

基于二代测序的RhD阳性个体RHD基因mRNA 选择性剪接分析

罗亚林¹, 魏玲², 姬艳丽², 罗广平², 付涌水^{1,2}, 温机智², 黎诚耀¹

(1. 南方医科大学检验与生物技术学院输血医学系, 广东广州 510515; 2. 广州血液中心临床输血研究所, 广东广州 510095)

摘要:【目的】本研究旨在对RhD阳性个体中红细胞特异性表达的RHD基因不同转录本进行分析,并对选择性剪接发生的机制进行初步探讨。【方法】利用体外有核红细胞培养体系分离并培养RhD阳性个体的有核红细胞,使用逆转录聚合酶链反应(RT-PCR)技术对RhD阳性个体RHD基因外显子6至3'非编码区扩增后进行二代测序,分析mRNA转录本,并采用生物信息学方法分析RHD基因所有外显子5'及3'剪接位点的最大熵值。【结果】RHD基因的转录本以正常全长mRNA为主,除此之外还存在其他8种异常转录本。按照表达量从多到少,分别为外显子7缺失、外显子9缺失、外显子8和9缺失、外显子7、8和9缺失、外显子7和9缺失、外显子8和9缺失合并外显子7与10间插入一段170 bp的内含子7片段、全长mRNA但插入170 bp的内含子7片段和外显子9缺失但插入一170 bp的内含子7片段,其中最后3种为新转录本。生物信息学分析提示外显子7和外显子9异常剪接可能与其5'-剪接位点(5'-splice site, 5' ss)或3' ss和剪接体结合能力下降有关,而内含子7片段异常表达于转录本中可能与其两端存在类似5' ss和3' ss剪接保守序列有关。【结论】RHD基因存在十分复杂的选择性剪接模式,使RhD阳性个体的RHD基因在mRNA水平上存在多种不同的剪接转录本。

关键词: RhD阳性; 二代测序; 转录本; 剪接

中图分类号: R331.1; R41

文献标志码: A

文章编号: 1672-3554(2021)01-0051-06

Alternative Splicing of RHD mRNA by the Next-generation Sequencing in RhD-positive Individuals

LUO Ya-lin¹, WEI Ling², JI Yan-li², LUO Guang-ping², FU Yong-shui^{1,2},
WEN Ji-zhi², LI Cheng-yao¹

(1. Department of Transfusion Medicine, Southern Medical University, Guangzhou 510515, China; 2. Institute of Clinical Blood Transfusion, Guangzhou Blood Center, Guangzhou, 510095, China)

Correspondence to: LI Cheng-yao, E-mail: chengyaoli@hotmail.com; WEN Ji-zhi, E-mail: wenjizhi@126.com

Abstract: 【Objective】 This study aimed to analyze different transcripts specifically expressed by red blood cells in RhD-positive individuals, and to explore the mechanism of alternative splicing. 【Methods】 Erythroblasts from RhD-positive individuals were isolated and cultured in an in vitro erythroblast culture system. The fragment from exon 6 to 3' untranslated coding regions of RHD gene of RhD-positive individuals was amplified by reverse transcription polymerase chain reaction (RT-PCR) method. The PCR products were then sequenced by next-generation sequencing and the mRNA transcripts were analyzed. The maximum entropy values of 5' splice site (ss) and 3' ss of all exons of RHD gene were analyzed

收稿日期: 2020-09-04

基金项目: 国家自然科学基金(81900184)

