

RESEARCH

Open Access



# Pan-cancer analysis of m<sup>5</sup>C regulator genes reveals consistent epigenetic landscape changes in multiple cancers

Yuting He<sup>1,2,3,4\*</sup> , Xiao Yu<sup>1,2,3,4</sup>, Menggang Zhang<sup>1,2,3,4</sup> and Wenzhi Guo<sup>1,2,3,4\*</sup>

## Abstract

**Background:** 5-Methylcytosine (m<sup>5</sup>C) 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<sup>5</sup>C 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<sup>5</sup>C regulatory genes were significantly correlated across many cancer types. We then assessed the correlation between the expression of individual m<sup>5</sup>C regulators and the activity of related hallmark pathways of cancers.

**Results:** After validating m<sup>5</sup>C regulators' expression based on their contributions to cancer development and progression, we observed their upregulation within tumor-specific processes. Notably, our research connected aberrant alterations to m<sup>5</sup>C regulatory genes with poor clinical outcomes among various tumors that may drive cancer pathogenesis and/or survival.

**Conclusion:** Our results offered strong evidence and clinical implications for the involvement of m<sup>5</sup>C regulators.

**Keywords:** m<sup>5</sup>C regulatory genes, Frequent network mining, Pan-cancer analysis, Survival, 5-Methylcytosine

## Background

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], oncogene activation, aberrant methylation modifications, alterations in epigenetic changes [6–8], aberrant expression of microRNAs [9], and alterations of signaling pathways

are crucial factors and contribute to cancer pathogenesis [10–12]. Methylation is an essential epigenetic modification and is closely related to the pathogenesis of cancers [13–17]. The 5-methylcytosine (m<sup>5</sup>C), N6-methyladenine (m<sup>6</sup>A), and N1-methyladenosine (m<sup>1</sup>A) have become the most common types of epigenetic modifications in eukaryotes [13]. Emerging evidence has demonstrated that m<sup>5</sup>C modification has the potential to serve as novel epigenetic markers with remarkable biological significance in biological processes [18–20].

m<sup>5</sup>C modification distributes in different types of RNAs and DNAs [21–23]. m<sup>5</sup>C modifications can even modify the destiny of cancer cells [24]. m<sup>5</sup>C 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

\*Correspondence: fcchey1@zzu.edu.cn; fccguowz@zzu.edu.cn

<sup>1</sup> Department of Hepatobiliary and Pancreatic Surgery, The First Affiliated Hospital of Zhengzhou University, No.1 Jianshe Road, Zhengzhou 450052, China

Full list of author information is available at the end of the article



© The Author(s) 2021. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

modifications [25, 26]. m<sup>5</sup>C could therefore be used as a biomarker for disease progression, including various types of cancers [27]. m<sup>5</sup>C maintains open and closed chromatin states to control gene expression, genome editing, organismal development, and cellular differentiation [23]. In this context, writers act within a methyltransferase complex to methylate targets, and erasers remove m<sup>5</sup>C methylation, while readers recognize and bind to m<sup>5</sup>C-methylated RNA and implement corresponding functions [25, 28, 29]. The anomalous interplay between writers and erasers, arising from alterations to their expression, has been linked to cancer pathogenesis and progression [23, 30]. However, pan-cancer effects of changes to m<sup>5</sup>C regulatory gene expression have not been fully defined. Next-generation sequencing (NGS) provides us effective tools to comprehensively view the m<sup>5</sup>C distribution landscape throughout the global transcriptome [27].

In this study, we identified the potential prognostic value of m<sup>5</sup>C regulators and provided a comprehensive understanding of m<sup>5</sup>C 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 kilobase transcripts based on fragments per kilobase of transcript per million (FPKM) gene expression from The Cancer Genome Atlas (TCGA, <https://www.cancer.gov/>) dataset among 33 different cancer types. m<sup>5</sup>C regulator patterns were investigated in 5480 samples among 33 different cancer types,

including somatic mutations, copy number variations (CNVs), gene expression, and RNA-seq data (Fig. 1B).

**Genomic data collection of m<sup>5</sup>C regulators**

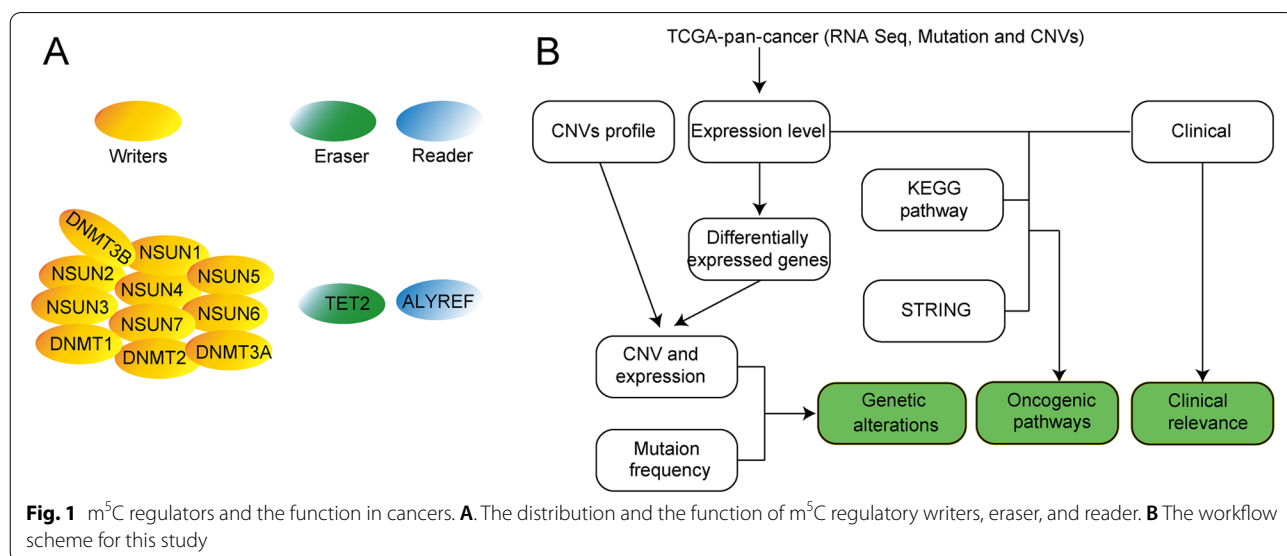
Thirteen m<sup>5</sup>C regulators were identified from published papers. Information of m<sup>5</sup>C regulators was collected from Gene Cards ([www.genecards.org](http://www.genecards.org)). Ensemble gene IDs and HUGO Gene Nomenclature Committee symbols were assigned to each m<sup>5</sup>C regulator-associated gene.

