

CancerMHL: the database of integrating key DNA methylation, histone modifications and lncRNAs in cancer

Pengyu Du¹, Yingli Chen^{1,2,*}, Qianzhong Li^{1,2,*}, Zhimin Gai¹, Hui Bai¹, Luqiang Zhang¹, Yuxian Liu¹, Yanni Cao¹, Yuanyuan Zhai¹ and Wen Jin¹

¹Laboratory of Theoretical Biophysics, School of Physical Science and Technology, Inner Mongolia University, 235 West Daxue Road, Hohhot 010021, China

²The State Key Laboratory of Reproductive Regulation and Breeding of Grassland Livestock, Inner Mongolia University, 235 West Daxue Road, Hohhot 010021, China

*Corresponding author: Yingli Chen. Tel: +86-0471-4992914; E-mail: stchenyl@imu.edu.cn

Correspondence may also be addressed to Qianzhong Li. Tel: +86-0471-4993145; E-mail: qzli@imu.edu.cn

Citation details: Du, P., Chen, Y., Li, Q. *et al.* CancerMHL: the database of integrating key DNA methylation, histone modifications and lncRNAs in cancer. *Database* (2024) Vol. 2024: article ID baae029; DOI: <https://doi.org/10.1093/database/baae029>

Abstract

The discovery of key epigenetic modifications in cancer is of great significance for the study of disease biomarkers. Through the mining of epigenetic modification data relevant to cancer, some researches on epigenetic modifications are accumulating. In order to make it easier to integrate the effects of key epigenetic modifications on the related cancers, we established CancerMHL (<http://www.positionprediction.cn/>), which provide key DNA methylation, histone modifications and lncRNAs as well as the effect of these key epigenetic modifications on gene expression in several cancers. To facilitate data retrieval, CancerMHL offers flexible query options and filters, allowing users to access specific key epigenetic modifications according to their own needs. In addition, based on the epigenetic modification data, three online prediction tools had been offered in CancerMHL for users. CancerMHL will be a useful resource platform for further exploring novel and potential biomarkers and therapeutic targets in cancer.

Database URL: <http://www.positionprediction.cn/>

Introduction

A large number of studies had found that epigenetic modifications play an important role in tumorigenesis (1–3). The abnormal changes of epigenetic modifications may lead to the alterations of genome structure and gene expression and tumorigenesis (4–6). Because epigenetic changes are reversible, the identification of key epigenetic modifications in cancer may be useful for the study of targeted epigenetic therapies (7).

In recent years, with the rapid increase of DNA methylation, histone modifications and lncRNA associated with cancer, some specific resources on DNA methylation, histone modifications, and lncRNAs in diseases have been made. For instance, MethHC 2.0 (8), DiseaseMeth 3.0 (9), MethyCancer (10), MethDB (11), Pancan-meQTL (12) and DDMGD (13) included DNA methylation data obtained from experiments. PubMeth (14) and MeInfoText (15) extracted information about cancer methylation by mining literature. MethMarkerDB focused on some especial genes and methylated regions in cancer (16). HHMD (17) and iHMS (18) were databases on histone modification data obtained from experiments. Lnc2Cancer 3.0 (19), LncRNADisease 3.0 (20), LncTarD 2.0 (21), Lnc2Meth (22), LncCell (23), LncSEA 2.0 (24) and

lncRNASNP2 (25) integrated the disease-related lncRNA data supported by experiment and computation.

In our previous works, the changes of DNA methylation and histone modification patterns in functional regions were explored by analyzing the DNA methylation and histone modification data in liver cancer, breast cancer, colon cancer, lung cancer and chronic myelogenous leukemia. Based on these changes, the key methylation sites and histone modification regions were discovered in these cancers and the effect of the key epigenetic modification changes on gene expression were deeply discussed.

Based on these databases and our previous studies, it will be useful to construct the platform of the integrated epigenetic modification data. Therefore, we established CancerMHL in which the information related key DNA methylation, histone modifications and lncRNAs in several cancers can be easily obtained. It will provide a useful platform for exploring novel epi-biomarkers and therapeutic targets in the study of these cancers. The current version of CancerMHL collects key methylation sites, histone modifications and their regions as well as lncRNAs related to cancer. Each item includes the detailed annotation information of epigenetic modification, such as regions, sites and their effects on gene expression.

