

Epigenetic-dysregulated lncRNAs in Prostate Cancer Based on Multi-omics Analysis

Jianxiao Zheng

Yueqing City People's Hospital: Wenzhou Medical College Affiliated Yueqing Hospital

Xuexiao Xiang

Guangxi Medical University

Zhipeng Xu

Central South University Third Xiangya Hospital

Yangjing Xu

Wenzhou Medical College First Affiliated Hospital: The First Affiliated Hospital of Wenzhou Medical University

Xiaowei Huang

Wenzhou Medical University First Affiliated Hospital: The First Affiliated Hospital of Wenzhou Medical University

Huafang Su (✉ suhuafang@wzhospital.cn)

Wenzhou Medical University <https://orcid.org/0000-0002-3775-1196>

Primary research

Keywords: epigenetics, long non-coding RNAs, prostate cancer, prognostic

Posted Date: October 1st, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-948350/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background: The purpose of this study was to investigate the relationship between Long non-coding RNAs (lncRNAs) expression and epigenetic changes, and explore the prognostic value of these findings in patients with prostate cancer.

Methods: In this study, we systematically compared the histone modification and methylation regions on lncRNAs promoter and enhancer elements by multiomics. We analyzed the distribution characteristics of histone modified apparent lncRNAs promoters and enhancers in the genome. Single sample Gene Set Enrichment Analysis (ssGSEA) and Kyoto Encyclopedia of Gene and Genomes (KEGG) were used to analyze the functional enrichment of epigenetic dysregulated lncRNAs. The prostate cancer associated lncRNAs were downloaded from lnc2cancer V3.0 and survival analysis was performed.

Results: We found 1124 epigenetic-dysregulated lncRNAs (epi-lncRNAs) and 9734 epigenetic-dysregulated protein coding genes (epi-PCGs) in prostate cancer, and the abnormal frequency of lncRNAs was much lower than that of PCGs. H3K4me3, H3K4me1, H3K27ac and H3K27me3 abnormally modified histones accounted for a large proportion and were mainly concentrated in the promoter region. Epigenetic abnormal lncRNAs affects the biological specific molecular function of prostate cancer by changing the abnormal modification of histone. The abnormal expression of lncRNAs caused by abnormal modification of H3K36me3 may be an important factor in the occurrence of prostate cancer. Finally, based on lncRNAs analysis of the prognosis of prostate cancer samples, three lncRNAs (HAR1A, SNHG12 and SNHG17) were identified as prognostic markers of prostate cancer.

Conclusions: These findings may help to better understand the abnormal epigenetic regulation of lncRNAs expression in prostate cancer.

Background

Prostate cancer (PCa) remains the second leading cause of deaths due to cancer in the United States in men [1]. Siegel et al. [2] demonstrated that there will have 164,690 newly diagnosed PCa patients and 29,430 deaths in 2018 in the United States. Thus, the diagnosis of PCa in early stage is vitally important [3]. Currently, the serum prostate-specific antigen (PSA) screening remains the primary way for early diagnosis of PCa. However, the sensitivity and specificity of PSA test remains low [4]. Therefore, it is important to find notable biomarkers for the diagnosis of PCa.

Previous studies have shown that epigenetic modifications plays a crucial role in the diagnosis and prognosis of PCa and proving additional options for PCa diagnosis and treatment strategies [5, 6]. Distinct PCa molecular subtypes have been characterized based on gene expression, clinical, and prognostic features of PCa. Long non-coding RNA (lncRNA) has been linked to many human cancers, including genetically heterogeneous PCa [7, 8]. lncRNAs are frequently differentially expressed in PCa subtypes [9]. Aberrant lncRNA expression could be caused by alteration of epigenetic patterns, such as changes in DNA methylation and post-translational histone modifications. For example, lncRNA DLX6-

AS1 promotes malignant phenotype and lymph node metastasis in PCa by inducing LARGE methylation [10]. Long non-coding RNA LINC00673 silencing inhibits proliferation and drug resistance of PCa cells via decreasing KLF4 promoter methylation [11]. LncRNA HOXD-AS1 recruited WDR5 to directly regulate the expression of target genes by mediating histone H3 lysine 4 tri-methylation (H3K4me3), Thus promotes proliferation, castration resistance, and chemo-resistance in PCa [12]. Although global epigenetic abnormalities have been identified as prominent cancer hallmarks, attempts to characterize the relationship between epigenetic alterations and expression of lncRNAs and their prognostic value in prostate cancer subtypes have been limited.

In this study, we screened out the peaks specific to prostate cancer based on the physical location of the histone modified peaks, and identification of lncRNAs and protein coding genes (PCGs) with epigenetic disorders. Subsequently, we extracted the expression profiles of lncRNAs caused by different histone modifications, and calculated the enrichment scores of each sample in these lncRNAs using single sample Gene Set Enrichment Analysis (ssGSEA). Collected prostate cancer-related lncRNAs gene were further loaded from Lnc2Cancer v3.0, and were performed survival analysis. Finally, three epigenetic-dysregulated lncRNA (epi-lncRNA) genes specific for prostate cancer were screened out, and a prognostic-related model was established.

Material And Methods

2.1 Data source

The gene expression profile of prostate cancer gene expression profile and its corresponding clinical information were downloaded from the Cancer Genome Atlas (TCGA) database (<https://confluence.broadinstitute.org/display/GDAC/Home/>). Through the gene annotation file of GENCODE[13], the expression profiles were divided into lncRNAs and PCGs. Based on the sample number of TCGA, the expression profiles of lncRNAs and PCGs were divided into 495 prostate cancer samples and 52 normal samples. The 450K array data of prostate adenocarcinoma (PRAD) were downloaded from TCGA database. The array data was split into 498 prostate cancer samples and 50 normal samples based on the PRAD sample number. The K-Nearest Neighbors (KNN) method was used to fill in missing values in the prostate cancer data. The replicated narrowPeak data of 6 types of histone modifications (H3K9me3, H3K36me3, H3K27me3, H3K27ac, H3K4me1, and H3K4me3) was downloaded from the hg38 version of the prostate cancer cell line PC-3 (https://www.ncbi.nlm.nih.gov/assembly/GCF_000001405.26/) and normal cell line RWPE2 in the Encyclopedia of DNA Elements (ENCODE) database (<https://www.encodeproject.org/>). The data of 218 prostate cancer samples with prognostic information were download from the cBioPortal database (<https://www.cbioportal.org/>, MSKCC, Nat Med 2017).

2.2 Identification of lncRNAs and PCGs with epigenetic disorders

To explore the correlation between epigenetic variation and prostate cancer, we converted the normalized gene expression profile count into integer Count, and then used the R library "DESeq2" [14] to identify differentially expressed lncRNAs and PCGs. The Benjamini-hochberg method was used to determine P value. lncRNAs or PCGs with false discovery rate (FDR) <0.05 were considered significant. We screened out the peaks specific to prostate cancer based on the physical location of the histone modified peaks, and only retained the peaks with p-value <0.05 as the differential peaks, and then we combined the GENCODE file to obtain the differential histone modifications genes. We downloaded the human enhancer database from FANTOM5(<https://fantom.gsc.riken.jp/5/>) to identify the enhancer of the gene; the promoter of the gene is defined as 2 kb upstream of the transcription start site (TSS) and 0.5 kb downstream. We used the R library "ChIPseeker" [15] and the human lncRNAs Transcripts database(<https://lncipedia.org/>) to identify gene promoters. Finally, we used the Bumhunter method of the R library 'CHAMP' [16, 17] to identify the differentially methylated region (DMR). The area where Bumhunter DMR p.value <0.001 is considered a significant DMR. Finally, we defined lncRNAs and PCGs as epigenetic abnormally regulated lncRNAs or PCGs. The criteria were as follows: (1) lncRNAs and PCGs are differentially expressed in prostate cancer and normal samples; (2) promoters or enhancers at least overlap a region of differential histone modification or differential methylation (named epi-lncRNA, non-epi-lncRNA, epi-PCG, non-epi-PCG). To compare the genomic characteristics of epigenetic and non-dysregulated lncRNA/PCG, we compared the exons and transcripts of epi-lncRNA, non-epi-lncRNA, epi-PCG, and non-epi-PCG genes, and the number and length of genes.