作者简介: 罗亚林, 硕士生, 研究方向: 输血相关免疫, E-mail: 812728336@qq.com; 黎诚耀, 通信作者, 博士, 教授, 研究方向: 输血传播疾病和输血相关免疫, E-mail: chengyaoli@hotmail.com; 温机智, 通信作者, 副主任技师, 研究方向: 输血相关免疫, E-mail: wenjizhi@126.com

using bioinformatics.【Results】The transcripts of the *RHD* gene were mainly normal full-length mRNA. In addition, eight other abnormal transcripts were also identified. According to the expression frequencies (high to low), these transcripts were exon 7 deletion, exon 9 deletion, exons 8 and 9 deletion, exons 7-9 deletion, exons 7 and 9 deletion, exons 8-9 deletion and 170 bp insertion between exon 7 and 10, full-length with 170 bp insertion between exon 7 and 8, exon 9 deletion and 170 bp insertion between exon 7 and 8. The last three transcripts were novel. Bioinformatic analysis suggested that the abnormal splicing of exon 7 and exon 9 might be due to the decreased binding ability of 5' ss or 3' ss to the spliceosome. The intron 7 was expressed in mRNA, which might be related to the sequences conserved with 5' ss and 3' ss splicing.【Conclusion】The alternative splicing patterns of *RHD* gene are very complicated, which leads to various transcripts at the mRNA level of the *RHD* gene from RhD-positive individuals.

Key words: RhD-positive; next-generation sequencing; transcripts; splicing

[J SUN Yat-sen Univ (Med Sci), 2021, 42(1):51-56]

Rh是除ABO之外最重要的血型系统之一,其中D抗原是Rh血型系统中免疫原性最强的抗原,其相应同种抗体可引起严重溶血性输血反应和胎儿新生儿溶血病。编码D抗原的*RHD*基因由10个外显子组成,并与编码RhCE抗原的*RHCE*基因具有高度同源性^[1]。大部分人类基因在剪接过程中能够通过选择性剪接产生多种mRNA转录本,使机体在不同组织和不同环境条件中能够利用一套固定的遗传物质表达出丰富的蛋白产物,以满足不同的生理需求^[2]。前期研究显示,*RHD*基因也同样存在选择性剪接现象,这些选择性剪接形成的转录本能否表达出不同C-端的RhD蛋白,且这些序列多样的RhD蛋白是否有完整的D抗原表位,迄今尚不清楚^[3]。以往对*RHD*基因可变剪接的研究多采用逆转录聚合酶链反应(reverse transcription polymerase chain reaction, RT-PCR)扩增后分子克隆测序,或RT-PCR后直接Sanger测序的方法进行,对于低丰度的转录本难以检出,并且无法进行准确定量分析。本研究首先从RhD阳性献血者的新鲜血液中分离出外周血单个核细胞(Peripheral blood mononuclear cell, PBMC),进行有核红细胞(Erythroblast)的体外扩增培养,提取红细胞特异性RNA,然后采用二代测序技术对*RHD*基因的cDNA扩增产物进行测序分析,以期获得*RHD*基因选择性剪接转录本的定性和定量分析结果,并使用生物信息学方法对*RHD*基因选择性剪接转录本形成机制进行初步探讨。

1 材料与amp方法

1.1 研究对象

随机选取D阳性献血者3名,分别留取新鲜

EDTA抗凝外周静脉血10 mL。本研究经广州血液中心伦理委员会同意,并获得研究对象知情同意。

1.2 前体有核红细胞的分离与体外培养

参照我们已经建立的从外周血中分离、培养并扩增前体有核红细胞的方法^[4]。采用Ficoll密度梯度法分离PBMC,使用包含1 ng/mL人白介素3(Interleukin-3, IL-3, Stem Cell公司)、2 U/mL促红细胞生成素(Erythropoietin, EPO, Stem Cell公司)、1 μmol/L地塞米松(Dexamethasone, Dex, Sigma公司)、40 ng/mL胰岛素样生长因子-1(Insulin-like growth factor-1, IGF-1, R&D公司)、40 μg/mL cholesterol-rich lipids(Sigma公司)和100 ng/mL干细胞因子(Stem cell factor, SCF, Stem Cell公司)的StemSpan SFEM培养基(Stem Cell公司)进行筛选培养,调节细胞浓度至 10^6 /mL。在第5天时使用Percoll分离液进行密度梯度离心,以纯化有核红细胞,并改用不含IL-3的相同浓度细胞因子(即2 U/mL EPO、1 μmol/L Dex、40 ng/mL IGF-1、40 μg/mL cholesterol-rich lipids和100 ng/mL SCF)的StemSpan SFEM培养基对有核红细胞进行培养,并将细胞浓度维持在 $1.5\sim 2.0\times 10^6$ /mL^[5]。

