

剪接突变与单基因遗传病

唐娇¹ 郝胜菊^{2,3} 杨蕊琼^{2,3} 王玉佩^{2,3*} 惠玲^{2,3*}

(¹甘肃中医药大学公共卫生学院, 兰州 730000; ²甘肃省妇幼保健院医学遗传中心, 兰州 730050;

³甘肃省出生缺陷与罕见病临床研究中心, 兰州 730050)

摘要 pre-mRNA剪接是真核基因表达中的关键步骤, 正确的pre-mRNA剪接高度依赖于剪接体对外显子的正确识别, 这一过程受参与剪接的“顺式”作用元件和“反式”作用因子的调控。剪接突变可造成mRNA中外显子的跳跃或片段丢失以及内含子片段或假外显子(Pseudo exon, PE)被包含等结果, 因此研究认为剪接突变与许多单基因遗传病相关。该文综述了剪接过程中顺式作用元件与反式作用因子的相互作用机制, 分析了不同类型的剪接突变导致单基因遗传病的致病原因, 总结了目前针对剪接突变所致单基因遗传病的验证方法与治疗手段。剪接突变致病机制的深入研究对单基因遗传病的临床诊断与针对性治疗药物的研发具有重要意义。

关键词 RNA剪接; 剪接突变; 单基因遗传病

Splicing Mutation and Single Gene Inheritance Disease

TANG Jiao¹, HAO Shengju^{2,3}, YANG Ruiqiong^{2,3}, WANG Yupei^{2,3*}, HUI Ling^{2,3*}

(¹School of Public Health, Gansu University of Chinese Medicine, Lanzhou 730000, China;

²Medical Genetics Center, Gansu Provincial Maternity and Child-Care Hospital, Lanzhou 730050, China;

³Gansu Provincial Clinical Research Center for Birth Defects and Rare Diseases, Lanzhou 730050, China)

Abstract pre-mRNA splicing is a key step in eukaryotic gene expression. Correct pre-mRNA splicing is highly dependent on the correct recognition of exons by spliceosomes, which is regulated by “cis” action elements and “trans” action factors involved in splicing. Splicing mutations can cause skipping or fragment loss of exons in mRNA and inclusion of intron fragments or pseudo exons. Therefore, studies suggest that splicing mutations are associated with many monogenic genetic diseases. This article reviews the interaction mechanism between cis-acting elements and trans-acting factors during splicing, analyzes the causes of monogenic genetic diseases caused by different types of splicing mutations, and summarizes the current verification methods and treatment methods for monogenic genetic diseases caused by splicing mutations. The in-depth study on the pathogenic mechanism of splicing mutation is of great significance for the clinical diagnosis of single gene inheritance disease and the research and development of targeted therapeutic drugs.

Keywords RNA splicing; splicing mutation; single gene inheritance disease

随着高通量测序技术的深入发展、生物信息学算法的不断升级, 以及临床基因诊断标准的不断完善, 越来越多的导致人类疾病的基因变异被

报道出来。但仍然有一部分遗传性疾病无法仅通过基因检测与目前的致病性评价标准来进行临床诊断, 发生在基因内含子与外显子上的剪接突变

收稿日期: 2022-03-10 接受日期: 2022-05-30

国家自然科学基金(批准号: 12005042)、甘肃省科技计划(批准号: 21JR7RA680、21JR1RA045)和甘肃卫生行业科研计划(批准号: GSWSKY2021-021)资助的课题

*通讯作者。Tel: 18919943232, E-mail: wang_yupei_jade@163.com; Tel: 0951-5188561, E-mail: zyhuil@hotmail.com

Received: March 10, 2022 Accepted: May 30, 2022

This work was supported by the National Natural Science Foundation of China (Grant No.12005042), the Gansu Provincial Science and Technology Program (Grant No.21JR7RA680, 21JR1RA045), and the Gansu Health Industry Scientific Research Program (Grant No.GSWSKY2021-021)

*Corresponding authors. Tel: +86-18919943232, E-mail: wang_yupei_jade@163.com; Tel: +86-951-5188561, E-mail: zyhuil@hotmail.com

