

RT-PCR 检测 *HUTP14A* mRNA 时排除假基因 *HUTP14C* 干扰的方法

许阁阁¹⁾, 刘小锋²⁾, 张春风³⁾, 杜晓娟^{1)*}

(¹⁾ 北京大学基础医学院细胞生物学系, 北京 100083; (²⁾ 北京大学肿瘤医院肝胆胰外一科, 北京 100142;
(³⁾ 北京大学基础医学院医学遗传学系, 北京 100083)

摘要 人类 U3 蛋白 14C 基因(*HUTP14C*)是人类 U3 蛋白 14A 基因(*HUTP14A*)的假基因。两者转录本序列同源性高达 95%。常规 RT-qPCR 技术在检测 *HUTP14A* mRNA 丰度时,*HUTP14C* 的存在会影响检测结果。本研究旨在建立检测 *HUTP14A* mRNA 时排除 *HUTP14C* 干扰的 RT-PCR 方法。本研究设计出能分别从多种肿瘤细胞 DNA 和 RNA 中特异性扩增 *HUTP14A* 和 *HUTP14C* 的引物,避免假基因 *HUTP14C* 对其同源基因 *HUTP14A* 检测的干扰。在检测细胞系 *HUTP14A* mRNA 时,通过 DNase I 消除 RNA 中污染的 *HUTP14C* DNA,用靶向 *HUTP14C* 3'-UTR 的 siRNA 沉默 *HUTP14C* mRNA 后,再用 RT-PCR 检测 *HUTP14A* mRNA 丰度,使结果更加准确。在 18 对肝癌及癌旁组织中,利用特异性引物进行 RT-PCR 检测,*HUTP14A* 和 *HUTP14C* mRNA 的表达略高于癌旁组织。本研究提示,针对有假基因存在的功能基因,对其 mRNA 丰度进行检测时,在提取细胞或组织总 RNA 后,用 DNase I 处理,再用 RNA 直接进行 PCR 扩增,排除 DNA 污染后,再进行 RT-PCR 或 RT-qPCR 扩增。大多假基因具有较长的 3'-UTR 区,在该区域设计 siRNA 特异性沉默假基因的 mRNA 后,用 RT-qPCR 检测功能基因的 mRNA 丰度,可以排除假基因 mRNA 的影响。在病理组织中检测功能基因的 mRNA 丰度时,可以根据假基因和其功能基因的序列差异设计出特异扩增功能基因的引物,从假基因的 3'-UTR 区设计特异扩增假基因的引物,通过 RT-qPCR 技术分别检测二者的 mRNA。

关键词 人类 U3 蛋白 14 A 基因; 人类 U3 蛋白 14C 基因(*HUTP14C*); 假基因; RT-qPCR; 肝癌中图分类号 Q26

An RT-PCR Method of Excluding the Interference of Pseudogene *HUTP14C* on the Detection of *HUTP14A* mRNA

XU Ge-Ge¹⁾, LIU Xiao-Feng²⁾, ZHANG Chun-Feng³⁾, DU Xiao-Juan^{1)*}

(¹⁾ Department of Cell Biology, School of Basic Medical Sciences, Peking University Health Science Center, Beijing 100083, China;
(²⁾ Hepatopancreatobiliary Surgery Department I, Peking University Cancer Hospital, Beijing 100142, China;
(³⁾ Department of Medical Genetics, School of Basic Medical Science, Peking University Health Science Center, Beijing 100083, China)

Abstract Human U three protein 14C (*HUTP14C*) is a pseudogene of human U three protein 14A (*HUTP14A*). The sequences of their transcripts are 95% conserved. Existence of *HUTP14C* may cause experimental error when the conventional RT-qPCR is used to evaluate the abundance of *HUTP14A* mRNA. The aim of this study is to establish an accurate RT-PCR procedure for the evaluation of *HUTP14A* mRNA. We have designed primers that can specifically amplify *HUTP14A* or *HUTP14C*

fragments from human genomic DNA or cDNA, respectively. An RT-PCR procedure was generated to exclude the interference of the pseudogene *HUTP14C* on the detection of *HUTP14A* mRNA. The mRNA of *HUTP14A* and *HUTP14C* was detected in 18 pairs of hepatocellular cancer tissues. When detecting the mRNA levels of functional genes that possess pseudogenes, PCR should be directly performed using the total RNA extracted from cells or tissues after DNase I treatment to exclude the possibility of DNA contamination. To exclude the influence of pseudogene mRNA on the evaluation of the mRNA of functional gene, specific siRNAs targeting the 3'-UTR region of the pseudogene can be used to silence the mRNA of the pseudogene before RT-PCR is performed. When detecting the mRNA levels of functional genes in pathological tissues, the specific primers of functional genes can be designed according to the sequence differences between the pseudogenes and their functional genes. The primers for pseudogenes can be designed according to the longer 3'-UTR region of pseudogenes. Thereby, the contaminations of pseudogenes can be excluded when detecting transcription levels of the functional genes by RT-qPCR.

Key words human U three protein 14 A (*HUTP14A*); human U three protein 14C (*HUTP14C*); pseudogene; RT-qPCR; liver cancer

假基因是基因进化过程中的无功能残留物,与正常基因相似,但丧失正常功能的 DNA 序列^[1]。它广泛存在于原核及真核生物中,在人类及小鼠基因组中约有 20 000 个假基因^[2-4]。假基因通常无内含子和启动子序列,且 3' 末端有多聚腺嘌呤尾巴等^[5-7]。假基因主要有 2 种:加工假基因和未加工假基因。加工假基因是通过逆转录转座形成。mRNA 反转录成 cDNA,然后整合进入基因组^[8]。未加工假基因又称为复制型假基因。在复制过程中发生点突变、插入、缺失和移码突变等,使其无法正常转录和翻译,丧失同源基因的正常功能进而沉默成为假基因^[9]。

近年来,大量研究证明,假基因具有生物学功能。有研究鉴定了 11 224 个人类假基因,其中 863 个可以转录^[10]。如磷酸酶和张力蛋白同源缺失基因 (phosphatase and tensin homology deleted on chromosome ten, *PTEN*) 的假基因 *PTENP1* 不仅能转录,且其转录产物有保护功能基因的作用^[11,12];又如常作为子宫内膜癌诊断标志物的肌球蛋白轻链激酶 (myosin light chain kinase, *MYLK*) 基因的假基因 *MYLKP1*^[13] 也能转录。研究发现,哺乳动物中约 2%~5% 的假基因能够表达蛋白质^[14]。如人类 U3 蛋白 14C 基因 (human U three protein 14 A, *HUTP14C*)^[15],细胞色素 P-450 超家族基因 (cytochrome P450)^[16]。

假基因功能主要有两点。部分假基因与功能基因序列高度相似,其反义链可能会和同源基因正义链的转录产物形成 RNA 复合体,从而影响亲本基因的表达^[17],假基因 *PTENP1* 可以编码 α 和 β 两种亚型的反义 RNA^[18]。一些假基因与其同源基因竞争性结合 miRNA 从而调节功能基因的表达。*PTENP1*

和 *PTEN* 相比有更长的 3'-UTR,其转录本和 *PTEN* 的 mRNA 竞争性结合 miR-19b 和 miR-20a,使 *PTEN* 转录水平升高^[12,19]。

癌症基因组学的研究提示,假基因与肿瘤相关。研究人员在肿瘤标本中检测出 2 082 个假基因的转录本,有 218 个假基因仅在肿瘤组织中表达,且部分有组织特异性^[20]。

由于假基因是由同源基因异常复制或逆转录转座形成的,所以两者基因序列有极高的相似性,且很多假基因可以被转录。所以,当用 RT-qPCR 技术定量检测含有假基因的功能基因转录水平时,假基因会干扰检测结果。如果是检测和医学诊断、肿瘤分级和预后相关的假基因,则会对诊断结果和治疗措施产生影响。细胞角蛋白 19 基因 (cytokeratin 19, *CK19*) 常作为检测上皮型肿瘤的标记物,实验技术多采用 PCR。然而有研究发现,PCR 扩增产物中可能包含 *CK19* 的假基因,因此该方法会影响肿瘤诊断准确性^[21]。