1.3 RNA的提取分离

采用QIA amp RNA Blood MiniKit试剂盒(QIAGEN GmbH, D-40724 Hilden, 德国),按照试剂盒操作说明,提取有核红细胞总RNA,测定浓度分装后,置-80℃冰箱保存备用。

1.4 RT-PCR

依据逆转录试剂盒(PrimeScript, II 1st strand cDNA Synthesis kit, TaKaRa)操作说明,采用Oligo(dT)20引物将RNA反转录为cDNA。通过引物设计软件(Premier Premier 5.0)设计1对特异性引物,

扩增RHD mRNA的外显子6到3'非编码区,即正向引物位于外显子6(引物序列:5'-TGGCTGGGCTGATCTCCG-3'),反向引物位于3'非编码区(引物序列:5'-TGCATAATAAATGGTGAGATTCTCCTC-3')。随后,用GoTaq DNA聚合酶(Promega公司,USA)进行PCR扩增。扩增条件为:94℃预变性5 min;然后94℃变性30 s,55℃退火30 s,72℃延伸1 min,共30个循环;最后72℃延伸10 min。PCR产物使用2%琼脂糖凝胶进行电泳分析。

1.5 PCR产物二代测序

将PCR产物送至成都柯莱博生物科技有限公司,建库后使用Illumina Novaseq测序仪进行二代测序。使用Hisat2-2.1.0软件对测序结果和RHD基因组序列进行比对分析,测序结果采用StringTie v2.1.1软件进行转录本拼接以及表达量计算,使用每千个碱基的转录和每百万映射读取的转录本(Transcripts Per Kilobase of exon model per Million mapped reads,TPM)对不同转录本相对表达量进行校正。

1.6 统计学方法

采用MaxEntScan软件(http://hollywood.mit.edu/burgelab/maxent/Xmaxentseq_scoreseq.html)对RHD基因各外显子的5'-剪接位点(5'-splice site,5'ss)和3'ss保守区域序列进行生物信息学分析,计算反映5'ss和3'ss与剪接体结合力的指标,即最大熵值,以推测不同RHD转录本形成的原因。根据文献^[6]报道,5'ss区域包括外显子最后3个碱基和内含子5'端的6个碱基,用该模型(Maximum En-

tropy Model,MaxENT)计算的最理想保守序列最大熵值为11.81;3'ss区域包括内含子3'端的6个碱基和外显子的前3个碱基,用该模型计算的最理想保守序列最大熵值为13.59。实际计算值越接近理想保守序列的最大熵值,表明在剪接时被剪接体识别且结合能力越强,更易被剪接。

