Pan-cancer analysis of m⁵C regulator genes reveals consistent epigenetic landscape changes in multiple cancers

Yuting He

Zhengzhou University First Affiliated Hospital

Xiao Yu

Zhengzhou University First Affiliated Hospital

Menggang Zhang

Zhengzhou University First Affiliated Hospital

Wenzhi Guo (✉ guowz66@163.com)

Zhengzhou University First Affiliated Hospital <https://orcid.org/0000-0002-0821-1318>

Research

Keywords: frequent network mining, m⁵C regulatory genes, pan-cancer analysis, survival, 5-methylcytosine

Posted Date: April 9th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-389804/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Background

5-methylcytosine (m^5C) is a reversible modification to both DNA and various cellular RNAs. However, its roles in developing human cancers are poorly understood, including the effects of mutant m^5C regulators and the outcomes of modified nucleobases in RNAs.

Methods

Based on The Cancer Genome Atlas (TCGA) database, we uncovered that mutations and copy number variations (CNVs) of m^5C regulatory genes were significantly correlated across 33 cancer types. We then assessed the correlation between the expression of individual m^5C regulators and the activity of related hallmark pathways of cancers. After validating m^5C regulators' expression based on their contributions to cancer development and progression, we observed their up-regulation within tumor-specific processes.

Results

Notably, our research connected aberrant alterations to m^5C regulatory genes with poor clinical outcomes among various tumors that may drive cancer pathogenesis and/or survival.

Conclusions

Our results offered strong evidence and clinical implications for the involvement of m^5C regulators.

Introduction

Cancers have become the second life-threatening malignancies, which contribute to almost 18.1 million people occurred, and 9.6 million death globally in 2018.[1] Lack of efficient diagnosis indicators at an early stage, and high rate of postoperative recurrence contribute to poor clinical prognosis and high mortality.[2,3] Growing evidence demonstrated that genomic instability,[4,5] oncogenes activation, aberrant methylation modifications, alterations in epigenetic changes,[6,7] aberrant expression of microRNAs,[8] and alterations of signaling pathways are crucial factors and contribute to cancer pathogenesis.[9,10] Methylation is an essential epigenetic modification and is closely related to the pathogenesis of cancers.[11,12] The 5-methylcytosine (m^5C), N6-methyladenine (m^6A), N1-methyladenosine (m^1A) have become the most common types of epigenetic modifications in eukaryotes.[11] Emerging evidence has demonstrated that m^5C modification has the potential to serve as novel epigenetic markers with remarkable biological significance in biological processes.[13,14,15]

m^5C modification distributes in different types of RNAs and DNAs.[16,17,18] m^5C modifications can even modify the destiny of cancer cells.[19] m^5C regulators contain writers, erasers, and readers, which function as common epigenetic modification and contribute to pre-mRNA splicing, gene expression, gene silencing, nuclear export, genomic maintenance, and translation initiation modifications.[20,21] m^5C could therefore be used as a biomarker for disease progression, including various types of cancers.[22] m^5C maintains open and closed chromatin states to control gene expression, genome editing, organismal development, and cellular differentiation.[18] In this context, writers act within a methyltransferase complex to methylate targets, and erasers remove m^5C methylation, while readers recognize and bind to m^5C -methylated RNA and implement corresponding functions.[20,23,24] The anomalous interplay between writers, erasers, arising from alterations to their expression, has been linked to cancer pathogenesis and progression.[18,25] However, pan-cancer effects of changes to m^5C regulatory gene expression have not been fully defined. [Next-generation sequencing](#) (NGS) provides us effective tools to comprehensively view the m^5C distribution landscape throughout the global [transcriptome](#). [22]

In this study, we identified the potential prognostic value of m^5C regulators and provided a comprehensive understanding of m^5C modifications in pan-cancers, which will help to find novel opportunities for cancer early detection, treatment, and prevention.

Materials And Methods

Study workflow

We downloaded fragments of kb transcripts based on Fragments Per Kilobase of transcript per Million (FPKM) gene expression from the TCGA dataset among 33 different cancer types. m^5C regulator patterns were investigated in 5,480 samples among 33 different cancer types, including somatic mutations, copy number variation (CNV), gene expression, and RNA-seq data (**Fig. 1a**).

Genomic data collection of m^5C regulators

Thirteen m⁵C regulators were identified from published papers. Information of m⁵C regulators was collected from Gene Cards (www.genecards.org). Ensemble gene IDs and HUGO Gene Nomenclature Committee symbols were assigned to each m⁵C regulator-associated gene.

Whole genomics data analysis of 33 pan-cancers

The integrated OMICS datasets based on the Cancer Genome Atlas (TCGA, <http://cancergenome.nih.gov/>) of 33 pan-cancers were applied in this study. We collected the mutation annotation format (MAF) profile from TCGA, which contains over 10000 cancer patient's information. The level-3 data of copy number alterations profiles in TCGA was acquired for secondary analysis. In addition, we downloaded pan-cancers RNAseq data of genomic variations profiles and corresponding clinical information from the GDC Data Portal using the R package "TCGAbiolinks".

Differently expression genes (DEGs) identification

The DESeq package in R language was utilized to validate the DEGs between 33 pan-cancer samples and adjacent samples. Genes with a mean value >0 were included in the screening of DEGs. To establish proper DEGs among 33 cancers, we settled the adjusted *p*-value less than 0.01 and |log₂ fold change (log₂FC)| no less than two as the statistical threshold value for differentially expressed genes. The results were screened as significant DEGs and methylated sites.