是其重要原因之一。研究表明, pre-mRNA剪接过程异常可导致多种单基因遗传疾病, 包括心脏病、神经肌肉病以及神经退行性疾病等^[1-3], 这提示剪接突变导致的剪接异常是相关单基因遗传病的发病机制。本文简要综述了目前剪接突变的类型及其对pre-mRNA剪接的影响, 结合已报到的单基因遗传病, 从分子生物学层面解释了剪接突变的致病机制, 总结了剪接突变目前的验证方法以及针对剪接突变所致疾病的药物的研发应用, 关注了分子生物学在临床应用中的难点, 以期提高此类疾病在临床中的诊断率, 并为研发相关治疗药物提供新方向。

1 pre-mRNA剪接概述

pre-mRNA剪接是真核基因表达的基本步骤, 它精确地去除非编码区, 并将编码区连接为成熟mRNA, 这个过程被称为剪接^[4]。pre-mRNA剪接一般包括组成性剪接和选择性剪接^[5]。真核基因转录产生的pre-mRNA如果是按组成性剪接这一种方式进行剪接, 则产生一种成熟的mRNA分子, 只翻译为一种蛋白质, 但这无法解释人类在蛋白质组水平上表现出的极高的复杂性, 因此人类的基因大多数都通过选择性剪接进行表达^[6-7]。选择性剪接是在转录后水平通过不同的剪接方式对基因表达进行调控的过程, 使得pre-mRNA产生多种不同的成熟mRNA剪接异构体, 从而翻译出具有不同结构和功能的蛋白质^[5]。其中剪接反式因子通过组织特异性及发育阶段特异性的表达与pre-mRNA上的特定剪接顺式元件结合, 使得选择性剪接和组成性剪接的相互调节成为可能^[8]。通过高通量测序技术研究发现, 选择性剪接这一过程存在于>95%的人类多外显子基因中^[9], 且在神经系统中尤为普遍。果蝇中的*Dscam1*基因是选择性剪接的一个极好的例子, 它可以通过4个外显子簇的互斥剪接产生38 016种不同的剪接异构体和蛋白质亚型^[10]。因此, 选择性剪接有效地增加了转录组和蛋白质组的多样性, 并显著影响了多种细胞过程, 同时选择性剪接具有组织特异性, 可参与胚胎发育和疾病的发展^[11]。选择性剪接事件包括盒式外显子包含或跳过、内含子保留、互斥外显子、5'选择性剪接、3'选择性剪接、可变起始外显子剪接和可变末端外显子剪接^[9,12]。

2 剪接的过程

剪接过程高度依赖于剪接体对外显子的正确识别。剪接体是由5个小核RNAs(small nuclear RNAs, snRNAs)和数百种不同的蛋白质组装而成的小核核糖核蛋白复合物(snRNPs, U1、U2、U4-U6), 可有效地识别pre-mRNA中的外显子-内含子边界。剪接过程的精确度是通过协调一系列的RNA-RNA、RNA-蛋白质和蛋白质-蛋白质相互作用来调控的^[13]。剪接体协调这些分子运动, 通过催化一系列复杂反应去除内含子并将外显子连接在一起^[13]。外显子与内含子边界有一高度保守的序列, 作为剪接识别信号, 即在内含子的5'端和3'端含有GT-AG经典的剪接位点(98.7%), 此外还存在1.20%的GC-AG和0.06%的AT-AC以及0.09%的非经典的剪接位点^[14]。内含子5'和3'末端的剪接位点的保守序列分别叫做剪接的供体位点(5'ss)与受体位点(3'ss), 是剪接的“顺式”元件, 可被snRNAs识别, 形成互补的RNA-RNA复合物^[15]。顺式元件除了经典剪接供体、受体位点外, 还包括剪接分支位点(branch point, BP)和多嘧啶序列(polypyrimidine tract, PPT)等保守序列, 可以指导剪接体成分与pre-mRNA之间的相互作用^[16]。BP是一个短的保守序列, 包含剪接受体位点上游10~50个核苷酸之间的腺苷残基, 而PPT由15~20个嘧啶核苷酸组成, 位于剪接受体位点上游的5~40个核苷酸之间^[17]。除了经典的剪接位点信号外, 还有外显子剪接增强子(exon splicing enhancer, ESE)、外显子剪接沉默子(exon splicing silencer, ESS)、内含子剪接增强子(intron splicing enhancer, ISE)和内含子剪接沉默子(intron splicing silencer, ISS)等辅助元件, 能通过结合反式作用因子调节剪接体选择性使用剪接位点的能力^[10]。调控组成性剪接和选择性剪接的主要剪接因子包括异质性核糖核蛋白(heterogeneous nuclear ribonucleoprotein, hnRNP)和富含丝氨酸和精氨酸的蛋白质(SR蛋白)。其中hnRNP是主要的负调控因子, 与ESS或ISS结合后, 两者相互作用可抑制剪接过程中剪接体对附近剪接位点的识别^[18]。ESE和ISE通常为一段保守的序列, 与SR蛋白结合后, 可促进剪接体对附近剪接位点的识别, 从而使增强子所在的外显子或内含子被包含^[17]。比如, 对人类基因组数据的分析和研究结果表明, 内含子3'端存在GGAGAA基序, 它可能是一个潜在的增强