Furthermore, based on our previous studies, CancerMHL collected the distribution patterns of differential DNA methylation and differential histone modifications in the various regions of genomes for different cancers. It also provides the several online tools of effective prediction algorithms and risk assessment models.

Data collection and database content

Data collection and processing

We extracted key DNA methylation and histone modification data in cancer based on the articles published by our research team (26–34). The main data include: (i) key differential histone modifications and their important regions in cancer, as well as their regulated target genes; (ii) key differential methylation sites in cancer and their regulated target genes; (iii) the effect of key epigenetic modifications on gene expression; (iv) the distribution changes of DNA methylation of up- and down-regulated genes in different functional regions; (v) the important differential expression genes; (vi) the important histone modification regions associated with the different types of gene expressions; (vii) the correlation of different histone modifications in up- and down-regulated genes. For their regulated target genes, we annotated gene types based on TSGene 2.0 (35), ONGene (36), COSMIC (37) and TAG (38) databases. Furthermore, we used the PubMed literature database to verify if these genes have been experimentally confirmed to influence cancer occurrence and development.

Many studies have found that DNA methylation, histone modifications and lncRNAs play a cooperative role in the regulation of gene expression (39–41). To facilitate such research, we searched the PubMed literature database using the keywords ‘cancer,’ ‘gene,’ and ‘lncRNA’. According to the following criteria: (i) lncRNAs related to human diseases and (ii) consistency in target genes regulated by lncRNAs with

those regulated by methylation and histone modifications, we retrieved information about lncRNAs and their regulatory mechanisms on target genes. Based on the cooperative regulatory mechanisms of lncRNAs with other regulatory factors on target genes, we categorized lncRNA regulation into the following types: (i) co-regulation with DNA methylation; (ii) co-regulation with histone modifications; (iii) co-regulation with transcription factors; (iv) co-regulation with DNA methylation and histone modifications; and (v) lncRNAs independently regulate target genes.

Database statistics

The current version of CancerMHL contained key epigenetic modifications from 10 types of cancers, including key DNA methylation sites in different functional regions, the key regions of different histone modifications, and important lncRNAs participated in different co-regulations. The statistical results are shown in Tables 1, 2 and 3, respectively. Furthermore, CancerMHL provided the distribution patterns of DNA methylation and histone modification in different regions by mining epigenetic modification data related to cancer. It also provided two online risk assessment tools and an early diagnosis tool for the convenience of medical researchers.

Database usage and access

Web interface

CancerMHL offers a user-friendly web interface that enables user to browse and retrieve the information of epigenetic modifications in cancer. The CancerMHL website comprises three main modules, including (i) ‘Search’ for accessing the information about key epigenetic modifications in cancer, (ii) ‘Conclusion’ for obtaining the distribution patterns of DNA methylation and histone modifications in cancer and (iii)

Table 1. The number of key methylation sites in different regions of the genome

Disease		Regions						
Breast cancer	Promoter 21	Enhancer 32	Exon 6	Intron 25	CGI 72	N_Shelf 5	N_Shore 5	S_Shore 1
Hepatocellular carcinoma	Promoter 12	5’UTR 2	exon 7		CGI 2	N_Shelf 1	N_Shore 8	S_Shore 2
Colon adenocarcinoma	TSS200 2	TSS1500 4	body 19		CGI 5	N_Shore 4	S_Shore 4	5’UTR 1
Pan-cancer		TSS200 1		TSS1500 2		Body 2		CGI 1

Abbreviation: TSS200, 200 bp in the upstream of the transcription start site (TSS). TSS1500, 200–1500 bp in the upstream of the TSS.

Table 2. The number of key regions in different histone modifications

Disease		Histone modifications					
Hepatocellular carcinoma	H3K4me1 2	H3K4me2 8	H3K4me3 22	H3K27me3 5	H3K79me2 5	H3K9ac 18	H3K27ac 28
Breast cancer	H3K36me3 3		H3K79me2 3		H3K9ac 6		H3K27ac 6
Disease	Histone modifications	Disease	Histone modifications	Disease	Histone modifications	Disease	Histone modifications
Colorectal cancer	H3K79me3 20	Lung adenocarcinoma	H3K79me2 12	Chronic myelogenous Leukemia			H3K36me3

Table 3. The number of important lncRNAs participated in different co-regulations