**Whole genomics data analysis of 33 pan-cancers**

The integrated OMICS datasets based on the TCGA database of 33 pan-cancers were applied in this study. We collected the mutation annotation format profile from TCGA, which contains over 10,000 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 Genomic Data Commons 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 |log<sub>2</sub> fold change (log<sub>2</sub>Fc) | no less than two as the statistical threshold value for differentially expressed genes. The results were screened as significant DEGs and methylated sites.



**Fig. 1** m<sup>5</sup>C regulators and the function in cancers. **A** The distribution and the function of m<sup>5</sup>C regulatory writers, eraser, and reader. **B** The workflow scheme for this study

### Functional annotations and pathway enrichment analysis

To evaluate the biological functions of each m<sup>5</sup>C 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 [31]. 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 m<sup>5</sup>C regulators. To further explore m<sup>5</sup>C 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 m<sup>5</sup>C regulators.

### The internships between m<sup>5</sup>C regulators

To visualize the intercorrelations among m<sup>5</sup>C regulators, we adopted the “CORRRAP” R package (<https://github.com/taiyun/corrplot>). Besides, the STRING database was also applied for the exploration and analysis of these associations between m<sup>5</sup>C regulators and 325 related genes [32]. The 325 genes were obtained from the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (<http://www.kegg.jp/> or <http://www.genome.jp/kegg/>). The correlations between m<sup>5</sup>C regulators and 325 genes were visualized through Cytoscape (<https://cytoscape.org/>).

### Clinical characteristic of m<sup>5</sup>C regulators

To explore the m<sup>5</sup>C 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 m<sup>5</sup>C regulators in cell growth

The CRISPR-CAS9 gene scale screening of cell lines from 33 cancer types was collected from previous study [32]. We calculated the proportion of every regulator as an essential gene in the cell lines.

## Results

### Results m<sup>5</sup>C regulators identification and its genomic extensive genetic changes

In this study, we identified 13 m<sup>5</sup>C regulators, as shown in Fig. 1A, eleven writers (NSUN1-7, DNMT1, DNMT2,

DNMT3A, and DNMT3B), one eraser (TET2), and one reader (ALYREF). This study validated the frequency of m<sup>5</sup>C regulator patterns among 33 cancers by integrating somatic mutations and 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 [33]. 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 m<sup>5</sup>C regulators was common. Regulators such as DNMT3B, ALYREF, and NSUN5 displayed extensive CNVs. On the contrary, TET2 and NSUN4 showed significant lack of m<sup>5</sup>C modification related CNV mutations among pan-cancers (Fig. 2B, and Table 2).

To further investigate whether the genomic mutations affect m<sup>5</sup>C regulators expression, we intensively detected the m<sup>5</sup>C 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 m<sup>5</sup>C regulator's expression (Fig. 3A). The m<sup>5</sup>C regulators with CNV amplification showed significantly increased expression in pan-cancers (such as DNMT3B), and m<sup>5</sup>C regulators with CNV deletion exhibited remarkably decreased expression, like TET2. In addition, we comparatively analyzed the m<sup>5</sup>C 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). These results uncovered that the m<sup>5</sup>C regulators among various cancers showed significant heterogeneity in gene expression and genetics. Collectively, our results demonstrated that aberrant m<sup>5</sup>C regulations were crucial for carcinogenesis and progression, which provided a clue for further functional detection.

### m<sup>5</sup>C regulators related pan-carcinogenic pathways

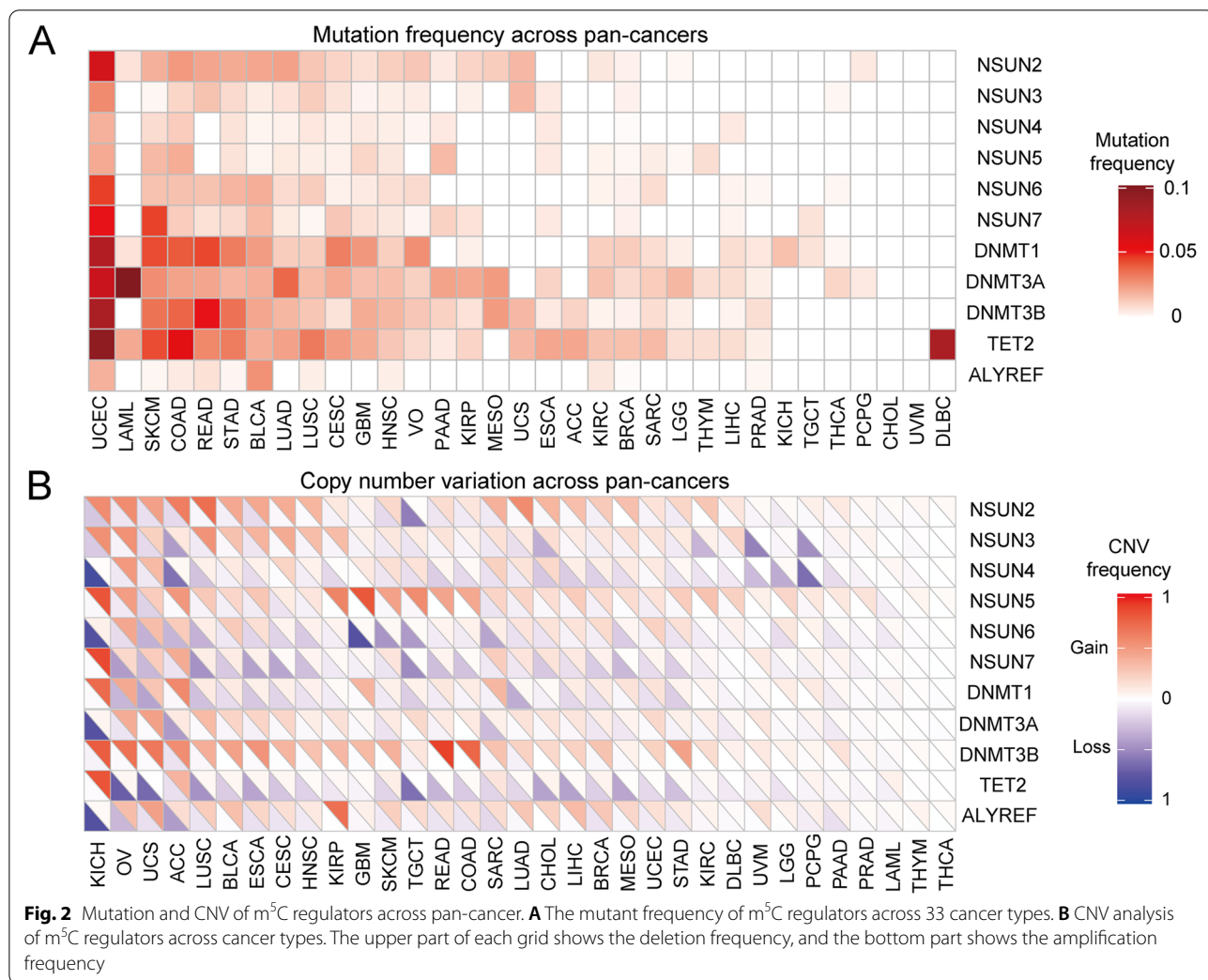
To comprehensively explore the molecular mechanism m<sup>5</sup>C regulators involved in cancers, we evaluated the correlations between m<sup>5</sup>C regulators' proteins expression and KEGG enrichment analysis-related activities. The results indicated that m<sup>5</sup>C regulator proteins have a close 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