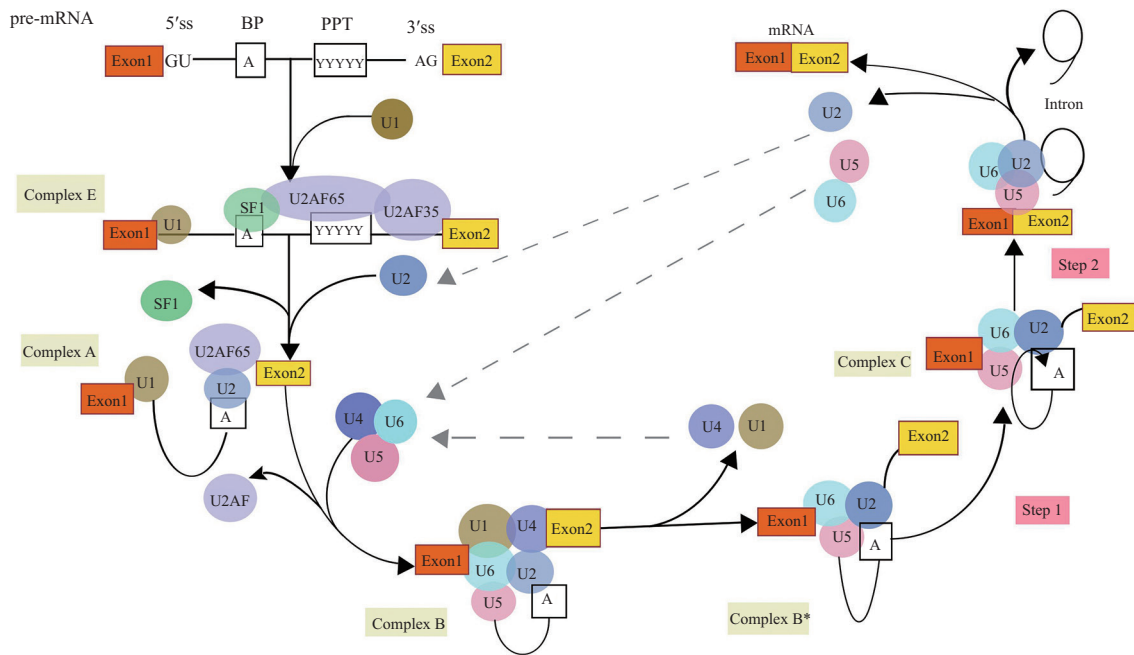


图1 pre-mRNA剪接的过程(根据参考文献[20]修改)

Fig.1 The process of pre-mRNA splicing (modified from the reference [20])