Functional annotations and pathways enrichment analysis

To evaluate the biological functions of each m⁵C modification-related gene, we transformed the RNA-seq data of all samples into transcripts per million (TPM) values. The methods have been described in a previous study.[26] The insufficient, duplicated, and zero expression genes will be eliminated. Furthermore, the gene set variation analysis (GSVA) was applied to determine transcriptomic activities and explore the biological processes of m⁵C regulators.

To further explore m⁵C regulators related inhibition and activation factors, we performed the Pearson Correlation coefficient (PCC) and defined the absolute value of the PCC greater than 0.5 and *p* value of less than 0.01 as the screen cut off. The results could be recognized as significantly correlated m⁵C regulators.

The internships between m⁵C regulators

To visualize the intercorrelations among m⁵C regulators, we adopted the "CORPrAP" R package (<https://github.com/taiyun/corplot>). Besides, the STRING database was also applied for the exploration and analysis of these associations between m⁵C regulators and 325 related genes. [27] The 325 genes were obtained from the KEGG database (<http://www.kegg.jp/> or <http://www.genome.jp/kegg/>). The correlations between m⁵C regulators and 325 genes were visualized through Cytoscape (<https://cytoscape.org/>).

Clinical characteristic of m⁵C regulators

To explore the m⁵C regulators' related clinical characteristics, we classified genes into high and low expression groups based on genes' median expression. Correlations between outcomes of the two groups were then analyzed through a log-rank test via R software (<https://cran.r-project.org/web/packages/survival/index.html>). The log-rank test was performed to weigh the overall survival rates that differ between the high and low expression groups. The CRAN Package survival (<https://cran.r-project.org/web/packages/survival/index.htm>) was performed, and we defined the *p* value of less than 0.05 as significant difference.

The roles of m⁵C regulators in cell growth

The CRISPR-CAS9 gene scale screening of cell lines from 33 cancer types was collected from previous study.[27] We calculated the proportion of every regulator as an essential gene in the cell lines.

Results

Resultsm⁵C regulators identification and its genomic extensive genetic changes

In this study, we identified 13 m⁵C regulators, as shown in **Fig. 1b**, eleven writers (NSUN1-7, DNMT1, DNMT2, DNMT3A, and DNMT3B), one eraser (TET2), and one reader (ALYREF). This study validated the frequency of m⁵C regulators patterns among 33 cancers by integrating somatic mutations and copy number variations (CNVs) data. Table 1 illustrated detailed information. The results indicated that the overall average mutation frequency of regulatory factors is low, ranging from 0 to 9% (**Fig. 2a**). The uterine corpus endometrial carcinoma (UCEC) is characterized as a high tumor mutation burden.[28] The UCEC showed significantly higher mutation frequency. Horizontal analysis indicated

that TET2, DNMT3B, DNMT3A, and DNMT1 demonstrated much higher mutation frequency among 33 cancers. Furthermore, we uncovered the CNV mutation frequency of m⁵C regulators was common. Regulators such as DNMT3B, ALYREF, and NSUN5 displayed extensive CNVs. On the contrary, TET2 and NSUN4 showed significant lack of m⁵C modification related CNVs mutations among pan-cancers (**Fig 2b and Table 2**).

To further investigate whether the genomic mutations affect m⁵C regulators expression, we intensively detected the m⁵C regulators' gene expression disturbances in thirty three pan-cancers and five standard control samples. The result implied that the CNV alterations (amplification and deletion) might profoundly affect the m⁵C regulator's expression (**Fig. 3a**). The m⁵C regulators with CNV amplification showed significantly increased expression in pan-cancers (such as DNMT3B), and m⁵C regulators with CNV deletion exhibited remarkably decreased expression, like TET2. In addition, we comparatively analyzed the m⁵C regulators' expression levels in cancers and corresponding normal tissues and found out that DNMT3B was significantly overexpressed in thirty-three tumor or cancer tissues compared with adjacent normal tissues (**Fig. 3b**). The results demonstrated that the m⁵C regulator gene expression has a close relationship with CNVs mutations. These results uncovered that the m⁵C regulators among various cancers showed significant heterogeneity in gene expression and genetics. Collectively, our results demonstrated that aberrant m⁵C regulations were crucial for carcinogenesis and progression, which provided a clue for further functional detection.

m⁵C regulators related pan-carcinogenic pathways

To comprehensively explore the molecular mechanism m⁵C regulators involved in cancers, we evaluated the correlations between m⁵C regulators' proteins expression and Kyoto Encyclopedia of Genes and Genomes pathway (KEGG) enrichment analysis-related activities. The results indicated that m⁵C regulator proteins have a close relationship with tumor-related pathways' activation and inactivation (**Fig. 4a and Table 3**). ALYREF, DNMT1, and TET2 were involved in the cell cycle, DNA replication, and prostate cancer-related pathways. Notably, ALYREF was involved in multiple pathways, including cell cycle, DNA replication, and prostate cancer-related pathways. DNMT1 was involved in drug metabolism, lipid metabolism, and nucleic acid biosynthesis signaling pathways. ALYREF, DNMT1, NSUN5, NSUN1, and TET2 showed active involvement in KEGG enrichment pathways (**Fig.4b**). In addition, genes will not function alone.[29] Growing evidence indicated that genes always co-effect with multiple genes and always have multiple functions.[29,30] We further explored the internal connections between m⁵C regulators gene expression. Results indicated that the readers, writers, and erasers also have high correlations with each other. The eraser TET2 was significant correlated with the writer NSUN3. Writers such as ALYREF and NSUN5 also showed obvious correlations ($R = 0.55$, $p < 0.01$, **Fig. 4c**). Additionally, to visualize the interactions between m⁵C regulators, we utilized the protein-protein interaction (PPI) analysis in m⁵C regulators related proteins. The results showed that writers, readers, and erase were particularly frequent (**Fig. 4d**). These results indicated that interactions among m⁵C regulators play crucial roles in the development and progression of cancers.