**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

cancer-related pathways. DNMT1 was involved in drug metabolism, lipid metabolism, and nucleic acid biosynthesis signaling pathways [34]. ALYREF, DNMT1, NSUN5, NSUN1, and TET2 showed active involvement in KEGG enrichment pathways (Fig. 4B). In addition, genes will not function alone [35]. Growing evidence indicated that genes always co-effect with multiple genes and always have multiple functions [35, 36]. We further explored the internal connections between m<sup>5</sup>C regulators gene expression. Results indicated that the readers, writers, and erasers also have high correlations

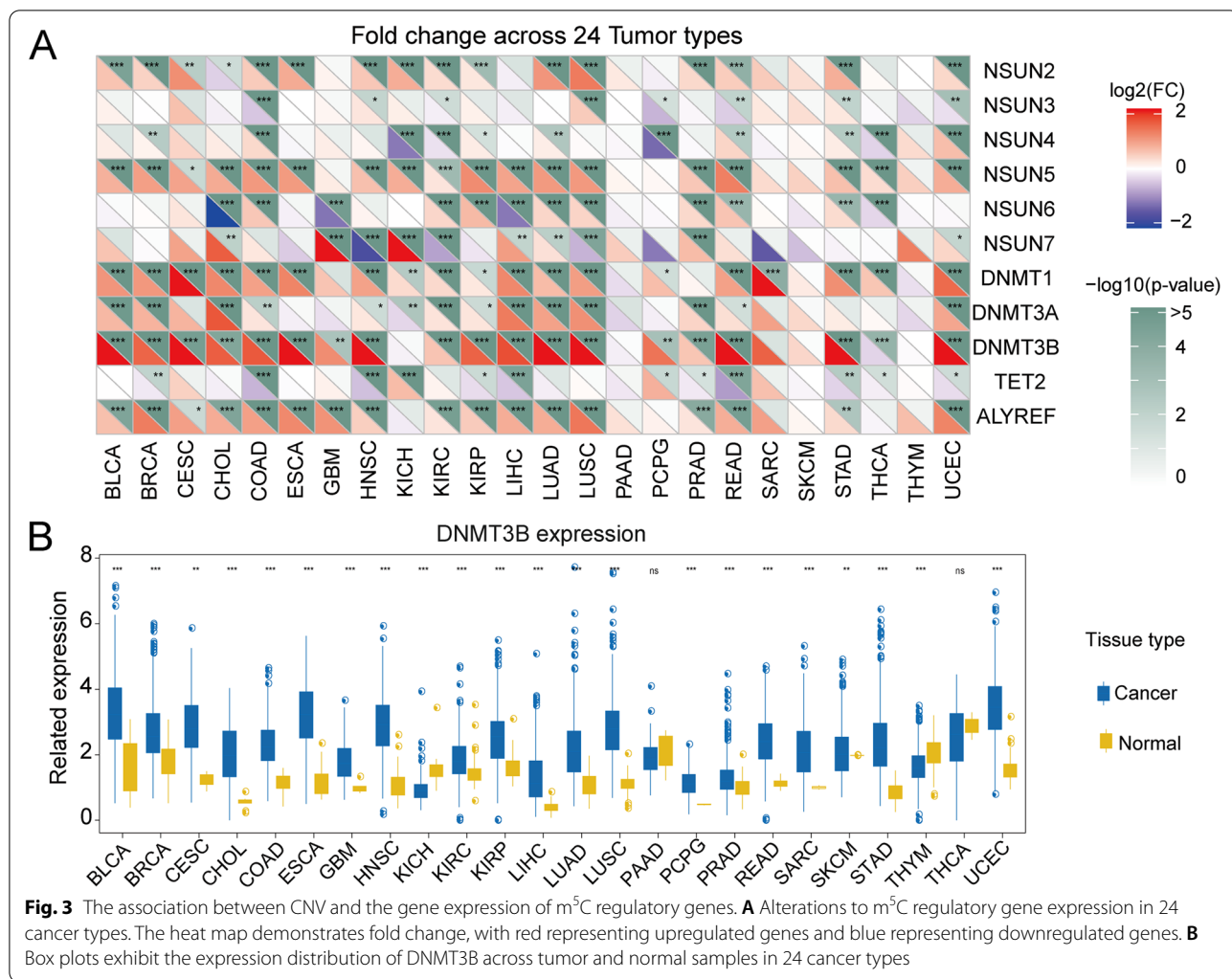
with each other. The eraser TET2 was significant correlated with the writer NSUN3. Writers such as ALYREF and NSUN5 also showed obvious correlations ( $R = 0.55$ ,  $P < 0.01$ ) (Fig. 4C). Additionally, to visualize the interactions between m<sup>5</sup>C regulators, we utilized the protein-protein interaction (PPI) analysis in m<sup>5</sup>C regulators related proteins. The results showed that writers, readers, and erase were particularly frequent (Fig. 4D). These results indicated that interactions among m<sup>5</sup>C regulators play crucial roles in the development and progression of cancers.



