-11$

Promoter methylation level of SETDB1 in BLCA

$p = 1.63E-11$

SETDB1-body-s_shelf-cg10444928

LR test p-value = 0.015
HR = 2.514

SETDB1-body-s_shelf-cg10444928

LR test p-value = 0.00063
HR = 0.916

(i) (j) (k) (l) (m) (n)

Figure 9: Continued.

SETDB1-body-s_shelf-cg10444928

LR test *p*-value = 0.0084
HR = 1.893

MESO

— Lower (n = 37)
— Higher (n = 50)

(o)

SETDB1-body-s_shelf-cg10444928

LR test *p*-value = 0.023
HR = 0.638

KIRC

— Lower (n = 161)
— Higher (n = 158)

(p)

SETDB1-body-s_shelf-cg10444928

LR test *p*-value = 0.021
HR = 2.673

KIRP

— Lower (n = 69)
— Higher (n = 206)

(q)

SETDB1-body-s_shelf-cg10444928

LR test *p*-value = 0.007
HR = 1.628

LGG

— Lower (n = 282)
— Higher (n = 233)

(r)

FIGURE 9: Continued.

(s)

Legend: PTM sites

- Phosphorylation   - Ubiquitylation
- Acetylation   - Other

(t)
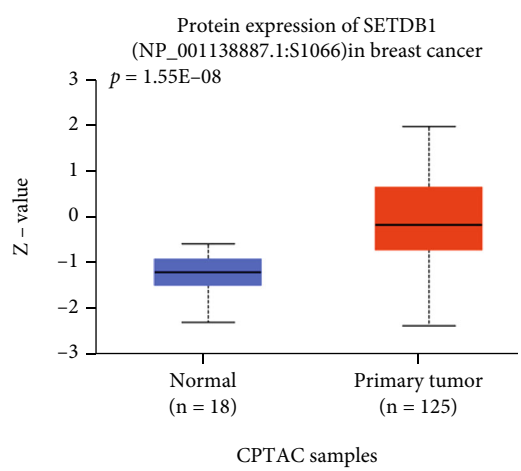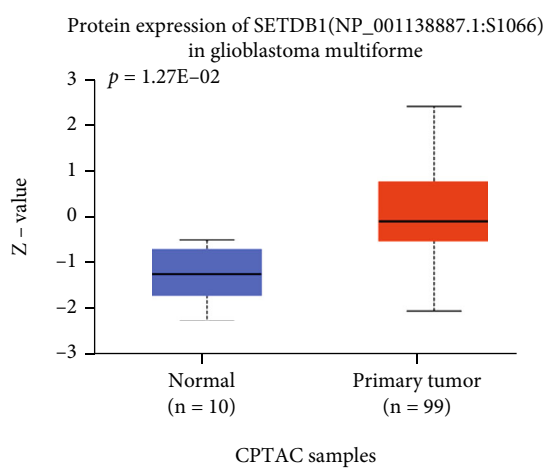
(u)   (v)

FIGURE 9: Continued.

(w)

(x)

(y)

(z)
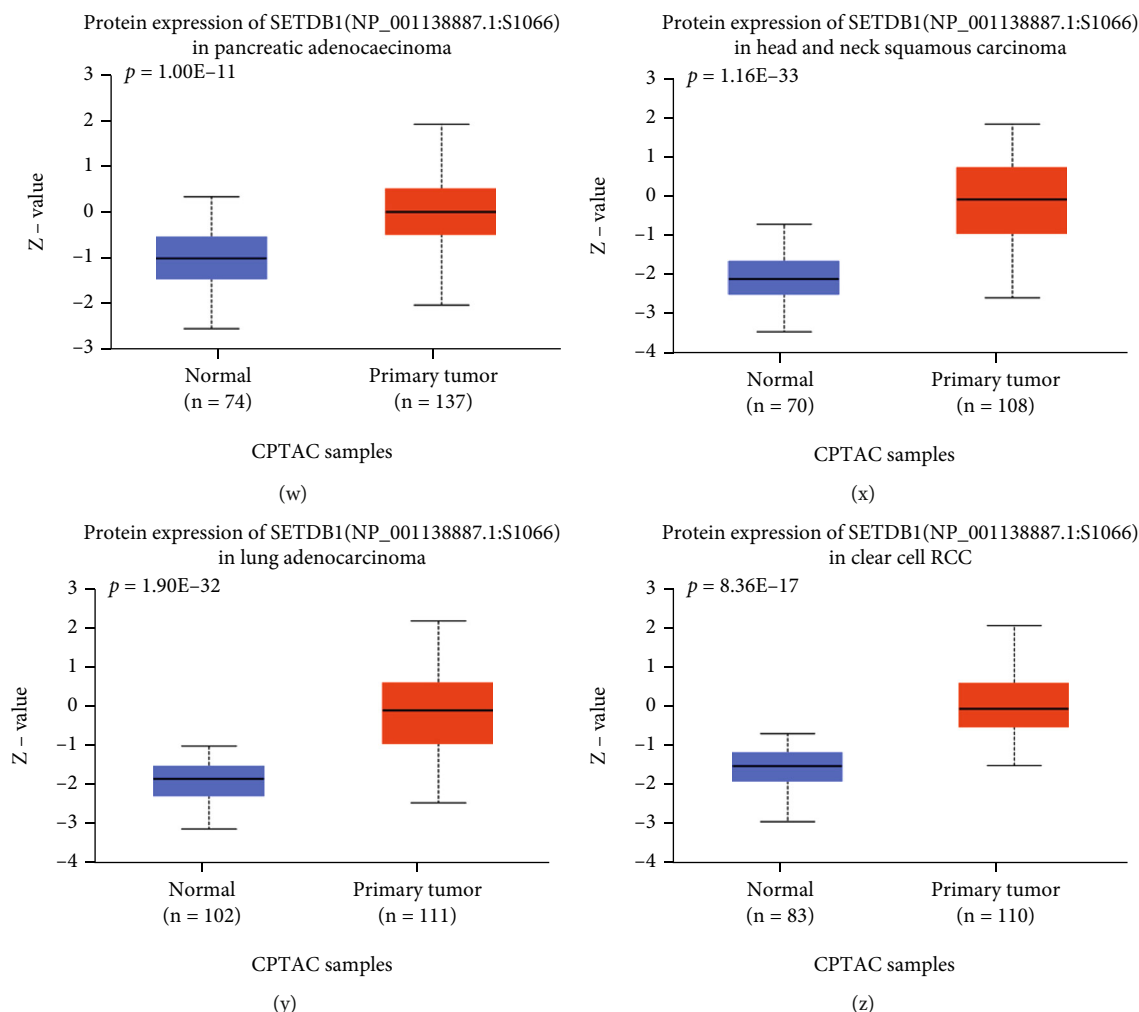
FIGURE 9: Methylation and protein phosphorylation analysis. (a–l) The differential DNA methylation level of SETDB1 promoter in twelve tumor types. (a) UCEC, (b) TGCT, (c) READ, (d) PRAD, (e) LUSC, (f) LUAD, (g) LIHC, (h) HNSC, (i) ESCA, (j) COAD, (k) BRCA, and (l) BLCA. (m–s) The prognosis value of single CpG (cg10444928) of SETDB1 in seven tumor types. (m) UCS, (n) UVM, (o) MESO, (p) KIRC, (q) KIRP, (r) LGG, and (s) CESC. (t) Phosphorylation site of SETDB1 protein. (u–z) The differential protein level of SETDB1 in S1006 site in six tumor types. (u) Breast cancer, (v) glioblastoma multiforme, (w) pancreatic adenocarcinoma, (x) head and neck squamous carcinoma, (y) lung adenocarcinoma, and (z) clear cell RCC.

These results implied that SETDB1 expression status was relevant to the immune subtypes and molecular subtypes of various cancers.

3.9. Immune Infiltration Analysis of SETDB1. Considering the importance of the immune microenvironment in tumorigenesis and cancer progression, we characterized immune infiltration levels of SETDB1 based on several databases. CAFs are the fibroblasts around tumor cells and the major stromal cells in the tumor microenvironment. They play a significant role in the initiation and progression of tumors [24, 25]. It has been demonstrated that targeting CAFs is an effective treatment strategy for various cancers [26]. The EPIC, MCPCOUNTER, and TIDE algorithms were applied to assess the relationship between the infiltration level of CAFs and SETDB1 gene expression in various human cancers. We observed a significantly positive correlation between SETDB1 expression and infiltration level of CAFs

in ACC (Rho = 0.358, $p = 1.88e - 03$), BRCA (Rho = 0.15, $p = 6.31e - 04$), CESC (Rho = 0.279, $p = 2.34e - 06$), COAD (Rho = 0.2, $p = 8.73e - 04$), HNSC (Rho = 0.204, $p = 5.07e - 06$), HNSC-HPV (Rho = 0.25, $p = 3.97e - 07$), KIRP (Rho = 0.306, $p = 5.57e - 07$), LIHC (Rho = 0.374, $p = 6.64e - 13$), and READ (Rho = 0.329, $p = 1.57e - 03$) but a strongly negative correlation in TGCT (Rho = −0.339, $p = 2.63e - 05$) (Figure 11). Furthermore, partial correlation analysis between SETDB1 expression and immune cell infiltration was conducted using the TIMER2.0 database. The results demonstrated a remarkable correlation between SETDB1 expression and CD8+ T cells, CD4+ T cells, Tregs, and B cells (Figure S7A-7D).

3.10. Correlation Analysis of Immune Checkpoint Inhibitor-Related Genes. Accumulating evidence suggests that immune checkpoint inhibitors are a class of biologics that interact with the immune system to encourage antitumor response
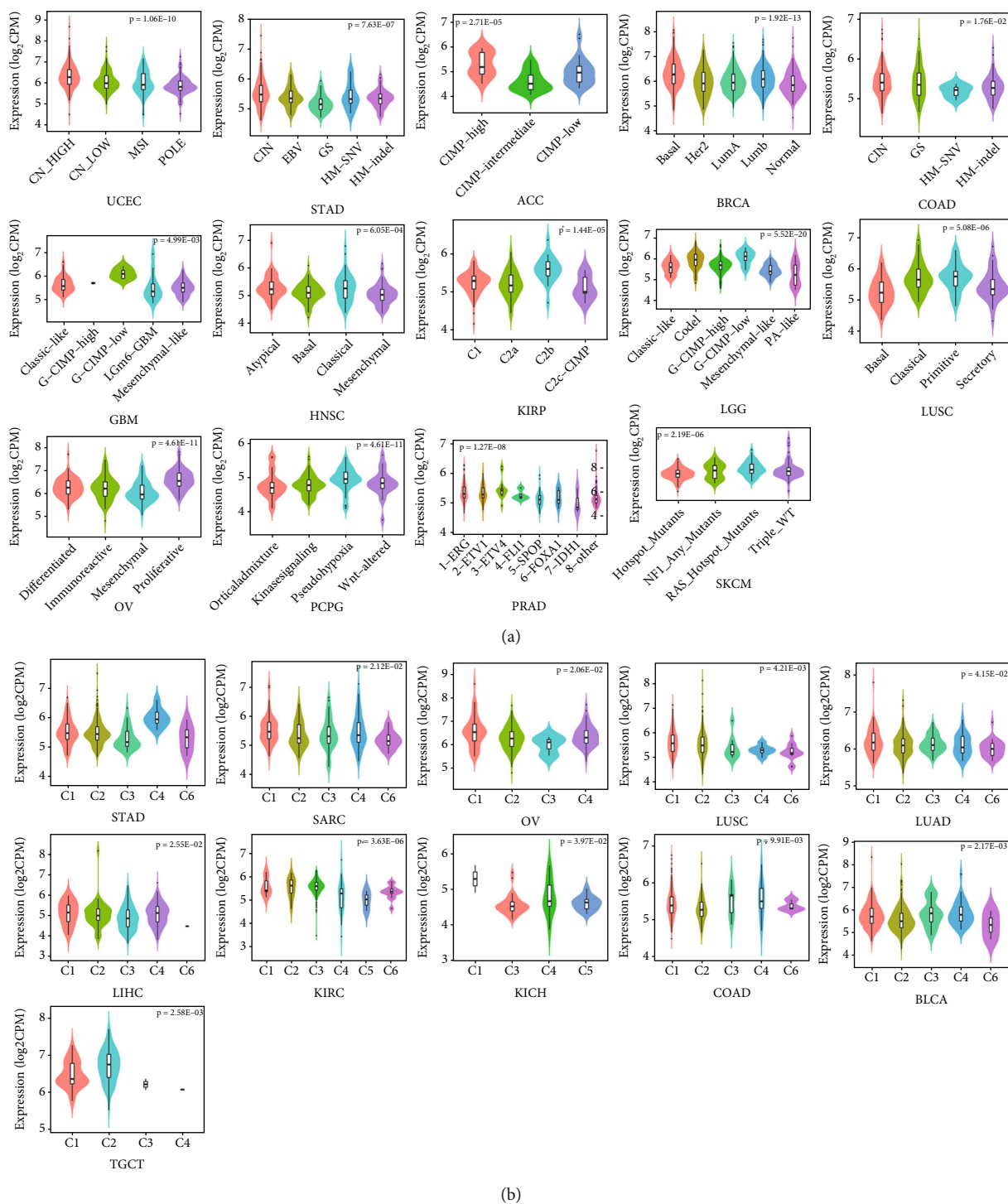
(a)