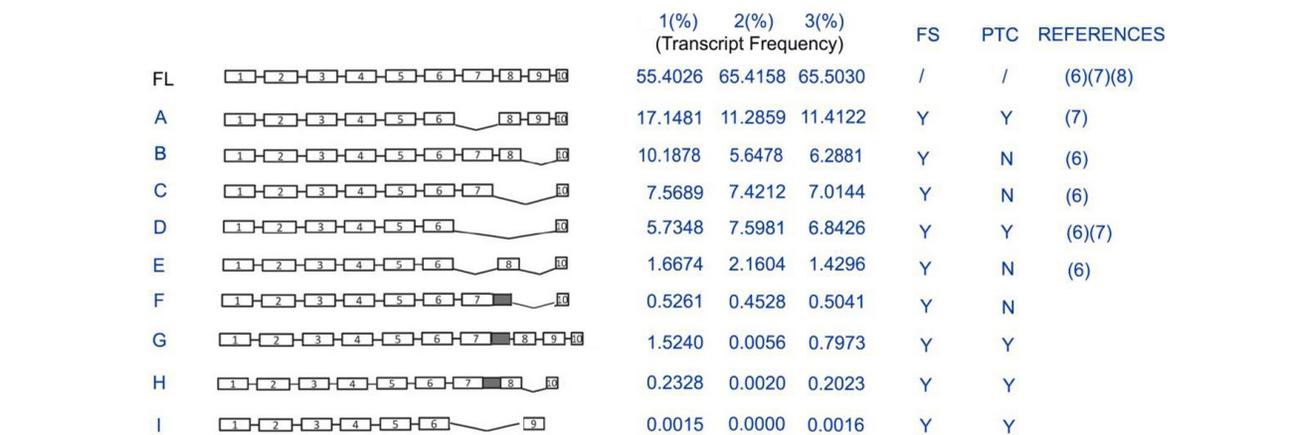
2 结果

2.1 使用二代测序对RHD基因mRNA转录本进行分析

为了解RhD阳性个体红细胞中RHD基因的各种转录本的表达情况,我们将RT-PCR产物进行二代测序。测序结果表明:在RhD阳性标本中除了包含10个外显子的RHD基因全长转录本外,还检测到了其它8种不同的RHD基因mRNA转录本(图1)。将结果进行校正后,每种转录本的相对表达量如图1所示。

2.2 生物信息学分析RHD基因各外显子5'ss与3'ss剪接区域剪接效能

对RhD阳性个体各外显子5'ss与3'ss剪接区域的序列进行了生物信息学分析,计算了10个外显子两端5'ss的9个碱基的保守序列(包括外显子最后3个碱基和紧邻内含子的前六个碱基序列)和3'ss的23个碱基长度的保守序列(包括内含子的最后20个碱基和紧邻外显子的前3个碱基序列)的最大熵值(图2),计算结果显示RHD基因外显子7的5'ss保守区域最大熵值(2.93)和外显子



FL: transcript of full-length mRNA; A-H: eight abnormal transcripts; White box with number: RHD exons1-10; Black box: 170 bp insertion from IVS7+918_1087 of intron 7; FS: frame shift; PTC: premature termination codon; Y: Yes; N: No

图1 三个RhD阳性个体中检出的不同RHD mRNA转录本及比例

Fig. 1 The expression frequencies of RHD mRNA transcripts in three RhD-positive individuals

2的3' ss保守区域最大熵值(2.83)明显低于最理想保守序列最大熵值,同时外显子7和9的3' ss保守序列的最大熵值也相对较低(分别为6.98和6.15)。这些结果提示在 *RHD* 基因剪接过程中,剪接体与这几个外显子5' ss剪接区域和3' ss剪接区域的保守序列相互识别结合和结合能力下降,导致外显子2、7和9无法被正确识别,从而出现 *RHD* 基因被异常剪接形成相应外显子缺乏的转录本。