子元件,而SRSF1作为具有代表性的SR蛋白之一,可以识别该增强子序列并促进剪接体对该增强子附近剪接位点的识别^[19]。

RNA剪接包括第一步识别外显子-内含子连接处的剪接位点,第二步内含子的去除和外显子末端的连接^[20]。首先,U1 snRNP识别并结合于供体剪接位点(5'ss),而U2 snRNP通过与BP结合来识别3'ss。U1 snRNP对5'ss识别的同时还可通过跨外显子的相互作用促进3'ss被识别^[21]。在高等真核生物中BP已经完全退化,因而对U2 snRNP的招募需要多种辅助因子,其中最重要的辅助因子之一是U2辅助因子(U2 auxiliary factor, U2AF),它由U2AF65和U2AF35亚基组成。其中U2AF65可与PPT相结合^[15],同时U2AF65可与SF1蛋白相互作用并协助SF1蛋白与BP结合。此外,剪接因子U2AF35与受体位点结合,形成了早期剪接体复合物E。随后,剪接体复合物E中的SF1被U2 snRNP替代而从分支位点移除,由此形成剪接体复合物A。U4/5/6 Tri-snRNP的加入使复合物A空间结构发生重排,将3个剪接位点(5'ss、3'ss和分支位点)聚集到一起,这时复合体B得以形成。之后剪接体通过一系列构象的改变,导致U1和U4 snRNP的释放,U6 snRNP与U2 snRNP通过snRNP配对,形成pre-mRNA环和C复合体。在RNA剪接过程中,发生了两次酯交换反应。第一次转酯反应将5'ss

与分支位点连接,切开RNA链形成套索,第二次转酯反应为第一步切下的外显子攻击3'ss,使得外显子连接形成成熟的mRNA,内含子以套索结构被释放降解(图1)。

3 不同类型的剪接突变与单基因遗传病

剪接过程极其复杂,且关键位点的突变会造成剪接位点的错误识别。剪接突变一般是指发生在顺式作用元件上的突变(如核苷酸替换、插入和缺失)。该突变可影响pre-mRNA外显子和内含子的正确识别,导致错误剪接事件,从而改变mRNA序列与蛋白质的结构,影响蛋白质的功能。通常,剪接突变既可以发生在内含子中也可以发生在外显子中,可破坏现有的剪接位点,产生新的剪接位点,或者激活隐蔽的剪接位点。某些点突变还可以影响剪接增强子和沉默子与mRNA特定区域的结合或改变mRNA的二级结构,从而阻止“顺式”剪接元件与“反式”剪接因子的结合,导致mRNA外显子整个或部分跳跃^[20]。如果跳跃未使mRNA开放阅读框移位,则mRNA翻译合成较短的蛋白质,若开放阅读框移位,则mRNA可能引入提前终止密码子(pre-mature termination codon, PTC),产生较短且氨基酸序列错误的蛋白质。但是,转录本中PTC的存在通常会导致无义介导的mRNA降解(nonsense-mediated mRNA

decay, NMD)这一过程出现,即异常的mRNA快速降解,从而阻断错误蛋白的合成^[17,20]。

剪接突变可包括发生在经典剪接位点的突变(图2A)、影响剪接的内含子突变(图2B)和影响剪接的外显子单核苷酸突变(图2C)。剪接突变可导致整个外显子跳跃或外显子片段丢失,以及内含子或假外显子包含等结果^[22]。BP或PPT序列中的基因突变可以影响剪接体对外显子的正确识别,导致外显子跳跃,如果BP或PPT中的基因突变导致新的3'ss产生,则新的3'ss被剪接体识别后参与剪接过程,导致内含子保留(完整或部分)^[23]。

3.1 经典剪接位点的突变

3.1.1 经典剪接位点突变的概述 经典供体和受体剪接位点的突变,会影响剪接体对外显子-内含子边界的识别,且大多数剪接突变都是在经典剪接位点上被发现的^[24],其中经典供体剪接位点突变比经典受体常见^[25-26]。外显子跳跃是经典剪接位点突变导致的最常见的结果,且相关病例表明,同一位点的剪接突变可以产生一个或多个外显子跳跃的异常剪接转录本^[27](图2A)。另外,如果经典剪接位点的突变激活了邻近外显子或内含子中“更强”的隐蔽的剪接位点,使得隐蔽外显子或内含子剪接位点成为替代位点用于剪接过程,则可能会导致内含子片段的包含或外显子片段的去除^[20,28]。通过相关研究分析