Clinical significance of m⁵C regulators in pan cancers

To evaluate the clinical prognosis of m⁵C regulators, we calculated the overall survival (OS), overall median progression-free interval (PFI), disease-specific survival (DSS), and disease-free interval (DFI) of m⁵C regulators. The OS analysis implied a significant correlation between m⁵C regulators and thirty-three pan-cancers. The heat map demonstrated that m⁵C regulators were significantly correlated with survival of patients, including OS, PFI, DSS, and DFI. In detail, the OS in ACC, KICH, KIRC, LGG, and LIHC showed significant correlations with m⁵C regulators (**Fig. 5a**). The highly expressed DNMT3A, DNMT3B, DNMT1, and ALYREF were significantly related to poor prognosis. Collectively, the DNMT3A, DNMT3B, DNMT1, and ALYREF might function as poor prognosis predictors in cancer progression. Moreover, we evaluated the prediction of PFI at fixed time points in patients with thirty-two solid tumors. PFIs of DNMT3B and DNMT1 showed significantly higher Hazard ratio (HR) values, indicating that they have the potential to be utilized as unfavorable prognosis prediction factors. The PFI in ACC, KICH, KIRC, LGG, and UVM showed significant correlation ships with most m⁵C regulators (**Fig. 5b**). Similarly, The DSS and DFI analysis indicated that ACC, LGG showed remarkably correlations with most m⁵C regulators (**Fig. 5c-d**). These results indicated that m⁵C regulators have crucial prognostic prediction values in a variety of cancer types.

Effect of m⁵C regulators in LIHC and CHOL

Studies have validated that m⁵C-related genes alterations have a close relationship with advanced tumor progression and advanced tumor stages.[31] Based on the above pieces of evidence, most m⁵C regulators are associated with patients' OS in LIHC and CHOL (**Fig. 5a**). Based on the overall expression patterns of m⁵C regulators, all patients in these cancer groups were categorized into two subgroups. The first subgroup consisted of 112 patients indicating high expression of m⁵C regulators (Reg-high), and the second subgroup consisted of 295 patients with low m⁵C regulators expression (Reg-low) (**Fig. 6a**). Compared with the Reg-high subgroup, the survival probability of patients in the Reg-low subgroup was significantly better ($p < 0.001$, **Fig.6b**). These results indicated that m⁵C regulators have a potential function as prognostic indicators in hepatocellular carcinoma and cholangiocarcinoma.

Discussion

Epigenetic variation is often related to human disease, especially cancers.[32,33,34] Remarkably progression has been made of various epigenetic-targeted therapies that have a broad application of malignancies and have exhibited detection and therapeutic potential for solid tumors in preclinical and clinical trials.[35,36,37,38,39] Aberrant methylation regulators process both in DNA and RNA play a critical role in epigenetic regulators, which are significantly associated with tumorigenesis.[40,41,42,43]

Original reports described that NSUN2 participates in catalyzing biological reactions of m⁵C formation in RNAs and regulating cell cycle,[44] linked to stem cell differentiation and involved in progression.[45] It is also reported that m⁵C regulators such as NSUN2 and binding partner ALYREF participant in promoting mRNA export coordinately,[46] and NSUN6, in complex with a full-length tRNA substrate targeting cytosine accessible to the enzyme for methylation.[12,47,48] The NSUN3 is required for the deposition of m⁵C at the anticodon loop in the mitochondria encoded transfer RNA methionine.[49] The NSUN5 demonstrates suppression characteristics in vivo glioma models.[50] The latest study found that the m⁵C levels of mRNA and circRNA in HCC are much higher than those in adjacent tissues.[51,52] NSUN5 gene mutation leads to an unmethylated condition at the C3782 position of 28S rRNA, which leads to a total depletion of protein synthesis and inducing an adaptive translational program under stress collectively,[53] as is illustrated that the m⁵C regulators may influence a wide variety of biological functions and metabolism.

In the present study, we applied certain methodological particularities to build a model and evaluated a catalog of genomic characteristics of tumors associated with m⁵C regulators. We obtained a total of 13 m⁵C regulators. The mutations and CNVs of m⁵C regulators are linked to several tumor developments. All cancers carry somatic mutations.[54] UCEC exhibited a significantly higher number of mutations across pan-cancers, analogously TET2, DNMT3B, DNMT3A, and DNMT1 have placed a moderate burden in m⁵C regulator genes. DNMT3B gene mutation was generally higher expression level among various cancers. Here we sequenced the m⁵C regulator genomes of pan-cancer and providing the first comprehensive remarkable insights into the forces that have shaped various cancer genomes. CNVs play an important role in tumor genesis and progression,[55] including amplification and deletion of oncogenes, which may significantly increase the risk of cancer.[56] In this research, the DNMT3B, ALYREF, NSUN5 showed extensive CNVs amplification. In contrast, CNVs such as TET2, NSUN4, etc. are generally deletion. These results indicated that CNV and the associated gene signatures are useful for early cancer detection and diagnosis, targeted therapeutics, and prediction of prognosis.

The genomic and transcriptomic parameters of various cancers are associated with m⁵C regulators gene expression and activity of the KEGG pathways.[57] We also investigate the gene expression perturbations of m⁵C regulators through 33 cancer types with parallel normal controls. The expression of ALYREF, DNMT1, NSUN2, TET2 are more positively correlated with the majority of pathways, such as the cell cycle, DNA replication, spliceosome, and nucleotide excision repair pathways. The DNMT1 expression is related to the activation of multiple metabolic pathways, including drug metabolism, lipid metabolism, and nucleic acid biosynthesis signaling pathways. The m⁵C regulators' pathways are significantly essential for a wide range of biological processes. At the same time, the precise molecular modifications mechanisms and cellular processes among pan-cancer need further study and deeper exploration for a better prognosis.