Disease		Regulation types (numbers)			
Hepatocellular carcinoma	lncRNA (28)	lncRNA-Meth (9)	lncRNA-HM (7)	lncRNA-TF (2)	lncRNA-Meth-HM (1)
Colorectal cancer	lncRNA (17)	lncRNA-Meth (4)	lncRNA-HM (2)	lncRNA-Meth-HM (2)	
Non-small cell lung carcinoma	lncRNA (20)	lncRNA-Meth (1)	lncRNA-HM (3)	lncRNA-TF (5)	

Disease		Regulation types (numbers)		Disease		Regulation types (numbers)	
Breast cancer	lncRNA (9)	lncRNA-Meth (1)		Colon adenocarcinoma	lncRNA (2)		
Lung adenocarcinoma	lncRNA (7)	lncRNA-TF (1)		Lung cancer	lncRNA (2)		
Chronic myeloid leukemia	lncRNA (6)	lncRNA-Meth (2)		Lung squamous cell carcinoma	lncRNA (1)		

Abbreviation: Meth, DNA methylation. HM, histone modification.

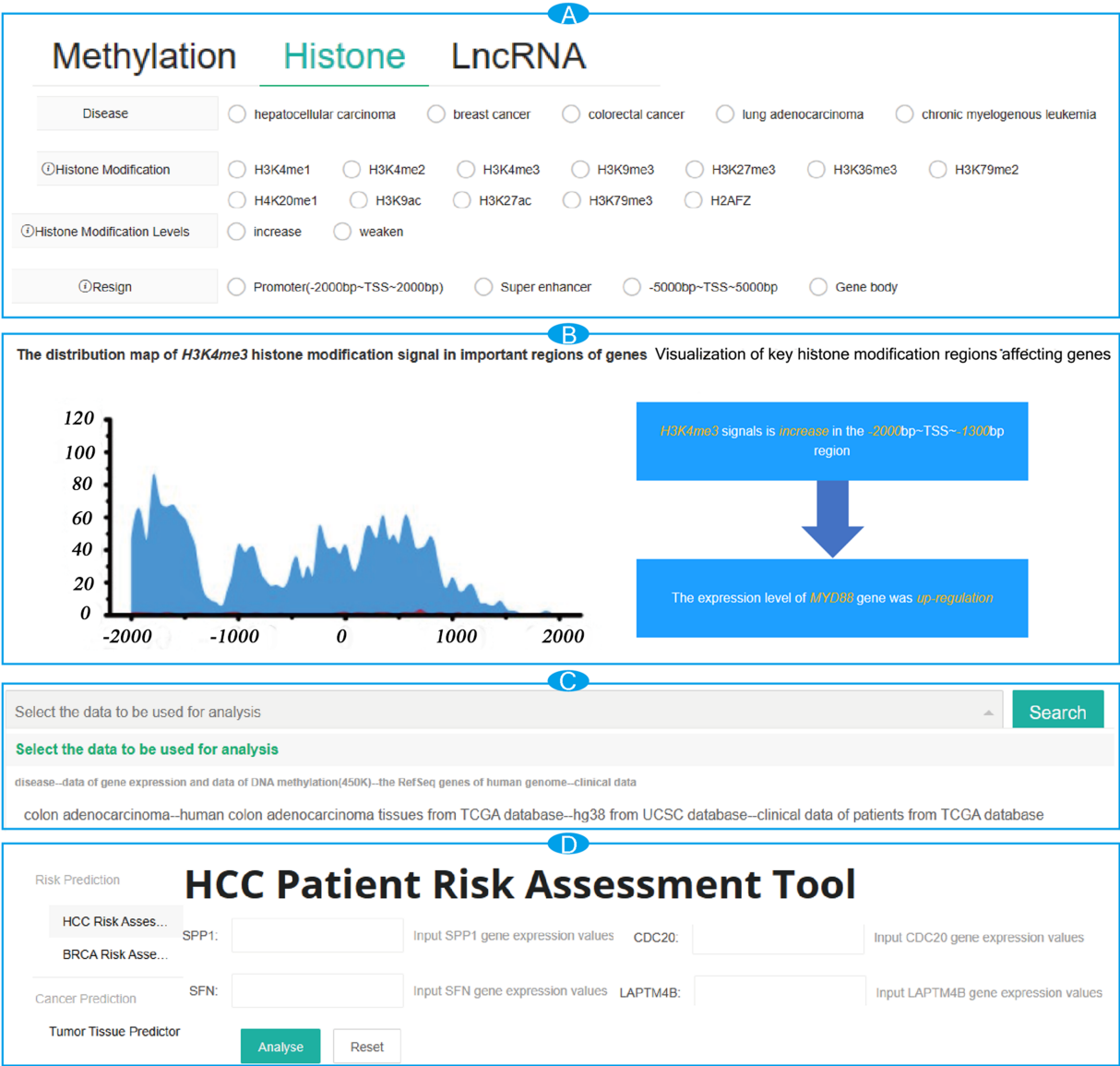


Figure 1. The interface of the browse module for CancerMHL. (A) The interface of the search modules; (B) search result page for MYD88 with detailed information; (C) module interface for analyzing results based on DNA methylation and histone modification data; (D) the interface of the tool modules.

Alt text: The figure was comprised of four main modules in the interfaces of CancerMHL. It includes the information about key epigenetic modifications, the distribution patterns of DNA methylation and histone modifications, the online tools for risk assessment and the early diagnosis of cancer patients.

'Tools' for using online tools for risk assessment and the early diagnosis of cancer patients. In addition, on the 'Help' page, CancerMHL provides a detailed tutorial for the usage of the database.

Using the search tool to retrieve the key epigenetic modifications in cancer

To access interesting key epigenetic modifications, users can query the database as follows. First, on the 'Search' page, users can filter data by using the type, disease and the location of the desired key epigenetic modifications (Figure 1A). Once selections are made, click the 'Search' button and the query results will be shown as a responsive table, with each item representing information about a key epigenetic modification. There is a 'Details' button at the end of each item, which will provide more comprehensive annotation information, including the visual representations of the effects of epigenetic modifications on gene expression, the location details of epigenetic modifications, information about genes and the distribution of signals in important histone modification regions (Figure 1B).

Retrieving the distribution patterns of DNA methylation and histone modifications in cancer

