Pharmacogenomics and Personalized Medicine

Dovepress

open access to scientific and medical research

Open Access Full Text Article

ORIGINAL RESEARCH

Genome-Wide Identification of m6A-Associated Single-Nucleotide Polymorphisms in Colorectal Cancer

Hongying Zhao[1]
Jinying Jiang[1]
Mingshan Wang[2]
Zixue Xuan[1,3]

[1]Clinical Pharmacy Center, Department of Pharmacy, Zhejiang Provincial People's Hospital, Affiliated People's Hospital, Hangzhou Medical College, Hangzhou, Zhejiang, People's Republic of China; [2]Departments of Infection Diseases, Zhejiang Provincial People's Hospital, Affiliated People's Hospital, Hangzhou Medical College, Hangzhou, Zhejiang, People's Republic of China; [3]Key Laboratory of Endocrine Gland Diseases of Zhejiang Province, Zhejiang Provincial People's Hospital, Affiliated People's Hospital, Hangzhou Medical College, Hangzhou, Zhejiang, People's Republic of China

Background: N6-methyladenosine (m6A)-associated single-nucleotide polymorphisms (SNPs) play important roles in cancers, with previous research suggesting potential associations between m6A-SNPs and cancer. However, the relationship between the genetic determinants of m6A modification and colorectal cancer (CRC) remains unclear.

Methods: An integrative method combining raw data and summary statistics of genome-wide association studies with expression quantitative trait loci (eQTL) and differential expression data was applied to screen potential candidate CRC-associated m6A-SNPs.

Results: A total of 402 m6A-SNPs were identified as being associated with CRC ($p < 0.001$), with 98 showing eQTL signals. In particular, three genes were found to harbor CRC-associated m6A-SNPs: rs178184 in *NOVA1*, rs35782901 in *HTR4*, and rs60571683 in *SLCO1B3*. These genes were differentially expressed in at least one publicly available dataset ($p < 0.05$), with *NOVA1* ($p = 3.41\times10^{-11}$) and *HTR4* ($p = 5.56\times10^{-7}$) being significantly downregulated in CRC (dataset: GSE89076), and *SLCO1B3* was significantly overexpressed (datasets: GSE32323 [$p = 3.27\times10^{-5}$], GSE21510 [$p = 1.09\times10^{-6}$], and GSE89076 [$p = 7.63\times10^{-6}$]).

Conclusion: This study identified three m6A-SNPs (rs178184, rs35782901, and rs60571683) that may be associated with CRC. However, the lack of analysis of primary CRC samples in order to further elucidate the underlying pathogenesis is a major limitation of this study. Future investigations are needed to validate these CRC-associated m6A-SNPs and explore the m6A-mediated pathogenic mechanism in CRC.

Keywords: colorectal cancer, genome-wide association study, N6-methyladenosine, single-nucleotide polymorphism

Background

Colorectal cancer (CRC) is the third most deadly cancer worldwide, after breast and lung cancer.[1] Etiological studies have shown that genetic factors play an important role in CRC, and several genes were associated with CRC pathogenic, including *MLH1, MSH2, PMS2, EPCAM, APC*, and *MUTYH*.[2] However, these common genetic variants account for only 6% of CRC cases, which suggests a higher genetic complexity of CRC beyond these genes. Previous genome-wide association studies (GWAS) have identified several single-nucleotide polymorphisms (SNPs) related to CRC.[3,4] For instance, one study reported susceptibility variants at 8q23.3 (rs16892766) and 8q24.21 (rs6983267) that were associated with advanced-stage tumors and familial history of CRC.[4]

Correspondence: Mingshan Wang; Zixue Xuan
Email wang1802@hotmail.com; xuanzixue0222@163.com

