

口腔鳞状细胞癌 m⁶A 相关差异表达基因的生物信息学分析

徐 莉, 南 茜

(华中科技大学同济医学院附属同济医院口腔医学中心, 武汉 430030)

【摘要】目的:筛选口腔鳞状细胞癌(oral squamous cell carcinoma, OSCC)m⁶A 相关的差异表达基因, 探寻相关信号通路, 为 OSCC 提供候选的诊断标记物和分子治疗靶点。**方法:**下载 TCGA 数据库中 OSCC 数据, 提取 m⁶A 相关基因的表达。R 软件筛选正常组织和肿瘤组织间的差异表达基因, 对 m⁶A 相关基因进行聚类分析与相关性分析。基因富集分析(gene set enrichment analysis, GSEA)研究差异基因的信号通路。**结果:**本研究筛选出 8 个差异表达基因($P < 0.05$), 包括 *HNRNPC*、*KIAA1429*、*METTL3*、*METTL14*、*RBM15*、*WTAP*、*YTHDF1* 和 *YTHDF2*, 差异基因在癌组织的表达量均高于正常组织($P < 0.05$)。m⁶A 相关基因之间存在复杂的相关性。GSEA 富集分析显示, 差异表达基因主要参与细胞周期、碱基切除修复、同源重组和剪接体等相关信号通路。**结论:***HNRNPC*、*KIAA1429*、*METTL3*、*METTL14*、*RBM15*、*WTAP*、*YTHDF1* 与 *YTHDF2* 可作为 OSCC 的候选诊断标记物和分子治疗靶点。

【关键词】口腔鳞状细胞癌; 差异表达基因; m⁶A 甲基化; 生物信息学

【中图分类号】R739.8

【文献标志码】A

【收稿日期】2020-05-11

Bioinformatics analysis of differentially expressed m⁶A associated genes of oral squamous cell carcinoma

Xu Li, Nan Xi

(Department of Stomatology, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology)

【Abstract】Objective: To identify differentially expressed m⁶A associated genes of oral squamous cell carcinoma (OSCC) and their related signaling pathways, providing candidate biomarkers for diagnosis and potential target genes for therapy of OSCC. **Methods:** After downloading data on OSCC from TCGA database, m⁶A related gene expression profiles were obtained. R software was used to select differentially expressed m⁶A genes between normal samples and tumor ones. Cluster analysis was performed for all m⁶A associated genes and the interactions among these genes were studied. Gene set enrichment analysis (GSEA) was done for signaling pathways of OSCC due to differentially expressed m⁶A genes. **Results:** Eight differentially expressed m⁶A genes were identified ($P < 0.05$), including *HNRNPC*, *KIAA1429*, *METTL3*, *METTL14*, *RBM15*, *WTAP*, *YTHDF1* and *YTHDF2*. They showed significantly higher expressions in tumor samples compared with normal ones ($P < 0.05$). There was a complex correlation between m⁶A related genes. GSEA revealed that the differentially expressed m⁶A genes were mainly involved in such signaling pathways as cell cycle, base excision repairing, homologous recombination and spliceosome. **Conclusion:** *HNRNPC*, *KIAA1429*, *METTL3*, *METTL14*, *RBM15*, *WTAP*, *YTHDF1* and *YTHDF2* are expected to be potential diagnostic markers and molecular therapeutic targets of OSCC patients.

【Key words】oral squamous cell carcinoma; differentially expressed genes; m⁶A methylation; bioinformatics

口腔鳞状细胞癌(oral squamous cell carcinoma, OSCC)是常见的头颈部恶性肿瘤, 在口腔癌中占90%

作者介绍: 徐 莉, Email: xlx1915@126.com,

研究方向: 口腔内科疾病及口腔肿瘤的研究和治疗。

优先出版: <https://kns.cnki.net/kcms/detail/50.1046.R.20201013.1355.006.html>
(2020-10-14)

以上^[1], 治疗方法主要包括手术、放疗、化疗和免疫治疗。对于复发、转移和晚期的 OSCC 患者, 治疗方法以放疗和化疗为主, 治疗效果不佳。OSCC 侵袭性强、易早期淋巴结转移, 影响治疗效果。早期诊断 OSCC, 对改善患者预后意义重大^[2]。目前, 研究作用