人类 U3 蛋白 14 A 基因 (human U three protein 14 A, *HUTP14A*) 位于人类染色体的 Xq25, cDNA 全长为 2 316 bp,编码 771 个氨基酸的蛋白质。本室前期研究发现, hUTP14a 是核仁蛋白质,与 U3 snoRNA 结合参与 18S rRNA 的加工,同时 hUTP14a 结合并降解 p53 蛋白,下调 hUTP14a 蛋白的表达抑制肿瘤细胞生长,促进凋亡^[22]。研究发现, hUTP14a 蛋白结合视网膜母细胞瘤 (retinoblastoma, RB) 蛋白,通过泛素依赖途径促进 RB 降解^[23]。hUTP14a 与 c-Myc 相互作用,并以泛素特异性蛋白酶 USP (ubiquitin specific protease, USP)、USP36 依赖的方式保护 c-Myc 不被泛素化和降解,从而稳定 c-Myc。c-Myc 激活 *HUTP14A* 的

转录,形成正反馈调节环路,促进结直肠癌进展^[24]。hUTP14a 在结直肠癌中表达升高,在肝细胞癌中的高表达与患者预后不良高度相关^[25]。治疗肿瘤时,hUTP14a 可保护肿瘤细胞免受化疗药物诱导的凋亡,该作用不依赖于 p53 蛋白,所以 hUTP14a 具有成为抗肿瘤治疗靶点的潜力^[26]。这些研究提示,hUTP14a 蛋白在肿瘤的发生、发展以及治疗过程均发挥着重要作用。

人类 U3 蛋白 14C 的基因 *HUTP14C* (human U three protein, 14C) 是 *HUTP14A* 的假基因,位于人类 13 号染色体,属于加工假基因。它是由位于 X 染色体上的 *HUTP14A* 的无内含子拷贝经转座产生^[27]。研究发现,hUTP14a 具有抗凋亡的作用,其假基因 *HUTP14C* 虽然编码蛋白质,但 hUTP14c 无抗凋亡的作用^[26]。这提示,假基因 *HUTP14C* 和功能基因 *HUTP14A* 的作用是不同的。此外,目前没有能区分 hUTP14a 和 hUTP14c 蛋白的特异性抗体,因此能否用 RT-PCR 技术准确区分两者就显得尤为重要。另有研究表明,hUTP14a 在肝细胞癌中的高表达与患者预后不良高度相关^[25]。然而,假基因 *HUTP14C* 是否存在于肝癌组织且是否转录是未知的。本研究围绕这两个问题进行探索。

1 材料与方法

1.1 细胞培养及转染

HepG2、H1299、HeLa、HCT116^{+/+} 及 BGC823 细胞系均用添加 10%胎牛血清的 DMEM 培养基培养。转染试剂 Lipofectamine 2000 购自于 Invitrogen 公司。siRNA 由上海吉玛公司合成,其序列如 Table 1 所示。

Table 1 Sequences of *HUTP14C*-siRNA

RNA	RNA sequences
Negative Control	5'-ACUACCGUUGUUAUAGGUG-3'
<i>HUTP14C</i> -siRNA-1	5'- GGUGCUGCCAGCAAUCCUUTT-3'
<i>HUTP14C</i> -siRNA-2	5'- GCAUAUAUAUCAAUCCUUTT-3'

1.2 肿瘤组织来源

收集 2003 年至 2006 年在北京大学第三医院接受肝癌切除手术的 18 例肝癌组织及癌旁非肿瘤组织标本,其入选标准为:(1)术前未接受任何放疗和化疗;(2)不合并身体其他部位的恶性肿瘤。本研究获得北京大学生物医学伦理委员会批准,并得到患者的知情同意,患者临床病理资料完整。组织标本冻存于液氮。

1.3 引物序列

本研究中所有引物由睿博公司合成,序列如 Table 2 所示。

Table 2 Sequences of primer

Primer	Primer sequences
Primer 1	5'-AGCCAGCCATTGCTCCCATG-3'
Primer 2	5'-GTGACTGGCTATCACATGACTGACA-3'
Primer 3	5'-GGTTCCTCTCTCTGGACCTG-3'
Primer 4	5'-CAATGAGAAGCGCAACATCC-3'
Primer 5	5'-TAATGATATGGCCTGGCTTGG-3'
Primer 6	5'-AATGACTGCCAACCGGCTTGCAGA-3'
Primer 7	5'-CAATCTCTTCTTTGTTTCAGAGGTAAGTCCA CTGT-3'
Primer 8	5'-CCTTTCATACTAGGTACTGGTGA-3'
Primer 9	5'-AGCTACTGGCCTAGTAACTAACA-3'

1.4 RT-PCR

采用 Trizol 法提取细胞总 RNA。再使用 Promega AMV 逆转录试剂盒,将细胞总 mRNA 逆转录为 cDNA 文库。反应条件如下:42 ℃ 1 h,75 ℃ 15 min。PCR 反应条件为:94 ℃ 预变性 5 min,94 ℃ 15 s,60 ℃ 30 s 和 68 ℃ 2 min。30 s 完成 25 个循环,68 ℃ 延伸 10 min。qPCR 使用 ABI 7500/7500 快速实时 PCR 系统。GAPDH 作为内参,数据采用 Ct 法进行分析。

1.5 DNA 凝胶电泳

采用 2% 琼脂糖凝胶及 100 bp 和 1 kb 的 DNA 标记物作为条带大小的参照。

1.6 DNase I 消化方法

在无 RNA 酶的反应管中,加入 Table 3 的组分。37 ℃ 孵育 30 min。然后加入 0.5 μL 200 mmol/L EDTA 作为反应终止液。置于 65 ℃ 孵育 10 min,失活 DNase I。

Table 3 Components of DNase I reaction

Reagent	Quantity
RNA	10 μg
10×DNase I reaction buffer	1 μL
DNase I	1 unit/μg RNA
RNase-free water	Up to 10 μL

1.7 统计学分析

采用 SPSS 17.0 软件进行数据录入和分析。采用 *t* 检验比较 DNase I 处理 RNA 前后 *HUTP14A* 和假基因 *HUTP14C* 表达的差异,以 *P*<0.05 为有统计学意义。