On the 'Conclusion' page, users can explore the distribution patterns of DNA methylation and histone modifications, as well as some relevant important results, by selecting the data used for computational analysis from the dropdown list (Figure 1C).

Using of risk assessment tools and tumor tissue prediction tools

CancerMHL provides three online tools: a risk assessment tool for liver cancer patients, a risk assessment tool for breast cancer patients and a tumor tissue prediction tool. On the 'Tools' page, users select their interested tool from the options on the left side of the page. Following the instructions provided on the page, the users enter patient data into the input boxes (Figure 1D). Subsequently, by clicking the 'Analyze' button, users can obtain the patient's risk grouping or predicted results.

Conclusions and future extensions

In this study, we established the integrated database of key epigenetic modifications in the several kinds of cancers. This platform provides a reference for further studying the effects of co-regulation with epigenetic modifications such as DNA methylation, histone modifications and lncRNAs on gene expression levels, including the effects of co-regulation with the different epigenetic modifications in same regions or same epigenetic modifications in different regions on gene expression, etc. Moreover, some of the important genes and epigenetic modification regions and sites can be used as potential epigenetic biomarkers. The establishment of this database will have significant guiding value for experimental researchers in exploring novel epigenetic biomarkers and therapeutic targets.

To make CancerMHL more comprehensive and useful, we will further improve and perfect the platform and database,

and increase more annotation information and practical analysis tools. We plan to supplement relationship between chromatin accessibility annotation and key epigenetic modifications and the effects of epigenetic modifications on 3D genome structure in future versions.

Data availability

All the data can be downloaded from <http://www.positionprediction.cn/>.

Contribution statement

Y.C. and Q.L. provided comprehensive guidance; P.D. designed the database; H.B., L.Z., Y.L., Y.C., Y.Z. and W.J. provided the data; Z.G. collected and analyzed the data; P.D., Y.C. and Q.L. wrote the paper; all authors read and agreed the final manuscript.

Funding

National Natural Science Foundation of China [Nos 32160216, 62361047, 62161033].

Conflict of interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We thank Zhiying Li and Yuanyuan Zhao for their contribution to some data collection.