2.3 Genomic map of epi-lncRNAs with different histone modifications

To explore the epigenetic characteristics of lncRNAs caused by histone modifications, we analyzed the distribution characteristics of the promoters and enhancers of epi-lncRNAs with different histone modifications on the genome.

2.4 Functional enrichment analysis of epigenetic-dysregulated lncRNAs

The "clusterProfiler"[18] package in R software was used to implement the Kyoto Encyclopedia of Gene and Genomes (KEGG) pathway [19]. The enrichment score of each sample in these lncRNAs was calculated by ssGSEA algorithm in the "GSVA" [20] package.

2.5 Genomic map of epi-lncRNAs with different histone modifications

We extracted the m6a, m5c, and m1A gene expression profiles from the TCGA prostate cancer expression profile, and calculated the correlation between these genes and 12 epi-lncRNA enrichment score.

2.6 Identification of candidate epigenetic disorders specific to prostate cancer lncRNA

To understand the relationship between epigenetic disorders lncRNA and cancer, we used Lnc2Cancer v3.0 (<http://www.bio-bigdata.net/lnc2cancer/>) to analyze the enrichment of epi-lncRNA and known cancer lncRNAs. We screened out epi-lncRNA related to prostate cancer. In addition, we combined the transcripts per kilobase million (TPM) expression profile data of TCGA to identify epi-lncRNAs with significant changes in expression in prostate cancer.

2.7 Correlation analysis of epi-lncRNA and prognosis of prostate cancer

To understand the potential prognostic value of epi-lncRNA dysregulation, we divided the samples into two subgroups of high expression and low expression based on the expression value of the above epi-lncRNA in cancer samples. The Kaplan Meier survival curve was used to estimate the survival difference between the two groups by “survminer” R package. The Receiver Operating Characteristic (ROC) curve was applied to evaluate the predictive accuracy of the risk model by the “timeROC” package in R [21]. The R software package max_stat was used to truncate genes to obtain high and low expression groups.

Results

3.1 Epigenetically dysregulated lncRNAs (epi-lncRNAs) and protein-coding genes (epi-PCGs) in prostate cancer

We aim to analyze the association between differential expression of lncRNA and epigenetic alterations in prostate cancer. Deseq2 identified significantly differentially expressed genes, including 4836 lncRNAs (Additional file1: Table S1) and 12410 PCGs (Additional file2: Table S2). Binding histone modification data and 450k methylation chip data, we identified 1124 epi-lncRNAs, 12130 non-epi-lncRNAs, 9734 epi-PCGs and 9626 non-epi-PCGs (Additional file3: Table S3). The aberrant frequency of lncRNAs (8%) was much lower than that of PCGs (50%) in prostate cancer (Fig. 1A).

To characterize genomic signatures of epi-lncRNAs, we compared the number and length of genes, exons of epi-lncRNAs, other lncRNAs (non-epi-lncRNAs), epi-PCGs, and other PCGs (non-epi-PCGs). Epi-lncRNAs had a larger number of transcripts but the transcript length was lower than that of non-epi-lncRNAs (Fig. 1B, C); epi-PCGs had a larger number of transcripts and the transcript length was higher than that of non-epi-PCGs; at the same time, epi-lncRNAs have longer exon length, epi-PCGs has more exon number and exon length (Fig. 1D, E).

3.2 The genomic landscape of epi-lncRNAs

We systematically analyzed epi-lncRNAs in prostate cancer, revealing the landscape of epi-lncRNAs caused by different histone modifications and differentially methylated regions (Fig. 2A). Abnormal histones modifications in lncRNA were mainly H3K4me3, H3K4me1, H3K27ac, and H3K27me3, and most epi-lncRNAs were accompanied by a variety of abnormal histones modifications (Fig. 2B). We also

observed that the abnormally modified regions of these histones were mainly concentrated in the promoter region (Fig. 2B).

3.3 The potential function analysis of epi-lncRNAs

To characterize the potential function of lncRNAs dysregulation caused by these histone modifications, we systematically analyzed the relationship between epi-lncRNAs expression and pathways in prostate cancer. The enrichment scores in adjacent samples were significantly higher than those in tumor samples, including H3K27ac_promoter, H3K4me1_enhancer, H3K4me3_enhancer, H3K4me3_promoter, H3K9me3_promoter, H3K27me3_enhancer, and H3K27me3_promoter (Fig. 3A). In addition, the enrichment scores of H3K9me3_enhancer, H3K36me3_enhancer, and H3K36me3_promoter in tumor samples were significantly higher than those in adjacent samples (Fig. 3A). This indicates that H3K9me3_enhancer, H3K36me3_enhancer, and H3K36me3_promoter might have cancer-promoting effects, while H3K27ac_promoter, H3K4me1_enhancer, H3K4me3_enhancer, H3K4me3_promoter, H3K9me3_promoter, H3K9me3_promoter, and H3K9me3_promoter might have protective effects.

In addition, we also evaluated the KEGG pathway score of each sample, and calculated the relationship between the enrichment score of each epi-lncRNA and KEGG pathway. A total of 39 KEGG pathways were related to 12 types of epi-lncRNA (Fig. 3B). This suggested that different types of epi-lncRNA-related pathways have a certain consistency, and were closely related to the occurrence, development and metabolism of tumors.

3.4 Dysregulated epi-lncRNAs and RNA modification

RNA modification is an important epigenetic feature, which is related to a variety of important biological processes. We have observed that these enrichment scores were significantly correlated with M6a, M5c and M1A modified genes (Fig. 4). Meantime, we found significant differences in the correlation between H3K36me3 _ enhancer, H3K36me3 _ promoter and M6A, M5c, M1A modified genes. The former was significantly correlated with M6A RNA modification related genes.

3.5 Screening of epi-lncRNA specific for prostate cancer

To further ensure the regulatory role of candidate lncRNAs in prostate cancer, we collected prostate cancer-related lncRNA gene set from Lnc2Cancer v3.0, and found that 14 epi-lncRNAs have been reported to be directly related to prostate cancer (Fig. 5A; Additional file4,5: Table S4,5).

Then we calculated the difference between 14 epi-lncRNAs in prostate cancer samples and normal samples. The expression of HOTTIP, LINC00663, LINC00844, and MIR222HG in tumors was significantly lower than that of normal samples, while the expression of the remaining 10 genes (GLIDR, HOXA11-AS, LINC00308, PCAT14, PCAT18, PCAT5, SNHG12, SNHG17, SNHG6, and TMPO-AS1) in tumors was significantly higher than that of normal samples (Fig. 5B). Among these 14 epi-lncRNAs, only HOXA11-AS was related to both histone-modified promoters and DMR promoters, and the remaining 13 epi-lncRNAs were related to histone-modified promoters (Fig. 5C, D; Additional file6: Fig S1).

Overall, the 78.57% of these 14 genes on the TCGA dataset are consistent with the performance of the Lnc2Cancer v3.0 database (Additional file5: Table S5). Among them, the expression of 11 genes (GLIDR, HOXA11-AS, LINC00308, LINC00663, LINC00844, PCAT14, PCAT18, PCAT5, SNHG17, SNHG6, and TMPO-AS1) in the tumors of the TCGA data was consistent with the performance in the Lnc2Cancer v3.0 database.

3.6 The correlation between epi-lncRNAs and prognosis of prostate cancer;

To gain insight into the potential prognostic value of epi-lncRNAs dysregulation, we used 91 epi-lncRNA genes in the above results to perform survival analysis on TCGA prostate cancer samples. In the univariate COX analysis, there were four genes (HAR1A, LINC00629, SNHG12 and SNHG17) related to the survival of prostate cancer patients (Fig. 6A). We found that except for the high and low expression survival curves of gene LINC00629, the survival curves of the other 3 genes were significant (Fig. 6B-E).

3.7 Three epi-lncRNA signatures could be used as prognostic markers

To establish a prognostic model of epi-lncRNAs dysregulation, the 3-EpiLncRNA signature was established using multivariate survival analysis in the TCGA data set. The higher 3-EpiLncRNA score has a higher mortality rate (Fig. 7A). ROC analysis showed that the 3-EpiLncRNA score of 1-, 3-,5-year AUC was 1, 0.78,0.79, which had good prognostic performance (Fig. 7B). The prognosis of high 3-EpiLncRNA score was significantly worse than that of low 3-EpiLncRNA score (Fig. 7C).