发现,隐蔽剪接位点具有与真实剪接位点高度相似的序列,而对人类5'ss序列的测序结果表明,在GU二核苷酸的-3至+5区域的核苷酸序列变异有9 000多个,剪接位点的周围序列元件非常短且保守性很差,这使得剪接位点识别和选择具有可塑性^[29]。剪接体如何区分那些真实剪接位点与隐蔽剪接位点?首先,5'ss剪接位点-3至+5位与U1 snRNA的+4至+11位形成典型碱基对,序列互补性越强,与U1 snRNP结合的亲和力越高,当出现其他隐蔽5'ss与之竞争时,由于隐蔽5'ss与U1 snRNA的5'端序列配对互补性相对较弱^[30],使得较强的5'ss被剪接体定义为真实的剪接位点。剪接过程中被激活的隐蔽5'ss,只需要遵循U1碱基配对规则且双链(5'ss:U1 snRNA)的稳定性高,就可以替代原有的经典5'ss参与剪接^[30-31]。其次,研究发现内含子隐蔽的5'ss的激活程度与其到上游内含子3'ss的距离或上游外显子的长度呈负相关,因此外显子长度本身可能是决定剪接位点选择的一个因素^[31]。而受体剪接位点的正确识别几乎取决于相应的供体剪接位点的正确识别^[32]。

在剪接过程中,当经典剪接位点发生突变,且附近没有可选择用的剪接位点时,该外显子通常会被跳过(剪接体选择下一个剪接位点),但在其附近至少有一个隐蔽的剪接位点存在的情况下,剪接体通常会识别该剪接位点^[33]。另外,剪接位点邻近的侧翼序列、

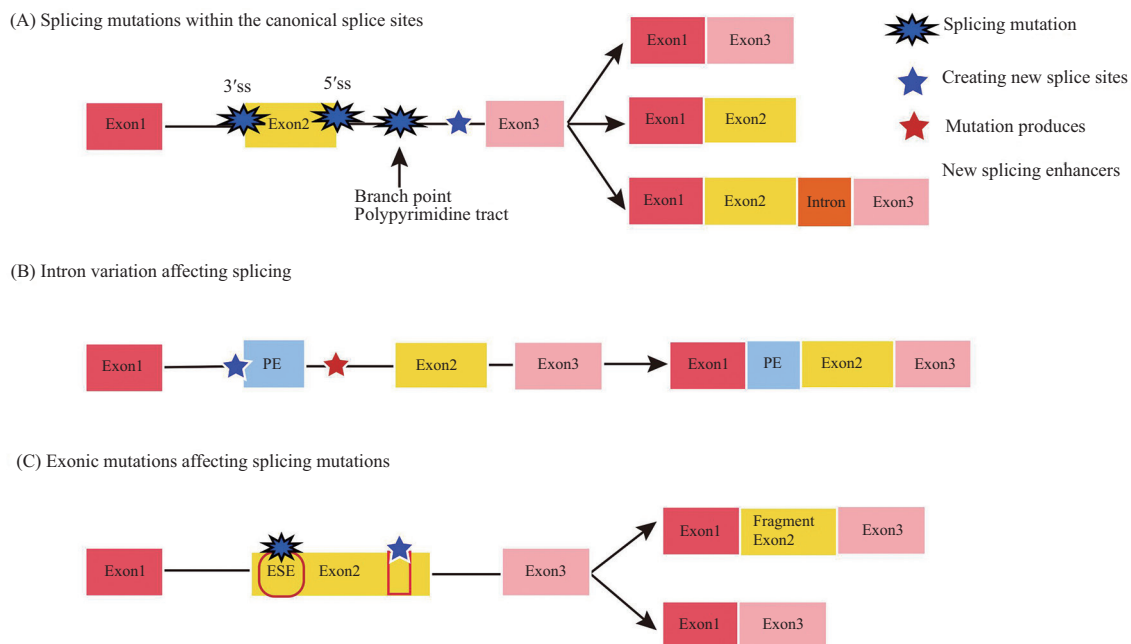


图2 剪接突变类型

Fig.2 Splicing mutation types

顺式调控元件和剪接因子,如SR和hnRNP蛋白、能稳定双链(5'ss:U1 snRNA)的U1C蛋白^[34]、在剪接体组装或催化的后期阶段发挥作用的U5 snRNP特异性蛋白Prp8^[35]等相关底物浓度的变化,也会影响剪接体对相互竞争中的5'ss的识别^[30]。