**Fig. 2** Mutation and CNV of m<sup>5</sup>C regulators across pan-cancer. **A** The mutant frequency of m<sup>5</sup>C regulators across 33 cancer types. **B** CNV analysis of m<sup>5</sup>C regulators across cancer types. The upper part of each grid shows the deletion frequency, and the bottom part shows the amplification frequency

**Table 2** The mutation frequency of m<sup>5</sup>C 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



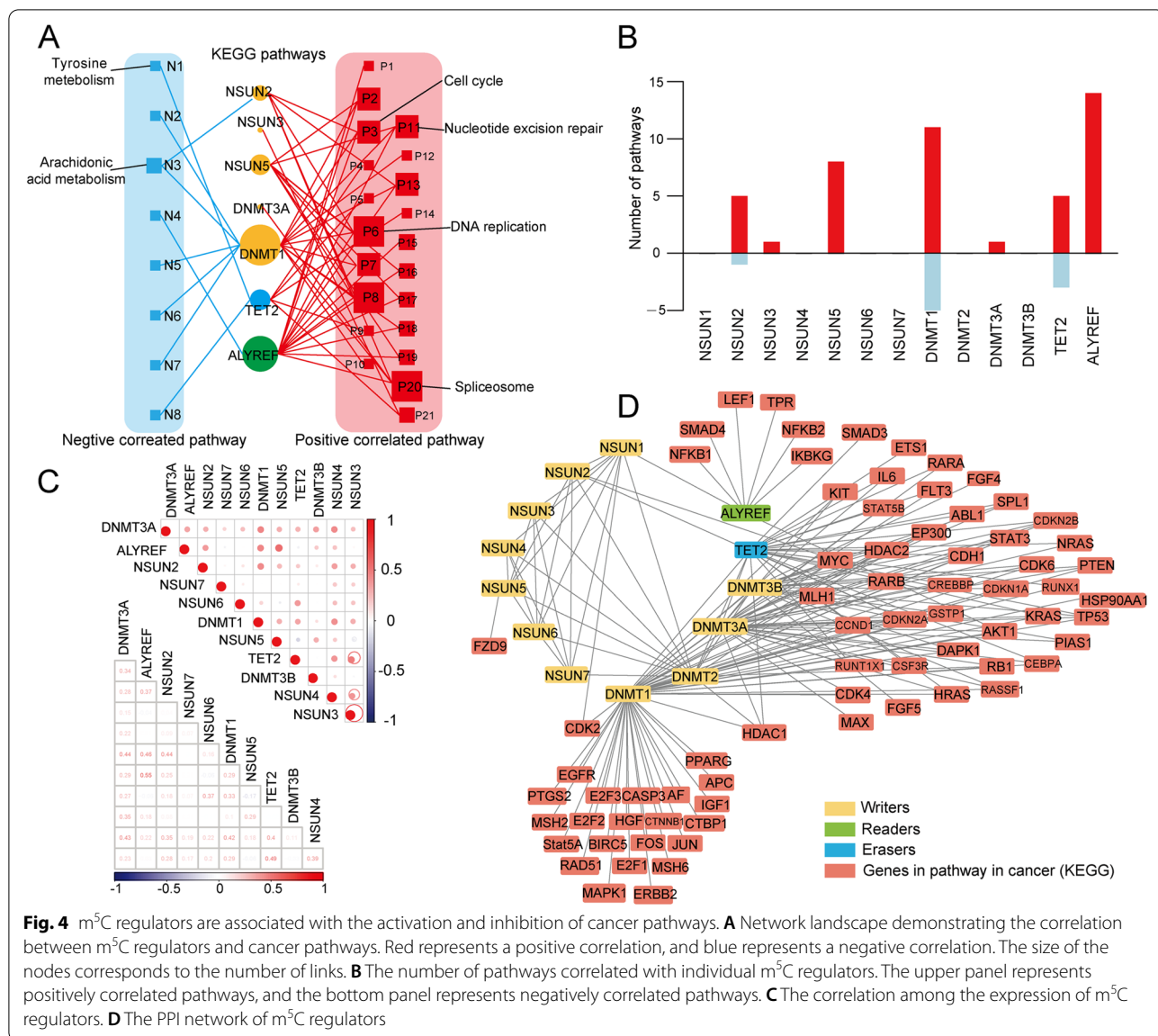
**Clinical significance of m<sup>5</sup>C regulators in pan cancers**

To evaluate the clinical prognosis of m<sup>5</sup>C regulators, we calculated the overall survival (OS), overall median progression-free interval (PFI), disease-specific survival (DSS), and disease-free interval (DFI) of m<sup>5</sup>C regulators. The OS analysis implied a significant correlation between m<sup>5</sup>C regulators and thirty-three pan-cancers. The heat map demonstrated that m<sup>5</sup>C regulators were significantly correlated with survival of patients, including OS, PFI, DSS, and DFI. In detail, the OS in adrenocortical carcinoma (ACC), kidney chromophobe (KICH), kidney renal clear cell carcinoma (KIRC), brain lower grade glioma (LGG), and liver hepatocellular carcinoma (LIHC) showed significant correlations with m<sup>5</sup>C 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 values, indicating that they have the potential to be utilized as unfavorable prognosis prediction factors. The PFI in ACC, KICH, KIRC, LGG, and uveal melanoma (UVM) showed significant correlations with most m<sup>5</sup>C regulators (Fig. 5B). Similarly, The DSS and DFI analyses indicated that ACC and LGG showed remarkably correlations with most m<sup>5</sup>C regulators (Fig. 5C, D). These results indicated that m<sup>5</sup>C regulators have crucial prognostic prediction values in a variety of cancer types.

**Effect of m<sup>5</sup>C regulators in LIHC and cholangiocarcinoma (CHOL)**

Studies have validated that m<sup>5</sup>C-related genes alterations have a close relationship with advanced tumor progression and advanced tumor stages [37], based on the above pieces of evidence that most m<sup>5</sup>C regulators



**Fig. 4** m<sup>5</sup>C regulators are associated with the activation and inhibition of cancer pathways. **A** Network landscape demonstrating the correlation between m<sup>5</sup>C 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 m<sup>5</sup>C regulators. The upper panel represents positively correlated pathways, and the bottom panel represents negatively correlated pathways. **C** The correlation among the expression of m<sup>5</sup>C regulators. **D** The PPI network of m<sup>5</sup>C regulators

are associated with patients' OS in LIHC and CHOL (Fig. 5A). Based on the overall expression patterns of m<sup>5</sup>C regulators, all patients in these cancer groups were categorized into two subgroups. The first subgroup consisted of 112 patients indicating high expression of m<sup>5</sup>C regulators (Reg-high), and the second subgroup consisted of 295 patients with low m<sup>5</sup>C 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 m<sup>5</sup>C regulators have a potential function as prognostic indicators in hepatocellular carcinoma and cholangiocarcinoma.

### Discussion

Epigenetic variation is often related to human disease, especially cancers [12, 38–40]. 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 [41–45]. Aberrant methylation regulators process both in DNA and RNA play a critical role in epigenetic regulators, which are significantly associated with tumorigenesis [17, 46–51].

Original reports described that NSUN2 participates in catalyzing biological reactions of m<sup>5</sup>C formation in RNAs and regulating cell cycle [52], linked to stem cell

**Table 3** The CNV-Gain and CNV-loss frequency of m<sup>5</sup>C 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

differentiation and involved in progression [53]. It is also reported that m<sup>5</sup>C regulators such as NSUN2 and binding partner ALYREF participant in promoting mRNA export coordinately [54], and NSUN6, in complex with a full-length tRNA substrate targeting cytosine accessible to the enzyme for methylation [14, 55, 56]. The NSUN3 is required for the deposition of m<sup>5</sup>C at the anticodon loop in the mitochondria encoded transfer RNA methionine [57]. The NSUN5 demonstrates suppression characteristics in vivo glioma models [58]. NSUN5 gene mutation leads to an un-methylated 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 [59], as is illustrated that the m<sup>5</sup>C regulators may influence a wide variety of biological functions and metabolism.