To evaluate the portability of 3-EpiLncRNA score, we obtained the data set MSKCC with prognostic information from the cBioPortal database, and extracted the expression profiles of 3 epi-lncRNAs (the gene SNHG12 does not exist in the data set MSKCC). The 3-EpiLncRNA score of the sample was evaluated by the above method, and ROC analysis showed that the 1-year AUC of 3-EpiLncRNA score was 0.69, which had good prognostic prediction performance (Fig. 7D). The prognosis of high 3-EpiLncRNA score was significantly worse than that of low 3-EpiLncRNA score (Fig. 7E), which was consistent with the previous results.

Discussion

In this study, we found a large number of epi-lncRNAs by using public databases such as TCGA, ENCODE and lncRNAs Transcripts, combined with the joint analysis of genomics, transcriptomics and epigenetic genomics. Many of these epi-lncRNAs have been reported to be associated with cancer in previous literature.

It can be clearly seen from our results that the rate of epi-lncRNA in prostate cancer (8.5%) is much lower than that of epi-PCG (50.2%) (Fig. 1), which is basically consistent with previous literature reports. Previous studies have shown that in specific tissues and cells, the overall occupancy of histone marks

across lncRNA genes is in the range 27% - 38%, while that on the transcriptional start sites of protein-coding is 65%-73% [22]. LncRNA transcription is generally regarded as a genome-wide monitoring mechanism and plays an important role in RNA quality control. The relatively low aberration rate of lncRNA in prostate cancer also indirectly reflects its high stability in the process of gene transcription or expression, which may be more suitable as a molecular marker of prostate cancer. In addition, from characterizing genomic signatures, it is also observed that epi-lncRNA had similar structural characteristics to epi-PCG except for the shorter length of transcripts. Compared with non-epi-lncRNA and non-epi-PCG, they have more transcripts, exons and longer exon length (Fig. 2). Based on the multiple relationships between epi-lncRNAs and PCGs (inter-gene, overlap, partial overlap, intron or exon) [23, 24], the more complex the splicing pattern of lncRNA is, the more likely it is to be regulated by abnormal epigenetic modification [25]. Our results also coincide with the above view.

In the landscape of epi-lncRNA, it is not difficult to find that H3K4me3, H3K4me1, H3K27ac and H3K27me3 account for a high proportion of abnormally modified histones in the causes of lncRNA epigenetic disorders, and these abnormally modified histones are mainly concentrated in the promoter region (Fig. 2). Combined with the ssGSEA analysis of epi-lncRNAs (Fig. 3), we further found that the abnormal expression of H3K4me3, H3K4me1, H3K27ac and H3K27me3, whether located in the enhancer or promoter region, they almost consistently showed protective effects. H3K9me3_enhancer, H3K36me3_enhancer and H3K36me3_promoter showed significant cancer promoting effect. Although the latter accounts for only a small proportion in the epi-lncRNA genome landscape, it eventually leads to tumorigenesis. It shows that abnormal modification of H3K36me3 plays an extremely important role in the development of prostate cancer. This seems to reflect a message that under the gene expression pattern of prostate cancer, the final outcome of epigenetic aberrant regulation does not depend on the number of histone aberrant modifications. On the contrary, the type, location or function of histone aberrant modification may be more important. Inhibition of abnormal modification of H3K36me3 may bring new hope for the treatment of prostate cancer.

We analyzed 39 KEGG pathways most related to 12 kinds of epi-lncRNA and found that different types of epi-lncRNA-related pathways had certain consistency and included a large number of metabolism-related pathways (HISTIDINE_METABOLISM, TYROSINE_METABOLISM, BETA_ALANINE_METABOLISM, PHENYLALANINE_METABOLISM) and tumor-related pathways (COLORECTAL_CANCER, ENDOMETRIAL_CANCER, PROSTATE_CANCER). Thus, it can be seen that epigenetic abnormal lncRNA affects the biological specific molecular function of prostate cancer by changing the abnormal modification of histone and plays an important role in the pathogenesis of prostate cancer.

Previously, RNA modification was not considered to be a cancer driver, but accumulated evidence gradually suggests that abnormal RNA modification contributes to cancer cell proliferation, self-renewal, migration, stress adaptation and survival, and was named epitranscriptome in 2015[26]. The functional study of these modifications is now on the rise and has shown great significance in human pathology [27, 28]. We visualized the correlation [between] 12 categories of epi-lncRNA enrichment scores and M6A, M5C and M1A modified genes. Among them, enhancer and promoter of H3K9me3, H3K27me3, H3K27ac,

H3K4me1 and H3K4me3 have both similar and unique correlations with these genes. This suggests that there may be different regulation modes of lncRNA disorders caused by histone modification in enhancer and promoter, and these epi-lncRNAs are closely related to RNA modification. These results have confirmed previous findings. These findings [29–31] emphasize that DNA, RNA and histone and their modification work together in a synergistic manner, leading to a special chromatin state that determines the important function and ultimate direction of the genome. m6A is an important post-transcriptional regulation mechanism of genes [32–34]. It is also the most abundant and well-characterized internal modification in mRNA. It plays an important role in a variety of normal and pathobiological processes, including the regulation of self-renewal of embryonic stem cells and cancer cells, and survival after DNA damage [35–37]. However, how N6-Methyladenosine m6A is accurately and dynamically deposited in the transcriptome has always been a mystery. Until a recent study found [38], H3K36me3 was regarded as a transcriptional extension marker that could guide m6A modification globally. M6A modification was enriched near the H3K36me3 peak. When intracellular H3K36me3 was exhausted, m6A modification decreased as a whole. H3K36me3 can be directly recognized and bound by RNA methyltransferase METTL14, promoting the binding of m6A-MTC to adjacent RNA polymerase II, thus transferring m6A-MTC to newly born RNA with active transcriptional activity, resulting in m6A co-transcriptional deposition. Our results further support the above findings.

One of the 14 prostate cancer epi-lncRNAs screened (HOXA11-AS) was not only related to the histone modified promoter but also related to the DMR promoter. The remaining 13 related epi-lncRNAs (GLIDR, LINC00308, PCAT14, PCAT18, PCAT5, SNHG12, SNHG17, SNHG6, MPo-AS1, HOTTIP, LINC00663, LINC00844, MIR222HG) was only associated with histone modified promoters. These results were consistent with our previous findings in the landscape of epigenetic-dysregulated lncRNAs of prostate cancer, that is, the areas with abnormal histone modification were mainly concentrated in the promoter region (Fig. 3B). By comparing the performance of the 14 genes on the TCGA dataset, 78.57% of them were consistent with those in the Lnc2CancerV3.0 database, which once again verified the reliability of the research results.

We identified three lncRNAs (HAR1A, SNHG12, SNHG17) to construct the prostate cancer 3-EpiLncRNA signature. These three lncRNAs are independent prognostic markers of prostate cancer and have been identified as risk factors for a variety of cancers. HAR1A, also known as LINC00064, is widely expressed in brain, testis, spleen, lymph nodes and prostate tissues. Previous studies have shown that HAR1 is related to the development and evolution of the brain [39, 40]. A recent study suggests that HAR1 is also associated with tumorigenesis, regulating the progression of oral cancer through α -kinase-1, bromodomain-7 and myosin IIA axis [41]. SNHG12 is usually overexpressed in tumor cells and may promote tumorigenesis and metastasis by acting as a sponge of microRNA [42, 43]. There is evidence that SNHG12 promotes prostate tumorigenesis and progression through AKT regulation [44], and promotes renal cell carcinoma progression and resistance to sunitinib through up-regulation of CDCA3 [45]. LncRNASNHG17 is a non-coding RNA widely expressed in human body, with the highest expression in bone marrow, followed by high expression in ovary, prostate, thyroid and other organs, which is closely related to the progression of osteosarcoma [46] and promotes the proliferation of gastric and colon

cancer cells through epigenetic silencing of P57[47–49]. In order to evaluate the predictive value of the model, we visualize the model and draw the survival curve and ROC curve (Fig. 8). The high 3-EpiLncRNA score of prostate cancer began to show a significant survival disadvantage in the fifth year, and the number of survivors in the 10th year was only 1/3 of the low 3-EpiLncRNA score. ROC analysis showed that the 1-year AUC, 3-year AUC and 5-year AUC of 3-EpiLncRNA signature were 1, 0.78 and 0.79, respectively, indicating that the model has a good ability to evaluate the prognosis and can be used to predict the survival of prostate cancer.