2 结果

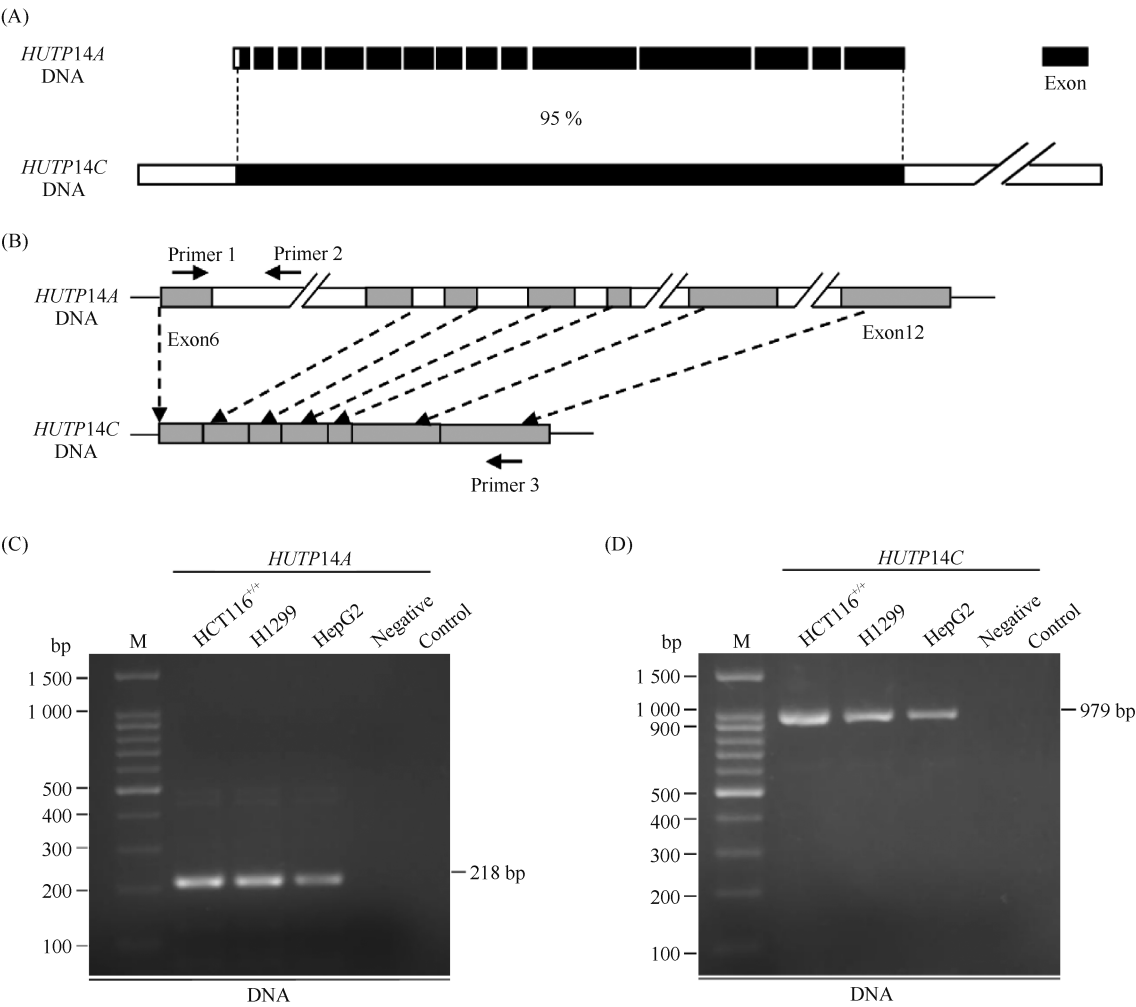
2.1 多种肿瘤细胞中存在功能基因 *HUTP14A* 与假基因 *HUTP14C*

有研究报道,*HUTP14A* 与其假基因 *HUTP14C* 在 DNA 序列上存在 95% 同源性 (Fig.1A) [27]。由于假基因 *HUTP14C* 中不含有内含子,因此采用不同的跨内含子的 PCR 引物可以区别两者。本研究在 *HUTP14A* 的第 6 个外显子处设计了上游引物 Primer 1,第 6 个内含子处设计了下游引物 Primer 2 (Fig.1B)。提取 HCT116^{+/+}、H1299 及 HepG2 细胞系的 DNA,用 Primer 1 和 2 通过 PCR 技术特异性扩增 218 bp 的 *HUTP14A* 片段,进行凝胶电泳 (Fig.1C)。3 个细胞系均有特异的 218 bp 产物条带,说明 *HUTP14A* 在 HCT116^{+/+}、H1299 及 HepG2 细胞系中都存在。为了区分假基因 *HUTP14C* 与 *HUTP14A*,在 *HUTP14C* 中与 *HUTP14A* 的第 12 个外显子序列相同处,设计了下游引物 Primer 3 (Fig.1B),其上游引物依然用 Primer 1,同样提取

HCT116^{+/+}、H1299 及 HepG2 细胞系 DNA,用 Primer1 和 3 通过 PCR 技术特异性扩增 979 bp 的 *HUTP14C* 片段,进行凝胶电泳 (Fig.1D)。同样的 3 个细胞系均有特异的 979 bp 产物条带,说明 *HUTP14C* 在 HCT116^{+/+}、H1299 及 HepG2 细胞系中均存在。以上结果说明,在不同组织中均存在 *HUTP14A* 与假基因 *HUTP14C*。将以上 PCR 产物进行测序 (Fig.1E,F)。序列比对后验证了 218 bp 产物是 *HUTP14A* 的片段,979 bp 产物则是 *HUTP14C* 片段。因此,本研究设计了 2 对可以特异性扩增 *HUTP14A* 与假基因 *HUTP14C* 的引物来将两者区分,并验证了两者同时存在于多种细胞系。

2.2 RNA 中的 *HUTP14C* DNA 污染影响 RT-PCR 对 *HUTP14A* mRNA 的检测

为了验证细胞总 RNA 中的 *HUTP14C* 基因污染是否会影响 *HUTP14A* mRNA 的检测,本研究从定性和定量两方面进行探索。首先,提取 HepG2、H1299、HeLa、HCT116^{+/+} 及 BGC823 细胞系的总 RNA,用 DNase I 处理消化 RNA 中可能污染的



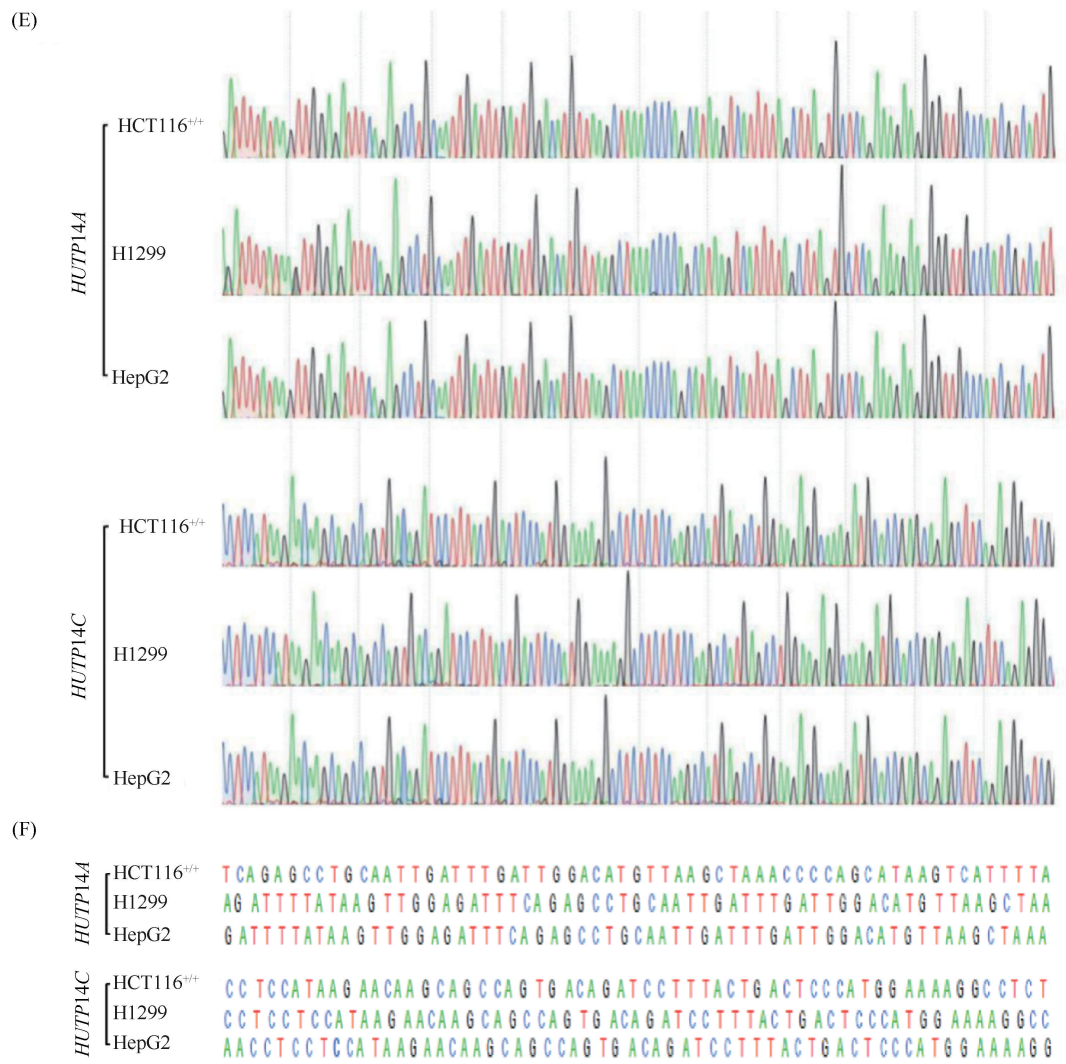


Fig.1 Specific amplification of *HUTP14A* and *HUTP14C* (A) Diagram of *HUTP14A* and *HUTP14C* gene structure. (B) Positions of primer 1 and 2 to specifically amplify *HUTP14A*. Positions of primer 1 and 3 to specifically amplify *HUTP14C*. (C) DNA was extracted from HCT116^{+/+}, H1299 and HepG2 cell lines. *HUTP14A* was amplified by PCR using primer 1 and 2. The product was subjected to DNA gel electrophoresis. (D) DNA was extracted from HCT116^{+/+}, H1299 and HepG2 cell lines. *HUTP14C* was amplified by PCR using primer 1 and 3. The product was subjected to DNA gel electrophoresis. (E) Amplification products were sequenced. (F) Partial sequencing results of amplification products