3.1.2 经典剪接位点突变致单基因遗传病 *PRUNE1* 突变是导致儿童神经退行性疾病的一种病因,患者通常表现为严重的神经发育障碍,并伴有小头畸形、低眼压和脑畸形等临床表现。测序证实先证者存在*PRUNE1* c.521-2A>G变异,该变异为内含子4的经典剪接受体位点突变,该*PRUNE1* c.521-2A>G可直接影响剪接或通过激活内含子隐蔽受体位点影响剪接过程。这种变异产生了两个异常的mRNA产物,剪接产物去除了外显子5或外显子4和5。外显子5的缺失维持了mRNA阅读框,mRNA翻译为截断蛋白。外显子4和外显子5的跳过会破坏mRNA阅读框,导致外显子6的阅读框移位并引入PTC^[36]。*SLC25A42* 基因编码线粒体辅酶A转运蛋白,该基因突变可导致患者临床表现为肌病、发育迟缓、乳酸酸中毒和脑病,先证者中发现*SLC25A42* c.380+2T>A,该变异通过影响内含子5的经典剪接供体位点,并激活内含子5中的隐蔽的剪接位点,产生了包含内含子5片段的异常转录本^[37]。遗传性压力易感性周围神经病(hereditary neuropathy with liability to pressure palsies, HNPP)是一种常染色体显性遗传性疾病,可由*PMP22*基因突变引起,常表现为无痛性、复发性、局灶性脱髓鞘神经病,可导致麻木、肌肉无力和萎缩^[38]。在一先证者身上发现*PMP22* c.179-1G>A,该变异为剪接受体位点突变,导致患者的*PMP22*基因的转录本缺少外显子4^[39]。此外,在另一个HNPP患者中检测到*PMP22*基因c.178+2T>C变异,该变异通过破坏剪接供体位点来影响剪接^[40]。遗传性痉挛性截瘫是一种具有临床及遗传异质性的神经系统遗传病,表现为缓慢进展的双下肢无力及痉挛性截瘫。在先证者中检测到*REEP1* c.303+2T>A变异为内含子4剪接供体位点突变,该突变导致外显子4的跳跃和移码,并在外显子5中引入PTC,通过NMD使得*REEP1*转录本减少^[41]。

3.2 发生在内含子中的剪接突变

3.2.1 内含子剪接突变的概述 深内含子突变同样也可以通过产生并激活隐蔽剪接位点,同时削弱经典剪接位点的识别,影响pre-mRNA的剪接。隐蔽剪

接位点的激活可发生在现有经典剪接位点附近,从而与之竞争^[26],导致成熟转录本中包含内含子片段,即隐蔽外显子或假外显子^[42]。此外,内含子上的基因突变可产生剪接增强子,或破坏剪接沉默子序列,从而促进剪接体对该假外显子的识别,导致错误剪接(图2B)。若转录本中的假外显子保留了正确的阅读框,假外显子就会成为蛋白质的一部分,相反若假外显子引入了PTC,则生成的转录本很可能会通过NMD被降解^[17]。

3.2.2 内含子中的剪接突变致单基因遗传病 内含子突变已成为人类某些单基因遗传病的病因之一,发现新的内含子突变并确定它们的致病性是至关重要的,因此需要对基因组进行测序并结合患者病变组织中产生的mRNA进行研究^[43]。位于*POLR3A*基因内含子深处的突变是遗传性痉挛性截瘫和小脑性共济失调的常见病因。在一先证者*POLR3A*的内含子中存在突变c.1909+22G>A,该突变可激活隐蔽的剪接位点,并导致内含子14的部分片段被包含在该基因的转录本中。分析发现,此内含子突变可激活隐蔽剪接位点,在分化成熟的神经上皮细胞中,隐蔽剪接位点激活程度比诱导多能干细胞更高,并且隐蔽剪接位点的激活程度与组织类型和发育阶段相关^[44]。以智力障碍、肌张力减退为特征的Cohen综合征(Cohen syndrome, CS)是常染色体隐性遗传性疾病,与其相关的*VPS13B*基因内含子中发生了c.3446-23T>G突变,导致位于上游的隐蔽剪接受体位点被激活使用,使得mRNA包含了该内含子22 bp的碱基片段^[45]。然而,大多数保留内含子的异常转录本会因引入PTC而通过NMD被降解^[46]。