(b)

Figure 10: The immune and Molecular Subtypes analysis of SETDB1. (a) The SETDB1 gene differential expression in different immune subtypes. (b) The SETDB1 gene differential expression in different molecular subtypes.

by immune cells [27]. We also demonstrated a significant correlation between SETDB1 expression and each immune checkpoint-related gene (immunoinhibitor and immunostimulator) in diverse cancers in TCGA (Figures 12(a) and 12(b)). For example, in KIRC, SETDB1 expression has a significantly positive correlation with the expression of TIGIT, PDCD1, CTLA4, CD96, CD244, CD160, BTLA, ADORA2A,

TGFBR1, LAG3, HHLA2, CXCR4, CD80, CD70, CD48, etc. (Figure 12(d)). The MHC is a human leukocyte antigen (HLA) that plays an important role in tumor immunotherapy by activating T cells [28, 29]. According to our results, the association between SETDB1 expression and HLA-related genes varies markedly among cancer types. The SETDB1 expression and HLA-related genes in ACC, CESC,
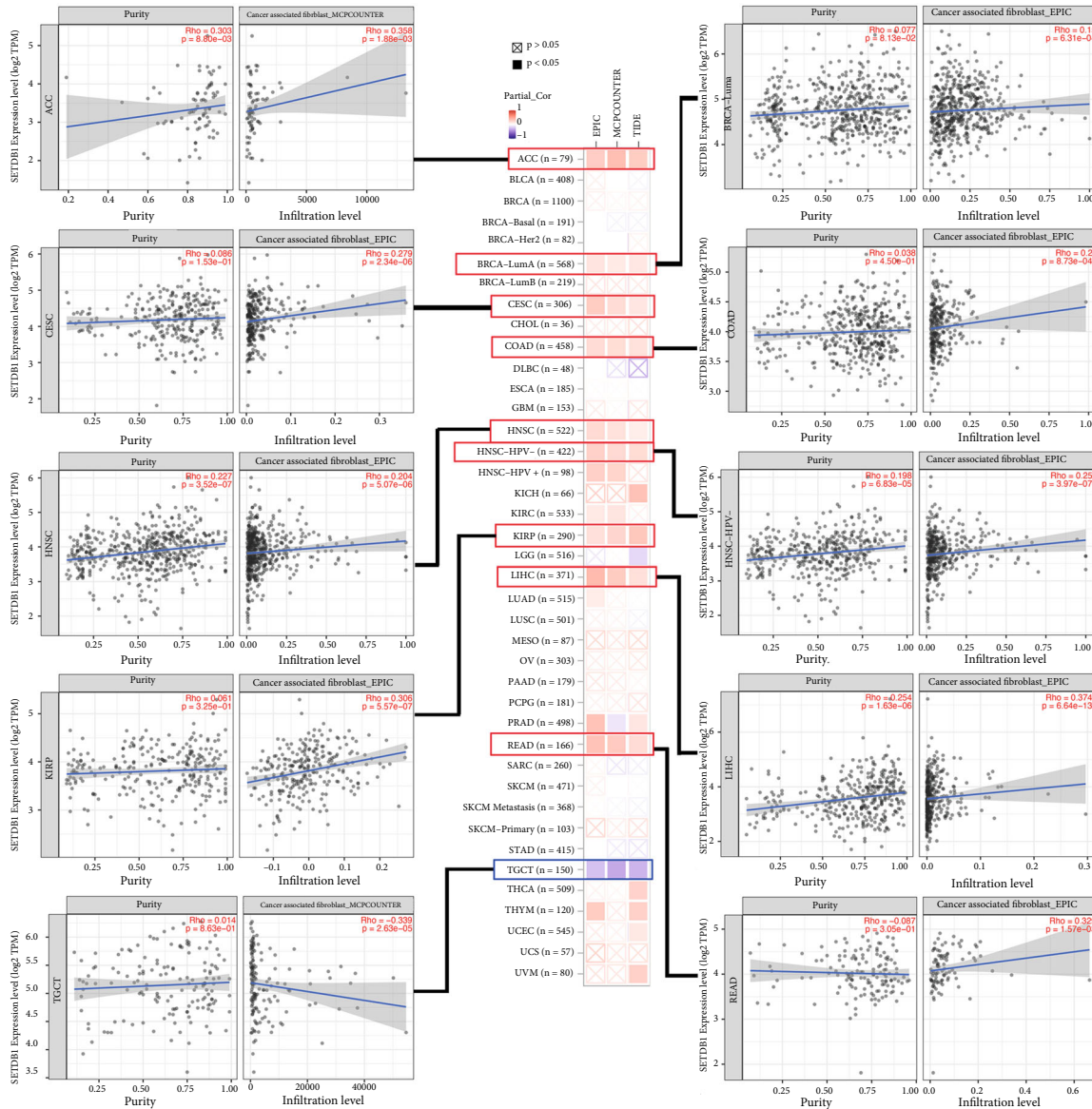
FIGURE 11: Correlation analysis between SETDB1 expression and immune infiltration of cancer-associated fibroblasts.

and KIRC were positively correlated but were negatively correlated in other cancers in TCGA (Figure 12(c)). TMB and MSI are emerging predictors associated with survival and response to immunotherapy [30, 31]. This study showed that SETDB1 was positively correlated with MSI in BLCA, CESC, LUAD, LUSC, READ, and SARC, while negatively correlated with MSI in DLBC, but did not show correlation with MSI in other cancers (Figure 12(e)). SETDB1 expression was positively correlated with TMB in BLCA, BRCA, LGG, LUAD, and STAD, while negatively correlated with TMB in THCA and UCS, but did not correlate with TMB in other cancers (Figure 12(f)).

3.11. Function and Pathway Analysis of SETDB1-Related Genes. To further elucidate the biological function and molecular mechanism of SETDB1 and provide theoretical support for the study of tumorigenesis, we identified the