In summary, our study analyzed the relationship between abnormal expression of lncRNA and epigenetic disorders and related genomic characteristics, and screened three specific epi-lncRNA for the construction of prostate cancer prognosis-related models. Mining data from a new perspective and supporting previous experimental research. At the same time, it is also found that the abnormal expression of lncRNA caused by the abnormal modification of enhancer-related H3K36me3 may be an important factor in the occurrence of prostate cancer. Reversing the abnormal modification of H3K36me3 may provide a new way for the treatment of prostate cancer, but further experimental research is needed.

At present, unexpected progress has been made in the analysis of the molecular mechanism of epigenetic regulation, which has a far-reaching impact on a better understanding of normal development and the treatment of human diseases. Different from genetic changes, the errors in epigenetic characteristics will be reversible, which has important guiding significance for disease treatment and drug research. With the deepening of epigenetic and other multi-group studies, the veil of prostate cancer will be lifted layer by layer, and it is expected to achieve an important breakthrough in future research.

Abbreviations

lncRNA: Long non-coding RNA

ssGSEA: Single sample Gene Set Enrichment Analysis

KEGG: Kyoto Encyclopedia of Gene and Genomes

epi-lncRNA: epigenetic-dysregulated lncRNA

epi-PCGs: epigenetic-dysregulated protein coding genes

PCa: Prostate cancer

PSA: prostate-specific antigen

H3K4me3: H3 lysine 4 tri-methylation

PCGs: protein coding genes

TCGA: the Cancer Genome Atlas

PRAD: prostate adenocarcinoma

KNN: K-Nearest Neighbors

ENCODE: Encyclopedia of DNA Elements

FDR: false discovery rate

TSS: transcription start site

DMR: differentially methylated region

TPM: transcripts per kilobase million

ROC: Receiver Operating Characteristic

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

The datasets generated and analysed during the current study are available in the Cancer Genome Atlas (TCGA) database [<https://confluence.broadinstitute.org/display/GDAC/Home/>] and the cBioPortal database [<https://www.cbioportal.org/>, MSKCC, Nat Med 2017].

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by the ZheJiang Province Public Welfare Technology Application Research Project [number 2021KY782] and the Wenzhou Municipal Science and Technology Bureau [number Y2020151].

Authors' contributions

Conceptualization, J.Z, X.X, H.S; methodology and software, J.Z, Z.X, Y.X, X,H; validation, J.Z, Y.X, X.H, H.S; formal analysis, X.X, Z.X, Y.X, X.H; data curation, all authors; writing—original draft preparation, J.Z, X.X,

H.S; writing—review and editing, J.Z, X.X, H.S; visualization, J.Z, Z.X; supervision, H.S; project administration and funding acquisition. All authors have read and agreed to the published version of the manuscript.

Acknowledgments

Not applicable

References

1. Wu Y, Wang A, Zhu B, Huang J, Lu E, Xu H, et al. KIF18B promotes tumor progression through activating the Wnt/ β -catenin pathway in cervical cancer. *Onco Targets Ther.* 2018;11:1707-20.
2. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2018. *CA Cancer J Clin.* 2018;68:7-30.
3. Zhang W, Shu P, Wang S, Song J, Liu K, Wang C, et al. ZNF154 is a promising diagnosis biomarker and predicts biochemical recurrence in prostate cancer. *Gene.* 2018;675:136-43.
4. Greene KL, Albertsen PC, Babaian RJ, Carter HB, Gann PH, Han M, et al. Prostate specific antigen best practice statement: 2009 update. *J Urol.* 2013;189:S2-2S11.
5. Chiam K, Ricciardelli C, Bianco-Miotto T. Epigenetic biomarkers in prostate cancer: Current and future uses. *Cancer Lett.* 2014;342:248-56.
6. Graça I, Pereira-Silva E, Henrique R, Packham G, Crabb SJ, Jerónimo C. Epigenetic modulators as therapeutic targets in prostate cancer. *Clin Epigenetics.* 2016;8:98.
7. Guo F, Parker Kerrigan BC, Yang D, Hu L, Shmulevich I, Sood AK, et al. Post-transcriptional regulatory network of epithelial-to-mesenchymal and mesenchymal-to-epithelial transitions. *J Hematol Oncol.* 2014;7:19.
8. Guo H, Ahmed M, Hua J, Soares F, He HH. Crucial role of noncoding RNA in driving prostate cancer development and progression. *Epigenomics.* 2017;9:1-3.
9. Crea F, Venalainen E, Ci X, Cheng H, Pikor L, Parolia A, et al. The role of epigenetics and long noncoding RNA MIAT in neuroendocrine prostate cancer. *Epigenomics.* 2016;8:721-31.
10. Zhao Z, Liang S, Sun F. LncRNA DLX6-AS1 Promotes Malignant Phenotype and Lymph Node Metastasis in Prostate Cancer by Inducing LARGE Methylation. *Front Oncol.* 2020;10:1172.
11. Jiang Z, Zhang Y, Chen X, Wu P, Chen D. Long non-coding RNA LINC00673 silencing inhibits proliferation and drug resistance of prostate cancer cells via decreasing KLF4 promoter methylation. *J Cell Mol Med.* 2020;24:1878-92.
12. Gu P, Chen X, Xie R, Han J, Xie W, Wang B, et al. lncRNA HOXD-AS1 Regulates Proliferation and Chemo-Resistance of Castration-Resistant Prostate Cancer via Recruiting WDR5. *Mol Ther.* 2017;25:1959-73.
13. Derrien T, Johnson R, Bussotti G, Tanzer A, Djebali S, Tilgner H, et al. The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res.* 2012;22:1775-89.

14. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 2014;15:550.
15. Yu G, Wang LG, He QY. ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics.* 2015;31:2382-3.
16. Tian Y, Morris TJ, Webster AP, Yang Z, Beck S, Feber A, et al. ChAMP: updated methylation analysis pipeline for Illumina BeadChips. *Bioinformatics.* 2017;33:3982-4.
17. Morris TJ, Butcher LM, Feber A, Teschendorff AE, Chakravarthy AR, Wojdacz TK, et al. ChAMP: 450k Chip Analysis Methylation Pipeline. *Bioinformatics.* 2014;30:428-30.
18. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS.* 2012;16:284-7.
19. Altermann E, Klaenhammer TR. PathwayVoyager: pathway mapping using the Kyoto Encyclopedia of Genes and Genomes (KEGG) database. *BMC Genomics.* 2005;6:60.
20. Hänzelmann S, Castelo R, Guinney J. GSEA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics.* 2013;14:7.
21. Blanche P, Dartigues JF, Jacqmin-Gadda H. Estimating and comparing time-dependent areas under receiver operating characteristic curves for censored event times with competing risks. *Stat Med.* 2013;32:5381-97.
22. Sati S, Ghosh S, Jain V, Scaria V, Sengupta S. Genome-wide analysis reveals distinct patterns of epigenetic features in long non-coding RNA loci. *Nucleic Acids Res.* 2012;40:10018-31.
23. Frankish A, Diekhans M, Ferreira AM, Johnson R, Jungreis I, Loveland J, et al. GENCODE reference annotation for the human and mouse genomes. *Nucleic Acids Res.* 2019;47:D766-766D773.
24. Du Z, Fei T, Verhaak RG, Su Z, Zhang Y, Brown M, et al. Integrative genomic analyses reveal clinically relevant long noncoding RNAs in human cancer. *Nat Struct Mol Biol.* 2013;20:908-13.
25. Zhao H, Liu X, Yu L, Lin S, Zhang C, Xu H, et al. Comprehensive landscape of epigenetic-dysregulated lncRNAs reveals a profound role of enhancers in carcinogenesis in BC subtypes. *Mol Ther Nucleic Acids.* 2021;23:667-81.
26. Nombela P, Miguel-López B, Blanco S. The role of m(6)A, m(5)C and Ψ RNA modifications in cancer: Novel therapeutic opportunities. *Mol Cancer.* 2021;20:18.
27. Barbieri I, Kouzarides T. Role of RNA modifications in cancer. *Nat Rev Cancer.* 2020;20:303-22.
28. Jonkhout N, Tran J, Smith MA, Schonrock N, Mattick JS, Novoa EM. The RNA modification landscape in human disease. *RNA.* 2017;23:1754-69.
29. Martienssen R, Moazed D. RNAi and heterochromatin assembly. *Cold Spring Harb Perspect Biol.* 2015;7:a019323.
30. Chalker DL, Meyer E, Mochizuki K. Epigenetics of ciliates. *Cold Spring Harb Perspect Biol.* 2013;5:a017764.
31. Reyes-Turcu FE, Zhang K, Zofall M, Chen E, Grewal SI. Defects in RNA quality control factors reveal RNAi-independent nucleation of heterochromatin. *Nat Struct Mol Biol.* 2011;18:1132-8.