For a deeper exploration of the relationship between m⁵C regulators and their clinical outcomes, we describe a comprehensive landscape of m⁵C regulator pathways activities across different cancer types and identify cancer characteristics in relation to clinical outcome. Collectively, we provide robust evidence for the close relationship between cancer-associated clinical relevance and m⁵C regulators. To determine the effect of methylation-based molecular for earlier detection diagnostics in patients with several types of cancer, we systematically analyzed the m⁵C regulators' pathway activities with the functional and clinically complication for estimating tumor development and progression with potential prognostic value.

Conclusion

This research reveals epigenetic landscape changes of m⁵C regulator genes expression, the crosslink pathways activities, and clinical outcomes in multiple cancers. We also provide a better understanding of the biology of m⁵C regulators that influences the onset and progression of cancers and in order to develop new interventions for early detection and treatments.

Declarations

Ethical Approval

The study has been approved by the Ethics Committee of the First Affiliated Hospital of Zhengzhou University, and the study was conducted in accordance with the Declaration of Helsinki.

Consent for publication

Not applicable.

Availability of data and materials

The dataset supporting the conclusions of this article is available in the TCGA repository, <https://portal.gdc.cancer.gov/>.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Authors' Contributions

All authors participated in the design, interpretation of the studies and analysis of the data and review of the manuscript. GWZ designed the study, HYT collected the data and analyzed it, YX and ZMG produced figures and tables, HYT and GWZ wrote the manuscript.

Funding

This work was supported by Key Scientific Research Project of Henan Higher Education Institutions of China (21A320026).

Acknowledgements

We thank the patients and investigators who participated in TCGA and GEO for providing data.

References

- 1 J. Cortes, J.M. Perez-García, A. Llombart-Cussac, G. Curigliano, N.S. El Saghir, F. Cardoso, C.H. Barrios, S. Wagle, J. Roman, N. Harbeck, A. Eniu, P.A. Kaufman, J. Taberner, L. García-Estévez, P. Schmid, and J. Arribas, Enhancing global access to cancer medicines. *CA Cancer J Clin* 70 (2020) 105-124.
- 2 Y. van der Pol, and F. Mouliere, Toward the Early Detection of Cancer by Decoding the Epigenetic and Environmental Fingerprints of Cell-Free DNA. *Cancer Cell* 36 (2019) 350-368.
- 3 T. Hu, J. Wolfram, and S. Srivastava, Extracellular Vesicles in Cancer Detection: Hopes and Hypes. *Trends Cancer* S2405-8033 (2020) 30257-0.
- 4 P.G. Pilié, C. Tang, G.B. Mills, and T.A. Yap, State-of-the-art strategies for targeting the DNA damage response in cancer. *Nat Rev Clin Oncol* 16 (2019) 81-104.
- 5 J. Mateo, C.J. Lord, V. Serra, A. Tutt, J. Balmaña, M. Castroviejo-Bermejo, C. Cruz, A. Oaknin, S.B. Kaye, and J.S. de Bono, A decade of clinical development of PARP inhibitors in perspective. *Ann Oncol* 30 (2019) 1437-1447.
- 6 Y. Okugawa, W.M. Grady, and A. Goel, Epigenetic Alterations in Colorectal Cancer: Emerging Biomarkers. *Gastroenterology* 149 (2015) 1204-1225.e12.
- 7 M. Acunzo, G. Romano, D. Wernicke, and C.M. Croce, MicroRNA and cancer—a brief overview. *Adv Biol Regul* 57 (2015) 1-9.
- 8 J.J. Chan, and Y. Tay, Noncoding RNA:RNA Regulatory Networks in Cancer. *Int J Mol Sci* 19 (2018) 1310.
- 9 J.N. Anastas, and R.T. Moon, WNT signalling pathways as therapeutic targets in cancer. *Nat Rev Cancer* 13 (2013) 11-26.
- 10 C. Vicente-Dueñas, J. Hauer, C. Coboleda, A. Borkhardt, and I. Sánchez-García, Epigenetic Priming in Cancer Initiation. *Trends Cancer* 4 (2018) 408-417.
- 11 H. Shi, P. Chai, R. Jia, and X. Fan, Novel insight into the regulatory roles of diverse RNA modifications: Re-defining the bridge between transcription and translation. *Molecular cancer* 19 (2020) 78.
- 12 R.J. Liu, T. Long, J. Li, H. Li, and E.D. Wang, Structural basis for substrate binding and catalytic mechanism of a human RNA:m5C methyltransferase NSun6. *Nucleic acids research* 45 (2017) 6684-6697.
- 13 R. García-Vílchez, A. Sevilla, and S. Blanco, Post-transcriptional regulation by cytosine-5 methylation of RNA. *Biochim Biophys Acta Gene Regul Mech* 1862 (2019) 240-252.