Becker肌营养不良症(Becker muscular dystrophy, BMD)是由编码肌营养不良蛋白(dystrophin)的*DMD*基因突变引起的X连锁隐性遗传病^[47]。在一例BMD患者中发现*DMD*基因26号内含子点突变c.3603+2053G>C形成了一个新的与外显子定义有关的ESE位点,该位点可以与SR蛋白结合,进而通过直接招募剪接体来促进假外显子的识别,从而在外显子26和27之间插入80 bp碱基,结果导致新的假外显子插入到mRNA中。对插入的突变序列分析显示,这80 bp碱基来源于*DMD*基因外显子26下游的位置(c.3603+2033_3603+2112)。在*DMD* c.3603+2031处有一个5'ss(GT),在*DMD* c.3603+2113处有一个3'ss(AG),从而造成插入的80 bp碱基被剪接体识别

为一个假外显子^[47]。

随着全基因组测序(whole-genome sequencing, WGS)的使用,已经发现了许多深内含子变异^[48]。对不明致病突变进行分析时,应考虑隐蔽剪接位点激活,可进一步对转录本进行检测,因为致病性深内含子变异并不罕见,且足以导致严重的、危及生命的疾病。因此,当全外显子组测序(whole-exome sequencing, WES)不能明确患者的致病变异时,应检测可疑疾病相关基因的内含子^[26]。因为非编码区(如深度内含子)的突变通过WES很难被检测到,需要通过WGS进一步检测^[15]。

3.3 发生在外显子中的剪接突变

3.3.1 外显子剪接突变的概述 外显子变异通常被认为通过破坏剪接增强子导致剪接体对典型剪接位点的识别显著降低^[49],造成外显子跳跃,从而影响pre-mRNA加工。然而,外显子单核苷酸变异也可以通过产生新的剪接位点或激活隐蔽的剪接位点而参与剪接过程,导致pre-mRNA的剪接异常和部分编码序列的丢失(图2C)。

3.3.2 外显子剪接突变致单基因遗传病 肢带型肌营养不良(limb girdle muscular dystrophy, LGMD)的特征是肩部和盆腔腰部肌肉呈进行性无力和萎缩。编码 α -肌聚糖的SGCA基因突变可使相关蛋白功能丧失,导致LGMD^[50]。在一例LGMD患者中检测到SGCA c.157G>A突变,分析其为外显子剪接突变,突变位置为外显子2的最后一个碱基,该变异导致外显子2的跳跃或外显子2和3的跳跃。MaxEntScan分析软件预测了SGCA外显子2的5'ss的强度,发现SGCA c.157G>A的突变大大降低了5'ss的强度^[51]。脊髓性肌萎缩症(spinal muscular atrophy, SMA)是以脊髓前角 α 运动神经元退化变性导致的肌无力和肌萎缩为主要临床特征的疾病,是儿童遗传肌肉功能障碍最常见的死亡原因。95%的SMA患者的SMN1基因存在外显子7纯合缺失,使得生存运动神经元蛋白(survival of motor neuron, SMN)缺乏^[52]。在SMN1基因第7外显子中的罕见变异(c.863G>T)可影响SMN1外显子7的剪接,该突变破坏了外显子7中的ESE,使得识别ESE的“反式”因子Tra2 β 1蛋白与ESE的相互作用减弱,导致mRNA中外显子7的跳跃^[53]。

4 剪接突变的验证

剪接受体和剪接供体位点的突变几乎总是导

致错误剪接,因此通常被认为是有害变异,无需进一步验证^[17]。然而,在经典剪接供体GU和受体位点AG的二核苷酸附近或者之外的突变的致病性仍不确定,所以需要进一步验证它们对剪接的影响。其中直接分析天然转录本是最可靠的方法,但仅限于该基因在易于获得的组织(如患者外周血、皮肤、肌肉组织等)中表达,或者这些组织或细胞类型在体外可以从诱导多能干细胞(induced pluripotent stem cells, iPSCs)中诱导出来。

4.1 直接分析转录本

确定所选变异是否影响剪接的最简单和最有效的方法是分析从相关患者表达目的基因的组织或细胞衍生的细胞系中提取mRNA,再通过对逆转录聚合酶链反应(RT-PCR)后的RNA/cDNA进行测序,可以验证所选变异是否影响mRNA序列,但这种方法的问题是存在NMD的可能性^[20]。这使得潜在的剪接突变的影响很容易被忽略,为了避免这一缺点,可以用NMD抑制剂处理,阻止RNA降解,再进行cDNA测序^[54]。