于 OSCC 发生和发展过程中的基因,以寻找诊断 OSCC 的基因表达标记物,是研究的热点,有利于阐明癌症的分子机制。

m⁶A 甲基化修饰是真核细胞中最常见且丰度最高的 mRNA 修饰方式,参与 RNA 转录、出细胞核、翻译与降解^[3-4],受甲基转移酶(KIAA1429、METTL3、METTL14、RBM15、WTAP 和 ZC3H13)、去甲基化酶(ALKBH5 和 FTO)及甲基化结合蛋白(HNRNPC、YTHDC1、YTHDC2、YTHDF1 和 YTHDF2)等调节因子的动态与可逆性调控。这些调节因子由 m⁶A 相关的基因所表达,可参与多种癌症的发生发展过程^[5-6],如乳腺癌和急性髓性白血病,可作为癌症诊断的候选分子标记物。近期研究发现,METTL3 与 YTHDF1 促进 OSCC 的发生^[7],其他 m⁶A 相关基因表达与 OSCC 的相关研究尚无报道,探讨 OSCC 组织中 m⁶A 相关基因的表达,以发现诊断 OSCC 的候选标记物,对临床和基础研究有重要意义。

生物信息学方法能用来筛选癌症的目标基因并分析其生物学功能^[8-9],但该方法在口腔鳞状细胞癌 m⁶A 相关基因研究中的应用较少。本研究通过该方法分析 TCGA 数据库中口腔鳞状细胞癌 level 3 的 RNA-seq 数据,筛选出 OSCC 中 m⁶A 甲基化的差异表达基因,探讨相关信号通路,为 OSCC 的诊断和治疗提供新的方向。

1 材料与方法

1.1 数据

选择 TCGA 数据库中的 OSCC 病例,选取头颈部肿瘤中的口腔部位肿瘤,包括颊、舌、唇、口底、牙龈、硬腭和牙槽嵴。下载病例中 level 3 的 RNA-seq 数据,其中一个样本生成一个文件,再将单个文件合并成矩阵文件。通过 ensembl 数据库,把基因名由 ensembl id 转换为 gene symbol 矩阵,从中提取癌组织和正常组织样本数据。然后提取所有样本中 m⁶A 相关基因的表达。

1.2 筛选差异表达基因

Rv3.5.2 软件的 Wilcoxon 秩和检验计算正常组织和肿瘤组织的基因差异表达情况,筛选出 m⁶A 相关的差异表达基因,绘制成小提琴图;用 pheatmap 包对 m⁶A 相关基因进行聚类分析,并绘制成热图。

1.3 m⁶A 相关基因表达的相关性分析

Rv3.5.2 软件的 corrplot 包对全部 m⁶A 相关基因进行相

关性分析,探讨它们之间的相关性。

1.4 差异表达基因的 KEGG 通路富集分析

对 m⁶A 相关的差异基因做基因集富集分析(gene set enrichment analysis, GSEA),选取基因集“c2.cp.kegg.v6.2.symbols.gmt”做 KEGG 富集分析,得到 KEGG 通路富集分析的结果。

1.5 统计学处理

基因的差异表达和基因表达的相关性分析应用 Rv3.5.2 软件进行统计学处理,Wilcoxon 秩和检验分析 m⁶A 相关基因在正常组织与 OSCC 组织间的差异表达情况,检验水准 $\alpha=0.05$ 。相关性分析的检验水准 $\alpha=0.001$ 。KEGG 通路分析采用 GSEA 4.0.0 软件处理,检验水准 $\alpha=0.05$ 。

2 结果

2.1 OSCC 中 m⁶A 相关差异表达基因与热图

拆分出 258 个样本,包括 241 个肿瘤样本和 17 个正常组织样本。提取样本中 m⁶A 相关基因的表达,得到 13 个 m⁶A 甲基化调节因子的表达量,包括 ALKBH5、FTO、HNRNPC、KIAA1429、METTL3、METTL14、RBM15、WTAP、YTHDF1、YTHDF2、YTHDC1、YTHDC2 和 ZC3H13。