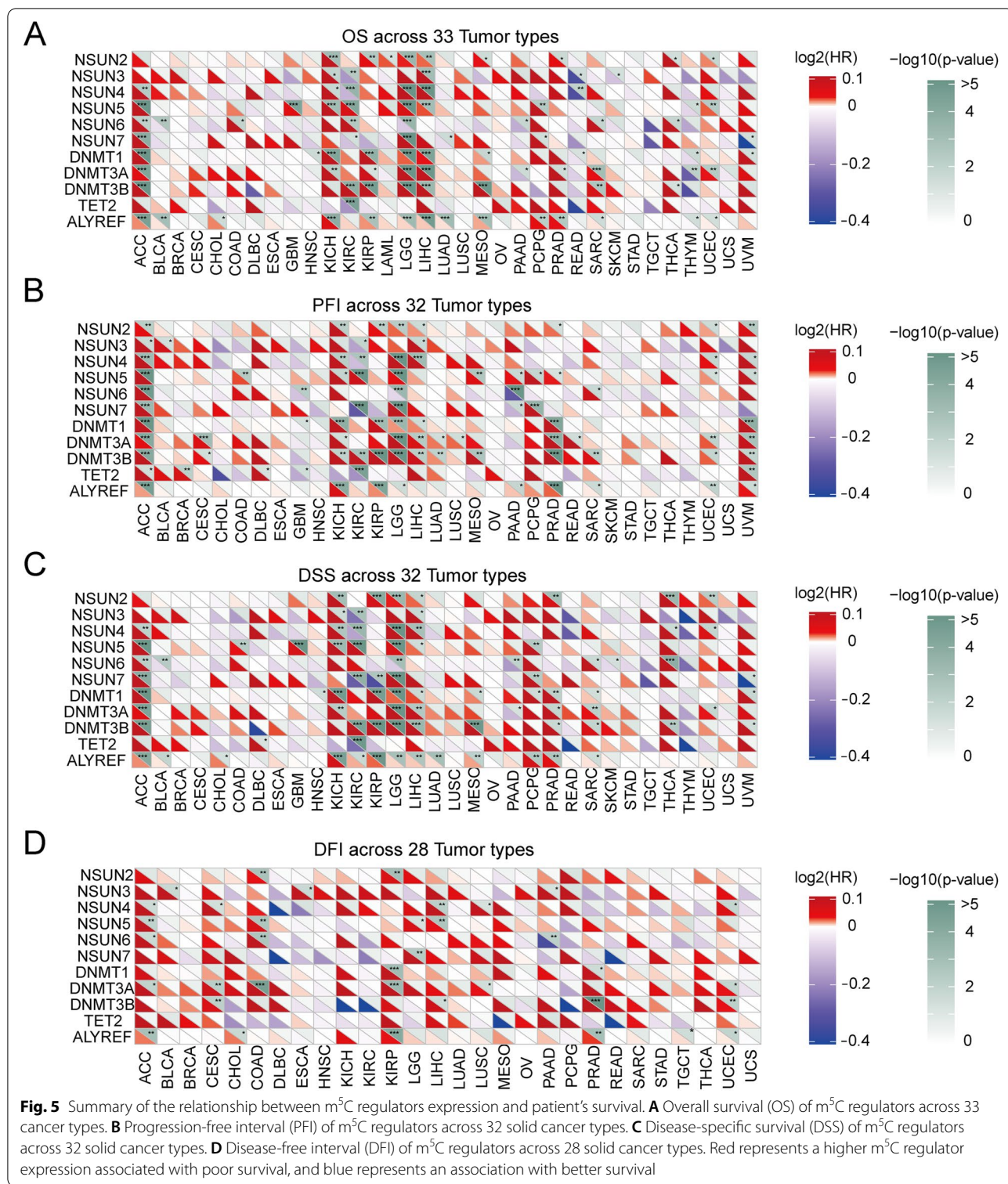
In the present study, we applied certain methodological particularities to build a model and evaluated a catalog of genomic characteristics of tumors associated with m<sup>5</sup>C regulators. We obtained a total of 13 m<sup>5</sup>C regulators. The mutations and CNVs of m<sup>5</sup>C regulators are linked to several tumor developments. All cancers carry somatic mutations [60]. UCEC exhibited a significantly higher number of mutations across pan-cancers, analogously TET2, DNMT3B, DNMT3A, and DNMT1 have placed a moderate burden in m<sup>5</sup>C regulator genes. DNMT3B gene mutation was generally higher expression level among various cancers. Here, we sequenced the m<sup>5</sup>C 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 [61], including amplification and deletion of oncogenes, which

may significantly increase the risk of cancer [62]. In this research, the DNMT3B, ALYREF, and NSUN5 showed extensive CNV amplification. In contrast, CNVs such as TET2 and NSUN4 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 m<sup>5</sup>C regulators gene expression and activity of the KEGG pathways [63]. We also investigate the gene expression perturbations of m<sup>5</sup>C regulators through 33 cancer types with parallel normal controls. The expression of ALYREF, DNMT1, NSUN2, and 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 m<sup>5</sup>C regulators' pathways are significantly essential for a wide range of biological processes. m<sup>5</sup>C regulators were also validated involved in malignant activities [64]. Recent studies have demonstrated that the m<sup>5</sup>C modification in pyruvate kinase muscle isozyme M2 was involved in bladder cancer proliferation and migration. M5C regulator Aly/REF export factor regulated pyruvate kinase muscle isozyme M2 promote the glucose metabolism of bladder cancer [64]. At the same time, the precise molecular modification mechanisms and cellular processes among pan-cancer need further study and deeper exploration for a better prognosis.