- 14 M. Frye, and S. Blanco, Post-transcriptional modifications in development and stem cells. *Development* 143 (2016) 3871-3881.
- 15 Z. Dong, and H. Cui, The Emerging Roles of RNA Modifications in Glioblastoma. *Cancers (Basel)* 12 (2020) 736.
- 16 L. Tixl, and A. Lusser, The dynamic RNA modification 5-methylcytosine and its emerging role as an epitranscriptomic mark. *Wiley Interdiscip Rev RNA* 10 (2019) e1510.
- 17 J. Catania, and D.S. Fairweather, DNA methylation and cellular ageing. *Mutat Res* 256 (1991) 283-93.
- 18 K.E. Bohnsack, C. Höbartner, and M.T. Bohnsack, Eukaryotic 5-methylcytosine (m⁵C) RNA Methyltransferases: Mechanisms, Cellular Functions, and Links to Disease. *Genes (Basel)* 10 (2019) 102.
- 19 T. Yang, J.J.A. Low, and E.C.Y. Woon, A general strategy exploiting m⁵C duplex-remodelling effect for selective detection of RNA and DNA m⁵C methyltransferase activity in cells. *Nucleic Acids Res* 48 (2020) e5.
- 20 X. Yang, Y. Yang, B.F. Sun, Y.S. Chen, J.W. Xu, W.Y. Lai, A. Li, X. Wang, D.P. Bhattarai, W. Xiao, H.Y. Sun, Q. Zhu, H.L. Ma, S. Adhikari, M. Sun, Y.J. Hao, B. Zhang, C.M. Huang, N. Huang, G.B. Jiang, Y.L. Zhao, H.L. Wang, Y.P. Sun, and Y.G. Yang, 5-methylcytosine promotes mRNA export - NSUN2 as the methyltransferase and ALYREF as an m⁵C reader. *Cell Res* 27 (2017) 606-625.
- 21 L. Dou, X. Li, H. Ding, L. Xu, and H. Xiang, Prediction of m⁵C Modifications in RNA Sequences by Combining Multiple Sequence Features. *Mol Ther Nucleic Acids* 21 (2020) 332-342.
- 22 M.A. Gama-Sosa, V.A. Slagel, R.W. Trewyn, R. Oxenhandler, K.C. Kuo, C.W. Gehrke, and M. Ehrlich, The 5-methylcytosine content of DNA from human tumors. *Nucleic acids research* 11 (1983) 6883-94.
- 23 H. Wu, and Y. Zhang, Mechanisms and functions of Tet protein-mediated 5-methylcytosine oxidation. *Genes Dev* 25 (2011) 2436-52.
- 24 R. Reid, P.J. Greene, and D.V. Santi, Exposition of a family of RNA m⁵C methyltransferases from searching genomic and proteomic sequences. *Nucleic Acids Res* 27 (1999) 3138-45.
- 25 S. Hua, Y. Quan, M. Zhan, H. Liao, Y. Li, and L. Lu, miR-125b-5p inhibits cell proliferation, migration, and invasion in hepatocellular carcinoma via targeting TXNRD1. *Cancer Cell Int* 19 (2019) 203.
- 26 Q. Zhou, L.Q. Zhou, S.H. Li, Y.W. Yuan, L. Liu, J.L. Wang, D.Z. Wu, Y. Wu, and L. Xin, Identification of subtype-specific genes signature by WGCNA for prognostic prediction in diffuse type gastric cancer. *Aging (Albany NY)* 12 (2020) 17418-17435.
- 27 D. Szklarczyk, J.H. Morris, H. Cook, M. Kuhn, S. Wyder, M. Simonovic, A. Santos, N.T. Doncheva, A. Roth, P. Bork, L.J. Jensen, and C. von Mering, The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. *Nucleic Acids Res* 45 (2017) D362-d368.
- 28 H. Zhou, L. Chen, Y. Lei, T. Li, H. Li, and X. Cheng, Integrated analysis of tumor mutation burden and immune infiltrates in endometrial cancer. *Curr Probl Cancer* (2020) 100660.
- 29 A.L. Barabási, N. Gulbahce, and J. Loscalzo, Network medicine: a network-based approach to human disease. *Nat Rev Genet* 12 (2011) 56-68.
- 30 S. Panneerdoss, V.K. Eedunuri, P. Yadav, S. Timilsina, S. Rajamanickam, S. Viswanadhapalli, N. Abdelfattah, B.C. Onyeagucha, X. Cui, Z. Lai, T.A. Mohammad, Y.K. Gupta, T.H. Huang, Y. Huang, Y. Chen, and M.K. Rao, Cross-talk among writers, readers, and erasers of m⁶A regulates cancer growth and progression. *Sci Adv* 4 (2018) eaar8263.
- 31 Y. He, X. Yu, J. Li, Q. Zhang, Q. Zheng, and W. Guo, Role of m⁵C-related regulatory genes in the diagnosis and prognosis of hepatocellular carcinoma. *American journal of translational research* 12 (2020) 912-922.
- 32 H.A. Abdel-Hafiz, and K.B. Horwitz, Role of epigenetic modifications in luminal breast cancer. *Epigenomics* 7 (2015) 847-62.
- 33 P. Costa-Pinheiro, D. Montezuma, R. Henrique, and C. Jerónimo, Diagnostic and prognostic epigenetic biomarkers in cancer. *Epigenomics* 7 (2015) 1003-15.
- 34 Y. Wang, R. Zhang, D. Wu, Z. Lu, W. Sun, Y. Cai, C. Wang, and J. Jin, Epigenetic change in kidney tumor: downregulation of histone acetyltransferase MYST1 in human renal cell carcinoma. *J Exp Clin Cancer Res* 32 (2013) 8.
- 35 Y. Cheng, C. He, M. Wang, X. Ma, F. Mo, S. Yang, J. Han, and X. Wei, Targeting epigenetic regulators for cancer therapy: mechanisms and advances in clinical trials. *Signal Transduct Target Ther* 4 (2019) 62.