targeting SETDB1-binding proteins with the STRING tool and conducted bioinformatics analyses. The interaction network analysis showed interaction information of SETDB1 and these proteins with 51 nodes and 715 edges (Figure S8A). These genes were considered the target genes to obtain enriched GO terms and significant KEGG pathways. In GO analysis, 364 GO categories were detected, including 251 biological process (BP), 40 cellular component (CC), and 73 molecular function (MF). In the GO-BP category, the target genes were mainly enriched in covalent chromatin modification (GO:0016569), histone modification (GO:0016570), and peptidyl-lysine modification (GO:0018205) (Figure 13(a)). In the GO-CC category, the genes were related to heterochromatin (GO:0000792), chromosomal region (GO:0098687), chromosome, and telomeric region (GO:0000781) (Figure 13(a)). In the GO-MF category,
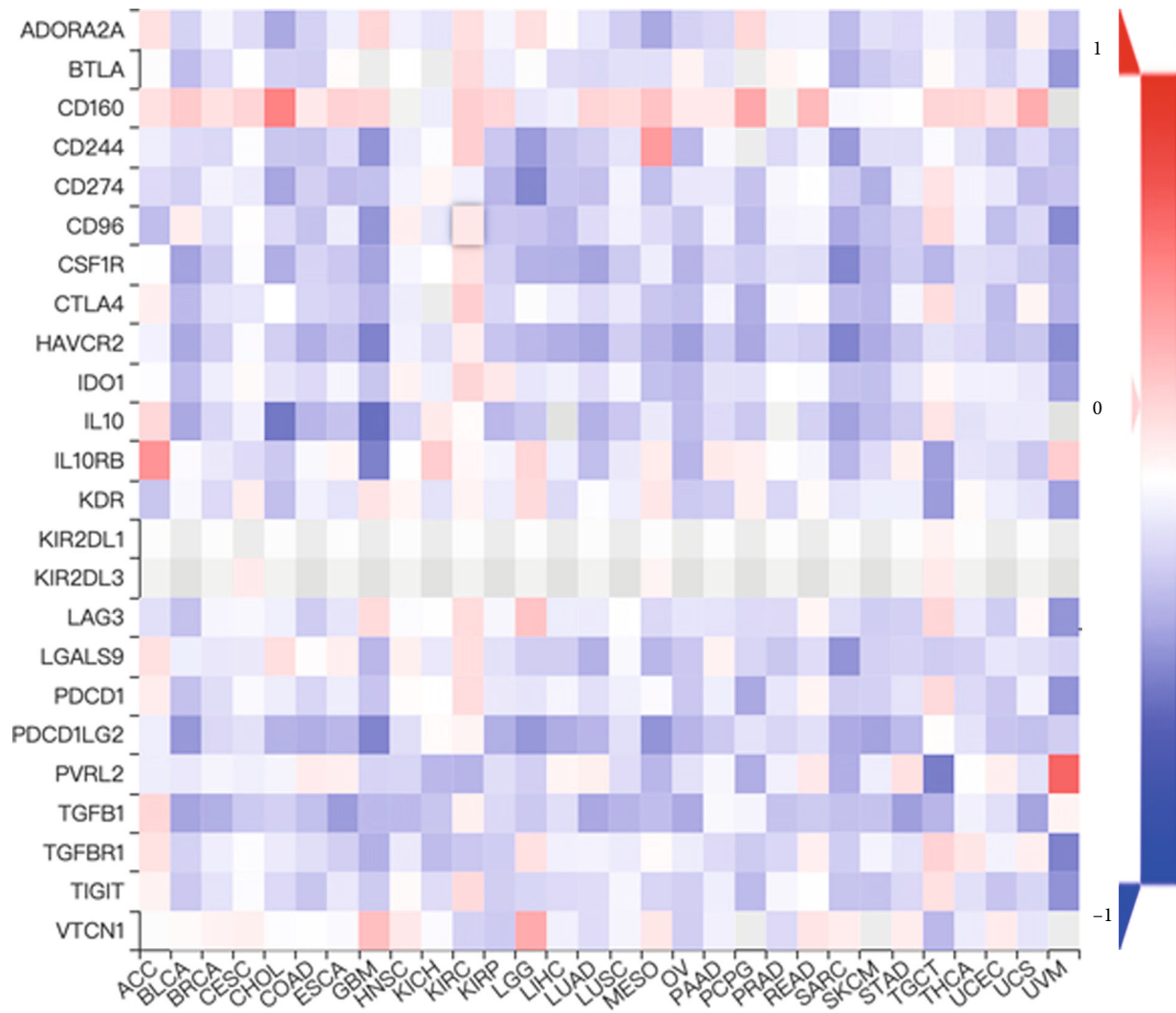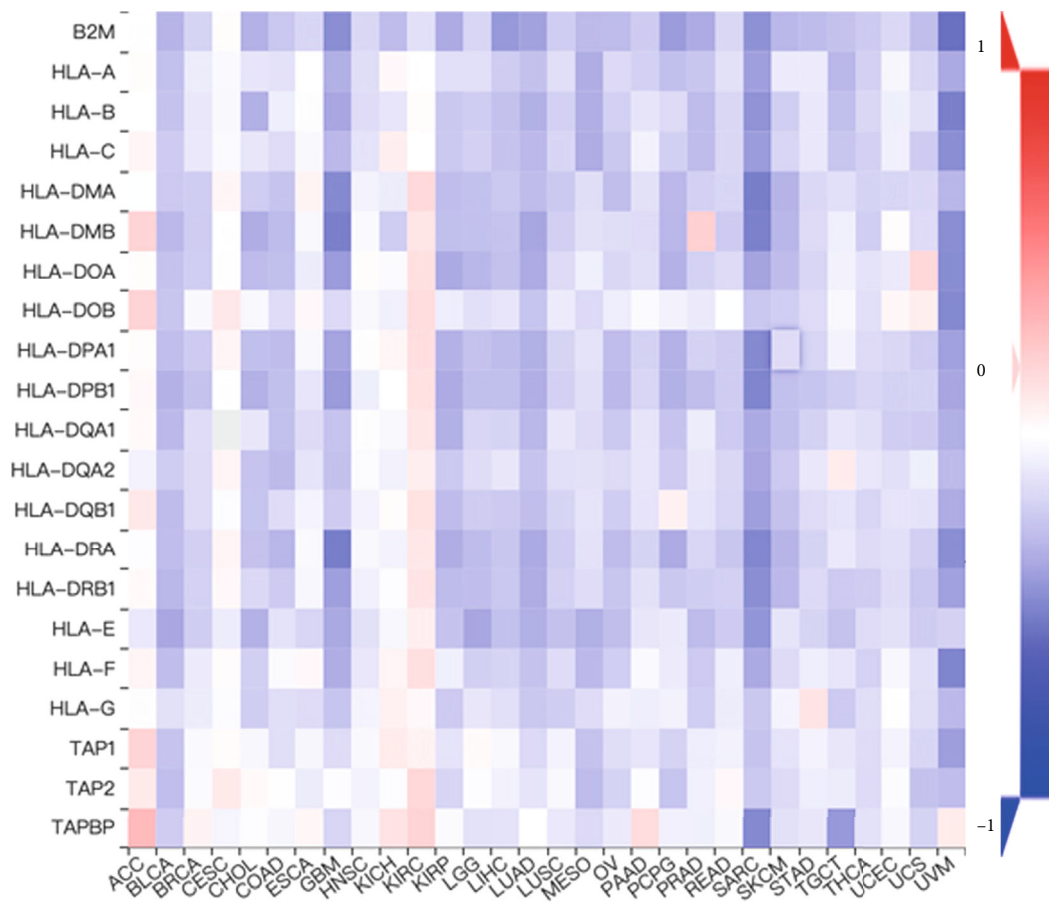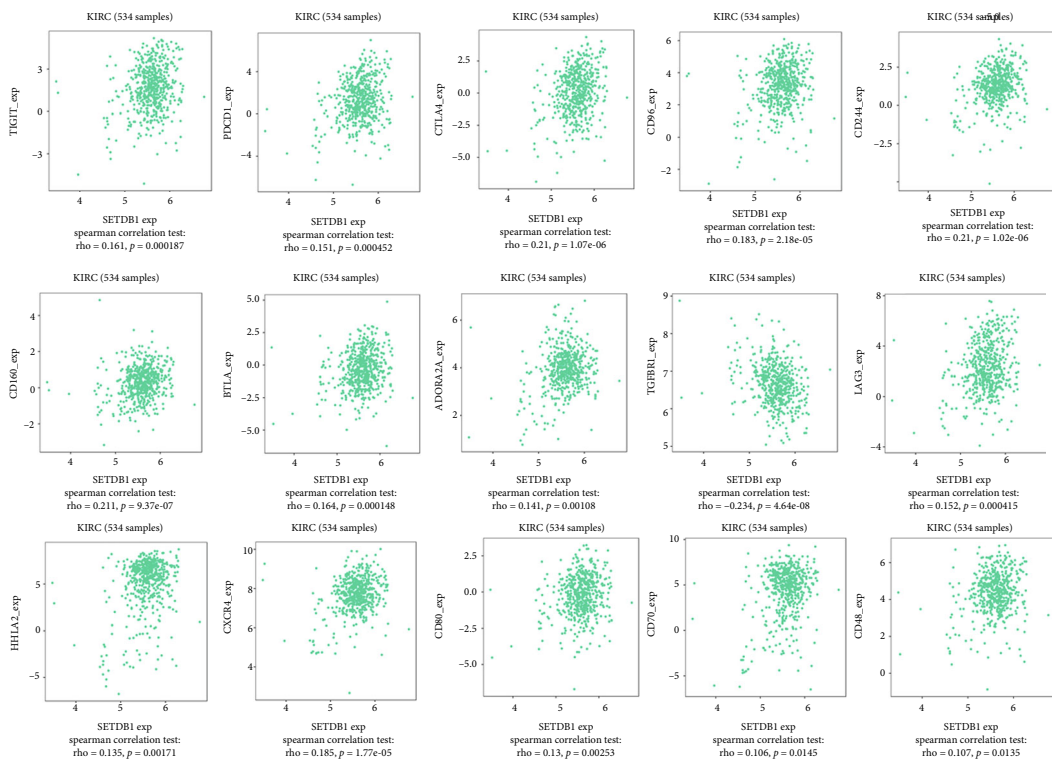
(a)

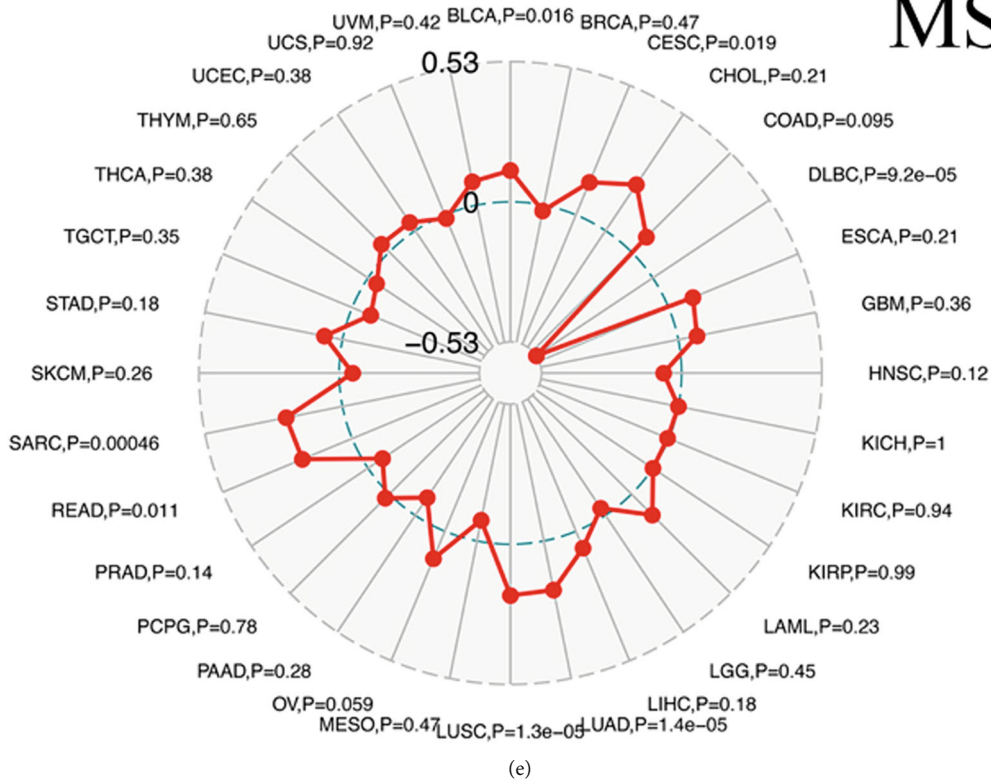FIGURE 12: Continued.
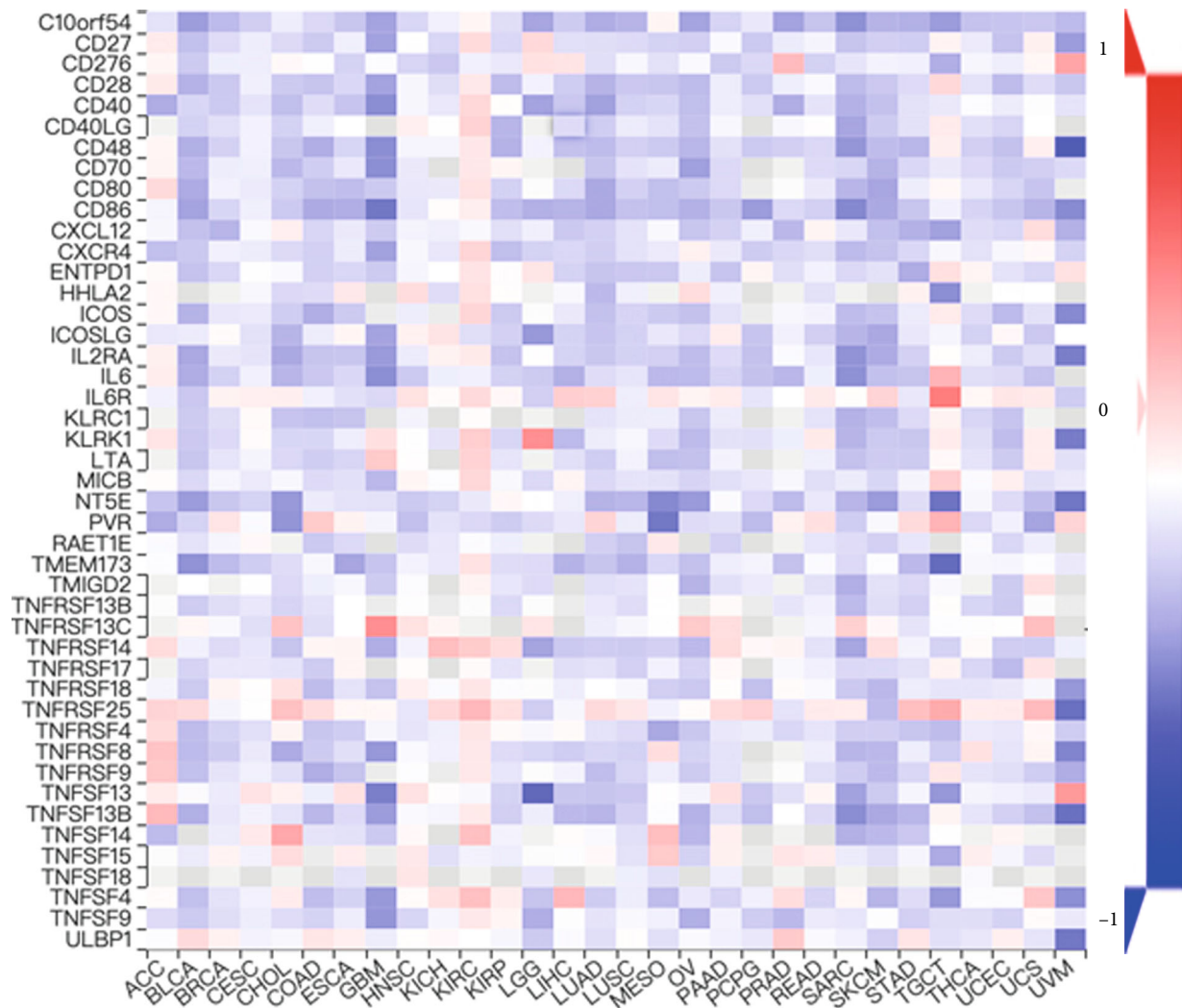
(b)

FIGURE 12: Continued.