32. Wang X, Lu Z, Gomez A, Hon GC, Yue Y, Han D, et al. N6-methyladenosine-dependent regulation of messenger RNA stability. *Nature*. 2014;505:117-20.
33. Wang X, Zhao BS, Roundtree IA, Lu Z, Han D, Ma H, et al. N(6)-methyladenosine Modulates Messenger RNA Translation Efficiency. *Cell*. 2015;161:1388-99.
34. Huang H, Weng H, Sun W, Qin X, Shi H, Wu H, et al. Recognition of RNA N(6)-methyladenosine by IGF2BP proteins enhances mRNA stability and translation. *Nat Cell Biol*. 2018;20:285-95.
35. Davalos V, Blanco S, Esteller M. SnapShot: Messenger RNA Modifications. *Cell*. 2018;174:498-498.e1.
36. Frye M, Harada BT, Behm M, He C. RNA modifications modulate gene expression during development. *Science*. 2018;361:1346-9.
37. Xiang Y, Laurent B, Hsu CH, Nachtergaele S, Lu Z, Sheng W, et al. RNA m(6)A methylation regulates the ultraviolet-induced DNA damage response. *Nature*. 2017;543:573-6.
38. Huang H, Weng H, Zhou K, Wu T, Zhao BS, Sun M, et al. Histone H3 trimethylation at lysine 36 guides m(6)A RNA modification co-transcriptionally. *Nature*. 2019;567:414-9.
39. Pollard KS, Salama SR, Lambert N, Lambot MA, Coppens S, Pedersen JS, et al. An RNA gene expressed during cortical development evolved rapidly in humans. *Nature*. 2006;443:167-72.
40. Walter Costa MB, Höner Zu Siederdisen C, Tulpan D, Stadler PF, Nowick K. Temporal ordering of substitutions in RNA evolution: Uncovering the structural evolution of the Human Accelerated Region 1. *J Theor Biol*. 2018;438:143-50.
41. Lee CP, Ko AM, Nithiyantham S, Lai CH, Ko YC. Long noncoding RNA HAR1A regulates oral cancer progression through the alpha-kinase 1, bromodomain 7, and myosin IIA axis. *J Mol Med (Berl)*. 2021;99:1323-34.
42. Li ZR, Yang Q, Zhou T, Huang YH, Zhang HZ, Meng R, et al. Prognostic Value of Long Noncoding RNA SNHG12 in Various Carcinomas: A Meta-Analysis. *Biomed Res Int*. 2020;2020:8847401.
43. Liu W, Che J, Gu Y, Song L, Jiao Y, Yu S. Silencing of lncRNA SNHG12 inhibits proliferation and migration of vascular smooth muscle cells via targeting miR-766-5p/EIF5A axis. *Adv Clin Exp Med*. 2021;30:591-8.
44. Chen Z, Qi T, Qin XP, Wang J, Huang ZS, Hu XY, et al. Long Noncoding RNA SNHG12 Promotes Prostate Tumor Occurrence and Progression via AKT Regulation. *Biomed Res Int*. 2020;2020:8812923.
45. Liu Y, Cheng G, Huang Z, Bao L, Liu J, Wang C, et al. Long noncoding RNA SNHG12 promotes tumour progression and sunitinib resistance by upregulating CDCA3 in renal cell carcinoma. *Cell Death Dis*. 2020;11:515.
46. Zhang YT, Yang GY. LncRNA SNHG17 acts as a ceRNA of miR-324-3p to contribute the progression of osteosarcoma. *J Biol Regul Homeost Agents*. 2020;34:1529-33.
47. Ma Z, Gu S, Song M, Yan C, Hui B, Ji H, et al. Long non-coding RNA SNHG17 is an unfavourable prognostic factor and promotes cell proliferation by epigenetically silencing P57 in colorectal cancer. *Mol Biosyst*. 2017;13:2350-61.

48. Chen LL, He J, Qiu XT, Yu J, Wang ZM. The prognostic roles of long non-coding RNA SNHG17 in the patients with gastric cancer. *Eur Rev Med Pharmacol Sci.* 2019;23:1063-8.
49. Zhang G, Xu Y, Wang S, Gong Z, Zou C, Zhang H, et al. LncRNA SNHG17 promotes gastric cancer progression by epigenetically silencing of p15 and p57. *J Cell Physiol.* 2019;234:5163-74.

Figure

Figure 8 is not available with this version.

Figures

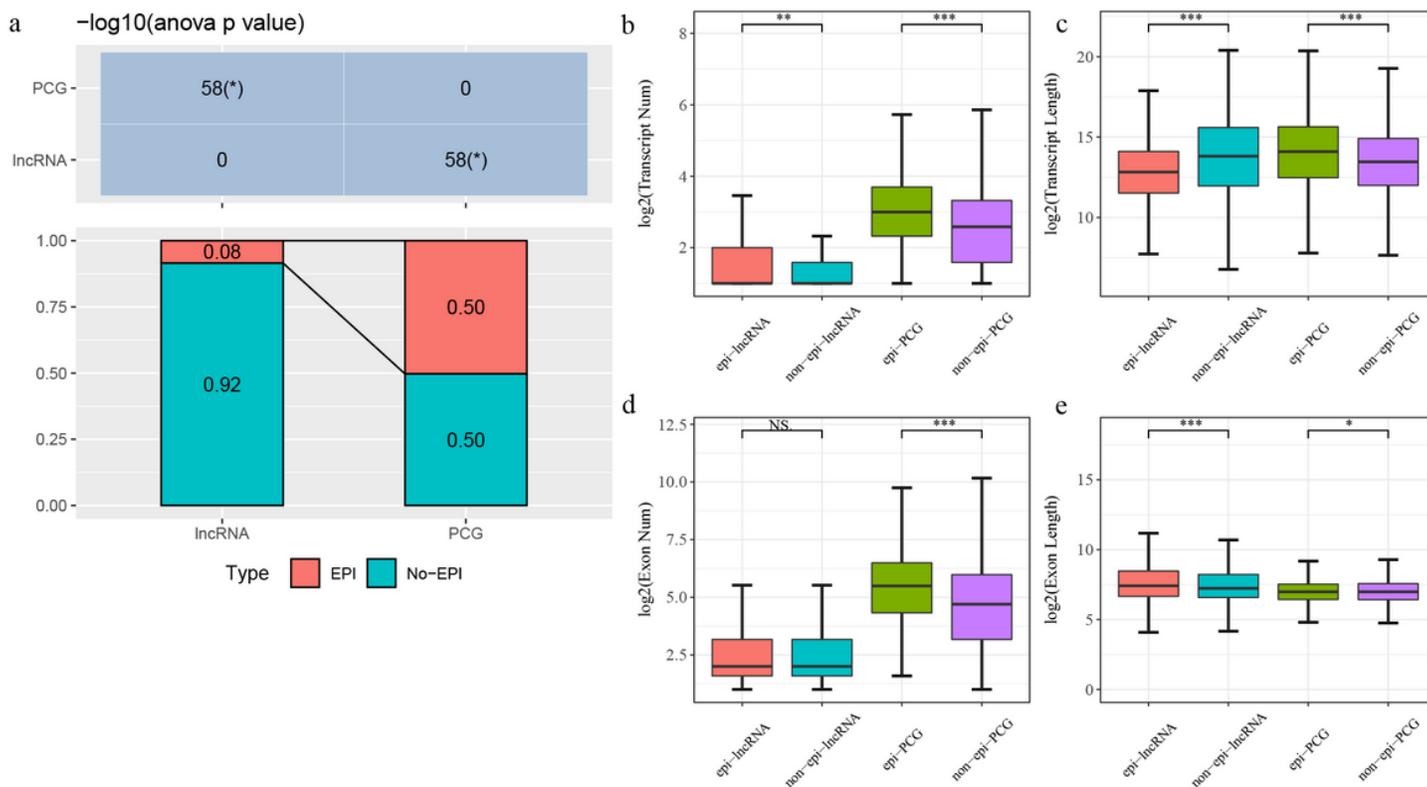


Figure 1

Epigenetically dysregulated lncRNAs (epi-lncRNAs) and protein-coding genes (epi-PCGs) in prostate cancer (A). The percentage of epi-lncRNAs and epi-PCGs to all lncRNAs and PCGs in the genome; (B-E). Comparison of genomic characteristics between apparently dysregulated lncRNA/PCG and non-apparently dysregulated lncRNA/PCG. * $p < 0.05$; ** $p < 0.01$, *** $p < 0.001$.