HUTP14C DNA 后,采用 Primer 1 和 3(Fig.2A) 直接进行 PCR 反应,特异性扩增 979 bp 的 *HUTP14C* 片段,未用 DNase I 处理消化组作为对照,将产物进行 2% DNA 凝胶电泳(Fig.2B)。结果显示,每组细胞系中,DNase I 酶消化处理组未扩增出产物,而对照组中则均有产物。由于假基因 *HUTP14C* 不包括内含子,Primer 1 和 3 跨越 *HUTP14A* 的 6 个内含子,因此 Primer 1 和 3 扩增出的产物是 *HUTP14C*。这个结果说明,所提取的细胞系总 RNA 中混有 *HUTP14C* 的 DNA,且 DNase I 酶消化处理后会排除 *HUTP14C* 的 DNA 污染。接下来,从定量角度验证细胞总 RNA 中的 *HUTP14C* 基因污染是否会影响

HUTP14A mRNA 的检测。采用能同时扩增 2 个基因的 Primer 4 和 5(Fig.2B)用上述经过 DNase I 处理的 RNA,通过 RT-qPCR 检测 *HUTP14A/C* 的 mRNA 丰度,分析与未经过 DNase I 处理的对照组 RT-qPCR 结果的差异(Fig.2C)。结果表明,DNase I 处理后 *HUTP14A/C* 的 mRNA 检测量显著降低($P<0.05$),有统计学意义。这表明,RNA 中的 *HUTP14C* 基因污染会影响 RT-qPCR 对 *HUTP14A* mRNA 的检测。本研究提示,对含有假基因且该假基因转录的基因来说,RT-qPCR 检测转录水平时需要通过 DNase I 将所提取的 RNA 中混有的假基因 DNA 消化干净来使结果更准确。

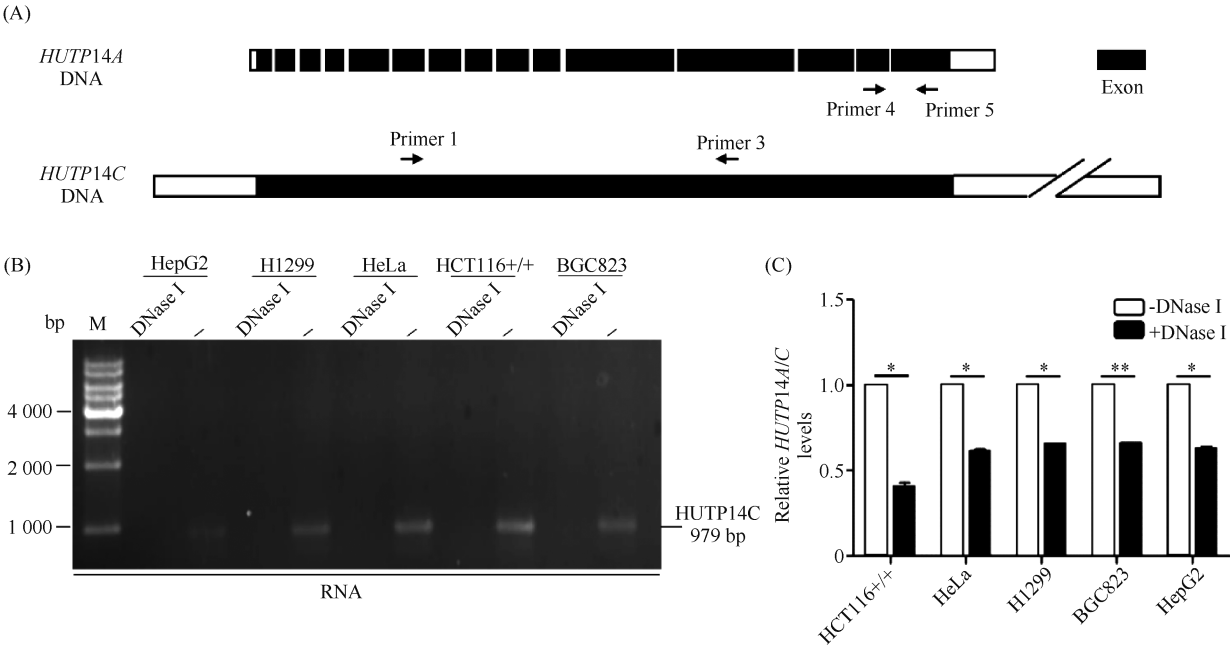


Fig.2 Effect of DNA contamination of *HUTP14C* on detection of *HUTP14A/C* mRNA by RT-PCR (A) Positions of primer 4 and 5 on *HUTP14A* and *HUTP14C*. (B) Total RNA was extracted from HepG2, H1299, HeLa, HCT116^{+/+} and BGC823 cell lines. Each cell line was treated with DNase I. The control group was treated with DEPC water. *HUTP14C* was amplified by PCR using primer 1 and 3. The product was subjected to agarose electrophoresis. (C) Total RNA was extracted from HepG2, H1299, HeLa, HCT116^{+/+} and BGC823 cell lines. Each cell line was treated with DNase I and the control group was treated with equal DEPC water before reverse transcription. mRNA abundance was performed by RT-PCR with primer 4 and 5

2.3 *HUTP14C* 的 mRNA 影响 *HUTP14A* mRNA 的检测

为了验证细胞总 RNA 中的 *HUTP14C* 的 mRNA 是否会影响 *HUTP14A* mRNA 的检测,本研究设计了从 RNA 水平区分两者的引物,分别是位于 *HUTP14A* 的 CDS 区特异性扩增 *HUTP14A* 的 Primer 6 和 7,以及位于 *HUTP14C* 的 3'-UTR 区特异性扩增 *HUTP14C* 的 Primer 8 和 9 (Fig.3 A)。由于两者转录本序列相似度高达 95%,所以在 CDS 区设计特异性扩增 *HUTP14A* 的引物时,应选择两者序列差异较大的区域。本研究在 NCBI 网站将它们的转录本序列进行比对后,选择 *HUTP14A* 的第 47~70 位核苷酸处设计上游引物 Primer 6,在 390~423 位核苷酸处设计下游引物 Primer 7 (Fig.3B),由于 *HUTP14C* 的 3'-UTR 区比 *HUTP14A* 更长,因此,在 *HUTP14C* 特有的序列处设计特异性扩增 *HUTP14C* 的 Primer 8 和 9。

为进一步验证这两对引物是否能准确区分 *HUTP14A* 和 *HUTP14C*,本文选择 HeLa 和 HepG2 细胞系提取 RNA,用 DNase I 消化处理后,采用 Primer 6,7, 8 和 9 进行反转录和 PCR。之后,选择 2%凝胶进行电泳 (Fig.3C)。两个细胞系分别有 377 bp

的 *HUTP14A* 片段,和 862 bp 的 *HUTP14C* 片段,且 Primer 6 和 7 的引物只能扩增出 377 bp 的条带,Primer 8 和 9 的产物只有 862 bp 的条带。说明这两对引物都是特异的。因此,可以直接定性检测 *HUTP14A* 和 *HUTP14C* 的表达。