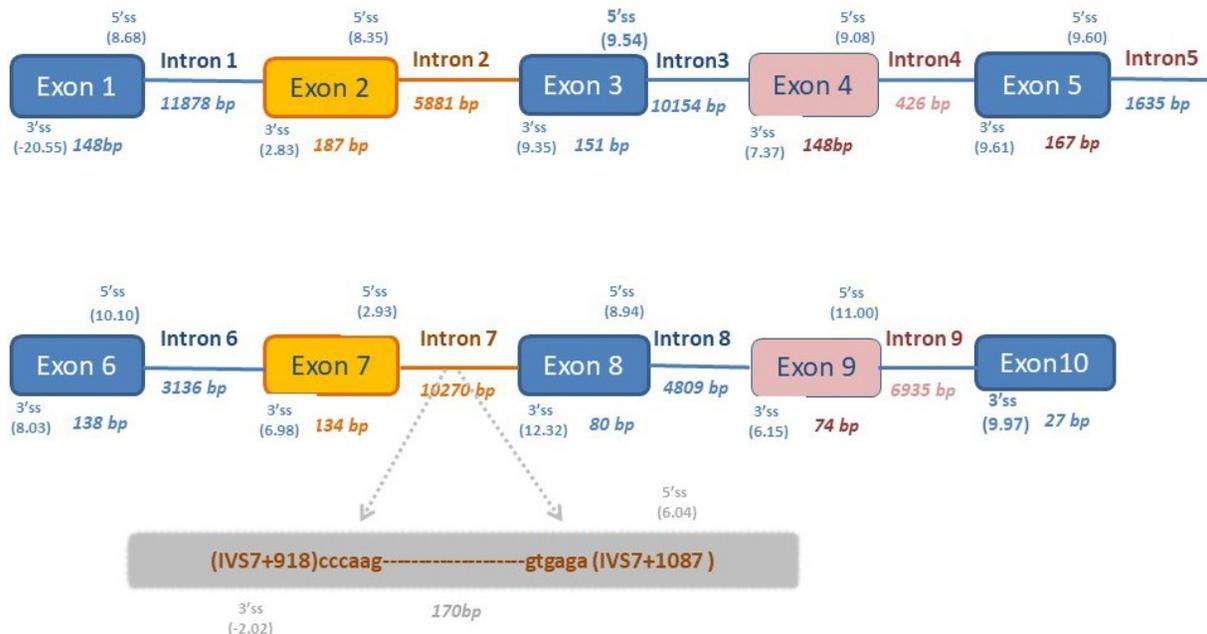
此外,我们分析了位于 IVS7+918_1087 区域(图2)的被异常插入 *RHD* 基因 mRNA 转录本的170 bp片段及其周围序列,发现其两端也存在GT-AG保守序列,即其可能会被识别为一个“外显子”。生物信息学分析结果显示这170 bp片段5' ss序列的最大熵值为6.04,高于上游紧邻的外显子7的5' ss序列的最大熵值2.93(图2),所以在剪接过程中有可能被剪接体识别并作为一个外显子被剪接到成熟的mRNA转录本中。

3 讨论

本研究采用二代测序技术,对RhD阳性个体红细胞表达的 *RHD* 基因 mRNA 转录本进行了分析,

检测到了9种不同的 *RHD* 基因 mRNA 转录本,其中包括3种新的低丰度表达转录本(图1F-H)。自从 Westhoff^[7]等明确鉴定了第一个 *RHD* 基因转录本以来,多项研究在RhD阳性个体红细胞中共发现了6种不同的 *RHD* mRNA 转录本^[3],从脐血中提取的总RNA中扩增检测到的 *RHD* mRNA 转录本可高达13种,种类更加丰富^[8]。以往研究在RhD阳性个体中报道过的转录本数目较少,这可能与这些研究采用一代Sanger测序难以检出低丰度转录本有关。本研究首次采用二代测序技术对 *RHD* 基因 RT-PCR 产物进行分析,一次能对几十万到几百万条DNA分子进行序列测序,不但提高了检测的灵敏度,而且可以计算出每种转录本的相对表达量。虽然我们采用的测序方法较传统方法在检测水平上有了较大提高,但我们推测RhD阳性个体可能仍然存在其它尚未发现的 *RHD* mRNA 转录本。

本研究仅扩增了 *RHD* 基因外显子6到3'非编码区的序列(即 *RHD* 基因 mRNA 后半部分)并进行了二代测序分析,而没有对 *RHD* 基因外显子1至外显子10进行全长扩增和二代测序分析。这是由于以往研究结果发现 *RHD* 基因的可变剪



5' ss: 5' splice sites; 3' ss: 3' splice sites

图2 *RHD* 基因各外显子5' ss端和3' ss端最大熵值计算结果示意图

Fig. 2 The maximum entropy values of 5' and 3' splice sites of all exons of *RHD* gene