(c)

(d)

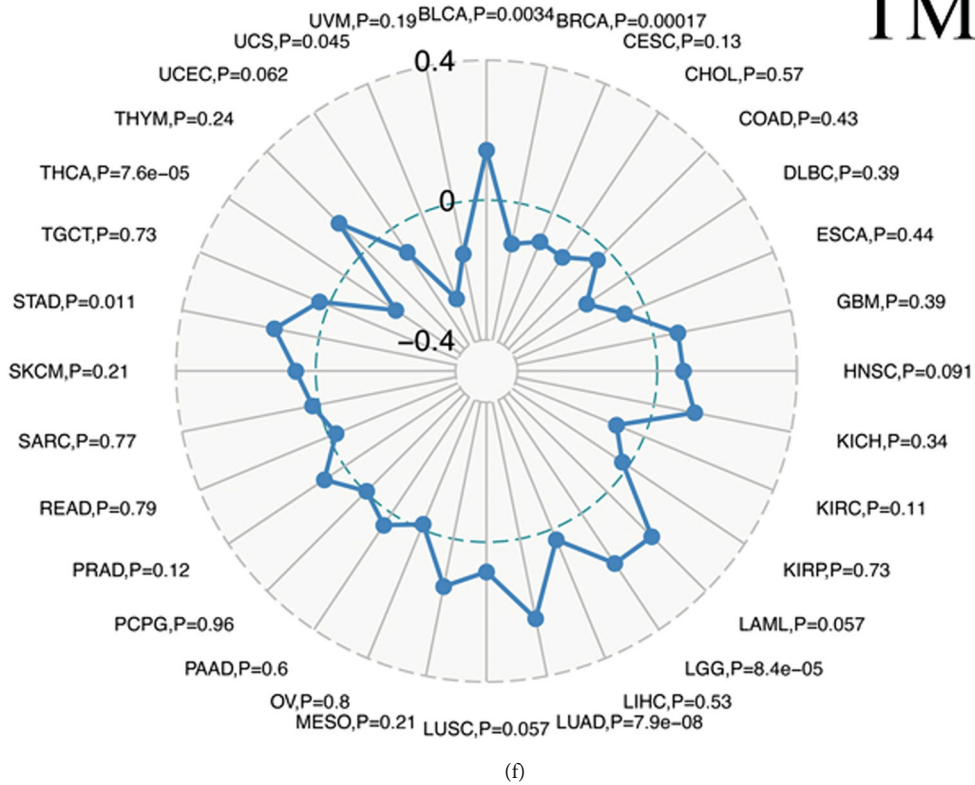Figure 12: Continued.

(e)

(f)

FIGURE 12: Correlation between SETDB1 expression and immunoinhibitor, immunostimulatory, human leukocyte antigen- (HLA-) associated genes, microsatellite instability, and tumor mutational burden. (a) Immunoinhibitor, (b) immunostimulator, and (c) human leukocyte antigen- (HLA-) associated genes. Red color indicates positive correlations; blue color indicates negative correlations. (d) SETDB1 expression significantly correlated with immunoinhibitor and immunostimulator in KIRC tissues. (e) Microsatellite instability and (f) tumor mutational burden.
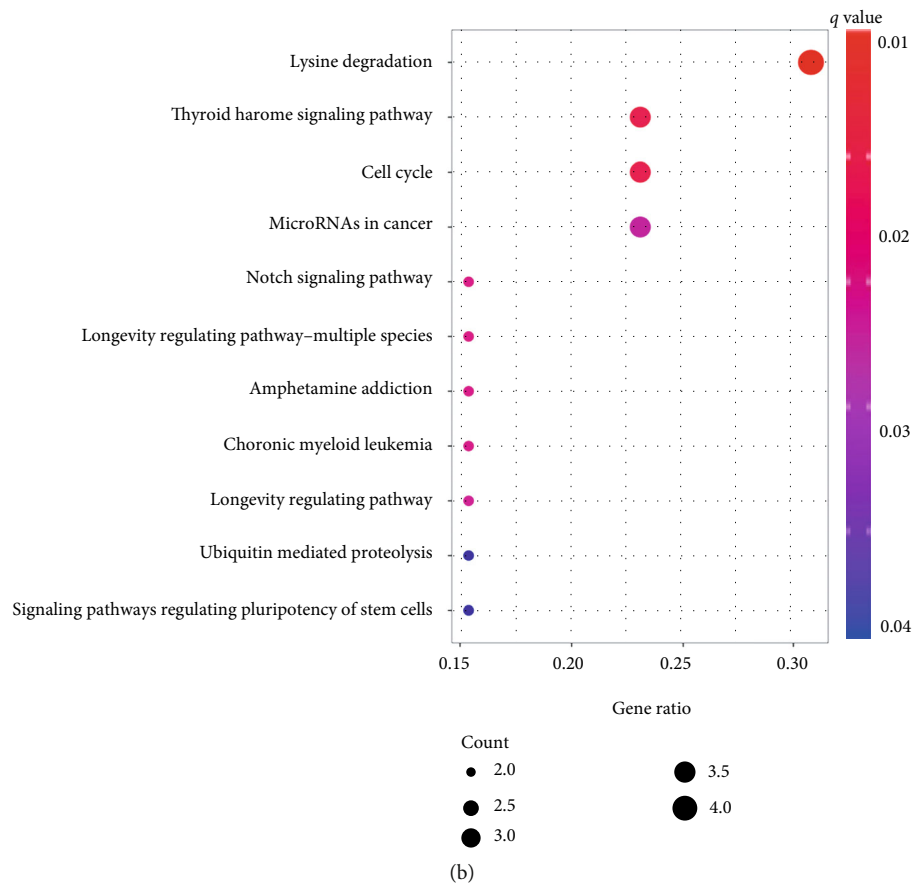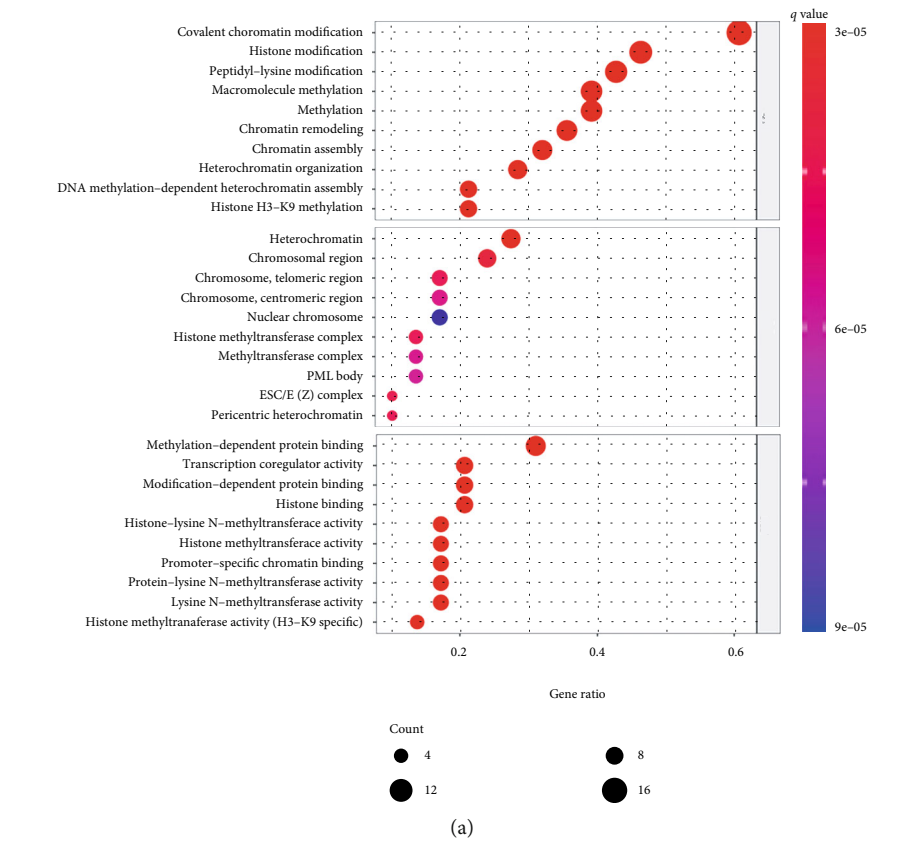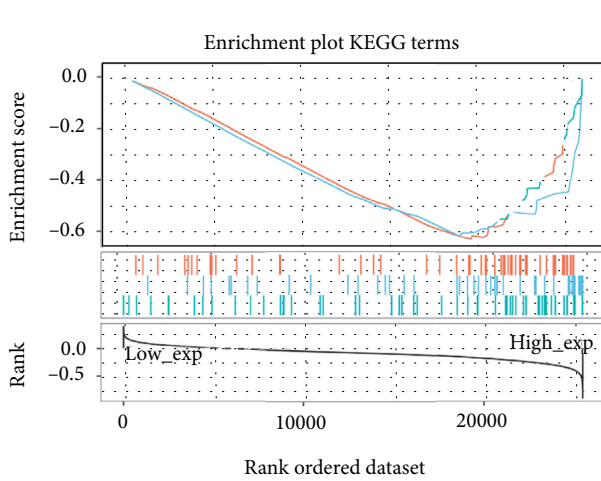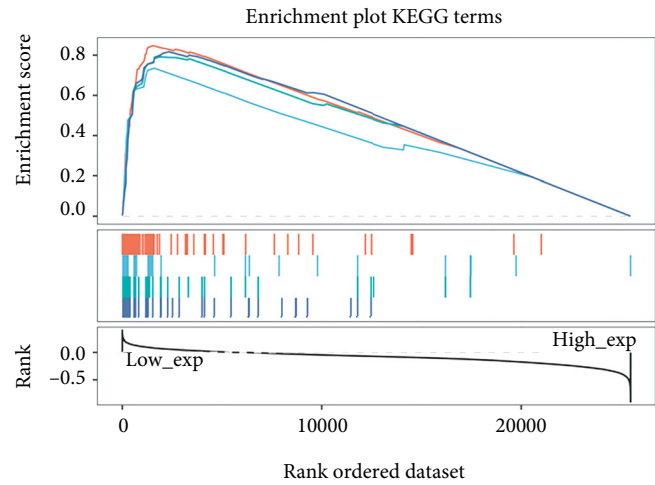
(a)

(b)

Figure 13: Continued.