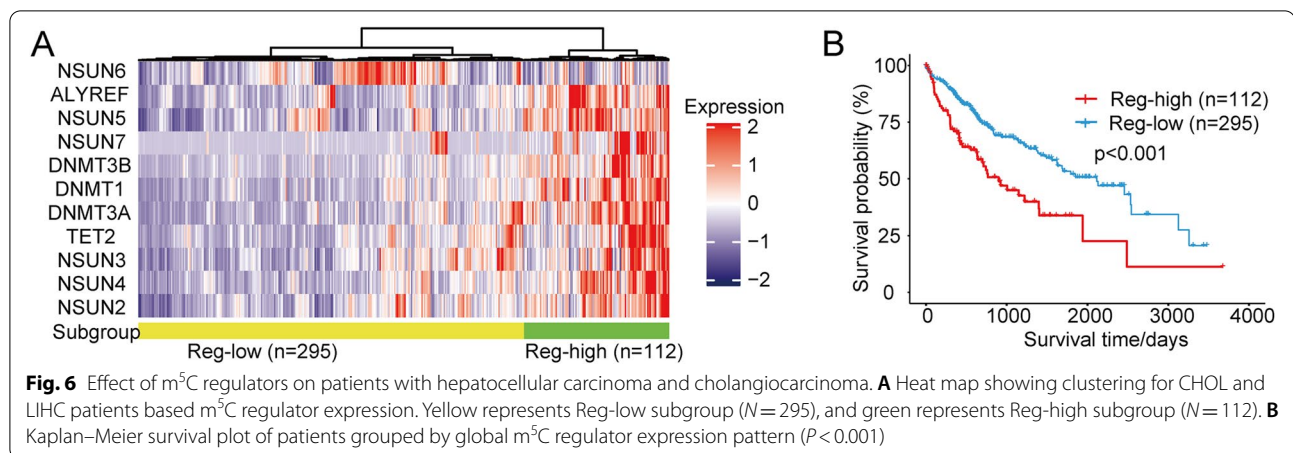
For a deeper exploration of the relationship between m<sup>5</sup>C regulators and their clinical outcomes, we describe





a comprehensive landscape of m<sup>5</sup>C 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 m<sup>5</sup>C regulators. To determine the effect of methylation-based molecular for earlier detection diagnostics in patients with several types of cancer, we systematically



analyzed the m<sup>5</sup>C regulators' pathway activities with the functional and clinically complication for estimating tumor development and progression with potential prognostic value.

## Conclusion

The m<sup>5</sup>C regulators were differently expressed and showed significantly different CNVs in pan-cancers, which also involved multiple oncogene pathways. In addition, m<sup>5</sup>C regulators also exhibited prognosis prediction value in pan-cancers. Therefore, our study provides a better understanding of the biology of m<sup>5</sup>C regulators in pan-cancers, indicating that m<sup>5</sup>C RNA methylation regulators have the potential to become novel biomarkers and therapeutic targets for various tumors.

## Abbreviations

m<sup>5</sup>C: 5-Methylcytosine; TCGA: The Cancer Genome Atlas; CNVs: Copy number variations; m<sup>6</sup>A: N<sup>6</sup>-methyladenine; m<sup>1</sup>A: N<sup>1</sup>-methyladenosine; NGS: Next-generation sequencing; FPKM: Fragments per kilobase of transcript per million; DEGs: Differently expression genes; TPM: Transcripts per million; GSVA: Gene set variation analysis; PCC: Pearson correlation coefficient; UCEC: Uterine corpus endometrial carcinoma; KEGG: Kyoto Encyclopedia of Genes and Genomes pathway; PPI: Protein–protein interaction; OS: Overall survival; PFI: Progression-free interval; DSS: Disease-specific survival; DFI: Disease-free interval; ACC: Adrenocortical carcinoma; KICH: Kidney chromophobe; KIRC: Kidney renal clear cell carcinoma; LGG: Brain lower grade glioma; LIHC: Liver hepatocellular carcinoma; HR: Hazard ratio; UVM: Uveal melanoma; CHOL: Cholangiocarcinoma.

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12957-021-02342-y>.

Additional file 1.

## Acknowledgements

Not applicable.

## Authors' contributions

YH and WG defined the research theme and discussed analyses, interpretation, and presentation. XY and MZ drafted the manuscript and analyzed the data. MZ helped with references collection. The authors read and approved the final manuscript.

## Funding

This work was supported by the National Natural Science Foundation of China (81902832) and the Youth Talent Lifting Project of Henan Province (2021HYTP059).

## Availability of data and materials

All of the data involved in this study are available in the public databases which are listed in the "Materials and methods" section.

## Declarations

### Ethics approval and consent to participate

This was not applicable to this manuscript.

### Consent for publication

Consent for publication was obtained from all participants.

### Competing interests

The authors declare that they have no competing interests.

### Author details

<sup>1</sup>Department of Hepatobiliary and Pancreatic Surgery, The First Affiliated Hospital of Zhengzhou University, No.1 Jianshe Road, Zhengzhou 450052, China. <sup>2</sup>Key Laboratory of Hepatobiliary and Pancreatic Surgery and Digestive Organ Transplantation of Henan Province, The First Affiliated Hospital of Zhengzhou University, Zhengzhou 450052, China. <sup>3</sup>Open and Key Laboratory of Hepatobiliary & Pancreatic Surgery and Digestive Organ Transplantation At Henan Universities, Zhengzhou 450052, China. <sup>4</sup>Henan Key Laboratory of Digestive Organ Transplantation, Zhengzhou 450052, China.

Received: 23 April 2021 Accepted: 21 July 2021

Published online: 29 July 2021

## References

- Cortes J, Perez-García JM, Llombart-Cussac A, Curigliano G, El Saghir NS, Cardoso F, Barrios CH, Wagle S, Roman J, Harbeck N, et al. Enhancing global access to cancer medicines. *CA Cancer J Clin*. 2020;70:105–24.
- van der Pol Y, Mouliere F. Toward the early detection of cancer by decoding the epigenetic and environmental fingerprints of cell-free DNA. *Cancer Cell*. 2019;36:350–68.