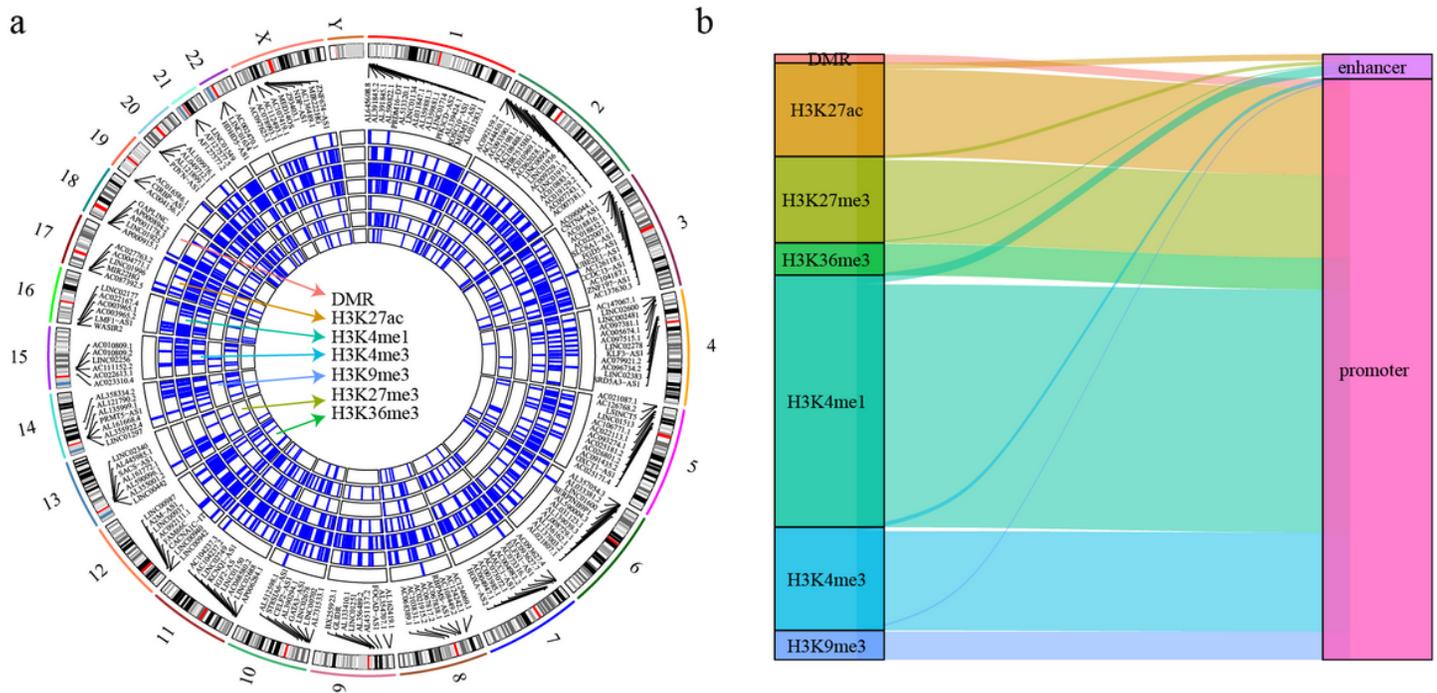


Figure 2

The genomic landscape of epi-lncRNAs (A). Genomic landscape of epi-lncRNAs caused by different histone modifications and differentially methylated regions; (B). distribution of epi-lncRNAs in different apparent disorders.

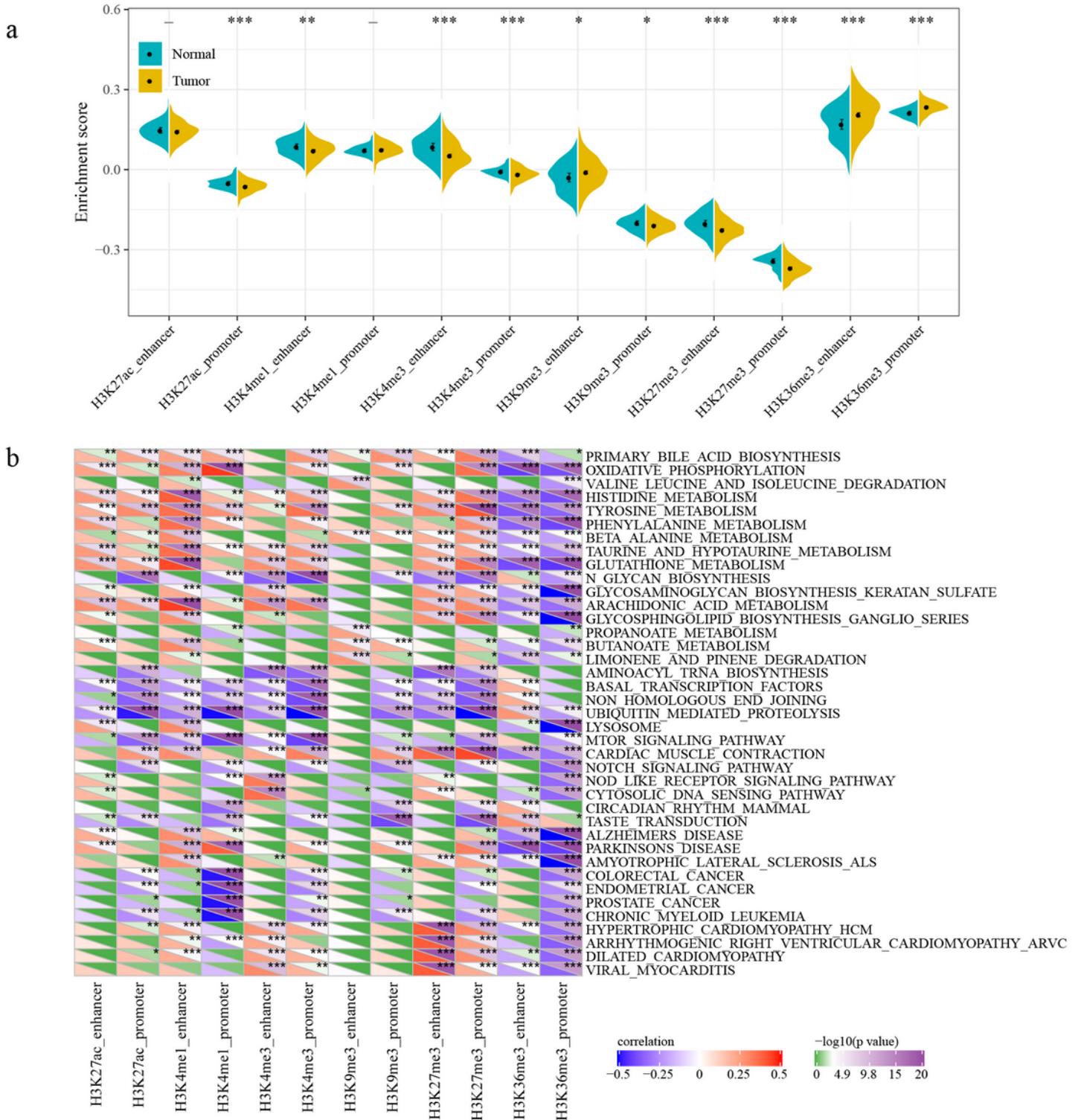


Figure 3

The potential function analysis of epi-lncRNAs (A). Functional analysis of epi-lncRNA; (B) The most relevant KEGG Pathway of 12 kinds of dysregulated lncRNAs. * $p < 0.05$; ** $p < 0.01$, *** $p < 0.001$.