正常组织样本和 OSCC 组织样本间比较后,筛选出 8 个差异表达的 m⁶A 相关基因,包括 HNRNPC、KIAA1429、METTL3、METTL14、RBM15、WTAP、YTHDF1 和 YTHDF2,它们在癌组织的表达量均高于正常组织。基因表达差异的结果见表 1 与图 1,这与热图中对所有 m⁶A 相关基因表达进行聚类分析所得的结果一致(图 2)。HNRNPC、KIAA1429、METTL3、RBM15 和 YTHDF1 的差异显著性最高,METTL14 和 WTAP 的差异显著性次之,YTHDF2 的差异显著性最低,ALKBH5、FTO、YTHDC1、YTHDC2 和 ZC3H13 的差异无统计学意义。

表 1 基因的表达差异

基因名	表达均值 (正常组)	表达均值 (癌症组)	差异倍数取 log 值	P 值
YTHDF1	14.463	22.793	0.656	0.000
KIAA1429	3.752	6.790	0.856	0.000
RBM15	2.114	3.021	0.516	0.000
HNRNPC	35.493	52.020	0.552	0.000
METTL3	2.738	4.427	0.693	0.000
WTAP	12.225	15.875	0.377	0.001
METTL14	2.061	2.544	0.303	0.006
YTHDF2	18.233	21.713	0.252	0.013
FTO	2.454	3.019	0.299	0.064
ZC3H13	4.973	5.705	0.198	0.207
YTHDC1	8.517	9.299	0.127	0.233
ALKBH5	27.374	29.057	0.086	0.468
YTHDC2	2.608	2.602	-0.003	0.840

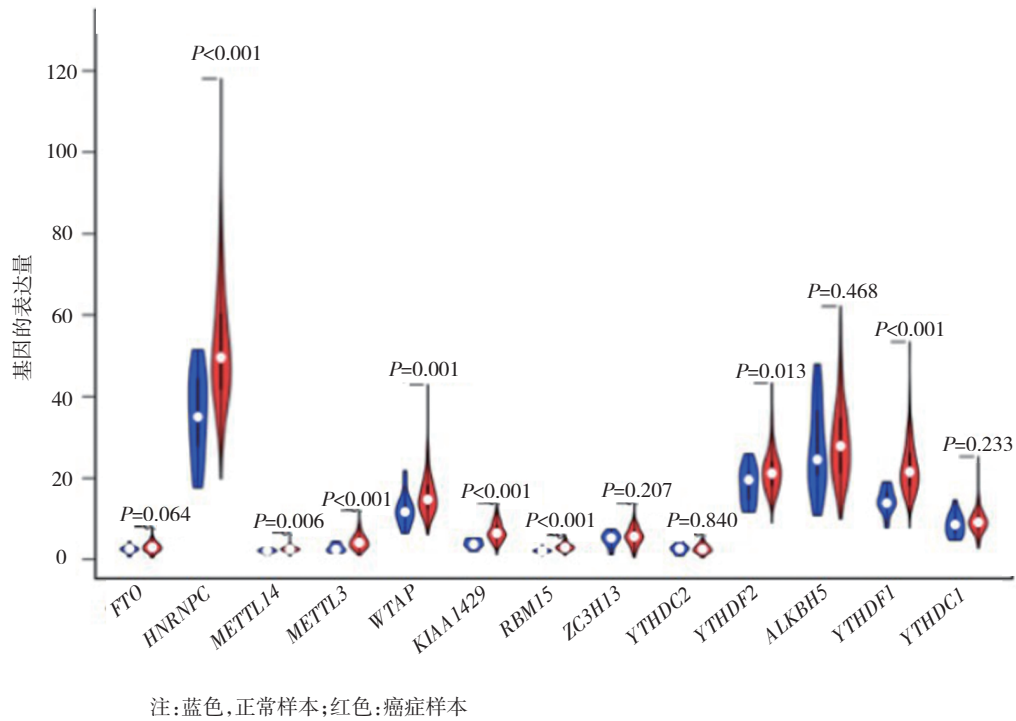
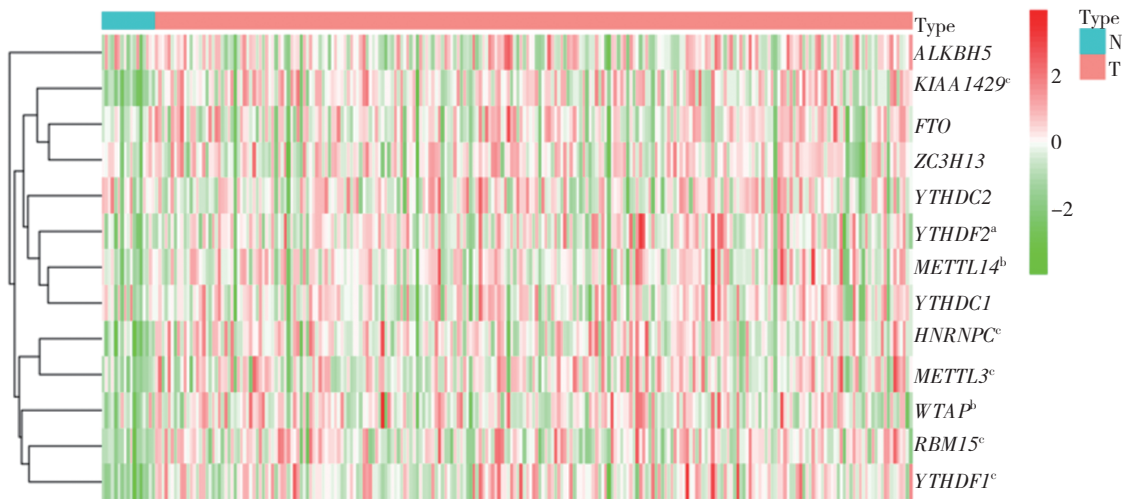