References

1. Momparler, R.L. (2003) Cancer epigenetics. *Oncogene*, **22**, 6479–6483.
2. Esteller, M. (2007) Cancer epigenomics: DNA methylomes and histone-modification maps. *Nat. Rev. Genet.*, **8**, 286–298.
3. Dawson, M.A. and Kouzarides, T. (2012) Cancer epigenetics: from mechanism to therapy. *Cell*, **150**, 12–27.
4. Egger, G., Liang, G., Aparicio, A. et al. (2004) Epigenetics in human disease and prospects for epigenetic therapy. *Nature*, **429**, 457–463.
5. Jones, P.A. and Baylin, S.B. (2002) The fundamental role of epigenetic events in cancer. *Nat. Rev. Genet.*, **3**, 415–428.
6. Sharma, S., Kelly, T.K. and Jones, P.A. (2009) Epigenetics in cancer. *Carcinogenesis*, **31**, 27–36.
7. Werner, R.J., Kelly, A.D. and Issa, J.J. (2017) Epigenetics and precision oncology. *Cancer J.*, **23**, 262–269.
8. Huang, H.Y., Li, J., Tang, Y. et al. (2021) MethHC 2.0: information repository of DNA methylation and gene expression in human cancer. *Nucleic Acids Res.*, **49**, D1268–d1275.
9. Xing, J., Zhai, R., Wang, C. et al. (2022) DiseaseMeth version 3.0: a major expansion and update of the human disease methylation database. *Nucleic Acids Res.*, **50**, D1208–d1215.
10. He, X., Chang, S., Zhang, J. et al. (2008) MethyCancer: the database of human DNA methylation and cancer. *Nucleic Acids Res.*, **36**, D836–841.
11. Grunau, C., Renault, E., Rosenthal, A. et al. (2001) MethDB—a public database for DNA methylation data. *Nucleic Acids Res.*, **29**, 270–274.