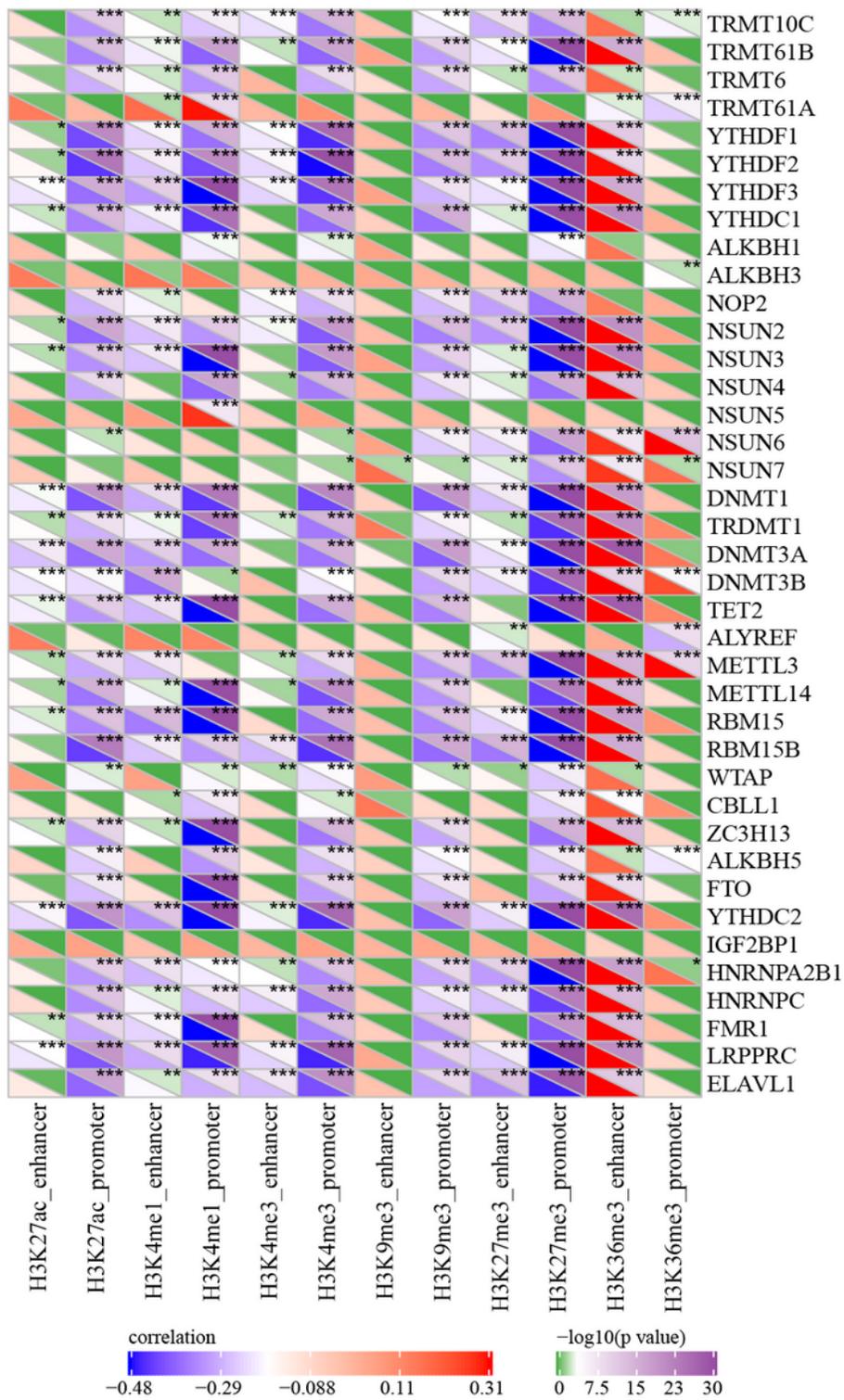


Figure 4

Dysregulated epi-lncRNAs and RNA modification. The correlation between the enrichment scores of 12 kinds of epi-lncRNAs and m6a, m5c, and m1A genes. * $p < 0.05$; ** $p < 0.01$, *** $p < 0.001$.

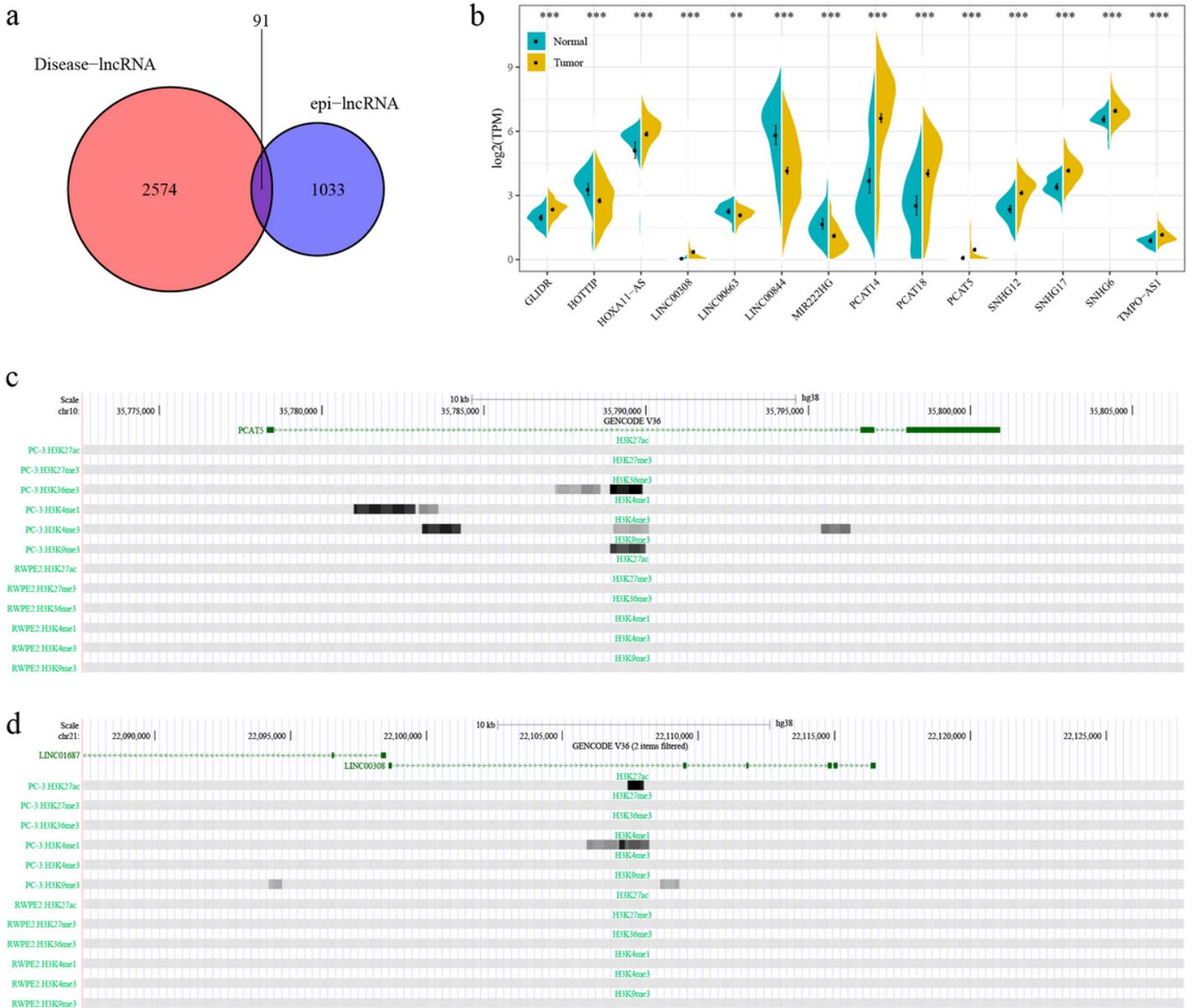


Figure 5

Screening of epi-lncRNA specific for prostate cancer (A). The intersection of disease-related lncRNA and prostate cancer epi-lncRNA; (B). The expression of epi-lncRNA specific to prostate cancer is different in tumor and normal samples; (C-D). PCAT5 and LINC00308 histone modification spectrum. * $p < 0.05$; ** $p < 0.01$, *** $p < 0.001$.