- 36 B.J. Wouters, and R. Delwel, Epigenetics and approaches to targeted epigenetic therapy in acute myeloid leukemia. *Blood* 127 (2016) 42-52.
- 37 A.B. Iniguez, G. Alexe, E.J. Wang, G. Roti, S. Patel, L. Chen, S. Kitara, A. Conway, A.L. Robichaud, B. Stolte, P. Bandopadhyay, A. Goodale, S. Pantel, Y. Lee, D.M. Cheff, M.D. Hall, R. Guha, M.I. Davis, M. Menard, N. Nasholm, W.A. Weiss, J. Qi, R. Beroukhi, F. Piccioni, C. Johannessen, and K. Stegmaier, Resistance to Epigenetic-Targeted Therapy Engenders Tumor Cell Vulnerabilities Associated with Enhancer Remodeling. *Cancer Cell* 34 (2018) 922-938.e7.
- 38 F. Atrian, and S.A. Lelièvre, Mining the epigenetic landscape of tissue polarity in search of new targets for cancer therapy. *Epigenomics* 7 (2015) 1313-25.
- 39 R. Hashizume, Epigenetic Targeted Therapy for Diffuse Intrinsic Pontine Glioma. *Neurol Med Chir (Tokyo)* 57 (2017) 331-342.
- 40 I. Estibariz, A. Overmann, F. Ailloud, J. Krebs, C. Josenhans, and S. Suerbaum, The core genome m5C methyltransferase JHP1050 (M.Hpy99III) plays an important role in orchestrating gene expression in *Helicobacter pylori*. *Nucleic Acids Res* 47 (2019) 2336-2348.
- 41 S. Saghafein, M. Mina, N. Riggi, D. Hanahan, and G. Ciriello, Pan-Cancer Landscape of Aberrant DNA Methylation across Human Tumors. *Cell Rep* 25 (2018) 1066-1080.e8.
- 42 Q. Zhang, Y. He, N. Luo, S.J. Patel, Y. Han, R. Gao, M. Modak, S. Carotta, C. Haslinger, D. Kind, G.W. Peet, G. Zhong, S. Lu, W. Zhu, Y. Mao, M. Xiao, M. Bergmann, X. Hu, S.P. Kerkar, A.B. Vogt, S. Pflanz, K. Liu, J. Peng, X. Ren, and Z. Zhang, Landscape and Dynamics of Single Immune Cells in Hepatocellular Carcinoma. *Cell* 179 (2019) 829-845.e20.
- 43 J.W.T. Tse, L.J. Jenkins, F. Chionh, and J.M. Mariadason, Aberrant DNA Methylation in Colorectal Cancer: What Should We Target? *Trends Cancer* 3 (2017) 698-712.
- 44 T. Sibbritt, H.R. Patel, and T. Preiss, Mapping and significance of the mRNA methylome. *Wiley Interdiscip Rev RNA* 4 (2013) 397-422.
- 45 M.C. Popis, S. Blanco, and M. Frye, Posttranscriptional methylation of transfer and ribosomal RNA in stress response pathways, cell differentiation, and cancer. *Curr Opin Oncol* 28 (2016) 65-71.
- 46 L. Lu, G. Zhu, H. Zeng, Q. Xu, and K. Holzmann, High tRNA Transferase NSUN2 Gene Expression is Associated with Poor Prognosis in Head and Neck Squamous Carcinoma. *Cancer Invest* 36 (2018) 246-253.
- 47 X. Cui, Z. Liang, L. Shen, Q. Zhang, S. Bao, Y. Geng, B. Zhang, V. Leo, L.A. Vardy, T. Lu, X. Gu, and H. Yu, 5-Methylcytosine RNA Methylation in *Arabidopsis thaliana*. *Mol Plant* 10 (2017) 1387-1399.
- 48 S. Haag, A.S. Warda, J. Kretschmer, M.A. Günnigmann, C. Höbartner, and M.T. Bohnsack, NSUN6 is a human RNA methyltransferase that catalyzes formation of m5C72 in specific tRNAs. *Rna* 21 (2015) 1532-43.
- 49 L. Van Haute, S.Y. Lee, B.J. McCann, C.A. Powell, D. Bansal, L. Vasiliauskaitė, C. Garone, S. Shin, J.S. Kim, M. Frye, J.G. Gleeson, E.A. Miska, H.W. Rhee, and M. Minczuk, NSUN2 introduces 5-methylcytosines in mammalian mitochondrial tRNAs. *Nucleic Acids Res* 47 (2019) 8720-8733.
- 50 M. Janin, V. Ortiz-Barahona, M.C. de Moura, A. Martínez-Cardús, P. Llinàs-Arias, M. Soler, D. Nachmani, J. Pelletier, U. Schumann, M.E. Calleja-Cervantes, S. Moran, S. Guil, A. Bueno-Costa, D. Piñeyro, M. Perez-Salvia, M. Rosselló-Tortella, L. Piqué, J.J. Bech-Serra, C. De La Torre, A. Vidal, M. Martínez-Iniesta, J.F. Martín-Tejera, A. Villanueva, A. Arias, I. Cuartas, A.M. Aransay, A.M. La Madrid, A.M. Carcaboso, V. Santa-Maria, J. Mora, A.F. Fernandez, M.F. Fraga, I. Aldecoa, L. Pedrosa, F. Graus, N. Vidal, F. Martínez-Soler, A. Tortosa, C. Carrato, C. Balañá, M.W. Boudreau, P.J. Hergenrother, P. Köttér, K.D. Entian, J. Hench, S. Frank, S. Mansouri, G. Zadeh, P.D. Dans, M. Orozco, G. Thomas, S. Blanco, J. Seoane, T. Preiss, P.P. Pandolfi, and M. Esteller, Epigenetic loss of RNA-methyltransferase NSUN5 in glioma targets ribosomes to drive a stress adaptive translational program. *Acta Neuropathol* 138 (2019) 1053-1074.
- 51 Q. Zhang, Q. Zheng, X. Yu, Y. He, and W. Guo, Overview of distinct 5-methylcytosine profiles of messenger RNA in human hepatocellular carcinoma and paired adjacent non-tumor tissues. *J Transl Med* 18 (2020) 245.
- 52 Y. He, Q. Zhang, Q. Zheng, X. Yu, and W. Guo, Distinct 5-methylcytosine profiles of circular RNA in human hepatocellular carcinoma. *Am J Transl Res* 12 (2020) 5719-5729.
- 53 P. Li, Y. Xu, Q. Zhang, Y. Li, W. Jia, X. Wang, Z. Xie, J. Liu, D. Zhao, M. Shao, S. Chen, N. Mo, Z. Jiang, L. Li, R. Liu, W. Huang, L. Chang, S. Chen, H. Li, W. Zuo, J. Li, R. Zhang, and X. Yang, Evaluating the role of RAD52 and its interactors as novel potential molecular targets for hepatocellular carcinoma. *Cancer Cell Int* 19 (2019) 279.