(c)

(d)

(e)

FIGURE 13: Continued.

Enrichment plot KEGG terms

Term

INFLAMMATORY_RESPONSE
ES = 0.34,NES = 1.2,P = 0.26,FDR = 0.81

ALLOGRAFT_REJECTION
ES = 0.47,NES = 1.5,P = 0.12,FDR = 0.29

COAGULATION
ES = 0.42,NES = 1.5,P = 0.038,FDR = 0.36

EPITHELIAL_MESENCHYMAL_TRANSITION
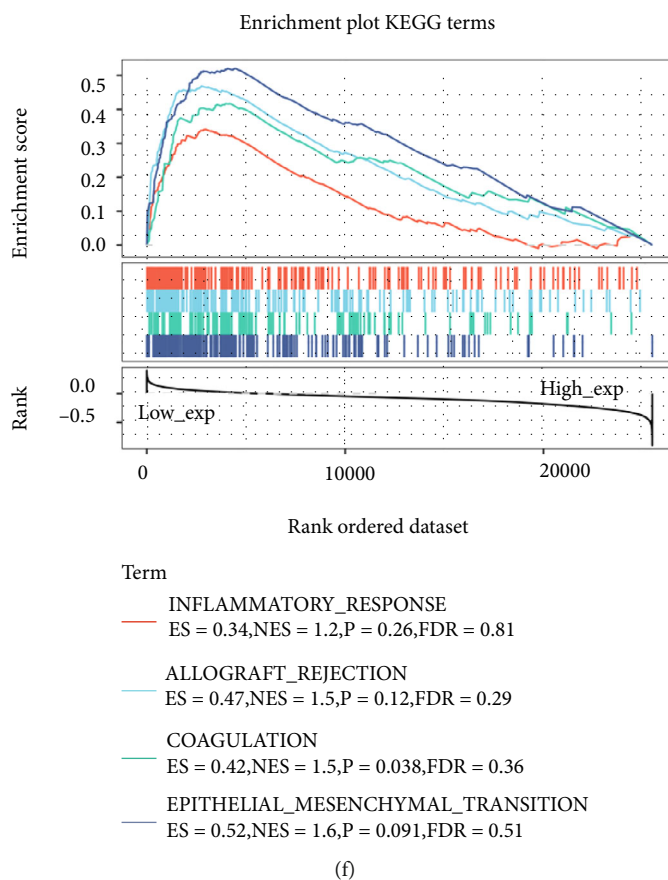ES = 0.52,NES = 1.6,P = 0.091,FDR = 0.51

(f)
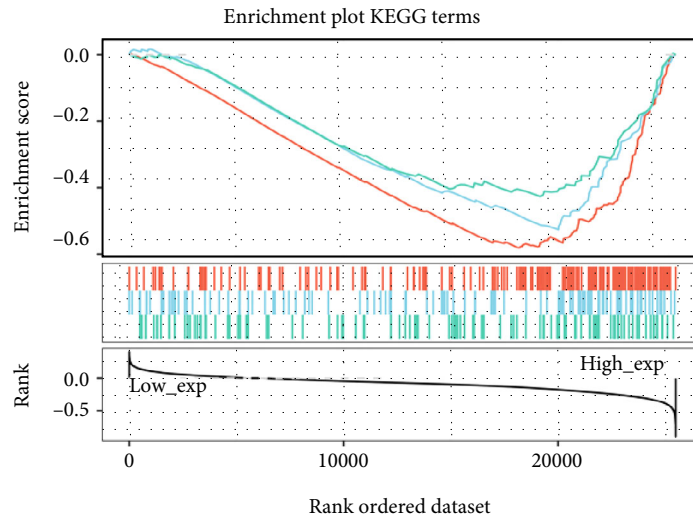
FIGURE 13: Gene Ontology (GO) functional enrichment and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway and Gene Set Enrichment Analysis (GSEA) for SETDB1 and its related genes. (a) GO analysis for SETDB1-related genes. (b) KEGG analysis for SETDB1-related genes. (c) The enriched gene sets in KEGG collection by the high SETDB1expression sample. (d) The enriched gene sets in KEGG by samples with low SETDB1 expression. (e) Enriched gene sets in HALLMARK by samples of high SETDB1 expression. (f) Enriched gene sets in HALLMARK by the low SETDB1 expression.

target genes were mainly enriched in transcription coregulator activity (GO:0003712), methylated histone binding (GO:0035064), and methylation-dependent protein binding (GO:0140034) (Figure 13(a)). In KEGG pathway analysis, 51 genes were categorized into 11 KEGG pathways. As a result, lysine degradation, thyroid hormone signaling pathway, and cell cycle were identified and marked as main KEGG pathways (Figure 13(b)). These results are consistent with the results of GSEA analysis (Figures 13(c) and 13(d)). Additionally, in terms of the HALLMARK, a high expression level of SETDB1 was significantly enriched in the mitotic spindle, unfolded protein response, and PI3K-AKT-MTOR signaling. In contrast, the low expression level of SETDB1 was significantly enriched in the inflammatory response, allograft rejection, coagulation, and epithelial-mesenchymal transition (Figures 13(e) and 13(f)). Furthermore, the results of GeneMANIA also revealed that SETDB1 and targeting SETDB1-binding proteins were mainly related to chromatin assembly, chromatin assembly or disassembly, DNA packaging, DNA conformation change, protein-DNA complex, nucleosome organization, and DNA packaging complex (Figure S8B). CeRNA is important for tumorigenesis by forming an extensive ceRNA network involving mRNA, miRNA, and ncRNA. We identified hsa-miR-29a-3p as the most vital miRNA regulator by overlapping predictions of three databases (Figure S8C). We then explored the complementary sequences between SETDB1 and hsa-miR-29a-3p using the TargetScanHuman database (Figure S8D). We also predicted the twelve target lncRNAs by interacting with miRNA and lncRNA sequences in the LncBase database. Then, the lncRNA-miRNA-mRNA network was constructed based on lncRNA-miRNA and miRNA-mRNA regulation pairs (Figure S8E).

3.12. Drug Sensitivity Analysis. Genetic mutations can influence the efficacy of chemotherapy and targeted therapy. Therefore, we evaluated the role of SETDB1 in chemotherapy or targeted therapy and investigated the drug sensitivity and drug resistance of cancer cell lines from the GDSC datasets. ANOVA analysis showed that drug sensitivity toward Arg-G (nelarabine), Nilotinib, and KIN001-042 was significantly correlated with the expression of SETDB1 (negative correlation with IC50). However, the drug resistance toward Bleomycin (50 μM), JAK3_7406, and FGFR_0939 was correlated with the expression of SETDB1 (positive correlation with IC50) (Figure 14(a)). The correlation between GDSC

(a)

(b)

FIGURE 14: Continued.

(c)

Nelarabine (Ara–G) (C11H15N505)

(d)

(e)
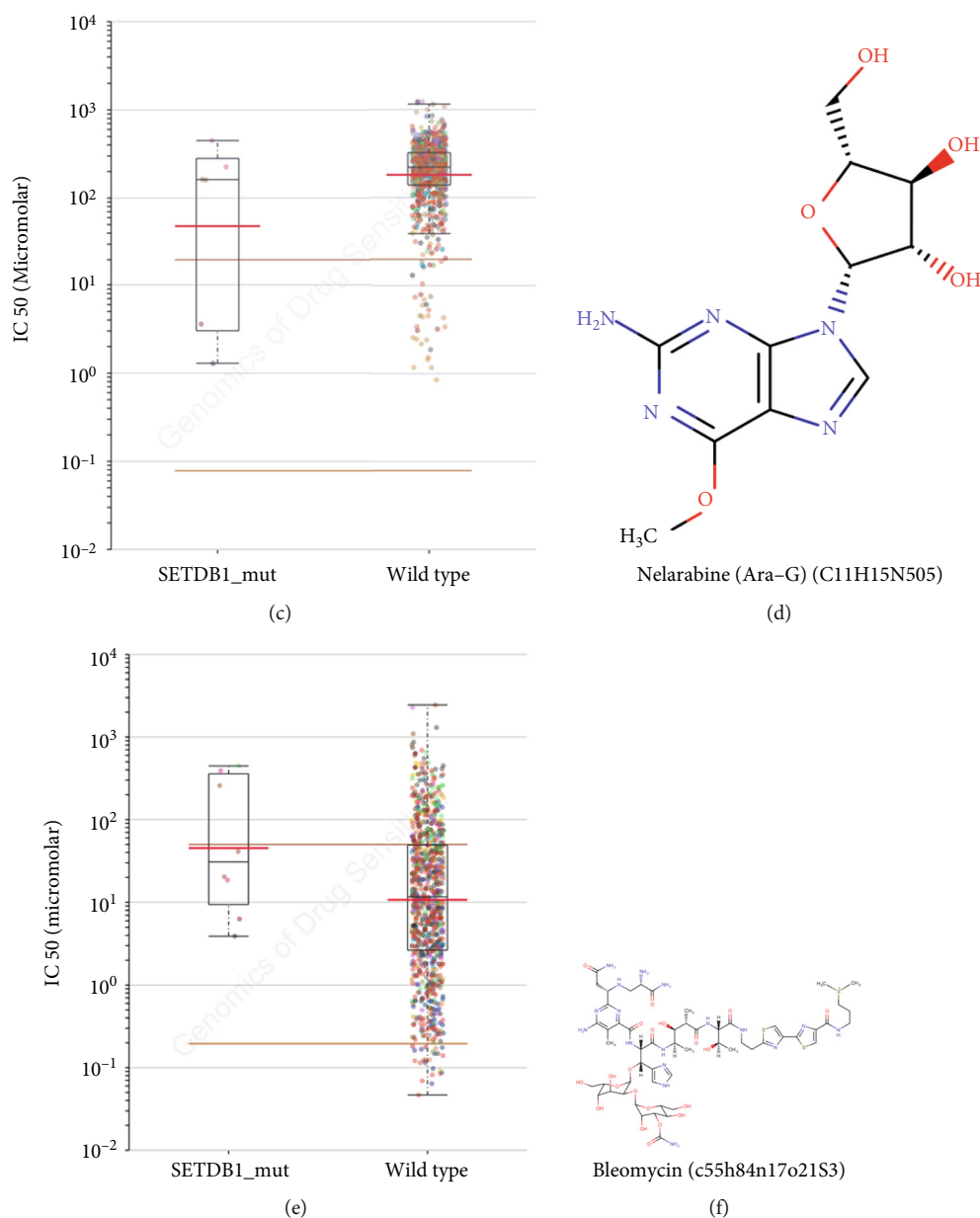

Bleomycin (c55h84n17o21S3)