Received: 5 April 2021
Accepted: 30 June 2021
Published: 17 July 2021

N6-methyladenosine (m6A) modification is a critical regulator of multiple cytopathological processes, including nuclear export, translation, splicing, and stability of mRNAs.[5] Emerging evidence has shown that m6A modification plays a critical role in CRC; for example, METTL14 regulates m6A-dependent primary miR-375 processing to inhibit CRC progression,[6] whereas METTL3 facilitates the progression of CRC via m6A-IGF2BP2/3-dependent mechanisms.[7,8] It was also confirmed that disease-associated genetic variants can influence m6A methylation by changing the RNA sequences or key flanking nucleotides of its target sites, suggesting that m6A-SNPs might affect mRNA stability, thereby contributing for the development of diseases.[9,10]

Recently, m6A-SNPs have attracted considerable attention, and a number of prioritized SNPs have been identified by integrative analysis of cancer-related GWAS summary data, including in pancreatic, bladder, and gastric cancers.[11–13] The aim of the present study was to shed light and explore the potential contribution of m6A-SNPs in CRC pathogenesis.

Methods

Identification of CRC-Associated m6A-SNPs

The m6AVar database can provide host variants associated with m6A and may boost further mechanistic studies of genetic variants affecting m6A modifications.[14] Therefore, CRC-associated m6A-SNPs were identified by integrating the data from GWAS and the m6Avar database. This study used publicly available GWAS data for CRC. The binary traits of GWAS datasets comprised 4562 CRC patients and 382,756 controls, relevant information was available from the link: ftp://share.sph.umich.edu/UKBB_SAIGE_HRC/PheCode_153_SAIGE_MACge20.txt.vcf.gz.The m6A-SNP list was downloaded from the m6Avar database (http://rmvar.renlab.org/download.html). Potential CRC-associated m6A-SNPs were identified through comparison of the SNPs in the GWAS datasets[15] and the list of m6A-SNPs,[16] which reached the genome-wide suggestive threshold ($p < 0.001$). Because SNP loci of genes encoding m6A regulators, including *METTL3, METTL14, METTL16, WTAP, VIRMA, RBM15, FTO, ALKBH5, YTHDC1, YTHDC2, YTHDF1, YTHDF2, YTHDF3, IGF2BP1, IGF2BP2, IGF2BP3, HNRNPA2B1*, and *EIF3A* were absent in the m6Var database, so these m6A-SNPs were identified, which reached the genome-wide suggestive threshold ($p < 0.05$).

Expression Quantitative Trait Loci (eQTLs) Analysis of CRC-Associated m6A-SNPs

Cis-acting eQTL (cis-eQTL) analysis can be used to evaluate the potential function of the m6A-SNPs that showed cis-eQTL signals in transcription regulation, such as altering protein binding, changing motifs, and affecting deoxyribonuclease.[17] Therefore, cis-eQTL analysis was performed to investigate which CRC-associated m6A-SNPs could affect the RNA modification using the HaploReg browser (https://www.encodeproject.org/software/haploreg/).

Differential Expression Analysis

The corresponding genes of the identified eQTL m6A-SNPs were further evaluated according to differential expression among CRC patients and controls.[18] Hence, three microarray gene expression CRC datasets (GSE89076, GSE21510, and GSE32323) publicly available in the Gene Expression Omnibus database (http://www.ncbi.nlm.nih.gov/geo) were used. Then, we used GEO2R online tool (https://www.ncbi.nlm.nih.gov/geo/geo2r/) to examine whether the CRC-associated genes (m6A-SNP with cis-eQTL) are differentially expressed between CRC and controls. A significance level of $p < 0.05$ was used for differential expression analysis.