54 E.D. Pleasance, R.K. Cheetham, P.J. Stephens, D.J. McBride, S.J. Humphray, C.D. Greenman, I. Varela, M.L. Lin, G.R. Ordóñez, G.R. Bignell, K. Ye, J. Alipaz, M.J. Bauer, D. Beare, A. Butler, R.J. Carter, L. Chen, A.J. Cox, S. Edkins, P.I. Kokko-Gonzales, N.A. Gormley, R.J. Grocock, C.D. Haudenschild, M.M. Hims, T. James, M. Jia, Z. Kingsbury, C. Leroy, J. Marshall, A. Menzies, L.J. Mudie, Z. Ning, T. Royce, O.B. Schulz-Trieglaff, A. Spiridou, L.A. Stebbings, L. Szajkowski, J. Teague, D. Williamson, L. Chin, M.T. Ross, P.J. Campbell, D.R. Bentley, P.A. Futreal, and M.R. Stratton, A comprehensive catalogue of somatic mutations from a human cancer genome. *Nature* 463 (2010) 191-6.

55 A.C. Krepischi, P.L. Pearson, and C. Rosenberg, Germline copy number variations and cancer predisposition. *Future Oncol* 8 (2012) 441-50.

56 X. Zhao, X. Liu, A. Zhang, H. Chen, Q. Huo, W. Li, R. Ye, Z. Chen, L. Liang, Q.A. Liu, J. Shen, X. Jin, W. Li, M. Nygaard, X. Liu, Y. Hou, T. Ni, L. Bolund, W. Gottschalk, W. Tao, J. Gu, X.L. Tian, H. Yang, J. Wang, X. Xu, M.W. Lutz, J. Min, Y. Zeng, and C. Nie, The correlation of copy number variations with longevity in a genome-wide association study of Han Chinese. *Aging (Albany NY)* 10 (2018) 1206-1222.

57 J. Fan, K. Wang, X. Du, J. Wang, S. Chen, Y. Wang, M. Shi, L. Zhang, X. Wu, D. Zheng, C. Wang, L. Wang, B. Tian, G. Li, Y. Zhou, and H. Cheng, ALYREF links 3'-end processing to nuclear export of non-polyadenylated mRNAs. *The EMBO journal* 38 (2019) e99910.

Tables

Table 1 The 33 cancer types in TCGA pan-cancer project

Cancer types	Abbr	Normal tissues	Cancer tissues	Mutation	CNV
Kidney Renal Clear Cell Carcinoma	KIRC	72	539	370	531
Kidney Renal Papillary Cell Carcinoma	KIRP	32	289	282	291
Kidney Chromophobe	KICH	24	65	66	69
Brain Lower Grade Glioma	LGG	0	529	526	516
Glioblastoma Multiforme	GBM	5	169	403	580
Breast Invasive Carcinoma	BRCA	113	1109	1026	1083
Lung Squamous Cell Carcinoma	LUSC	49	502	485	504
Lung Adenocarcinoma	LUAD	59	535	569	519
Rectum Adenocarcinoma	READ	10	167	151	168
Colon Adenocarcinoma	COAD	41	480	408	454
Uterine Carcinosarcoma	UCS	0	56	57	59
Uterine Corpus Endometrial Carcinoma	UCEC	35	552	531	542
Ovarian Serous Cystadenocarcinoma	OV	0	379	412	582
Head and Neck Squamous Carcinoma	HNSC	44	502	509	525
Thyroid Carcinoma	THCA	58	510	500	502
Prostate Adenocarcinoma	PRAD	52	499	498	495
Stomach Adenocarcinoma	STAD	32	375	439	444
Skin Cutaneous Melanoma	SKCM	1	471	468	370
Bladder Urothelial Carcinoma	BLCA	19	414	411	411
Liver Hepatocellular Carcinoma	LIHC	50	374	365	373
Cervical Squamous Cell Carcinoma and Endocervical Adenocarcinoma	CESC	3	306	291	298
Adrenocortical Carcinoma	ACC	0	79	92	93
Pheochromocytoma and Paraganglioma	PCPG	3	183	184	165
Sarcoma	SARC	2	263	239	260
Acute Myeloid Leukemia	LAML	0	151	141	194