HUTP14A mRNA 的检测不仅会受到假基因 *HUTP14C* DNA 的干扰,且会受到 *HUTP14C* mRNA 的影响。为了排除该干扰因素,可以利用 *HUTP14C* 较长的 3'-UTR 这一特点来设计实验。由于假基因 *HUTP14C* 具有 2 504 bp 的 3'-UTR,*HUTP14A* 的 3'-UTR 只有 172 bp,可以在 3'-UTR 区设计 *HUTP14C* 特异的 siRNA 来沉默假基因。本研究在 HeLa 细胞中,利用 *HUTP14C* 特异的 siRNA 沉默 *HUTP14C* 的 mRNA,对所提取的细胞总 RNA 用 DNase I 酶消化处理后,用 Primer 6 和 7 通过 RT-PCR 特异性扩增 377 bp 的 *HUTP14A* 片段;用 Primer 8 和 9 通过 RT-PCR 特异性扩增 862 bp 的 *HUTP14C* 片段,将所获得的产物选择 2%凝胶共同进行电泳 (Fig.3D)。空白对照组和转染了 Control siRNA 组的产物是 377 bp 的 *HUTP14A* 片段,和 862 bp 的 *HUTP14C* 片段,说明 cDNA 中含有两个基因的 mRNA。然而,转染了 *HUTP14C* siRNA-1 和 siRNA-2 组只有 377 bp 的

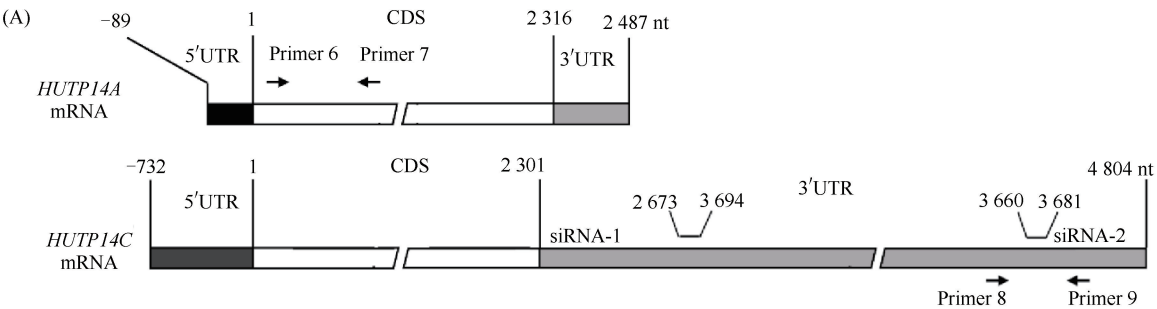
HUTP14A 片段,一方面说明 DNase I 将 RNA 中污染的 *HUTP14C* DNA 消化完全,另一方面说明特异性沉默了 *HUTP14C* 的 mRNA,这时所提取的总 RNA 中不包括 *HUTP14C* 的 mRNA。

为了明确 377 bp 条带是否包含 *HUTP14C* 片段,本文将产物纯化进行测序 (Fig.3E)。序列比对结果说明,377 bp 条带只包含 *HUTP14A* 片段,因此 Primer 6 和 7 是 *HUTP14A* 的特异性产物 (Fig.3F)。同时,为了证明 862 bp 的片段只含 *HUTP14C*,本文将产物纯化进行测序 (Fig.3E)。序列比对结果说明,862 bp 条带只包含 *HUTP14C* 片段,因此 Primer 8 和 9 是 *HUTP14C* 的特异性产物 (Fig.3G)。上述结果说明,*HUTP14C* 的 mRNA 会影响 *HUTP14A* 的

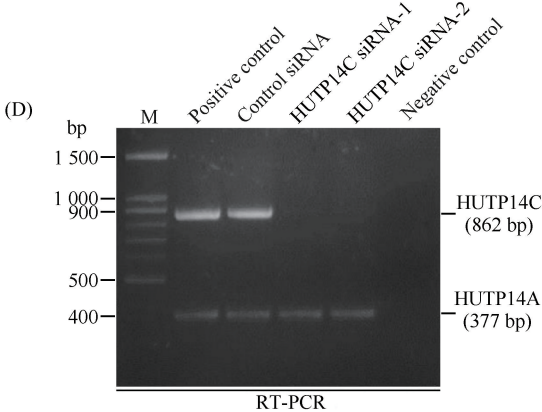
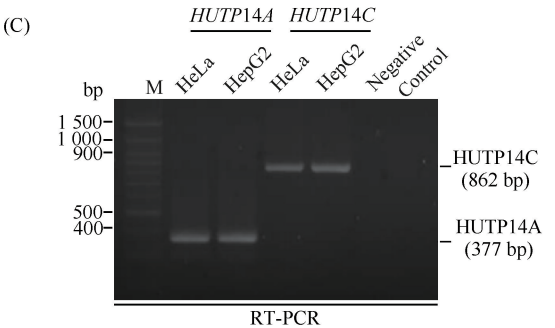
mRNA 检测。同时,本研究设计了从 RNA 水平区分肿瘤细胞系中 *HUTP14A* 和 *HUTP14C* 的引物。

2.4 假基因 *HUTP14C* 在肝癌中的表达

已有研究发现,*HUTP14A* 在肝癌组织中较癌旁组织高表达,且它的高表达和患者的不良预后相关^[25]。为了明确假基因 *HUTP14C* 是否在肝癌及癌旁组织被转录,本研究分别采用上述特异性扩增 *HUTP14A* 和 *HUTP14C* 的引物 Primer 6 和 7 及 Primer 8 和 9 进行 PCR。选取 18 对肝癌及癌旁组织提取 RNA,用 DNase I 消化处理后进行 RT-PCR,产物进行凝胶电泳 (Fig.4A)。结果显示,这 18 对肝癌组织中,约有 72% 标本呈现出 *HUTP14A* 在癌组织的表达高于癌旁的现象 (Fig.4B)。这与之前所报



(B)	<i>HUTP14A</i> ¹	CCTTCGGCTTCCGTTCTTGGTCCATGTGAGAGAAGCTGGCTGCTGAATGACTGCGAACC	60
	<i>HUTP14C</i> ⁶⁸⁷	CCTTCGGCTTCCATTTCTTGGTATACATGAGAGAGGCTGGCTGCTGATGAATGTGAACC	746
	<i>HUTP14A</i> ⁶¹	GGCTTGCAGAGACCTTCTGGCTTTGAGCCAAACAGGAAGAACTAGCGGATTGGCCAAAAG	120
	<i>HUTP14C</i> ⁷⁴⁷	AGGTTCGAGAGA—ATCTGGCTTTGAGCCACCAGGAAGAACTAGTGGATTGGCCAAAAG	803
	<i>HUTP14A</i> ¹²¹	ACTACCTCTTGAGTGAAGTGAAGATGAGGGGACAAATGATGGAGAGAGAAACATCAAA	180
	<i>HUTP14C</i> ⁸⁰⁴	ACTACCTCTTGAGTGAAAATGAAGATGAGGGGACAGTGAAGAGAGAGAAACATCAAA	863
	<i>HUTP14A</i> ¹⁸¹	AGCTTCTGGAAGCAATCAGTTCCTTGATGGAAAGAAATAGCGGAAATGGCTGAGAGGT	240
	<i>HUTP14C</i> ⁸⁶⁴	AGCTTCTGGAAGCAATCATTTCCTTGATGGAAAGAAATAGCGGAAATGGCTGAGAGGT	923
	<i>HUTP14A</i> ²⁴¹	CTGAGGCTAGTCTGAAGTGTGAGAGTTCATGTCAGTTCGAAGGATCAGGAGAAAAGC	300
	<i>HUTP14C</i> ⁹²⁴	CTGAGGCTAGTCTGAAGTGTGAGAGTTCAGTTCAGTTCGAAGGATCAGGAGAAAAGC	983
	<i>HUTP14A</i> ³⁰¹	TGGTCCTTGCAGATCTGCTTGAGCCGTGTAACAACTTCATCTCTTTGGCCACTGTGAAAA	360
	<i>HUTP14C</i> ⁹⁸⁴	TGGCCCTTGCAGATCTGCTTGAGCCGTGTAACAACTTCATCTCTTTGGCCACTGTGAAAA	1043
	<i>HUTP14A</i> ³⁶¹	AGCAACTGAGTAGAGTCAAATCAAGAAACAGTGGAGTTACCTCTGAACAAGAGAGAGA	420
	<i>HUTP14C</i> ¹⁰⁴⁴	AGCAACTGAATAGAGTCAAATCAAGAAACGTGGAGTTACCTCTTAACAAGAGAAAAA	1103
	<i>HUTP14A</i> ⁴²¹	TTGACGGATCCACAGAGAAAGTGCATTCAATAAAACCGCACAGTCCTCTCCAAATGGG	480
	<i>HUTP14C</i> ¹¹⁰⁴	TTGACAGATCCACAGAGAAAGTGCATTCAATAAAACCTCACAGTCTCTCCAAATGGG	1163