图 1 m⁶A 相关基因表达的小提琴图



注: 红色, 基因高表达; 绿色: 基因低表达; 基因后面的上标 “a” 代表 $0.01 \leq P < 0.05$, “b” 代表 $0.001 \leq P < 0.01$, “c” 代表 $P < 0.001$, 无上标代表 $P \geq 0.05$

图 2 m⁶A 相关基因表达的热图

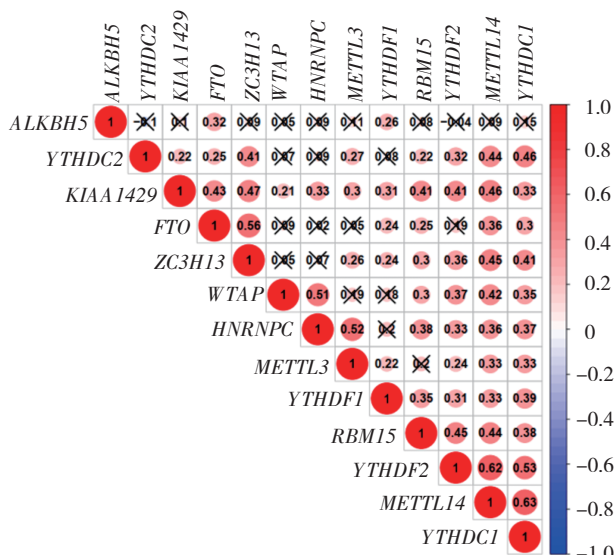
2.2 m⁶A 基因表达间的相关性

m⁶A 相关基因表达之间的相关性分析结果如图 3 所示。所有基因表达蛋白间的相关性均为正相关。YTHDF2 与除 ALKBH5 和 FTO 外的都相关。HNRNPC、METTL3、WTAP 和 YTHDF2 不受 FTO 影响。ALKBH5 仅与 FTO 和 YTHDF1 相关。差异表达基因 HNRNPC、KIAA1429、METTL14、RBM15、WTAP 和 YTHDF2 间互相相关; YTHDF1 与 KIAA1429、MET-

TL3、METTL14、RBM15 和 YTHDF2 分别互相相关; METTL3 与 HNRNPC、KIAA1429、METTL14、YTHDF1 和 YTHDF2 分别互相相关。它们之间的作用关系复杂。

2.3 KEGG 富集的通路

m⁶A 相关差异基因所富集的通路共有 27 个, 根据校正后 P 值 ($P < 0.05$) 选取富集程度最高的前 10 条 (表 2)。主要包括细胞周期、碱基切除修复、同源重组和剪接体等相关的信号通路。



注:蓝色,负相关;红色,正相关;打叉的圆圈,相关性不显著($P>0.001$)