Pancreatic Adenocarcinoma	PAAD	4	178	178	187
Esophageal Carcinoma	ESCA	11	162	185	187
Testicular Germ Cell Tumors	TGCT	0	156	151	153
Thymoma	THYM	2	119	123	126
Mesothelioma	MESO	0	86	82	90
Uveal Melanoma	UVM	0	80	80	83
Lymphoid Neoplasm Diffuse Large B-cell Lymphoma	DLBC	0	48	37	51
Cholangiocarcinoma	CHOL	9	36	36	39
In total		730	10,363	10,295	10,944

Table 2 The mutation frequency of m5C regulators across 33 cancer types (Top 5)

Function	Genes	UCEC	SKCM	COAD	READ	STAD
writers	NSUN1	0	0	0	0	0
	NSUN2	0.0622642	0.0192719	0.0250627	0.0218978	0.020595
	NSUN3	0.0283019	0.0021413	0.0100251	0.0145985	0.0091533
	NSUN4	0.0188679	0.0085653	0.0125313	0	0.006865
	NSUN5	0.0207547	0.0171306	0.0200501	0	0.006865
	NSUN6	0.045283	0.0149893	0.0150376	0.0145985	0.0183066
	NSUN7	0.0509434	0.0449679	0.0125313	0.0072993	0.0091533
	DNMT1	0.0773585	0.0428266	0.0401003	0.0437956	0.0320366
	DNMT2	0	0	0	0	0
	DNMT3A	0.0660377	0.0278373	0.0225564	0.0218978	0.0183066
	DNMT3B	0.0811321	0.0342612	0.037594	0.0510949	0.0343249
eraser	TET2	0.0943396	0.0428266	0.0551378	0.0291971	0.0320366
reader	ALYREF	0.0188679	0.0021413	0.0050125	0.0072993	0.0022883

Table 3 The CNV-Gain and CNV-loss frequency of m5C regulators across 33 cancer types (Top 5)

Genes	CNV Gain					CNV loss				
	KICH	OV	ACC	UCS	LUSC	KICH	ACC	TGCT	UCS	OV
NSUN1	0	0	0	0	0	0	0	0	0	0
NSUN2	0.560606	0.558678	0.633333	0.464286	0.699801	0.227273	0.144444	0.544872	0.125	0.099174
NSUN3	0.545455	0.540496	0.122222	0.232143	0.526839	0.212121	0.411111	0.102564	0.142857	0.044628
NSUN4	0.015152	0.499174	0.033333	0.339286	0.073559	0.863636	0.6	0.128205	0.089286	0.087603
NSUN5	0.818182	0.477686	0.511111	0.25	0.26839	0.030303	0.022222	0.032051	0.178571	0.082645
NSUN6	0.075758	0.428099	0.255556	0.321429	0.089463	0.787879	0.3	0.416667	0.321429	0.135537
NSUN7	0.863636	0.195041	0.411111	0.25	0.083499	0.015152	0.111111	0.49359	0.285714	0.408264
DNMT1	0.727273	0.418182	0.577778	0.285714	0.101392	0.015152	0.033333	0.198718	0.375	0.295868
DNMT2	0	0	0	0	0	0	0	0	0	0
DNMT3A	0.045455	0.403306	0.111111	0.464286	0.335984	0.772727	0.422222	0.012821	0.017857	0.102479
DNMT3B	0.787879	0.689256	0.555556	0.660714	0.39165	0.045455	0.111111	0.038462	0.017857	0.019835
TET2	0.818182	0.044628	0.366667	0	0.037773	0.015152	0.122222	0.608974	0.642857	0.694215
ALYREF	0.030303	0.320661	0.166667	0.464286	0.252485	0.787879	0.411111	0.038462	0.125	0.292562

Figures

Figure 1

m5C regulators and the function in cancers. (a) The distribution and the function of m5C regulatory writers, eraser, and reader. (b) The workflow scheme for this study.

Figure 2

Mutation and CNV of m5C regulators across pan-cancer. (a) The mutant frequency of m5C regulators across 33 cancer types. (b) CNV analysis of m5C regulators across cancer types. The upper part of each grid shows the deletion frequency, and the bottom part shows the amplification frequency.

Figure 3

The association between CNV and the gene expression of m5C regulatory genes. (a) Alterations to m5C regulatory gene expression in 24 cancer types. The heat map demonstrates fold change, with red representing up-regulated genes and blue representing down-regulated genes. (b) Box plots exhibit the expression distribution of DNMT3B across tumor and normal samples in 24 cancer types.

Figure 4

m5C regulators are associated with the activation and inhibition of cancer pathways. (a) Network landscape demonstrating the correlation between m5C regulators and cancer pathways. Red represents a positive correlation, and blue represents a negative correlation. The size of the nodes corresponds to the number of links. (b) The number of pathways correlated with individual m5C regulators. The upper panel represents positively correlated pathways, and the bottom panel represents negatively correlated pathways. (c) Correlation among the expression of m5C regulators. (d) The protein-protein interaction (PPI) network of m5C regulators.

Figure 5

Summary of the relationship between m5C regulators expression and patient's survival. (a) Overall survival (OS) of m5C regulators across 33 cancer types. (b) Progression-free interval (PFI) of m5C regulators across 32 solid cancer types. (c) Disease-specific survival (DSS) of m5C regulators across 32 solid cancer types. (d) Disease-free interval (DFI) of m5C regulators across 28 solid cancer types. Red represents a higher m5C regulator expression associated with poor survival, and blue represents an association with better survival.

Figure 6

Effect of m5C regulators on patients with hepatocellular carcinoma and cholangiocarcinoma. (a) Heat map showing clustering for CHOL and LIHC patients based m5C regulator expression. Yellow represents Reg-low subgroup (n = 295), and green represents Reg-high subgroup (n = 112). (b) Kaplan-Meier survival plot of patients grouped by global m5C regulator expression pattern (p <0.001).