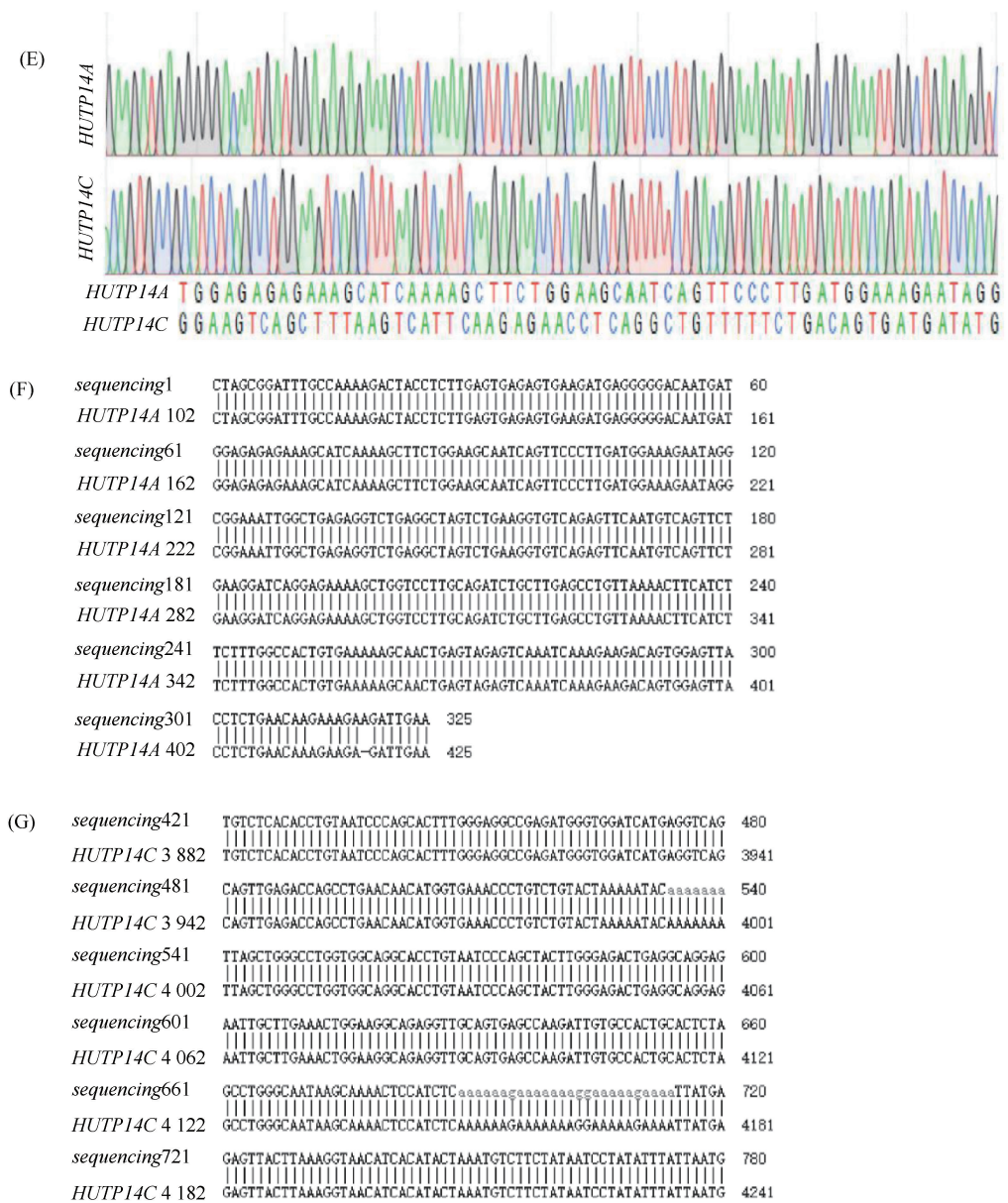


Fig.3 Differentiation of HUTP14A and HUTP14C in tumor cell lines (A) Positions of primer 6 and 7 to specifically amplify HUTP14A. Positions of primer 8 and 9 to specifically amplify HUTP14C. Positions of siRNA-1 and siRNA-2 to specifically knock down HUTP14C. (B) Specific positions of primer 6 and 7. The red boxes represent the positions of primer 6 and 7. 47-70 refers to the primer 6 located at nucleotides 47-70 of HUTP14A. 390-423 refers to the primer 7 located at nucleotides 390-423 of HUTP14A. (C) Total RNA was extracted from HeLa and HepG2 cell lines. The RNA was digested with DNase I before reverse transcription. PCR were performed using primer 6/7 and primer 8/9. The product was subjected to gel electrophoresis. (D) Total RNA was extracted from HeLa cells transfected with HUTP14C siRNAs for 72 hours. RNA was treated with DNase I before reverse transcription. PCR was respectively performed with primer 6 / 7 and primer 8/9. (E) Diagram of PCR product sequencing results. (F) Sequence alignment of HUTP14A. (G) Sequence alignment of HUTP14C

导的 HUTP14A 在肝癌中的高表达结果是一致的^[25]。假基因 HUTP14C 在这 18 对标本中,约有 55% 的标本呈现出癌组织中的表达有略高于癌旁的趋势 (Fig.4C),和 HUTP14A 的表达类似。但为了更准确地判断 HUTP14C 在癌组织中的表达及两者的表达是否相关,应选取更多的样本进行研究。这样可以准确检测两者的表达。

3 讨论

在检测肿瘤细胞中的某些标志基因进行诊断时会遇到假基因的干扰。KRASP1 是鼠肉瘤病毒癌基因 (kirsten rat sarcoma viral oncogene, KRAS) 的假基因,检测 KRAS 突变时,如何避免同源性干扰取决于实验设计^[28]。