(f)
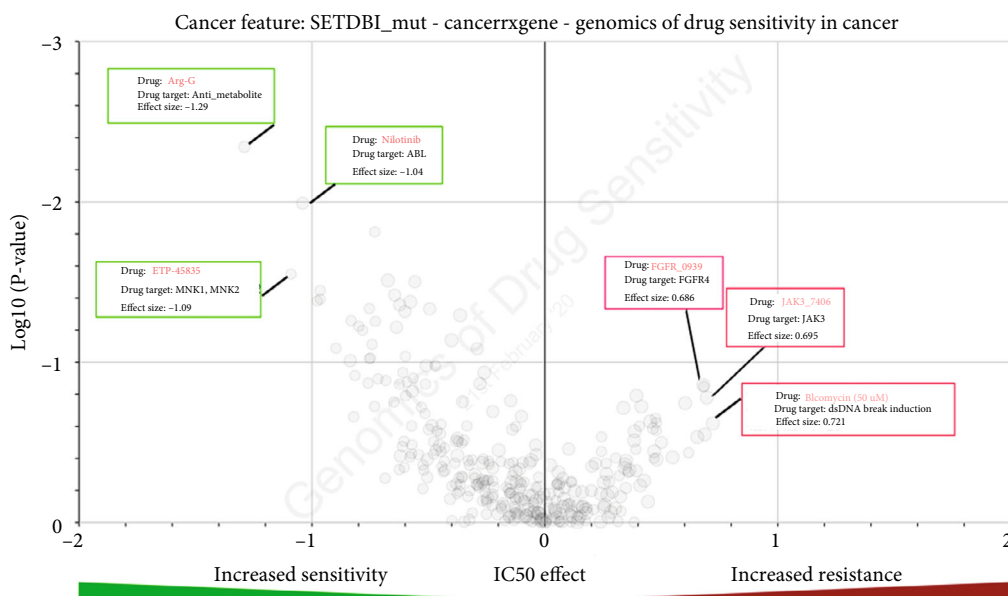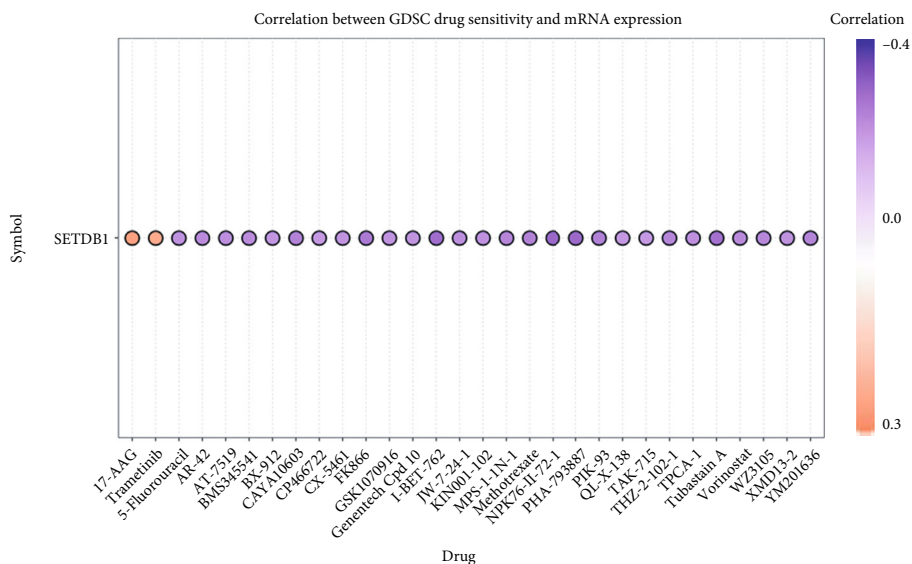
FIGURE 14: Drug sensitivity analysis based on SETDB1 mutation. (a) Genomics of drug sensitivity in cancer. (b) Correlation between GDSC drug sensitivity, drug resistance, and SETDB1 mutation. (c) Nelarabine (Ara-G) IC50 values for SETDB1 mutation. (d) The structural formulas of Nelarabine (Ara-G). (e) Bleomycin IC50 values for SETDB1 mutation. (f) The structural formulas of Bleomycin.

drug sensitivity and SETDB1 expression showed that most drugs were negatively correlated with SETDB1 expression, while 17-AAG and trametinib showed a positive correlation with SETDB1 expression (Figure 14(b)). Furthermore, the IC50 values of nelarabine and Bleomycin ($50 \mu M$) for SETDB1 (Mut and wild type) were displayed. The chemical formulas of nelarabine and Bleomycin are $C_{11}H_{15}N_5O_5$ and $C_{55}H_{84}N_{17}O_{21}S_3$, respectively, and their structural formulas are shown in Figures 14(c) and 14(d).

3.13. Validating Expression of SETDB1 by IHC. For further validation, the expression level of SETDB1 gene was analyzed by IHC. Since gastrointestinal malignancies are the most frequent primary tumors, we focus on the expression level of SETDB1 gene in the most common gastrointestinal malignancies including LIHC, CHOL, COAD, ESCA, PAAD, and STAD. The results showed that SETDB1 was highly expressed in tumor tissues and rarely expressed in normal tissues (Figure 15). These results are similar to those we have previously reported for different database.

4. Discussion

The SETDB1 gene, first identified in 1999 by Harte et al. [11], with a length of 38.6 kilobases (kb), is located on the human chromosome 1q21.3. Chromosome numbers and
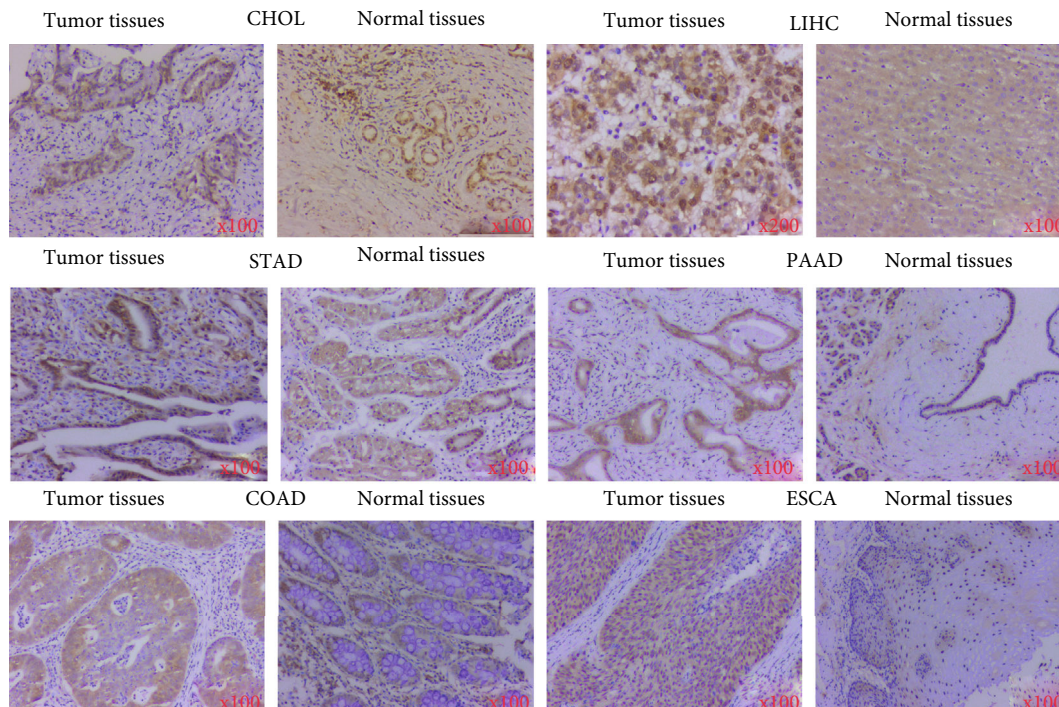
FIGURE 15: Immunohistochemistry results of SETDB1 performed in normal and tumor tissues of human.

structural abnormalities are important factors for tumorigenesis and the therapeutic response [32, 33]. Tumor tissues have many chromosomal variants. Chromosome 1q gains occurred in various human cancers, such as LUAD, LIHC, OV, BRCA, and multiple myeloma [32, 34, 35]. Chromosome 1q21.3 abnormalities are related to breast cancer recurrence, and they can promote cell proliferation and DNA damage response in metastatic melanoma [35, 36]. SETDB1 is located in the 1q21.3 region that encodes a histone methyltransferase which regulates transcriptional repression, histone methylation, and gene silencing [37, 38]. This study has demonstrated that the SETDB1 is differentially expressed in most tumors and normal tissues, indicating that it also plays an oncogenic role in these tumors. The amplification of SETDB1 in human tumors is significantly associated with immune exclusion and tumor progression, but its biological and functional role or contribution to tumor prognosis is unknown [12]. This paper is the first pan-cancer analysis of SETDB1 across 33 different tumors based on the data of TCGA, CPTAC, and GEO databases. The results show that SETDB1 is significantly correlated with tumorigenesis and clinical outcomes.

SETDB1 has specific domains [39], such as two Tudors, MBD, pre-SET, SET, and post-SET, and this result is consistent with our finding. The most biological function of SETDB1 is ascribed to the SET domain, which is highly conserved across species and originally identified in the Drosophila Trithorax (TRX) and human MLL proteins [40]. SETDB1 is a member of SET family and is an H3K9 methyltransferase that modulates gene activity. The pre-SET, SET, and post-SET domains are crucial for histone methyltransferase activity. Furthermore, SETDB1 protein has a canonical CpG DNA methyl binding domain (MBD) at the N-terminus, which can bind methylated DNA at one site [41]. Growing evidence suggests the involvement of MBD genes in cancers [42]. MBD is involved in various signaling pathways and cellular functions, including DNA damage repair, chromatin remodeling, histone methylation, and X chromosome inactivation [42]. MBD can also potentially coordinate the functions of DNA methyl-CpG binding and H3K9 methylation, both of which can promote epigenetic marks [42, 43]. SETDB1 also contains a unique tandem Tudor domain that recognizes histone H3 sequences containing acetylated lysines and methylated [44]. SETDB1 biological function is a two-edged sword. On the one hand, it may downregulate antioncogenes through histone methylation. On the other hand, it may inhibit tumor-intrinsic immunogenicity, enabling cancer cells to evade immune responses [12, 45].

A recent study indicated that tumor cell-intrinsic epigenetic alterations drive tumorigenesis and cancer progression [46]. The epigenetic characters reflect the heterogeneity of tumors and indicate potential epigenetic changes, which lead to cancer cell invasion during tumor progress [46, 47]. As an important player in tumor epigenetics, SETDB1 expression is significantly differential in most cancerous tissues and adjacent healthy tissues [8, 48–51], which is consistent with our findings. It is demonstrated that SETDB1 is an oncogene and an important prognostic factor in some tumors. SETDB1 expression is upregulated in LIHC tissues and is associated with tumor size, enhanced stage, and TNM classification [52]. Similarly, for TCGA-LIHC patients, we observed that the expression level of SETDB1 is significantly elevated in tumor tissues compared to that in paracancerous

tissues. LIHC tissues from patients with advanced-stage tumors show significantly higher expression levels of SETDB1 compared with those from patients with early-stage tumors. We also observed that the expression levels of SETDB1 were significantly lower in the stage 4 tumor than those in early-stage tumors. However, this result may be inaccurate due to the sample size limitation (six samples with stage 4 tumor). Therefore, a large sample size study is needed to further verify the conclusion.