图 3 m⁶A 相关基因间的相关性

表 2 KEGG 信号通路富集最高的前 10 条

信号通路名	富集得分	富集得分 (校正后)	校正后 P 值
KEGG_SPLICEOSOME	0.817	2.413	0.000
KEGG_RNA_DEGRADATION	0.700	2.394	0.000
KEGG_PYRIMIDINE_METABOLISM	0.736	2.393	0.000
KEGG_CELL_CYCLE	0.770	2.387	0.000
KEGG_NUCLEOTIDE_EXCISION_REPAIR	0.612	2.338	0.000
KEGG_HOMOLOGOUS_RECOMBINATION	0.859	2.318	0.000
KEGG_BASE_EXCISION_REPAIR	0.815	2.243	0.000
KEGG_PURINE_METABOLISM	0.765	2.224	0.000
KEGG_UBIQUITIN_MEDIATED_PROTEOLYSIS	0.765	2.188	0.000
KEGG_ONE_CARBON_POOL_BY_FOLATE	0.609	2.168	0.001

3 讨论

OSCC由遗传和表观遗传的变异导致原癌基因活化或多个肿瘤抑制基因失活而致病^[10],其分子机制未被阐明,深入研究其分子机制,寻找 OSCC 生物标记物,有利于早期诊断,提供新的肿瘤药物治疗靶点。TCGA 数据库用大规模的基因组测序方法,绘制出人类 50 种肿瘤基因组变异图谱。本研究通过生物信息学方法对 TCGA 数据库中 OSCC 相关数据进行分析,筛选出癌组织与正常组织间 m⁶A 相关的差异表达基因,对差异基因进行 KEGG 信号通路富集分析,并对 m⁶A 相关基因表达做聚类分析与相关性分析,以发现用于 OSCC 诊断的标记物和治疗的新靶点。

在正常组织和 OSCC 组织间,对 13 个 m⁶A 甲基化相关基因的表达量进行比较,共筛选出 8 个差异表达基因,包括 *HNRNPC*、*KIAA1429*、*METTL3*、*METTL14*、*RBM15*、*WTAP*、*YTHDF1* 和 *YTHDF2*。研究证实,抑制 *HNRNPC* 会抑制乳腺癌细胞的增殖和肿瘤的生长^[11]; *KIAA1429* 在胃癌中通过稳定 c-Jun 信使 RNA 发挥癌基因的作用^[12]。 *METTL3* 在体外促进 OSCC 细胞的增殖、浸润和转移,通过由 *YTHDF1* 调节的 m⁶A 甲基化修饰过程,增加甲基化后 c-Myc 信使 RNA 的稳定性,使 OSCC 发生;在体内敲除 *METTL3* 后抑制肿瘤的生长^[7]。 *METTL14* 在急性髓性白血病中作为癌基因,调控癌症的发生^[13]。 *RBM15* 经表观遗传学机制,作用于 c-Myc 抑制巨核细胞增殖,参与急性巨核细胞白血病的发展^[14];在 Kaposi 肉瘤中,它帮助信使 RNA 从细胞核进入细胞质^[15]。 *WTAP* 在胃癌中高表达,表达越高胃癌预后越差^[16]。 *YTHDF2* 被敲除后,急性髓性白血病好转^[17]。这些 m⁶A 相关的差异表达基因在恶性肿瘤中的作用已被部分证实。本研究发现,它们在 OSCC 组织的表达量均明显高于正常组织,可能对 OSCC 的发生起促进作用,有望成为 OSCC 的诊断标记物和治疗靶点,这需要后续体内或体外实验研究进一步验证。

m⁶A 甲基化修饰在多种肿瘤的发生发展中发挥重要作用,由 m⁶A 甲基化调节因子调控,甲基转移酶催化,甲基化结合蛋白识别,进而发挥生物功能,而去甲基化酶氧化可以消除 m⁶A 甲基化修饰作用^[18-20]。本研究所发现的差异表达基因中没有去甲基化酶,推测 OSCC 的发生过程主要与 m⁶A 甲基化修饰增强相关;另外, m⁶A 相关基因表达间具有相关性,形成复杂的分子调控网络。值得注意的是,差异表达基因 *HNRNPC*、*KIAA1429*、*METTL14*、*RBM15*、*WTAP* 和 *YTHDF2* 间两两互相相关,联系紧密,它们间的关系可能在 OSCC 的发生过程中起关键作用。相关性分析还发现,OSCC 组织中所有 m⁶A 相关表达基因间均为正相关,推测若靶向下调其中某一个基因的表达,可能下调所有 m⁶A 相关基因的表达。