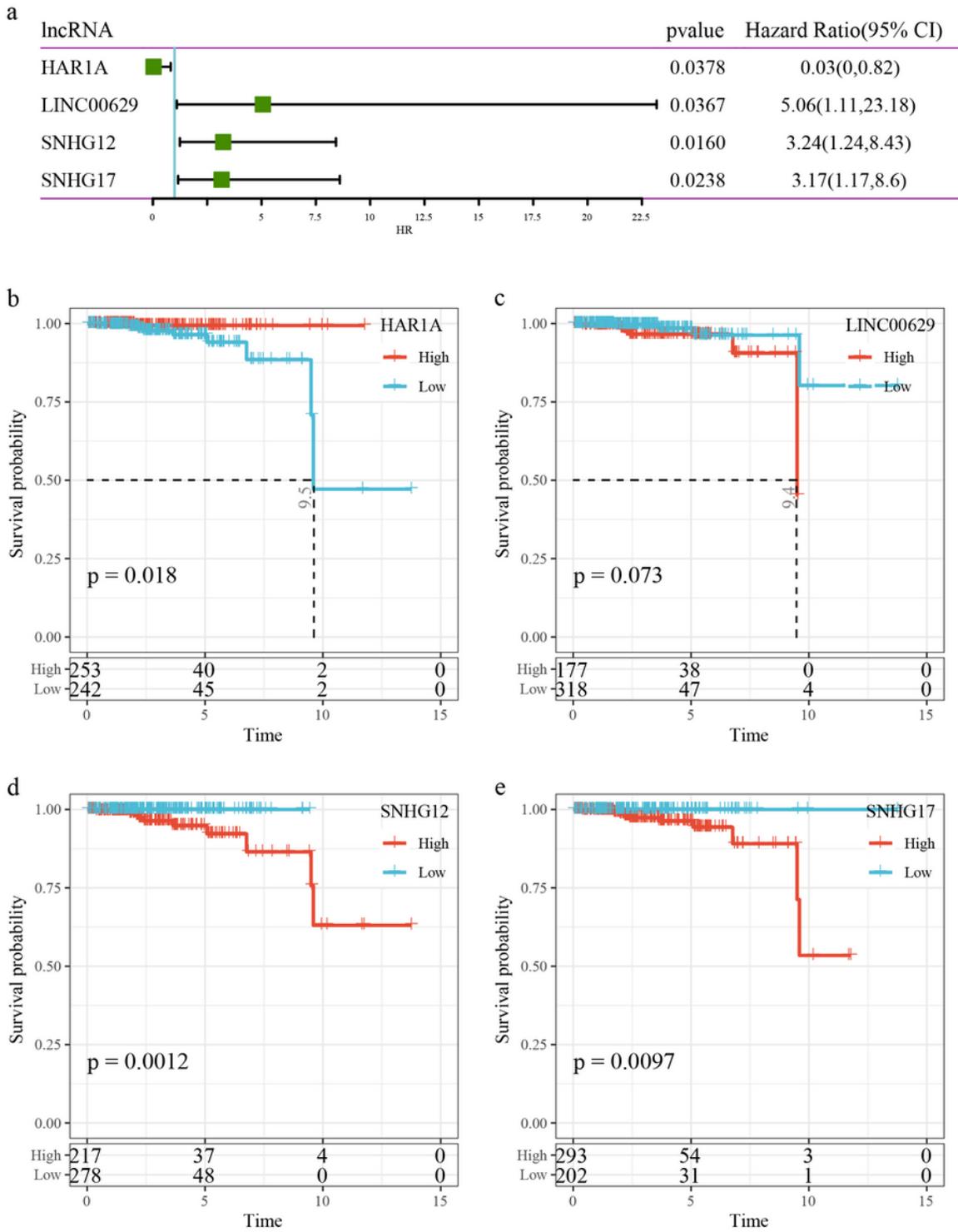


Figure 6

The correlation between epi-lncRNAs and prognosis of prostate cancer (A). Forest plot of single factor analysis of 4 prostate cancer-related epi-lncRNA genes; (B-E). KM curve of high and low expression samples of 4 prostate cancer-related epi-lncRNA genes.

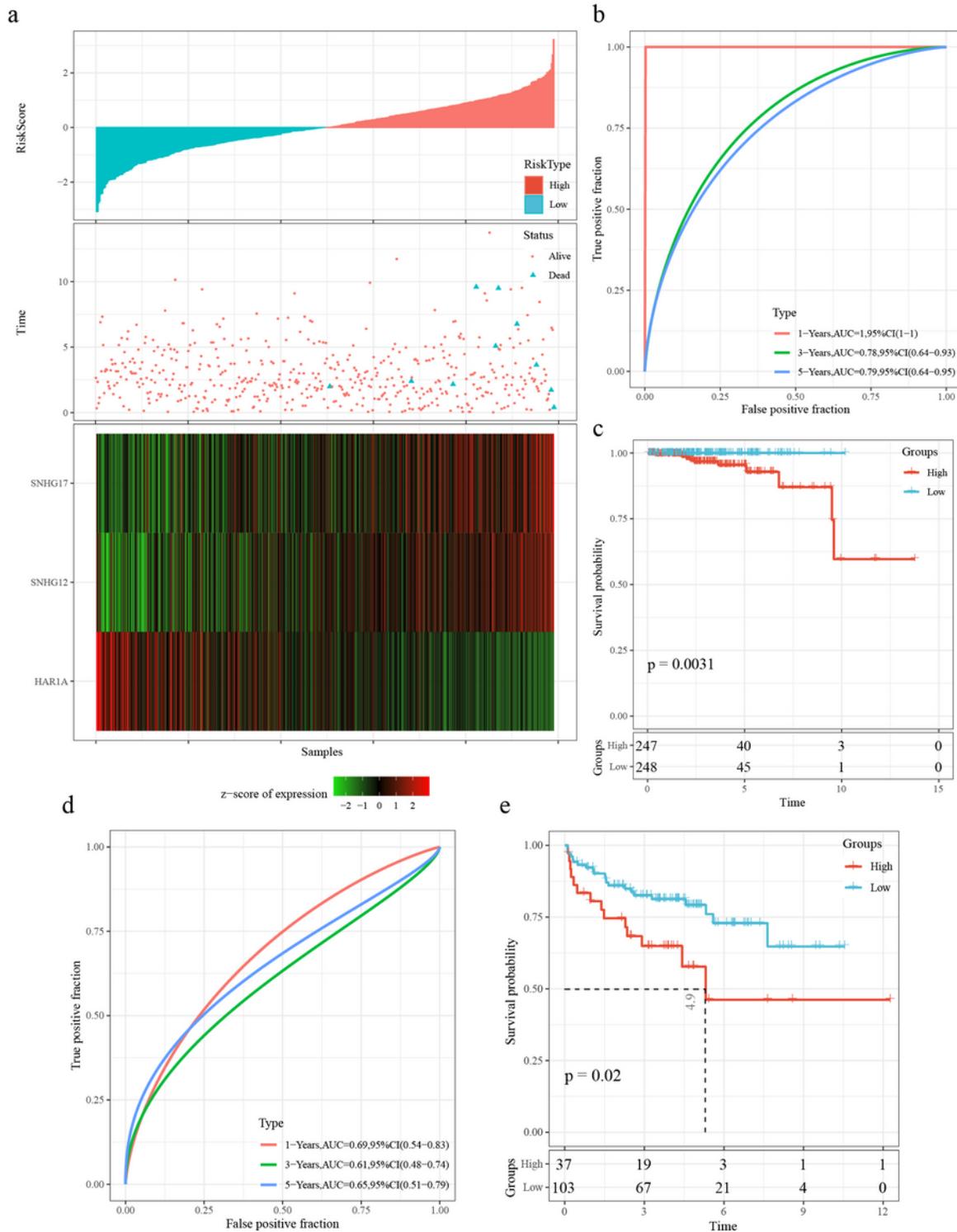


Figure 7

Three epi-lncRNA signatures could be used as prognostic markers (A). The relationship between 3-EpiLncRNA score and patient death and SNHG17, SNHG12, HAR1A gene expression in TCGA data set; (B). ROC analysis of 3-EpiLncRNA score in TCGA data set; (C). In TCGA data set, the prognosis of patients with high and low 3-EpiLncRNA score is different; (D). ROC analysis of 3-EpiLncRNA score in MSKCC data set; (E). Prognostic difference of patients with high and low 3-EpiLncRNA score in MSKCC data set.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Fig.S1.pdf](#)
- [Tables.S1S5.xlsx](#)