Results

CRC-Associated m6A-SNPs

A total of 402 m6A-SNPs were extracted from the raw GWAS data by comparing the SNPs identified from the GWAS datasets and the m6A-SNPs from the m6AVar database (Figure 1). Next, the SNPs within genes encoding m6A regulators, including *METTL3, METTL14, METTL16, WTAP, VIRMA, RBM15, FTO, ALKBH5, YTHDC1, YTHDC2, YTHDF1, YTHDF2, YTHDF3, IGF2BP1, IGF2BP2, IGF2BP3, HNRNPA2B1*, and *EIF3A*, were collected as these SNP loci were absent in the m6Var database. This analysis revealed that rs112126539 (in *IGF2BP1*) could be a CRC-associated m6A-SNP ($p = 0.018$) (Table 1).

eQTL Analysis

eQTL analysis was performed on the 402 m6A-SNPs, which revealed that 98 of these m6A-SNPs had eQTL signals, with 76 and 22 m6A-SNPs having lost and gained modification functions, respectively (Table S1). Moreover,

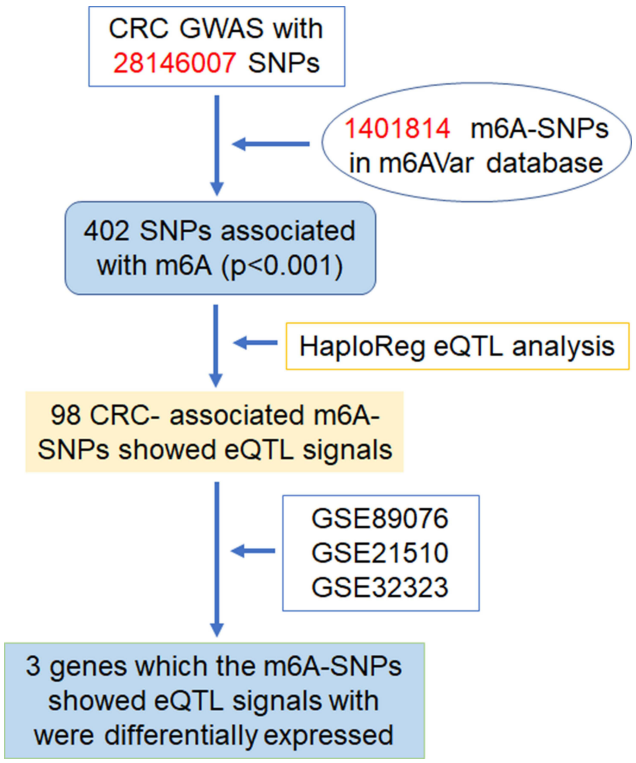

Figure 1 Flow chart of study design and analysis.

the rs2037844 polymorphism showed a stronger eQTL signals ($p = 3.60\times10^{-8}$) as compared to the other SNPs (the second was rs2957748; $p = 1.17\times10^{-4}$) (Figure 2A). Additionally, 98 CRC-associated m6A-SNPs showed eQTL signals displaying a unique distribution pattern, with most found in the intron and exon regions, and with few being in the coding sequences, 3′/5′-untranslated regions, or stop codon regions (Figure 2B).

Differential Expression Analysis

Three microarray datasets containing gene expression signals in CRC were analyzed: GSE89076, comprising paired normal and tumor tissues from 275 CRC patients; GSE21510, comprising 148 microarray datasets obtained from CRC tissue samples; and GSE32323, comprising 17 pairs of cancer and non-cancerous tissues from CRC patients.

Among the 98 CRC-associated m6A-SNPs that showed eQTL signals, three genes harboring m6A-SNPs (rs178184 in neuro-oncological ventral antigen 1 [*NOVA1*], rs35782901 in hydroxytryptamine receptor 4 (*HTR4*), and rs60571683 in solute carrier organic anion transporter family member 1B3 (*SLCO1B3*) were found to be differentially expressed in at least one CRC dataset ($p < 0.05$). *NOVA1* ($p = 3.41\times10^{-11}$) and *HTR4* ($p = 5.56\times10^{-7}$) were significantly downregulated in GSE89076, whereas *SLCO1B3* was significantly overexpressed in GSE32323 ($p = 3.27\times10^{-5}$), GSE21510 ($p = 1.09\times10^{-6}$), and GSE89076 ($p = 7.63\times10^{-6}$) (Figure 3 and Table 2). Therefore, these findings suggest that these three m6A-SNPs (rs178184, rs35782901, and rs60571683) may alter the expression of *NOVA1, HTR4*, and *SLCO1B3*, and subsequently impact on CRC pathogenesis.