接主要集中在外显子7、外显子8和外显子9^[9]。此外,考虑到目前二代测序技术测序长度有限,仅扩增RHD基因mRNA后半部分也可减少二代测序分析过程中序列拼接的次数,从而增加测序结果的可靠性。

本研究中发现多种不同RHD基因mRNA转录本,其变异形式存在外显子缺失以及内含子保留两种情况,具体表现为外显子7、外显子8和外显子9存在不同组合的缺失,以及内含子7部分序列(170 bp的片段)的保留,本研究发现了三种包含部分内含子7序列(位于RHD基因IVS7+918_1087区域)的RHD基因mRNA转录本,这3种转录本所占的比例均较低,其中3种转录本(图1F、G和H)发生移码突变,两种转录本提前出现终止密码子(图1G和H)。此外,我们进一步对RHD基因这个内含子7的170 bp片段及其周围序列进行分析,发现这170 bp的序列和既往报道的DEL个体的异常插入序列一致^[10]。另外,在长度为10 270 bp的整个内含子7中,我们关注的经常被异常插入RHD基因mRNA转录本的170 bp片段的两端碱基序列符合剪接位点的GT-AG原则。我们采用生物信息学方法模拟计算两端剪接序列的最大熵值,结果显示该170 bp片段模拟的5' ss剪接序列的最大熵值(6.04)明显大于RHD基因正常外显子7的5' ss最大熵值(2.93)。因此我们推测在RHD基因剪接过程中,与5' ss保守序列互补结合的U1 snRNP (small nuclear ribonucleoproteins)剪接体,在识别外显子7和内含子7的170 bp片段的5' ss保守序列时,U1 snRNP剪接体与内含子7的170 bp片段的5' ss相似序列结合能力更强,可能会将这170 bp的内含子7序列识别为一个“外显子”进行剪接,从而将这170 bp的内含子7序列保留在RHD基因mRNA中(图2)。

此外,以往不同研究得到的RHD基因mRNA转录本种类和数量上存在一定差异(图1),这可能是由于不同研究使用的红细胞种类不同。由于RhD蛋白在红细胞中表达而不在白细胞中表达,但成熟红细胞已经脱核无法提取到RNA,故需要从尚未脱核的红细胞中提取RNA进行RHD基因mRNA转录本的分析研究。如Shao^[11]等先从新鲜全血中提取网织红细胞后再抽提RNA,然后再进行后续分析研究;许先国^[8]等从脐带血中抽取总RNA并进行后续分析研究。与以往研究不同,

本研究中首次使用前期建立的有核红细胞的培养方法,从新鲜外周血分离、培养并扩增尚未脱核的有核红细胞,然后再提取RNA并进行后续实验,检测到的常见转录本也与以往报道一致。与前期研究相比,本研究使用的方法能大量获得红细胞总RNA,提高了检测的灵敏度,并检测到3种新的含量较少的RHD基因mRNA转录本,该方法也可以应用于其他红细胞特异性表达基因的转录分析的研究中。

为分析外显子7和9容易发生剪接异常从而形成外显子7和9缺失转录本的原因,我们对RHD基因所有外显子的5' ss和3' ss保守区域序列进行了生物信息学分析。5' ss和3' ss剪接保守序列位于内含子和外显子的交界处,是在进化过程中形成的相对保守区域,至少包括5' ss的9个碱基及3' ss的23个碱基序列,该区域中最保守的是内含子的5' GT和3' AG位点,但也存在一定的变异。MaxENT模型可对该保守区域进行分析,5' ss和3' ss区域最理想的保守序列的最大熵值分别为11.81和13.59。如果一个基因某个外显子剪接序列的数值越接近最大值,即表明其序列为一个真正的剪接位点几率越大,更易被剪接体识别而发生剪接。分析结果显示外显子7的5' ss序列最大熵值相对于其他外显子较低,即外显子7的5' ss序列在剪接起始阶段与U1 snRNP剪接体的相互结合能力可能出现显著下降,导致其不易被U1 snRNP剪接体识别并结合,从而发生异常剪接。此外,外显子9的5' ss序列最大熵值正常,仅存在3' ss序列最大熵值中度下降的情况,外显子9异常剪接是否由3' ss与其互补结合的U2 snRNP剪接体结合减弱相关,还需要进一步的研究。