KEGG 富集分析显示 m⁶A 甲基化差异表达基因在 27 个信号通路中富集,主要涉及细胞周期、碱基切除修复、同源重组和剪接体等相关生物进程的信号通路。细胞周期包括 G₁、S、G₂、M 与 G₀ 期,主要由肿瘤蛋白 53 (tumor protein 53, P53) 和细胞周期蛋白依赖性激酶 (cyclin-dependent kinases, CDKs) 家族调节。大约 50% 的肿瘤存在 P53 功能缺陷^[21]; CDKs 发生异常时,细胞增殖和凋亡间的平衡失控,细胞周期发生紊乱,肿瘤发生^[22]。差异表达基因可能抑制

P53 和 CDKs 的产生,使细胞周期发生异常,促进 OSCC 的发生。碱基切除修复主要用于修复微小的碱基损伤,这些损伤表现为去氨基化、氧化或甲基化,它们不会严重影响 DNA 双螺旋结构,若碱基切除修复功能异常,会伴发癌症^[23]。同源重组主要用于修复 DNA 双链断裂损伤,双链断裂必须得到及时且准确的修复,一旦有相关的分子发生突变或者缺失极易导致癌症^[24]。剪接体突变后,无法识别内含子,可导致肿瘤抑制因子失活^[25],发生癌变。推测 m⁶A 甲基化差异表达基因抑制碱基切除修复、破坏同源重组并诱发剪接体突变,促使 OSCC 发生。差异表达基因在这些富集通路中的作用机制需要继续深入研究。

综上所述,本研究应用生物信息学方法筛选出了 OSCC 的 m⁶A 相关差异表达基因,有望成为 OSCC 的新诊断标记物和治疗靶点。它们参与癌症发生发展的通路,对其深入研究有助于阐明 OSCC 的分子机制。