Discussion

Given the evidence that m6A contributes to CRC,[19,20] the relationship between candidate SNPs in 20 m6A regulators and the risk of CRC was also investigated.[18] However, a broader analysis method, integrating independent GWAS summary statistics with eQTL data to identify potential functional genetic variants of CRC-associated m6A-SNPs, has more research value.[21,22] Therefore, we used this method combining raw data and summary statistics of genome-wide association studies with eQTL and differential expression data to screen potential candidate CRC-associated m6A-SNPs.

In this study, three CRC-associated m6A-SNPs, specifically rs178184 in *NOVA1*, rs35782901 in *HTR4*, and rs60571683 in *SLCO1B3*, were found to be associated with altered gene expression in CRC. As we known, m6A-SNPs not only affect gene expression level, also involve in disease progression by influencing the ratio between different RNA isoforms and the translation level of their protein products. However, we only explored the potential effect of m6A-SNPs on gene expression level in this study, because other data are currently publicly limited.

The role of *NOVA1, HTR4*, and *SLCO1B3* polymorphisms has not been previously explored in CRC, and genetic variants found near m6A sites have more possibilities to influence the pathogenesis of CRC. We found the m6A-SNP rs178184 is located in the *NOVA1* coding gene on chromosome 14, which is essential for growth and invasion-related signaling in cancer

Table 1 Rs112126539 (in *IGF2BP1*) Could Be a CRC-Associated m6A-SNP

SNP ID	Chr	Position	m6A_ID	Gene	Confidence_level	Gene_region	Modification_function	Pvalue
rs112126539	17	49050005	RMVar_ID_1070422	*IGF2BP1*	Prediction:(Low)	3′UTR	Loss	0.018

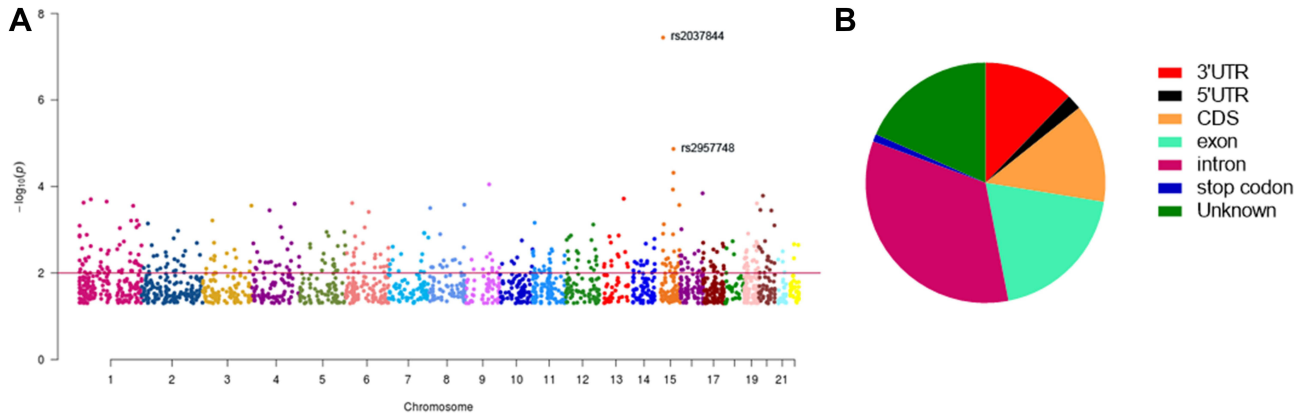

Figure 2 Manhattan plot of genome-wide identified CRC-associated m6A-SNPs (**A**) and 98 CRC-associated m6A-SNPs showed eQTL signals displaying a unique distribution pattern (**B**).