Cancer develops as a result of genetic mutational events that lead either to the overexpression of growth-promoting oncogenes or the inactivation of cell cycle-controlling tumor suppressor genes [53]. Growing evidence implies that SETDB1 is a potential oncogene for tumorigenesis [8]. Therefore, comprehensively understanding the biological functions of SETDB1 mutations can help to inhibit tumorigenesis and develop effective antitumor agents. It has been reported that mutated SETDB1 is widespread and occurs in most malignant pleural mesothelioma [54, 55]. The frequent SETDB1 mutation indicates that there may be a potential therapeutic target for malignant pleural mesothelioma [55]. We first used the cBioPortal online database to explore genetic mutation levels of SETDB1 in various cancers. The pan-cancer mutation spectra showed that a high mutant frequency of SETDB1 occurred in most human tumors, with the highest frequency in UCEC (15.09% of 517 cases). These results also demonstrated that SETDB1 mutation played a significant role in tumorigenesis.

Cancer is an increasingly health-threatening disease that has a poor prognosis due to the lack of effective treatment. The progression and recurrence of the tumor challenge the effectiveness of therapies [56, 57]. Due to the therapeutic resistance and tumor relapse after therapy, the paradigms of cancer-centric therapeutics are not sufficient to eradicate the malignancy [58]. Targeting tumor microenvironment (TME) is a novel tumor treatment strategy in recent years. CAFs are the most abundant stromal cells in the TME and play significant roles in tumor development. Our results also revealed the significant association between SETDB1 expression and tumor-related immune cell infiltration level of CAFs in certain tumors, including ACC, BRCA, CESC, COAD, HNSC, HNSC-HPV, KIRP, LIHC, READ, and TGCT. Furthermore, we used the online databases to explore the correlation between SETDB1 expression and immune cell infiltration level in human cancer and found that tumor-related immune cells significantly increased in tumor tissues with high SETDB1 expression levels. These results also demonstrated that the expression levels of SETDB1 influenced tumor growth, metastasis, and prognosis.

Immunotherapy has emerged as a new pillar of cancer treatment in recent years. The introduction of PD-1, PD-L1, and CAR-T cell immunotherapy into the therapeutic strategy of advanced cancer leads to unprecedentedly prolonged survival for patients [59]. According to our findings, increased expression of SETDB1 has a significantly negative correlation with immunoinhibitor and immunostimulator in most cancers. Therefore, we speculated that decreasing SETDB1 expression in tumor cells might enhance immunotherapeutic responses.

5. Conclusion

In summary, we conducted the pan-cancer analysis of SETDB1 oncogenes for the first time. The omics analysis, prognostic analysis, methylation and phosphorylation analysis, immune analysis, and enrichment analysis of SETDB1 were performed. The mRNA and protein expression levels and gene alteration levels were analyzed. It is expected that the investigation and characterization of SETDB1 biological function can help to identify the key targets and regulatory pathways and promote human cancer treatment in the future.

Abbreviations

SETDB1:	SET domain bifurcated histone lysine methyltransferase 1
ACC:	Adrenocortical carcinoma
BLCA:	Bladder urothelial carcinoma
CESC:	Cervical squamous cell carcinoma and endocervical adenocarcinoma
ESCA:	Esophageal carcinoma
HNSC:	Head and neck squamous cell carcinoma
KIRC:	Kidney renal clear cell carcinoma
KIRP:	Kidney renal papillary cell carcinoma
LAML:	Acute myeloid leukemia
LGG:	Lower grade glioma
LIHC:	Liver hepatocellular carcinoma
MESO:	Mesothelioma
PAAD:	Pancreatic adenocarcinoma
READ:	Rectum adenocarcinoma
SARC:	Sarcoma
SKCM:	Skin cutaneous melanoma
UCE:	Uterine corpus endometrial carcinoma
UCS:	Uterine carcinosarcoma
UVM:	Uveal melanoma
CHOL:	Cholangiocarcinoma
COAD:	Colon adenocarcinoma
GBM:	Glioblastoma multiforme
KICH:	Kidney chromophobe
LUAD:	Lung adenocarcinoma
LUSC:	Lung squamous cell carcinoma
DLBC:	Lymphoid neoplasm diffuse large B-cell lymphoma
OV:	Ovarian serous cystadenocarcinoma
PCPG:	Pheochromocytoma and paraganglioma
PRAD:	Prostate adenocarcinoma
STAD:	Stomach adenocarcinoma
TGCT:	Testicular germ cell tumors
THYM:	Thymoma THCA thyroid carcinoma
K-M:	Kaplan-Meier
OS:	Overall survival
H3K9:	Histone H3 lysine 9
CDS:	Coding sequence
NCBI:	National Center for Biotechnology Information
IBS:	Illustrator for Biological Sequences
HPA:	Human Protein Atlas
UCSC:	University of California, Santa Cruz
TCGA:	The Cancer Genome Atlas

IHC: Immunohistochemical
H&:   Hematoxylin-eosin
DFS: Disease-free survival
DSS: Disease-specific survival
PFI: Progression-free interval
GEO: Gene Expression Omnibus
CNV: Copy number variation
SNV: Single nucleotide variation
COSMIC: Catalogue Of Somatic Mutations In Cancer
TMB: Tumor mutational burden
MSI: Microsatellite instability
GO: Gene Ontology
KEGG: Kyoto Encyclopedia of Genes and Genomes.

Data Availability

The data comes from the public database including TCGA, GEO, GTEx, and CPTAC.

Ethical Approval

Not applicable as this is not a study.

Conflicts of Interest

The authors declare that they have no competing interests.

Authors' Contributions

All authors read and approved the final manuscript. Xin Lin and Min Xiao collected and analyzed the data. Zhitao Chen and Xin Lin designed the study. Zhitao Chen, Chenchen Ding, and Xin Lin drafted the manuscript. Ting Zhang, Qiyong Li, and Min Xiao reviewed and revised the manuscript. Chenchen Ding and Ting Zhang edited the figures and tables of the article. All authors made substantial contributions to conception and design, acquisition of data, or analysis and interpretation of data; took part in drafting the article or revising it critically for important intellectual content; agreed to submit to the current journal; gave final approval of the version to be published; and agree to be accountable for all aspects of the work. Xin Lin and Min Xiao contributed equally to this work.

Supplementary Materials

The supplementary material for this article can be found at the module of Supplementary figure and table legends. Figure S1: SETDB1 protein, location, and conservation analysis. (A) The main location of SETDB1 protein in cells. (B) Comparison of protein sequences encoded by SETDB1 among ten different species. (C) The phylogenetic tree of SETDB1 in different species. (D) The sites of SETDB1 protein in the U-251 MG cell, U-2 OS cells, and A-431 cells. Figure S2: the expression levels of SETDB1 in normal and tumor tissues as well as different single cell types. (A) The expression levels of SETDB1 in normal and tumor tissues in human body. (B) The mRNA expression levels of SETDB1 in normal tissues (data from GTEx and release V6). (C) The mRNA expression levels of SETDB1 in different single cell types. (D) The SETDB1 expression patterns in testis tissues (data from published RNA-sequencing information). Figure S3: the differential expression level of SETDB1 between normal tissues and tumor tissues. (A) Data from TCGA+GTEx. (B) Data from TCGA. Figure S4: SETDB1 mutation analysis. (A) The main mutation type of SETDB1. (B) The primary single nucleotide variation (SNV) class type of SETDB1. (C) Correlation between SETDB1 expression and different mutated types. (D) The mutation spectrum of SETDB1. (E) The copy number variation (CNV) percentage of SETDB1 in each cancer. (F) Correlation between SETDB1 expression and CNV. Figure S5: correlation between SETDB1 expression and RNA modifications related genes. (A) m6A, (B) m1A, and (C) m5C. Figure S6: the heatmap of DNA methylation level of SETDB1 in different cancers. High-expression (red) and low-expression (blue). Figure S7: correlation of SETDB1 expression with immune infiltration level. Figure S8: protein-protein interaction (PPI) network and mRNA-miRNA-lncRNA network. (A) PPI network for SETDB1 and SETDB1-binding proteins using STRING. (B) The PPI network of SETDB1 and SETDB1-binding proteins using GeneMANIA. (C) The result of predicted miRNAs using 3 different databases based on Venn. (D) mRNA-miRNA-lncRNA network for SETDB1. Table S1: the SETDB1 expression level in normal tissues and tumor tissues based on the GEPIA2. Table S2: subgroup analysis on the correlation of SETDB1 expression and prognosis of gastric cancer cases. Table S3: subgroup analysis on the correlation of SETDB1 expression and prognosis of liver cancer cases. Table S4: subgroup analysis on the correlation of SETDB1 expression and prognosis of lung cancer cases. Table S5: subgroup analysis on the correlation of SETDB1 expression and prognosis of ovarian cancer cases. Table S6: subgroup analysis on the correlation of SETDB1 expression and prognosis of breast cancer cases. (*Supplementary Materials*)

References

[1] J. M. Kocarnik, K. Compton, F. E. Dean et al., "Cancer Incidence, Mortality, years of life lost, years lived with disability, and disability-adjusted life years for 29 cancer groups from 2010 to 2019: a systematic analysis for the global burden of disease study 2019," *JAMA Oncology*, vol. 8, no. 3, pp. 420–444, 2022.

[2] B. A. Helmink, M. A. W. Khan, A. Hermann, V. Gopalakrishnan, and J. A. Wargo, "The microbiome, cancer, and cancer therapy," *Nature Medicine*, vol. 25, no. 3, pp. 377–388, 2019.

[3] J. D. Mizrahi, R. Surana, J. W. Valle, and R. T. Shroff, "Pancreatic cancer," *Lancet*, vol. 395, no. 10242, pp. 2008–2020, 2020.

[4] R. D. Neal, F. Sun, J. D. Emery, and M. E. Callister, "Lung cancer," *BMJ*, vol. 365, p. l1725, 2019.

[5] N. Harbeck, F. Penault-Llorca, J. Cortes et al., "Breast cancer," *Nature Reviews. Disease Primers*, vol. 5, no. 1, article 2, 2019.

[6] J. Liu, C. Tokheim, J. D. Lee et al., "Genetic fusions favor tumorigenesis through degron loss in oncogenes," *Nature Communications*, vol. 12, no. 1, p. 6704, 2021.

[7] U. Harjes, "SETDB1, a new target for immunotherapy," *Nature Reviews. Cancer*, vol. 21, no. 7, p. 412, 2021.

[8] A. V. Karanth, R. R. Maniswami, S. Prashanth et al., "Emerging role of SETDB1 as a therapeutic target," *Expert Opinion on Therapeutic Targets*, vol. 21, no. 3, pp. 319–331, 2017.

[9] R. L. Bennett, A. Swaroop, C. Troche, and J. D. Licht, "The role of nuclear receptor-binding SET domain family histone lysine methyltransferases in cancer," *Cold Spring Harbor Perspectives in Medicine*, vol. 7, no. 6, article a026708, 2017.

[10] A. Sorrentino, A. Federico, M. Rienzo et al., "PR/SET domain family and cancer: novel insights from the Cancer Genome Atlas," *International Journal of Molecular Sciences*, vol. 19, no. 10, article 3250, 2018.

[11] P. J. Harte, W. Wu, M. M. Carrasquillo, and A. G. Matera, "Assignment of a novel bifurcated SET domain gene, SETDB1, to human chromosome band 1q21 by in situ hybridization and radiation hybrids," *Cytogenetics and Cell Genetics*, vol. 84, no. 1-2, pp. 83–86, 1999.

[12] G. K. Griffin, J. Wu, A. Iracheta-Vellve et al., "Epigenetic silencing by SETDB1 suppresses tumour intrinsic immunogenicity," *Nature*, vol. 595, no. 7866, pp. 309–314, 2021.

[13] R. Hachiya, T. Shiihashi, I. Shirakawa et al., "The H3K9 methyltransferase Setdb1 regulates TLR4-mediated inflammatory responses in macrophages," *Scientific Reports*, vol. 6, no. 1, article 28845, 2016.