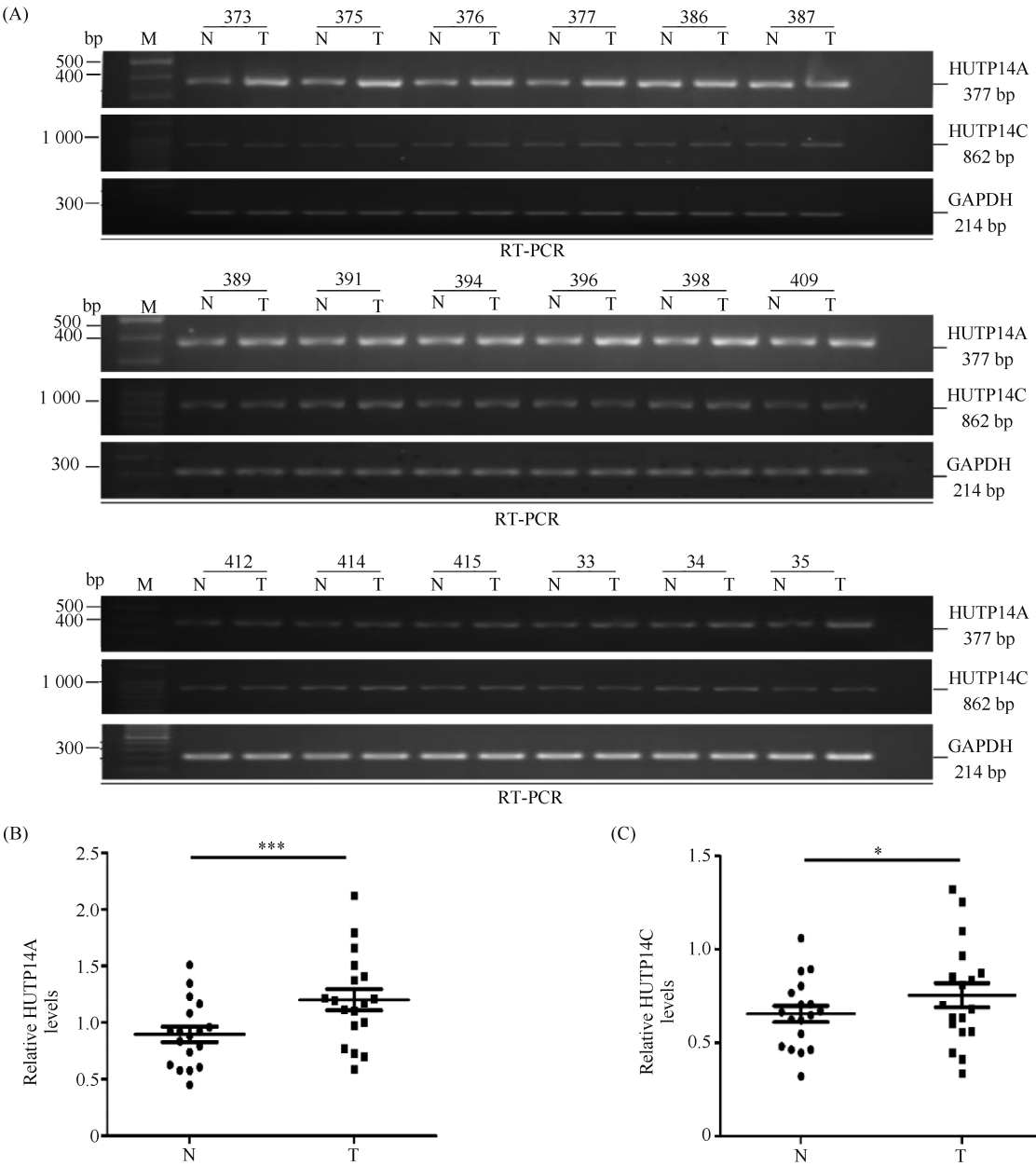


Fig.4 Expression of *HUTP14A* and *HUTP14C* in liver cancer (A) Total RNA was extracted from 18 pairs of liver cancer and adjacent tissues. The RNA was digested with DNase I before reverse transcription and then was subjected to PCR. The results were subjected to gel electrophoresis. N: non-tumor tissue; T: tumor tissue. (B,C) Grayscale scanning and data analysis. The gray value of GAPDH is used as an internal parameter. The ratio of gray value of *HUTP14A* and *HUTP14C* to GAPDH were calculated, and they were performed with statistical analysis

HUTP14A 基因全长 23 641 bp,有 15 个外显子, 14 个内含子, mRNA 全长 2 461 bp。假基因 *HUTP14C* 编码蛋白质,基因全长 8 881 bp,无内含子, mRNA 全长 5 521 bp。两者转录本差异仅为 5%,给检测表达带来困扰。hUTP14c 蛋白没有同源基因抗凋亡的作用,生物信息学分析发现,两者结合的蛋白质不同,这提示二者功能不同,但目前仍无能区分两者的抗体^[26]。因此能否用 RT-qPCR 技术准确检测两者的表达至关重要。由于假基因

HUTP14C 无内含子且有较长的 3'-UTR,本文从实验设计上避免相互干扰,结合测序技术验证本文结果。

一方面,本文用 DNase I 和假基因特异性 siRNA 排除检测干扰。本研究验证了 RNA 中 *HUTP14C* 基因 DNA 污染对 RT-qPCR 技术检测 *HUTP14A* mRNA 存在影响。提示:对于有假基因且假基因为发生转录的基因,如果用 RT-qPCR 技术定量检测 mRNA 丰度,应在反转录前用 DNase I 消化处理 RNA 中可能

混有的 DNA,用跨内含子的引物直接以 RNA 为模板通过 PCR 和凝胶电泳证明已排除可能污染的 DNA。有研究报道,假基因的 3'-UTR 通过和其功能基因竞争性结合 miRNA,从而保护基因使其不被降解。高迁移率族蛋白 (high mobility group protein, HMG) 家族的 *HMGA1* 基因有 2 个假基因 *HMGA1P6* 和 *HMGA1P7*。靶向 *HMGA1* 的 mRNA 的 miR-15, miR-16, miR-26a, miR-214, miR-548c-3p 和 miR-761 也靶向 *HMGA1P7*^[29]。这提示我们,用 RT-qPCR 技术定量研究有假基因且发生转录的基因时,要在假基因 3'-UTR 的序列非保守区设计特异性 siRNA 来排除假基因 mRNA 的干扰。

另一方面,本研究利用两者序列的差异,直接通过引物区分 2 个基因。但是,对于差异序列不集中分布的假基因,从引物上直接区分两者比较困难,这时可通过本文验证的 DNase I 和假基因特异性 siRNA 来排除假基因的干扰。本文提供了从 DNA 角度区分两者的引物设计思路,即跨越内含子设计引物。本研究中,由于 Primer 6、7 能够特异扩增 *HUTP14A*,可以在 Primer 6、7 附近寻找差异大的核苷酸进行 qPCR 引物设计,达到较为准确的定量,用于临床检测的目的,我们将在以后的研究中进行验证。

此外,本文首先在 HeLa 和 HepG2 细胞中验证了引物的特异性,然后在 18 对肝癌及癌旁组织中利用本研究设计的特异性引物定性检测了 *HUTP14A* 和 *HUTP14C* 的表达。初步发现,*HUTP14C* 在肝癌组织的表达略高于癌旁,由于样本量较小,不能得出确定的结论。已有研究发现,hUTP14a 在肝细胞癌中的高表达指示预后不良^[25]。

对于假基因的功能研究表明,一些假基因对功能基因发挥保护作用。比如,靶向 *PTEN* 的 miR19-b 和 miR20-a 同样会靶向假基因 *PTENP1*,且由于假基因 *PTENP1* 具有更长的 3'-UTR,存在 miR19-b 和 miR20-a 的靶序列,因此会抑制 miR19-b 和 miR20-a 对 *PTEN* 的降解,使其 mRNA 丰度升高^[12]。然而,假基因 *HUTP14C* 的 3'-UTR 是否会对 *HUTP14A* 的 mRNA 有保护作用有待进一步探讨,需要探索靶向 *HUTP14A* mRNA 的 miRNA 是否会同时靶向假基因 *HUTP14C* 的 3'-UTR。通过生物信息学分析发现,能同时靶向 *HUTP14A* 和 *HUTP14C* 的 miRNA,如 miR-6504-5p 和 miR-3064-5p。

另一方面,假基因 *HUTP14C* 在肝癌中的高表达是否与肿瘤相关,以及与 *HUTP14A* 在肝癌中的功

能是否相同仍是未知的,有待深入探索。研究发现,假基因可以作为肿瘤的预后及分型标记^[30]。POU 转录因子家族 OCT4 的假基因 *OCT4pg1* 在胃癌中的高表达指示预后不良^[31];类泛素蛋白修饰分子 (small ubiquitin-like modifier, SUMO) 家族的 *SUMO1* 基因的假基因 *SUMO1P3*,在胃癌中的高表达指示预后不良^[17]。本文在 18 对肝癌及癌旁组织中对 *HUTP14A* 和 *HUTP14C* 的检测也提示,*HUTP14C* 或许有和 *HUTP14A* 同样的预后不良作用,但样本量偏小,还无法得出更准确的结论。