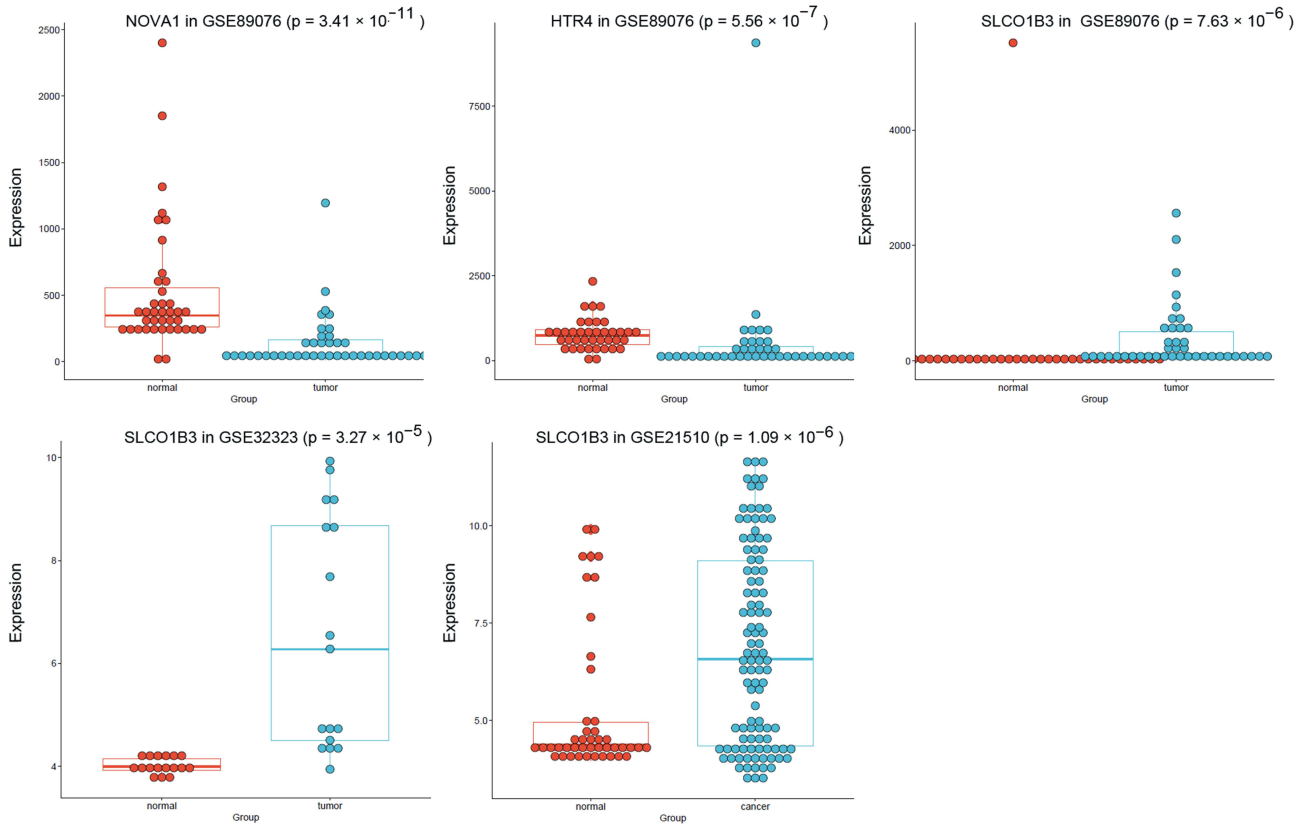

Figure 3 Expression levels of selected genes were displayed among controls and CRC of GSE32323, GSE21510, and GSE89076 datasets.