[14] W. Liu, Y. Xie, J. Ma et al., "IBS: an illustrator for the presentation and visualization of biological sequences," *Bioinformatics*, vol. 31, no. 20, pp. 3359–3361, 2015.

[15] Z. Tang, C. Li, B. Kang, G. Gao, C. Li, and Z. Zhang, "GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses," *Nucleic Acids Research*, vol. 45, no. W1, pp. W98–w102, 2017.

[16] Z. Tang, B. Kang, C. Li, T. Chen, and Z. Zhang, "GEPIA2: an enhanced web server for large-scale expression profiling and interactive analysis," *Nucleic Acids Research*, vol. 47, no. W1, pp. W556–w560, 2019.

[17] A. Lánczky and B. Győrffy, "Web-based survival analysis tool tailored for medical research (KMplot): development and implementation," *Journal of Medical Internet Research*, vol. 23, no. 7, article e27633, 2021.

[18] V. Modhukur, T. Iljasenko, T. Metsalu, K. Lokk, T. Laisk-Podar, and J. Vilo, "MethSurv: a web tool to perform multivariable survival analysis using DNA methylation data," *Epigenomics*, vol. 10, no. 3, pp. 277–288, 2018.

[19] B. Ru, C. N. Wong, Y. Tong et al., "TISIDB: an integrated repository portal for tumor-immune system interactions," *Bioinformatics*, vol. 35, no. 20, pp. 4200–4202, 2019.

[20] H. Kuipers, T. J. de Bitter, M. T. D. Boer et al., "Gallbladder cancer: current insights in genetic alterations and their possible therapeutic implications," *Cancers*, vol. 13, no. 21, article 5257, 2021.

[21] S. Sakata, K. Otsubo, H. Yoshida et al., "Real-world data on NGS using the Oncomine DxTT for detecting genetic alterations in non-small-cell lung cancer: WJOG13019L," *Cancer Science*, vol. 113, no. 1, article 221, 2022.

[22] R. Heery and M. H. Schaefer, "DNA methylation variation along the cancer epigenome and the identification of novel epigenetic driver events," *Nucleic Acids Research*, vol. 49, no. 22, pp. 12692–12705, 2021.

[23] M. F. Perez and P. Sarkies, "Malignancy and NF-κB signalling strengthen coordination between expression of mitochondrial and nuclear-encoded oxidative phosphorylation genes," *Genome Biology*, vol. 22, no. 1, p. 328, 2021.

[24] S. J. Cowman and M. Y. Koh, "Revisiting the HIF switch in the tumor and its immune microenvironment," *Trends Cancer*, vol. 8, no. 1, pp. 28–42, 2022.

[25] M. Desbois and Y. Wang, "Cancer-associated fibroblasts: key players in shaping the tumor immune microenvironment," *Immunological Reviews*, vol. 302, no. 1, pp. 241–258, 2021.

[26] X. Mao, J. Xu, W. Wang et al., "Crosstalk between cancer-associated fibroblasts and immune cells in the tumor microenvironment: new findings and future perspectives," *Molecular Cancer*, vol. 20, no. 1, p. 131, 2021.

[27] F. G. Dall'Olio, A. Marabelle, C. Caramella et al., "Tumour burden and efficacy of immune-checkpoint inhibitors," *Nature Reviews Clinical Oncology*, vol. 19, no. 2, pp. 75–90, 2022.

[28] A. M. Johnson, J. M. Boland, J. Wrobel et al., "Cancer cell-specific major histocompatibility complex II expression as a determinant of the immune infiltrate organization and function in the NSCLC tumor microenvironment," *Journal of Thoracic Oncology*, vol. 16, no. 10, pp. 1694–1704, 2021.

[29] P. Karpiński, Ł. Łaczmański, and M. M. Sąsiadek, "Major histocompatibility complex genes as therapeutic opportunity for immune cold molecular cancer subtypes," *Journal of Immunology Research*, vol. 2020, Article ID 8758090, 9 pages, 2020.

[30] B. Zhou and S. Gao, "Construction and validation of a novel immune and tumor mutation burden-based prognostic model in lung adenocarcinoma," *Cancer Immunology, Immunotherapy*, vol. 71, no. 5, pp. 1183–1197, 2022.

[31] M. Sorokin, E. Rabushko, V. Efimov et al., "Experimental and meta-analytic validation of RNA sequencing signatures for predicting status of microsatellite instability," *Frontiers in Molecular Biosciences*, vol. 8, p. 737821, 2021.

[32] A. M. Taylor, J. Shih, G. Ha et al., "Genomic and functional approaches to understanding cancer aneuploidy," *Cancer Cell*, vol. 33, no. 4, pp. 676–689, 2018.

[33] J. Guan, B. Hallberg, and R. H. Palmer, "Chromosome imbalances in neuroblastoma-recent molecular insight into chromosome 1p-deletion, 2p-gain, and 11q-deletion identifies new friends and foes for the future," *Cancers*, vol. 13, no. 23, article 5897, 2021.

[34] S. M. Tirier, J. P. Mallm, S. Steiger et al., "Subclone-specific microenvironmental impact and drug response in refractory multiple myeloma revealed by single-cell transcriptomics," *Nature Communications*, vol. 12, no. 1, p. 6960, 2021.

[35] O. Sipos, H. Tovey, J. Quist et al., "Assessment of structural chromosomal instability phenotypes as biomarkers of carboplatin response in triple negative breast cancer: the TNT trial," *Annals of Oncology*, vol. 32, no. 1, pp. 58–65, 2021.

[36] X. Zhang, M. A. Bustos, R. Gross et al., "Interleukin enhancer-binding factor 2 promotes cell proliferation and DNA damage response in metastatic melanoma," *Clinical and Translational Medicine*, vol. 11, no. 10, article e608, 2021.

[37] L. Yu, F. Ye, Y. Y. Li et al., "Histone methyltransferase SETDB1 promotes colorectal cancer proliferation through the STAT1-

CCND1/CDK6 axis," *Carcinogenesis*, vol. 41, no. 5, pp. 678–688, 2020.

[38] K. Fukuda and Y. Shinkai, "SETDB1-mediated silencing of retroelements," *Viruses*, vol. 12, no. 6, article 596, 2020.

[39] M. Markouli, D. Strepkos, and C. Piperi, "Structure, activity and function of the SETDB1 protein methyltransferase," *Life*, vol. 11, no. 8, article 817, 2021.

[40] C. Cardoso, S. Timsit, L. Villard, M. Khrestchatisky, M. Fontès, and L. Colleaux, "Specific interaction between the XNP/ATR-X gene product and the SET domain of the human EZH2 protein," *Human Molecular Genetics*, vol. 7, no. 4, pp. 679–684, 1998.

[41] V. J. Lazaro-Camp, K. Salari, X. Meng, and S. Yang, "SETDB1 in cancer: overexpression and its therapeutic implications," *American Journal of Cancer Research*, vol. 11, no. 5, pp. 1803–1827, 2021.

[42] H. N. Cukier, J. M. Lee, D. Ma et al., "The expanding role of MBD genes in autism: identification of a MECP2 duplication and novel alterations in MBD5, MBD6, and SETDB1," *Autism Research*, vol. 5, no. 6, pp. 385–397, 2012.

[43] Y. K. Kang, "SETDB1 in early embryos and embryonic stem cells," *Current Issues in Molecular Biology*, vol. 17, pp. 1–10, 2015.

[44] Y. Guo, X. Mao, L. Xiong et al., "Structure-guided discovery of a potent and selective cell-active inhibitor of SETDB1 Tudor domain," *Angewandte Chemie (International Ed. in English)*, vol. 60, no. 16, pp. 8760–8765, 2021.

[45] J. Lin, D. Guo, H. Liu et al., "The SETDB1-TRIM28 complex suppresses antitumor immunity," *Cancer Immunology Research*, vol. 9, no. 12, pp. 1413–1424, 2021.

[46] S. J. Hogg, P. A. Beavis, M. A. Dawson, and R. W. Johnstone, "Targeting the epigenetic regulation of antitumour immunity," *Nature Reviews. Drug Discovery*, vol. 19, no. 11, pp. 776–800, 2020.

[47] C. L. Miranda Furtado, M. C. Dos Santos Luciano, R. D. Silva Santos, G. P. Furtado, M. O. Moraes, and C. Pessoa, "Epidrugs: targeting epigenetic marks in cancer treatment," *Epigenetics*, vol. 14, no. 12, pp. 1164–1176, 2019.

[48] L. Yuan, B. Sun, L. Xu, L. Chen, and W. Ou, "The updating of biological functions of methyltransferase SETDB1 and its relevance in lung cancer and mesothelioma," *International Journal of Molecular Sciences*, vol. 22, no. 14, article 7416, 2021.

[49] G. Wang, J. Long, Y. Gao et al., "SETDB1-mediated methylation of Akt promotes its K63-linked ubiquitination and activation leading to tumorigenesis," *Nature Cell Biology*, vol. 21, no. 2, pp. 214–225, 2019.

[50] D. Strepkos, M. Markouli, A. Klonou, A. G. Papavassiliou, and C. Piperi, "Histone methyltransferase SETDB1: a common denominator of tumorigenesis with therapeutic potential," *Cancer Research*, vol. 81, no. 3, pp. 525–534, 2021.

[51] S. Sharma, T. K. Kelly, and P. A. Jones, "Epigenetics in cancer," *Carcinogenesis*, vol. 31, no. 1, pp. 27–36, 2010.

[52] C. Wang, Z. Xia, Z. Li et al., "Expression of SET domain bifurcated histone lysine methyltransferase 1 and its clinical prognostic significance in hepatocellular carcinoma," *Journal of Clinical Laboratory Analysis*, vol. 36, article e24090, 2021.

[53] I. Martincorena, K. M. Raine, M. Gerstung et al., "Universal patterns of selection in cancer and somatic tissues," *Cell*, vol. 171, no. 5, pp. 1029–1041, 2017.

[54] J. Hmeljak, F. Sanchez-Vega, K. A. Hoadley et al., "Integrative molecular characterization of malignant pleural mesothelioma," *Cancer Discovery*, vol. 8, no. 12, pp. 1548–1565, 2018.

[55] H. C. Kang, H. K. Kim, S. Lee et al., "Whole exome and targeted deep sequencing identify genome-wide allelic loss and frequent SETDB1 mutations in malignant pleural mesotheliomas," *Oncotarget*, vol. 7, no. 7, pp. 8321–8331, 2016.

[56] Y. Zhang, S. H. Ho, B. Li, G. Nie, and S. Li, "Modulating the tumor microenvironment with new therapeutic nanoparticles: a promising paradigm for tumor treatment," *Medicinal Research Reviews*, vol. 40, no. 3, pp. 1084–1102, 2020.

[57] E. Dhimolea, R. de Matos Simoes, D. Kansara et al., "An embryonic diapause-like adaptation with suppressed Myc activity enables tumor treatment persistence," *Cancer Cell*, vol. 39, no. 2, pp. 240–256, 2021.

[58] N. Vasan, J. Baselga, and D. M. Hyman, "A view on drug resistance in cancer," *Nature*, vol. 575, no. 7782, pp. 299–309, 2019.

[59] R. S. Riley, C. H. June, R. Langer, and M. J. Mitchell, "Delivery technologies for cancer immunotherapy," *Nature Reviews. Drug Discovery*, vol. 18, no. 3, pp. 175–196, 2019.