12. Gong,J., Wan,H., Mei,S. *et al.* (2018) Pancan-meQTL: a database to systematically evaluate the effects of genetic variants on methylation in human cancer. *Nucleic Acids Res.*, **47**, D1066–D1072.
13. Raies,A.B., Mansour,H., Incitti,R. *et al.* (2014) DDMGD: the database of text-mined associations between genes methylated in diseases from different species. *Nucleic Acids Res.*, **43**, D879–D886.
14. Ongenaert,M., Van Neste,L., De Meyer,T. *et al.* (2007) PubMeth: a cancer methylation database combining text-mining and expert annotation. *Nucleic Acids Res.*, **36**, D842–D846.
15. Xuan Lin,Q.X., Sian,S., An,O. *et al.* (2018) MethMotif: an integrative cell specific database of transcription factor binding motifs coupled with DNA methylation profiles. *Nucleic Acids Res.*, **47**, D145–D154.
16. Zhu,Z., Zhou,Q., Sun,Y. *et al.* (2023) MethMarkerDB: a comprehensive cancer DNA methylation biomarker database. *Nucleic Acids Res.*, **52**, D1380–D1392.
17. Zhang,Y., Lv,J., Liu,H. *et al.* (2010) HHMD: the human histone modification database. *Nucleic Acids Res.*, **38**, D149–154.
18. Gan,Y., Tao,H., Guan,J. *et al.* (2017) iHMS: a database integrating human histone modification data across developmental stages and tissues. *BMC Bioinf.*, **18**, 103.
19. Gao,Y., Shang,S., Guo,S. *et al.* (2021) Lnc2Cancer 3.0: an updated resource for experimentally supported lncRNA/circRNA cancer associations and web tools based on RNA-seq and scRNA-seq data. *Nucleic Acids Res.*, **49**, D1251–d1258.
20. Lin,X., Lu,Y., Zhang,C. *et al.* (2023) LncRNADisease v3.0: an updated database of long non-coding RNA-associated diseases. *Nucleic Acids Res.*, **52**, D1365–D1369.
21. Zhao,H., Yin,X., Xu,H. *et al.* (2022) LncTarD 2.0: an updated comprehensive database for experimentally-supported functional lncRNA–target regulations in human diseases. *Nucleic Acids Res.*, **51**, D199–D207.
22. Zhi,H., Li,X., Wang,P. *et al.* (2017) Lnc2Meth: a manually curated database of regulatory relationships between long non-coding RNAs and DNA methylation associated with human disease. *Nucleic Acids Res.*, **46**, D133–D138.
23. Wang,P., Guo,Q., Hao,Y. *et al.* (2020) LnCeCell: a comprehensive database of predicted lncRNA-associated ceRNA networks at single-cell resolution. *Nucleic Acids Res.*, **49**, D125–D133.
24. Zhang,G., Song,C., Fan,S. *et al.* (2023) LncSEA 2.0: an updated platform for long non-coding RNA related sets and enrichment analysis. *Nucleic Acids Res.*, **52**, D919–D928.
25. Miao,Y.R., Liu,W., Zhang,Q. *et al.* (2017) lncRNASNP2: an updated database of functional SNPs and mutations in human and mouse lncRNAs. *Nucleic Acids Res.*, **46**, D276–D280.
26. Bai,H., Li,Q.Z., Qi,Y.C. *et al.* (2022) The prediction of tumor and normal tissues based on the DNA methylation values of ten key sites. *Biochim. Biophys. Acta Gene. Regul. Mech.*, **1865**, 194841.
27. Cao,Y.N., Li,Q.Z. and Liu,Y.X. (2022) Discovered key CpG sites by analyzing DNA methylation and gene expression in breast cancer samples. *Front. Cell Dev. Biol.*, **10**, 815843.
28. Jin,W., Li,Q.Z., Zuo,Y.C. *et al.* (2019) Relationship between DNA methylation in key region and the differential expressions of genes in human breast tumor tissue. *DNA Cell Biol.*, **38**, 49–62.
29. Liu,Y.X., Li,Q.Z., Cao,Y.N. *et al.* (2020) Identification of key genes and important histone modifications in hepatocellular carcinoma. *Comput. Struct. Biotechnol. J.*, **18**, 2657–2669.
30. Liu,Y.X., Li,Q.Z. and Cao,Y.N. (2022) The effect of key DNA methylation in different regions on gene expression in hepatocellular carcinoma. *Mol. Omics*, **18**, 57–70.
31. Zhai,Y.Y., Li,Q.Z., Chen,Y.L. *et al.* (2022) Identification of key histone modifications and hub genes for colorectal cancer metastasis. *Curr. Bioinform.*, **17**, 206–216.
32. Zhang,L.Q., Fan,G.L., Liu,J.J. *et al.* (2020) Identification of key histone modifications and their regulatory regions on gene expression level changes in chronic myelogenous leukemia. *Front. Cell Dev. Biol.*, **8**, 621578.
33. Zhang,L.Q., Yang,H., Liu,J.J. *et al.* (2022) Recognition of driver genes with potential prognostic implications in lung adenocarcinoma based on H3K79me2. *Comput. Struct. Biotechnol. J.*, **20**, 5535–5546.
34. Bai,H., Yan,D.S., Chen,Y.L. *et al.* (2024) Potential biomarkers: The hypomethylation of cg18949415 and cg22193385 sites in colon adenocarcinoma. *Comput. Biol. Med.*, **169**, 107884.
35. Zhao,M., Kim,P., Mitra,R. *et al.* (2016) TSGene 2.0: an updated literature-based knowledgebase for tumor suppressor genes. *Nucleic Acids Res.*, **44**, D1023–1031.
36. Liu,Y., Sun,J. and Zhao,M. (2017) ONGene: A literature-based database for human oncogenes. *J. Genet. Genomics*, **44**, 119–121.
37. Tate,J.G., Bamford,S., Jubb,H.C. *et al.* (2019) COSMIC: the catalogue of somatic mutations in cancer. *Nucleic Acids Res.*, **47**, D941–d947.
38. Chen,J.S., Hung,W.S., Chan,H.H. *et al.* (2013) In silico identification of oncogenic potential of fyn-related kinase in hepatocellular carcinoma. *Bioinformatics*, **29**, 420–427.
39. Xu,M., Xu,X., Pan,B. *et al.* (2019) LncRNA SATB2-AS1 inhibits tumor metastasis and affects the tumor immune cell microenvironment in colorectal cancer by regulating SATB2. *Mol. Cancer.*, **18**, 135.
40. Tang,J., Xie,Y., Xu,X. *et al.* (2017) Bidirectional transcription of Linc00441 and RB1 via H3K27 modification-dependent way promotes hepatocellular carcinoma. *Cell Death Dis.*, **8**, e2675–e2675.
41. Li,Y., Gan,Y., Liu,J. *et al.* (2022) Downregulation of MEIS1 mediated by ELFN1-AS1/EZH2/DNMT3a axis promotes tumorigenesis and oxaliplatin resistance in colorectal cancer. *Signal Transduct. Target. Ther.*, **7**, 87.