cells and is a master regulator of alternative splicing. *NOVA1* is a crucial regulator of alternative splicing in pancreatic beta cells,[23] acts as an oncogene in the development of melanoma,[24] and regulates hTERT splicing and promotes cell growth in non-small cell lung cancer.[25] In CRC, Hong et al showed that *NOVA1* is involved in cancer progression, suggesting that *NOVA1* might be a valuable prognostic biomarker and a target for CRC treatment. The m6A-SNP rs35782901 is located in the *HTR4* coding gene on chromosome 5. Studies have shown that *HTR4* variants are associated with chronic obstructive pulmonary disease.[26] Moreover, *5-HTR4* was found predominantly in high-grade tumors, with *5-HTR4* inhibition reducing the proliferative activity of androgen-independent prostate cancer cell lines.[27] The m6A-SNP

Table 2 Three Genes of the m6A-SNPs (rs178184, rs35782901, rs60571683) Were Differentially Expressed Between Controls and CRC in at Least One Data Set

SNP ID	Chr	Position	m6A_ID	Gene	DEG	eQTL	Confidence_Level	Gene_Region	Modification_Function
rs178184	14	26509987	RMVar_ID_684700	*NOVA1*	Yes	Yes	m6A-Label-seq:(High)	Intron	Loss
rs35782901	5	148629028	RMVar_ID_1344219	*HTR4*	Yes	Yes	Prediction:(Low)	Intron	Gain
rs60571683	12	20916115	RMVar_ID_945462	*SLCO1B3*	Yes	Yes	Prediction:(Low)	CDS	Loss

rs60571683 is located in the *SLCO1B3* coding gene on chromosome 12. *SLCO1B3* is a functional transporter that is normally expressed in the liver but that was also detected in different cancers and reported to be involved in cancer.[28] For example, *SLCO1B3* inhibits tumorigenesis and progression of breast cancer.[29] Genotypic variants of *SLCO1B3* affect docetaxel pharmacokinetics.[30] In addition, the *SLCO1B3* GG/AA haplotype is associated with impaired testosterone transport and improved survival in patients with prostate cancer.[31]

Conclusion

This study reports the first comprehensive analysis of GWAS raw data and summary statistics combined with eQTL and differential gene expression data to identify candidate CRC-associated m6A-SNPs. Three m6A-SNPs (rs178184, rs35782901, and rs60571683) were found to be potentially associated with CRC, as demonstrated by their high eQTL signals and altered the expression of their coding genes (*NOVA1, HTR4*, and *SLCO1B3*, respectively). However, this study has certain limitations. For example, the identified m6A-SNPs have not been validated in tissue samples. Despite these limitations, the finding will provide the opportunity of further research to elucidate the practical impact of m6A-SNPs on the pathogenesis of CRC.

Abbreviations

cis-eQTL, cis-acting eQTL; CRC, colorectal cancer; eQTL, expression quantitative trait loci; GWAS, genome-wide association study; *HTR4*, hydroxytryptamine receptor 4; m6A, N6-methyladenosine; *NOVA1*, neuro-oncological ventral antigen 1; *SLCO1B3*, solute carrier organic anion transporter family member 1B3; SNP, single-nucleotide polymorphism.

Data Sharing Statement

The GWAS dataset used in this study was downloaded from ftp://share.sph.umich.edu/UKBB_SAIGE_HRC/PheCode_153_SAIGE_MACge20.txt.vcf.gz. The expression datasets supporting the conclusions of this article are publicly available in the Gene Expression Omnibus repository (http://www.ncbi.nlm.nih.gov/gds/GSE89076; http://www.ncbi.nlm.nih.gov/gds/GSE21510; and http://www.ncbi.nlm.nih.gov/gds/GSE32323).

Acknowledgments

This study was supported by the Natural Science Foundation of Zhejiang Province (Grant Number LYY21H310008; LYY19H310009).

Author Contributions

All authors made a significant contribution to the work reported, such as the conception, study design, execution, acquisition of data, analysis and interpretation; took part in drafting, revising or critically reviewing the article; gave final approval of the version to be published; have agreed on the journal to which the article has been submitted; and agree to be accountable for all aspects of the work.

Disclosure

The authors report no conflicts of interest in this work.