3. Hu T, Wolfram J, Srivastava S. Extracellular vesicles in cancer detection: hopes and hypes. *Trends Cancer*. 2020;S2405–8033:30257–30250.
4. Pilié PG, Tang C, Mills GB, Yap TA. State-of-the-art strategies for targeting the DNA damage response in cancer. *Nat Rev Clin Oncol*. 2019;16:81–104.
5. Mateo J, Lord CJ, Serra V, Tutt A, Balmaña J, Castroviejo-Bermejo M, Cruz C, Oaknin A, Kaye SB, de Bono JS. A decade of clinical development of PARP inhibitors in perspective. *Ann Oncol*. 2019;30:1437–47.
6. Okugawa Y, Grady WM, Goel A. Epigenetic alterations in colorectal cancer: emerging biomarkers. *Gastroenterology*. 2015;149:1204–1225.e1212.
7. Acunzo M, Romano G, Wernicke D, Croce CM. MicroRNA and cancer—a brief overview. *Adv Biol Regul*. 2015;57:1–9.
8. Shen Z, Lin L, Cao B, Zhou C, Hao W, Ye D. LZTS2 promoter hypermethylation: a potential biomarker for the diagnosis and prognosis of laryngeal squamous cell carcinoma. *World J Surg Oncol*. 2018;16:42.
9. Chan JJ, Tay Y. Noncoding RNA:RNA regulatory networks in cancer. *Int J Mol Sci*. 2018;19:1310.
10. Anastas JN, Moon RT. WNT signalling pathways as therapeutic targets in cancer. *Nat Rev Cancer*. 2013;13:11–26.
11. Vicente-Dueñas C, Hauer J, Cobaleda C, Borkhardt A, Sánchez-García I. Epigenetic priming in cancer initiation. *Trends Cancer*. 2018;4:408–17.
12. Zhao H, Wang S, Song C, Zha Y, Li L. The prognostic value of MGMT promoter status by pyrosequencing assay for glioblastoma patients' survival: a meta-analysis. *World J Surg Oncol*. 2016;14:261.
13. Shi H, Chai P, Jia R, Fan X. Novel insight into the regulatory roles of diverse RNA modifications: re-defining the bridge between transcription and translation. *Mol Cancer*. 2020;19:78.
14. Liu RJ, Long T, Li J, Li H, Wang ED. Structural basis for substrate binding and catalytic mechanism of a human RNA:m5C methyltransferase NSun6. *Nucleic Acids Res*. 2017;45:6684–97.
15. Lei L, Jiang Z, Zhang G, Cheng Q, Lu H. MGMT promoter methylation and 1p/19q co-deletion of surgically resected pulmonary carcinoid and large-cell neuroendocrine carcinoma. *World J Surg Oncol*. 2018;16:110.
16. Dixit R, Singh G, Pandey M, Basu S, Bhartiya SK, Singh KK, Shukla VK. Association of methylenetetrahydrofolate reductase gene polymorphism (MTHFR) in patients with gallbladder cancer. *J Gastrointest Cancer*. 2016;47:55–60.
17. Wu J, Zhang Y, Li M. Identification of methylation markers and differentially expressed genes with prognostic value in breast cancer. *J Comput Biol*. 2019;26:1394–408.
18. García-Vílchez R, Sevilla A, Blanco S. Post-transcriptional regulation by cytosine-5 methylation of RNA. *Biochim Biophys Acta Gene Regul Mech*. 2019;1862:240–52.
19. Frye M, Blanco S. Post-transcriptional modifications in development and stem cells. *Development*. 2016;143:3871–81.
20. Dong Z, Cui H. The emerging roles of RNA modifications in glioblastoma. *Cancers (Basel)*. 2020;12:736.
21. Trixl L, Lusser A. The dynamic RNA modification 5-methylcytosine and its emerging role as an epitranscriptomic mark. *Wiley Interdiscip Rev RNA*. 2019;10:e1510.
22. Catania J, Fairweather DS. DNA methylation and cellular ageing. *Mutat Res*. 1991;256:283–93.
23. Bohnsack KE, Höbartner C, Bohnsack MT. Eukaryotic 5-methylcytosine (m5C) RNA methyltransferases: mechanisms, cellular functions, and links to disease. *Genes (Basel)*. 2019;10:102.
24. Yang T, Low JJA, Woon EY. A general strategy exploiting m5C duplex-remodelling effect for selective detection of RNA and DNA m5C methyltransferase activity in cells. *Nucleic Acids Res*. 2020;48:e5.
25. Yang X, Yang Y, Sun BF, Chen YS, Xu JW, Lai WY, Li A, Wang X, Bhattarai DP, Xiao W, et al. 5-methylcytosine promotes mRNA export - NSUN2 as the methyltransferase and ALYREF as an m(5)C reader. *Cell Res*. 2017;27:606–25.
26. Dou L, Li X, Ding H, Xu L, Xiang H. Prediction of m5C Modifications in RNA sequences by combining multiple sequence features. *Mol Ther Nucleic Acids*. 2020;21:332–42.
27. Gama-Sosa MA, Slagel VA, Trewhin RW, Oxenhandler R, Kuo KC, Gehrke CW, Ehrlich M. The 5-methylcytosine content of DNA from human tumors. *Nucleic Acids Res*. 1983;11:6883–94.
28. Wu H, Zhang Y. Mechanisms and functions of Tet protein-mediated 5-methylcytosine oxidation. *Genes Dev*. 2011;25:2436–52.
29. Reid R, Greene PJ, Santi DV. Exposition of a family of RNA m(5)C methyltransferases from searching genomic and proteomic sequences. *Nucleic Acids Res*. 1999;27:3138–45.
30. Hua S, Quan Y, Zhan M, Liao H, Li Y, Lu L. miR-125b-5p inhibits cell proliferation, migration, and invasion in hepatocellular carcinoma via targeting TXNRD1. *Cancer Cell Int*. 2019;19:203.
31. Zhou Q, Zhou LQ, Li SH, Yuan YW, Liu L, Wang JL, Wu DZ, Wu Y, Xin L. Identification of subtype-specific genes signature by WGCNA for prognostic prediction in diffuse type gastric cancer. *Aging (Albany NY)*. 2020;12:17418–35.
32. Szklarczyk D, Morris JH, Cook H, Kuhn M, Wyder S, Simonovic M, Santos A, Doncheva NT, Roth A, Bork P, et al. The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. *Nucleic Acids Res*. 2017;45:D362–d368.
33. Zhou H, Chen L, Lei Y, Li T, Li H, Cheng X. Integrated analysis of tumor mutation burden and immune infiltrates in endometrial cancer. *Curr Probl Cancer*. 2021;45(2):100660.
34. Du Z, Ma K, Sun X, Li A, Wang H, Zhang L, Lin F, Feng X, Song J. Methylation of RASSF1A gene promoter and the correlation with DNMT1 expression that may contribute to esophageal squamous cell carcinoma. *World J Surg Oncol*. 2015;13:141.
35. Barabási AL, Gulbahce N, Loscalzo J. Network medicine: a network-based approach to human disease. *Nat Rev Genet*. 2011;12:56–68.
36. Panneerross S, Eedunuri VK, Yadav P, Tumulilina S, Rajamanickam S, Viswanadhappalli S, Abdelfattah N, Onyeagucha BC, Cui X, Lai Z, et al. Cross-talk among writers, readers, and erasers of m(6)A regulates cancer growth and progression. *Sci Adv*. 2018;4:eaar8263.
37. He Y, Yu X, Li J, Zhang Q, Zheng Q, Guo W. Role of m(5)C-related regulatory genes in the diagnosis and prognosis of hepatocellular carcinoma. *Am J Transl Res*. 2020;12:912–22.
38. Abdel-Hafiz HA, Horwitz KB. Role of epigenetic modifications in luminal breast cancer. *Epigenomics*. 2015;7:847–62.
39. Costa-Pinheiro P, Montezuma D, Henrique R, Jerónimo C. Diagnostic and prognostic epigenetic biomarkers in cancer. *Epigenomics*. 2015;7:1003–15.
40. Wang Y, Zhang R, Wu D, Lu Z, Sun W, Cai Y, Wang C, Jin J. Epigenetic change in kidney tumor: downregulation of histone acetyltransferase MYST1 in human renal cell carcinoma. *J Exp Clin Cancer Res*. 2013;32:8.
41. Cheng Y, He C, Wang M, Ma X, Mo F, Yang S, Han J, Wei X. Targeting epigenetic regulators for cancer therapy: mechanisms and advances in clinical trials. *Signal Transduct Target Ther*. 2019;4:62.
42. Wouters BJ, Delwel R. Epigenetics and approaches to targeted epigenetic therapy in acute myeloid leukemia. *Blood*. 2016;127:42–52.
43. Iniguez AB, Alexe G, Wang EJ, Roti G, Patel S, Chen L, Kitara S, Conway A, Robichaud AL, Stolte B, et al. Resistance to epigenetic-targeted therapy engenders tumor cell vulnerabilities associated with enhancer remodeling. *Cancer Cell*. 2018;34:922–938.e927.
44. Atrian F, Lelièvre SA. Mining the epigenetic landscape of tissue polarity in search of new targets for cancer therapy. *Epigenomics*. 2015;7:1313–25.
45. Hashizume R. Epigenetic targeted therapy for diffuse intrinsic pontine glioma. *Neurol Med Chir (Tokyo)*. 2017;57:331–42.
46. Estibariz I, Overmann A, Ailloud F, Krebs J, Josenhans C, Suerbaum S. The core genome m5C methyltransferase JHP1050 (M.Hpy99III) plays an important role in orchestrating gene expression in *Helicobacter pylori*. *Nucleic Acids Res*. 2019;47:2336–48.
47. Saghafinia S, Mina M, Riggi N, Hanahan D, Ciriello G. Pan-cancer landscape of aberrant DNA methylation across human tumors. *Cell Rep*. 2018;25:1066–1080.e1068.
48. Zhang Q, He Y, Luo N, Patel SJ, Han Y, Gao R, Modak M, Carotta S, Haslinger C, Kind D, et al. Landscape and dynamics of single immune cells in hepatocellular carcinoma. *Cell*. 2019;179:829–845.e820.
49. Tse JWT, Jenkins LJ, Chionh F, Mariadason JM. Aberrant DNA methylation in colorectal cancer: what should we target? *Trends Cancer*. 2017;3:698–712.
50. Chen L, Wang D. Identification of potential CpG sites for oral squamous cell carcinoma diagnosis via integrated analysis of DNA methylation and gene expression. *World J Surg Oncol*. 2021;19:16.
51. Yoshinobu Takahashi HN. Keishi Makino: Prognostic value of isocitrate dehydrogenase 1, O6-methylguanine-DNA methyltransferase promoter methylation, and 1p19q co-deletion in Japanese malignant glioma patients. *World J Surg Oncol*. 2013;11:284.