总之,RHD基因存在十分复杂的选择性剪接模式,使RhD阳性个体的RHD基因在mRNA水平上存在多种不同的剪接转录本。本研究首次采用二代测序技术,检测到RhD阳性个体红细胞9种不同的RHD基因mRNA转录本,其中包括三种新的低丰度转录本,从而为进一步了解RhD蛋白的抗原表位及其在临床同种免疫中的意义奠定了基础,亦将为探讨剪接异常所致RhD变异型的分子机制研究提供帮助。这些不同的RHD基因转录本的形成机制以及其能否翻译出含正常D抗原表位的RhD蛋白,目前尚不清楚,仍需进一步研究。

参考文献

- [1] Tounsi WA, Madgett TE, Avent ND. Complete RHD next-generation sequencing: establishment of reference RHD alleles [J]. *Blood Adv*, 2018, 2(20): 2713-2723.
- [2] Auboeuf D. Alternative mRNA processing sites decrease genetic variability while increasing functional diversity [J]. *Transcription*, 2018, 9(2): 75-87.
- [3] Chen DP, Sun CF, Ning HC, et al. Comprehensive analysis of RHD splicing transcripts reveals the molecular basis for the weak anti-D reactivity of Del-red blood cells [J]. *Transfus Med*, 2016, 26(2): 123-129.
- [4] Ji Y, Veldhuisen B, Ligthart P, et al. Novel alleles at the Kell blood group locus that lead to Kell variant phenotype in the Dutch population [J]. *Transfusion*, 2015, 55(2): 413-421.
- [5] 温机智,魏玲,王贞,等.用于研究RHD等位基因的有核红细胞体外表达系统的建立[J].*中国输血杂志*, 2018, 31(11): 1240-1242.
- Wen JZ, Wei L, Wang Z, et al. Establishment of a erythroblast expression system in vitro for RHD allele studies [J]. *Chin J Blood Transfusion*, 2018, 31(11): 1240-1242.
- [6] Eng L, Coutinho G, Nahas S, et al. Nonclassical splicing mutations in the coding and noncoding regions of the ATM gene: maximum entropy estimates of splice junction strengths [J]. *Hum Mutat*, 2004, 23(1): 67-76.
- [7] Westhoff CM, Wylie DE. Identification of a new RHD-specific mRNA from K562 cells [J]. *Blood*, 1994, 83(10): 3098-3100.
- [8] 许先国,吴俊杰,洪小珍,等.鉴定9个新的RHD基因 mRNA 可变剪接体[J].*遗传*, 2006, 28(10): 1213-1218.
- Xu XG, Wu JJ, Hong XZ, et al. Identification of nine novel alternative splicing isoforms of RHD mRNA [J]. *Hereditas (Beijing)*, 2006, 28(10): 1213-1218.
- [9] Sakata N, Yamazaki K, Kogure T, et al. Alternative splicing of Rh blood group polypeptide mRNA produces a novel transcript containing a short nucleotide insertion on human erythroleukemia K562 cells [J]. *Cell Biol Int*, 2001, 25(7): 697-703.
- [10] Richard M, Perreault J, Constanzo-Yanez J, et al. A new DEL variant caused by exon 8 deletion [J]. *Transfusion*, 2007, 47(5): 852-857.
- [11] Shao CP, Xiong W, Zhou YY. Multiple isoforms excluding normal RhD mRNA detected in Rh blood group Del phenotype with RHD 1227A allele [J]. *Transfus Apher Sci*, 2006, 34(2): 145-152.

(编辑 余菁)