References

1. Sung H, Ferlay J, Siegel RL, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2021;71:209–249. doi:10.3322/caac.21660
2. Yurgelun MB, Kulke MH, Fuchs CS, et al. Cancer susceptibility gene mutations in individuals with colorectal cancer. *J Clin Oncol*. 2017;35(10):1086–1095. doi:10.1200/JCO.2016.71.0012
3. Bian S, Hou Y, Zhou X, et al. Single-cell multiomics sequencing and analyses of human colorectal cancer. *Science*. 2018;362(6418):1060–1063. doi:10.1126/science.aao3791
4. Abulí A, Bessa X, González JR, et al. Susceptibility genetic variants associated with colorectal cancer risk correlate with cancer phenotype. *Gastroenterology*. 2010;139(3):786. doi:10.1053/j.gastro.2010.05.072
5. He L, Li H, Wu A, Peng Y, Shu G, Yin G. Functions of N6-methyladenosine and its role in cancer. *Mol Cancer*. 2019;18(1):176. doi:10.1186/s12943-019-1109-9
6. Chen X, Xu M, Xu X, et al. METTL14 suppresses CRC progression via regulating N6-methyladenosine-dependent primary miR-375 processing. *Mol Ther*. 2020;28(2):599–612. doi:10.1016/j.ymthe.2019.11.016
7. Li T, Hu PS, Zuo Z, et al. METTL3 facilitates tumor progression via an m(6) A-IGF2BP2-dependent mechanism in colorectal carcinoma. *Mol Cancer*. 2019;18(1):112. doi:10.1186/s12943-019-1038-7
8. Shen C, Xuan B, Yan T, et al. m(6)A-dependent glycolysis enhances colorectal cancer progression. *Mol Cancer*. 2020;19(1):72. doi:10.1186/s12943-020-01190-w