52. Sibbritt T, Patel HR, Preiss T. Mapping and significance of the mRNA methylome. *Wiley Interdiscip Rev RNA*. 2013;4:397–422.
53. Popis MC, Blanco S, Frye M. Posttranscriptional methylation of transfer and ribosomal RNA in stress response pathways, cell differentiation, and cancer. *Curr Opin Oncol*. 2016;28:65–71.
54. Lu L, Zhu G, Zeng H, Xu Q, Holzmann K. High tRNA transferase NSUN2 gene expression is associated with poor prognosis in head and neck squamous carcinoma. *Cancer Invest*. 2018;36:246–53.
55. Cui X, Liang Z, Shen L, Zhang Q, Bao S, Geng Y, Zhang B, Leo V, Vardy LA, Lu T, et al. 5-Methylcytosine RNA methylation in *Arabidopsis thaliana*. *Mol Plant*. 2017;10:1387–99.
56. Haag S, Warda AS, Kretschmer J, Günnigmann MA, Höbartner C, Bohnsack MT. NSUN6 is a human RNA methyltransferase that catalyzes formation of m5C72 in specific tRNAs. *RNA*. 2015;21:1532–43.
57. Van Haute L, Lee SY, McCann BJ, Powell CA, Bansal D, Vasiliauskaitė L, Garone C, Shin S, Kim JS, Frye M, et al. NSUN2 introduces 5-methylcytosines in mammalian mitochondrial tRNAs. *Nucleic Acids Res*. 2019;47:8720–33.
58. Janin M, Ortiz-Barahona V, de Moura MC, Martínez-Cardús A, Llinàs-Arias P, Soler M, Nachmani D, Pelletier J, Schumann U, Calleja-Cervantes ME, et al. Epigenetic loss of RNA-methyltransferase NSUN5 in glioma targets ribosomes to drive a stress adaptive translational program. *Acta Neuropathol*. 2019;138:1053–74.
59. Li P, Xu Y, Zhang Q, Li Y, Jia W, Wang X, Xie Z, Liu J, Zhao D, Shao M, et al. Evaluating the role of RAD52 and its interactors as novel potential molecular targets for hepatocellular carcinoma. *Cancer Cell Int*. 2019;19:279.
60. Pleasance ED, Cheetham RK, Stephens PJ, McBride DJ, Humphray SJ, Greenman CD, Varela I, Lin ML, Ordóñez GR, Bignell GR, et al. A comprehensive catalogue of somatic mutations from a human cancer genome. *Nature*. 2010;463:191–6.
61. Krepschi AC, Pearson PL, Rosenberg C. Germline copy number variations and cancer predisposition. *Future Oncol*. 2012;8:441–50.
62. Zhao X, Liu X, Zhang A, Chen H, Huo Q, Li W, Ye R, Chen Z, Liang L, Liu QA, et al. The correlation of copy number variations with longevity in a genome-wide association study of Han Chinese. *Aging (Albany NY)*. 2018;10:1206–22.
63. Zhou Y, Wang Y, Zhou W, Chen T, Wu Q, Chutturghoon VK, Lin B, Geng L, Yang Z, Zhou L, Zheng S. YAP promotes multi-drug resistance and inhibits autophagy-related cell death in hepatocellular carcinoma via the RAC1-ROS-mTOR pathway. *Cancer Cell Int*. 2019;19:179.
64. Wang JZ, Zhu W, Han J, Yang X, Zhou R, Lu HC, Yu H, Yuan WB, Li PC, Tao J, et al. The role of the HIF-1alpha/ALYREF/PKM2 axis in glycolysis and tumorigenesis of bladder cancer. *Cancer Commun (Lond)*. 2021;41(7):560–75.

### Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