除了采用 RT-qPCR 检测基因的转录水平外,从 mRNA 角度分析基因的表达可以有很多种不同的方法,如 Northern 印迹和基因芯片等。Northern 印迹不仅可以检测目的基因的表达,而且能识别目的片段的大小以及是否有可变剪切和重复序列。虽然假基因转录本序列和其功能基因高度相似,但其 3'-UTR 往往更长。*HUTP14A* mRNA 全长 2 461 bp,3'-UTR 只有 172 bp。而假基因 *HUTP14C* 全长有 5 521 bp,3'-UTR 却有 2 504 bp。因此,在用 Northern 印迹区分两者的 mRNA 时,一方面可以通过片段大小来判断,另一方面可以在两者序列差异较大的地方设计探针,这是一个非常准确的办法,且它较高的特异性可以有效的减少实验结果的假阳性。RT-qPCR 的优点是灵敏度较高。Northern 印迹实验的不足之处在于对实验室条件要求较高,一般临床实验室不宜采取此方法。基因芯片常与 Northern 印迹联合应用,能够大规模与高效率地一次性检测数千个基因的表达。

综上所述,本研究从引物设计和实验过程优化两个角度提出针对含有假基因的基因 mRNA 丰度检测的解决办法,同时通过以上建立的方法,在 18 对肝癌和癌旁组织中初步检测 *HUTP14A* 及 *HUTP14C* 的 mRNA 丰度,为深入探讨两者的调控及其机制奠定了基础。

参考文献 (References)

- [1] Sur S, Saha S, Tisa LS, *et al.* Characterization of pseudogenes in members of the order frankineae [J]. J Biosci, 2013, **38**(4): 727-732
- [2] Bischof JM, Chiang AP, Scheetz TE, *et al.* Genome-wide identification of pseudogenes capable of disease-causing gene conversion [J]. Hum Mutat, 2006, **27**(6): 545-552
- [3] Torrents D, Suyama M, Zdobnov E, *et al.* A genome-wide survey of human pseudogenes [J]. Genome Res, 2003, **13**(12): 2559-2567
- [4] Zhang Z, Harrison PM, Liu Y, *et al.* Millions of years of evolution preserved: a comprehensive catalog of the processed pseudogenes in the human genome [J]. Genome Res, 2003, **13**

- (13): 2541-2558
- [5] Esnault C, Maestre J, Heidmann T. Human LINE retrotransposons generate processed pseudogenes [J]. *Nat Genet*, 2000, **24**(4): 363-367
 - [6] Goncalves I, Duret L, Mouchiroud D. Nature and structure of human genes that generate retropseudogenes [J]. *Genome Res*, 2000, **10**(5): 672-678
 - [7] Krahling J, Graveley BR. The origins and implications of alternative splicing [J]. *Trends Genet*, 2004, **20**(1): 1-4
 - [8] Goncalves I, Duret L, Mouchiroud D. Nature and structure of human genes that generate retropseudogenes [J]. *Genome Res*, 2000, **10**(5): 672-678
 - [9] Li W, Yang W, Wang XJ. Pseudogenes: pseudo or real functional elements? [J]. *J Genet Genomics*, 2013, **40**(4): 171-177
 - [10] Pei B, Sisu C, Frankish A, *et al.* The GENCODE pseudogene resource [J]. *Genome Biol*, 2012, **13**(9): R51
 - [11] Fujii GH, Morimoto AM, Berson AE, *et al.* Transcriptional analysis of the PTEN/MMAC1 pseudogene, psiPTEN [J]. *Oncogene*, 1999, **18**(9): 1765-1769
 - [12] Poliseno L, Salmena L, Zhang J, *et al.* A coding-independent function of gene and pseudogene mRNAs regulates tumour biology [J]. *Nature*, 2010, **465**(7301): 1033-1038
 - [13] Han YJ, Ma SF, Yourek G, *et al.* A transcribed pseudogene of MYLK promotes cell proliferation [J]. *FASEB J*, 2011, **25**(7): 2305-2312
 - [14] Harrison PM, Zheng D, Zhang Z, *et al.* Transcribed processed pseudogenes in the human genome: an intermediate form of expressed retrosequence lacking protein-coding ability [J]. *Nucleic Acids Res*, 2005, **33**(8): 2374-2383
 - [15] Ma T, Lu C, Guo Y, *et al.* Human U3 protein 14a plays an anti-apoptotic role in cancer cells [J]. *Biol Chem*, 2017, **398**(11): 1247-1257
 - [16] Pai HV, Kommaddi RP, Chinta SJ, *et al.* A frameshift mutation and alternate splicing in human brain generate a functional form of the pseudogene cytochrome P4502D7 that demethylates codeine to morphine [J]. *J Biol Chem*, 2004, **279**(26): 27383-27389
 - [17] Mei D, Song H, Wang K, *et al.* Up-regulation of SUMO1 pseudogene 3 (SUMO1P3) in gastric cancer and its clinical association [J]. *Med Oncol*, 2013, **30**(4): 709
 - [18] Johnsson P, Ackley A, Vidarsdottir L, *et al.* A pseudogene long noncoding-RNA network regulates PTEN transcription and translation in human cells [J]. *Nat Struct Mol Biol*, 2013, **20**(4): 440-446
 - [19] Fujii GH, Morimoto AM, Berson AE, *et al.* Transcriptional analysis of the PTEN/MMAC1 pseudogene, psiPTEN [J]. *Oncogene*, 1999, **18**(9): 1765-1769
 - [20] Kalyana-Sundaram S, Kumar-Sinha C, Shankar S, *et al.* Expressed pseudogenes in the transcriptional landscape of human cancers [J]. *Cell*, 2012, **149**(7): 1622-1634
 - [21] Lopez-Girona A, Fumari B, Mondesert O, *et al.* Nuclear localization of Cdc25 is regulated by DNA damage and a 14-3-3 protein [J]. *Nature*, 1999, **397**(6715): 172-175
 - [22] Hu L, Wang J, Liu Y, *et al.* A small ribosomal subunit (SSU) processome component, the human U3 protein 14A (hUTP14A) binds p53 and promotes p53 degradation [J]. *J Biol Chem*, 2011, **286**(4): 3119-3128
 - [23] Liu H, Wang J, Liu Y, *et al.* Human U3 protein14a is a novel type ubiquitin ligase that binds RB and promotes RB degradation depending on a leucine-rich region [J]. *Biochim Biophys Acta Mol Cell Res*, 2018, **1865**(11 Pt A): 1611-1620
 - [24] Zhang J, Ren P, Xu D, *et al.* Human UTP14a promotes colorectal cancer progression by forming a positive regulation loop with c-Myc [J]. *Cancer Lett*, 2019, **440-441**: 106-115
 - [25] Zhang JY, Xu D, Liu ZZ, *et al.* Human U three protein 14a expression is increased in hepatocellular carcinoma and associated with poor prognosis [J]. *Chin Med J (Engl)*, 2017, **130**(4): 470-476
 - [26] Ma T, Lu C, Guo Y, *et al.* Human U3 protein 14a plays an anti-apoptotic role in cancer cells [J]. *Biol Chem*, 2017, **398**(11): 1247-1257
 - [27] Bradley J, Baltus A, Skaletsky H, *et al.* An X-to-autosome retrogene is required for spermatogenesis in mice [J]. *Nat Genet*, 2004, **36**(8): 872-876
 - [28] Nollau P, Jung R, Neumaier M, *et al.* Tumour diagnosis by PCR-based detection of tumour cells [J]. *Scand J Clin Lab Invest Suppl*, 1995, **221**: 116-121
 - [29] Esposito F, De Martino M, Petti MG, *et al.* HMGA1 pseudogenes as candidat proto-oncogenic competitive endogenous RNAs [J]. *Oncotarget*, 2014, **5**(18): 8341-8354
 - [30] Han L, Yuan Y, Zheng S, *et al.* The Pan-Cancer analysis of pseudogene expression reveals biologically and clinically relevant tumour subtypes [J]. *Nat Commun*, 2014, **5**: 3963
 - [31] Hayashi H, Arao T, Togashi Y, *et al.* The OCT4 pseudogene POU5F1B is amplified and promotes an aggressive phenotype in gastric cancer [J]. *Oncogene*, 2015, **34**(2): 199-208