9. Lin W, Xu H, Wu Y, Wang J, Yuan Q. In silico genome-wide identification of m6A-associated SNPs as potential functional variants for periodontitis. *J Cell Physiol*. 2020;235(2):900–908. doi:10.1002/jcp.29005
10. Liu Z, Zhang J. Most m6A RNA modifications in protein-coding regions are evolutionarily unconserved and likely nonfunctional. *Mol Biol Evol*. 2018;35(3):666–675. doi:10.1093/molbev/msx320
11. Ying P, Li Y, Yang N, et al. Identification of genetic variants in m(6)A modification genes associated with pancreatic cancer risk in the Chinese population. *Arch Toxicol*. 2021;95(3):1117–1128. doi:10.1007/s00204-021-02978-5
12. Liu H, Gu J, Jin Y, et al. Genetic variants in N6-methyladenosine are associated with bladder cancer risk in the Chinese population. *Arch Toxicol*. 2021;95(1):299–309. doi:10.1007/s00204-020-02911-2
13. Wang X, Guan D, Wang D, et al. Genetic variants in m(6)A regulators are associated with gastric cancer risk. *Arch Toxicol*. 2021;95(3):1081–1088. doi:10.1007/s00204-020-02958-1
14. Luo X, Li H, Liang J, et al. RMVar: an updated database of functional variants involved in RNA modifications. *Nucleic Acids Res*. 2021;49(D1):D1405–d1412. doi:10.1093/nar/gkaa811
15. Zhou W, Nielsen JB, Fritsche LG, et al. Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. *Nat Genet*. 2018;50(9):1335–1341. doi:10.1038/s41588-018-0184-y
16. Lin W, Xu H, Yuan Q, Zhang S. Integrative genomic analysis predicts regulatory role of N (6)-methyladenosine-associated SNPs for adiposity. *Front Cell Dev Biol*. 2020;8:551. doi:10.3389/fcell.2020.00551
17. Larson NB, McDonnell S, French AJ, et al. Comprehensively evaluating cis-regulatory variation in the human prostate transcriptome by using gene-level allele-specific expression. *Am J Hum Genet*. 2015;96(6):869–882. doi:10.1016/j.ajhg.2015.04.015
18. Wang X, Fu X, Zhang J, Xiong C, Zhang S, Lv Y. Identification and validation of m(6)A RNA methylation regulators with clinical prognostic value in Papillary thyroid cancer. *Cancer Cell Int*. 2020;20:203. doi:10.1186/s12935-020-01283-y
19. Yang X, Zhang S, He C, et al. METTL14 suppresses proliferation and metastasis of colorectal cancer by down-regulating oncogenic long non-coding RNA XIST. *Mol Cancer*. 2020;19(1):46. doi:10.1186/s12943-020-1146-4
20. Li J, Liang L, Yang Y, Li X, Ma Y. N(6)-methyladenosine as a biological and clinical determinant in colorectal cancer: progression and future direction. *Theranostics*. 2021;11(6):2581–2593. doi:10.7150/thno.52366
21. Zhu R, Tian D, Zhao Y, Zhang C, Liu X. Genome-wide detection of m(6) A-associated genetic polymorphisms associated with ischemic stroke. *J Mol Neurosci*. 2021. doi:10.1007/s12031-021-01805-x
22. Zhang Z, Luo K, Zou Z, et al. Genetic analyses support the contribution of mRNA N(6)-methyladenosine (m(6)A) modification to human disease heritability. *Nat Genet*. 2020;52(9):939–949. doi:10.1038/s41588-020-0644-z
23. Villate O, Turatsinze JV, Mascali LG, et al. Nova1 is a master regulator of alternative splicing in pancreatic beta cells. *Nucleic Acids Res*. 2014;42(18):11818–11830. doi:10.1093/nar/gku861
24. Yu X, Zheng H, Chan MTV, Wu WKK. NOVA1 acts as an oncogene in melanoma via regulating FOXO3a expression. *J Cell Mol Med*. 2018;22(5):2622–2630. doi:10.1111/jcmm.13527
25. Ludlow AT, Wong MS, Robin JD, et al. NOVA1 regulates hTERT splicing and cell growth in non-small cell lung cancer. *Nat Commun*. 2018;9(1):3112. doi:10.1038/s41467-018-05582-x
26. Soler Artigas M, Wain LV, Repapi E, et al. Effect of five genetic variants associated with lung function on the risk of chronic obstructive lung disease, and their joint effects on lung function. *Am J Respir Crit Care Med*. 2011;184(7):786–795. doi:10.1164/rccm.201102-0192OC
27. Dizeyi N, Bjartell A, Hedlund P, Taskén KA, Gadaleanu V, Abrahamsson PA. Expression of serotonin receptors 2B and 4 in human prostate cancer tissue and effects of their antagonists on prostate cancer cell lines. *Eur Urol*. 2005;47(6):895–900. doi:10.1016/j.eururo.2005.02.006
28. Sekine S, Ogawa R, Ojima H, Kanai Y. Expression of SLCO1B3 is associated with intratumoral cholestasis and CTNNB1 mutations in hepatocellular carcinoma. *Cancer Sci*. 2011;102(9):1742–1747. doi:10.1111/j.1349-7006.2011.01990.x
29. Tang T, Wang G, Liu S, et al. Highly expressed SLCO1B3 inhibits the occurrence and development of breast cancer and can be used as a clinical indicator of prognosis. *Sci Rep*. 2021;11(1):631.
30. Chew SC, Sandanaraj E, Singh O, et al. Influence of SLCO1B3 haplotype-tag SNPs on docetaxel disposition in Chinese nasopharyngeal cancer patients. *Br J Clin Pharmacol*. 2012;73(4):606–618. doi:10.1111/j.1365-2125.2011.04123.x
31. Hamada A, Sissung T, Price DK, et al. Effect of SLCO1B3 haplotype on testosterone transport and clinical outcome in caucasian patients with androgen-independent prostatic cancer. *Clin Cancer Res*. 2008;14(11):3312–3318. doi:10.1158/1078-0432.CCR-07-4118