参 考 文 献

- [1] Haddad RI,Shin DM. Recent advances in head and neck cancer[J]. *N Engl J Med*,2008,359(11):1143-1154.
- [2] Chi AC,Day TA,Neville BW. Oral cavity and oropharyngeal squamous cell carcinoma;an update[J]. *CA Cancer J Clin*,2015,65(5):401-421.
- [3] Ke SD,Pandya-Jones A,Saito Y,et al. m⁶A mRNA modifications are deposited in nascent pre-mRNA and are not required for splicing but do specify cytoplasmic turnover[J]. *Genes Dev*,2017,31(10):990-1006.
- [4] Liu N,Pan T. N⁶-methyladenosine-encoded epitranscriptomics[J]. *Nat Struct Mol Biol*,2016,23(2):98-102.
- [5] Fry NJ,Law BA,Ilkayeva OR,et al. N⁶-methyladenosine contributes to cellular phenotype in a genetically-defined model of breast cancer progression[J]. *Oncotarget*,2018,9(58):31231-31243.
- [6] Kwok CT,Marshall AD,Rasko JE,et al. Genetic alterations of m⁶A regulators predict poorer survival in acute myeloid leukemia[J]. *J Hematol Oncol*,2017,10(1):39-44.
- [7] Zhao W,Cui YM,Liu LN,et al. METTL3 facilitates oral squamous cell carcinoma tumorigenesis by enhancing c-Myc stability via YTHDF1-mediated m⁶A modification[J]. *Mol Ther Nucleic Acids*,2020,20:1-12.
- [8] 肖仕全,王利,刘太行,等. 基于生物信息学技术挖掘番茄红素对前列腺癌细胞的作用靶点[J]. *重庆医科大学学报*,2019,44(7):878-884.
- [9] Xiao SQ,Wang L,Liu TH,et al. Mining targets of lycopene on prostate cancer cells based on bioinformatics technique[J]. *J Chongqing Med Univ*,2019,44(7):878-884.
- [10] 郭秋云,张孟贤,胡广原,等. 胶质母细胞瘤基因表达谱芯片的生物信息学分析[J]. *华中科技大学学报(医学版)*,2018,47(1):38-43.
- [11] Guo QY,Zhang MX,Hu GY,et al. Bioinformatics analysis of differentially expressed genes in glioblastoma[J]. *Acta Med Univ Sci Technol Huazhong*,2018,47(1):38-43.
- [12] Asokan GS,Jeelani S,Gnanasundaram N. Promoter hypermethylation profile of tumour suppressor genes in oral leukoplakia and oral squamous cell carcinoma[J]. *J Clin Diagn Res*,2014,8(10):ZC09-ZC12.
- [13] Wu YS,Zhao WW,Liu Y,et al. Function of HNRNPC in breast cancer cells by controlling the dsRNA-induced interferon response[J]. *EMBO J*,2018,37(23):e99017.
- [14] Miao R,Dai CC,Mei L,et al. KIAA1429 regulates cell proliferation by targeting c-Jun messenger RNA directly in gastric cancer[J]. *J Cell Physiol*,2020,235(10):7420-7432.
- [15] Weng HY,Huang HL,Wu HZ,et al. METTL14 inhibits hematopoietic stem/progenitor differentiation and promotes leukemogenesis via mRNA m⁶A modification[J]. *Cell Stem Cell*,2018,22(2):191-205.e9.
- [16] Descot A,Rex-Haffner M,Courtois G,et al. OTT-MAL is a deregulated activator of serum response factor-dependent gene expression[J]. *Mol Cell Biol*,2008,28(20):6171-6181.
- [17] Majerciak V,Uranishi H,Kruhlik M,et al. Kaposi's sarcoma-associated herpesvirus ORF57 interacts with cellular RNA export cofactors RBM15 and OTT3 to promote expression of viral ORF59[J]. *J Virol*,2011,85(4):1528-1540.
- [18] Li HF,Su Q,Li B,et al. High expression of WTAP leads to poor prognosis of gastric cancer by influencing tumour-associated T lymphocyte infiltration[J]. *J Cell Mol Med*,2020,24(8):4452-4465.
- [19] Paris J,Morgan M,Campos J,et al. Targeting the RNA m⁶A reader YTHDF2 selectively compromises cancer stem cells in acute myeloid leukemia[J]. *Cell Stem Cell*,2019,25(1):137-148.e6.
- [20] Yang Y,Hsu PJ,Chen YS,et al. Dynamic transcriptomic m⁶A decoration;writers,erasers,readers and functions in RNA metabolism[J]. *Cell Res*,2018,28(6):616-624.
- [21] Chen B,Li Y,Song RF,et al. Functions of RNA N⁶-methyladenosine modification in cancer progression[J]. *Mol Biol Rep*,2019,46(2):2567-2575.
- [22] Tuncel G,Kalkan R. Importance of m⁶A-methyladenosine(m⁶A) RNA modification in cancer[J]. *Med Oncol*,2019,36(4):1-6.
- [23] Hainaut P,Hollstein M. p53 and human cancer;the first ten thousand mutations[J]. *Adv Cancer Res*,2000,77:81-137.
- [24] Santo L,Siu KT,Raje N. Targeting cyclin-dependent kinases and cell cycle progression in human cancers[J]. *Semin Oncol*,2015,42(6):788-800.
- [25] 刘博雅,杨鑫,任梦梦,等. DNA 损伤修复机制:解读 2015 年诺贝尔化学奖[J]. *中国生物化学与分子生物学报*,2015,31(12):1322-1329.
- [26] Liu BY,Yang X,Ren MM,et al. The Repair Mechanism for DNA damage understanding—the 2015 Nobel prize in chemistry[J]. *Chin J Biochem Mol Biol*,2015,31(12):1322-1329.
- [27] Meindl A,Hellebrand H,Wiek C,et al. Germline mutations in breast and ovarian cancer pedigrees establish RAD51C as a human cancer susceptibility gene[J]. *Nat Genet*,2010,42(5):410-414.
- [28] Jung H,Lee D,Lee J,et al. Intron retention is a widespread mechanism of tumor-suppressor inactivation[J]. *Nat Genet*,2015,47(11):1242-1248.