4.2 计算机预测工具

通过计算机预测工具来验证剪接变异是一种辅助方法。一般来说,剪接位点预测工具在预测剪接位点的危害性时,其敏感性相对高于特异性^[55]。这里我们总结部分常用的工具及其功能区域(表1)。不同的程序基于不同的假设和算法,建议使用者在分析候选剪接突变的有害性时,使用多种预测工具,从而提高预测的准确性。需要强调的是突变的确切影响应该在进一步的功能研究中进行验证。

4.3 剪接突变的功能分析

minigene是一种外显子捕获的体外杂交系统,可用来模拟并分析难以取材或无法获得的在目的组织细胞中表达的相关基因变异体对剪接过程的影响。在体外,利用minigene技术构建含有基因变异位点的外显子及内含子的质粒载体,将其转染至真核细胞,再运用RT-PCR及基因测序技术分析基因变异对mRNA序列的影响。临床应用,常对白血细胞或成纤维细胞中低表达的基因进行剪接突变分析^[56]。minigene结果可以分析剪接过程中隐蔽剪接位点有无激活,有无导致相关外显子跳跃或内含子片段的包含等。虽然minigene系统的结果能用于分析相关临床疾病的发病机制,但该最终结果也不能完全等同于目的组织的剪接异常情况。

表1 部分预测相关工具及网址
Table 1 Some prediction tools and websites

工具 Tools	功能或区域 Function or region	网址 Websites
SpliceView	5'ss, 3'ss	http://bioinfo.itb.cnr.it/oriel/splice-view.html
MaxEntScan	5'ss, 3'ss	http://hollywood.mit.edu/burgelab/maxent/Xmaxentscan_scoreseq.html
SVM-BP Finder	BP, PPT	http://regulatorygenomics.upf.edu/Software/SVM_BP/
IntSplice	BP, PPT	https://www.med.nagoya-u.ac.jp/neurogenetics/IntSplice_v1.0/index.php
ESE Finder	ESE	http://krainer01.cshl.edu/cgi-bin/tools/ESE3/esefinder.cgi?process=home
FAS-ESS	ESS	http://hollywood.mit.edu/fas-ess
EX-SKIP	ESE, ESS	https://ex-skip.img.cas.cz/
SpliceAid	ESE, ESS, ISE, ISS	http://www.introni.it/splicing.html
CRYP-SKIP	Exons and their flanking intron sequences	https://cryp-skip.img.cas.cz/
Spliceman	Mutation sites and flanking nucleotides	http://fairbrother.biomed.brown.edu/spliceman/
pFold	RNA sequence	http://daimi.au.dk/~compbio/pfold/
UNAFold	RNA sequence	http://unafold.rna.albany.edu/?q=mfold
Human Splicing Finder	Splice element, splice site	http://www.umd.be/HSF3/
Sroogle	Splice element, splice site	http://sroogle.tau.ac.il/
Alamut Visual Software	Prediction of variant pathogenicity	http://www.interactive-biosoftware.com
FSPLICE	Prediction of splice isomers generated by variation	http://www.softberry.com/berry.phtml?topic=fsplce&group=programs&subgroup=gfind
SpliceGrapher	Identify the authenticity of splicing sites	http://SpliceGrapher.sf.net
NetGene2	Neural networks	http://www.cbs.dtu.dk/services/NetGene2
SPANR	Effect of predicted variation on splicing	http://tools.genes.toronto.edu/

5 剪接突变与单基因遗传病的检测与治疗

5.1 检测

测序技术的发展不仅是临床诊断剪接突变所致疾病的关键,而且对疾病新的治疗方案的发展至关重要^[16]。目前针对全基因组序列的二代测序技术可以在基因的编码区和非编码区发现大量的变异,其在遗传性疾病中的应用有助于识别影响剪接的外显子或内含子突变,可提高诊断率^[57]。目前,WES和WGS已经成为孟德尔疾病诊断的常规方法^[58],变异解读成为遗传性单基因病诊断的主要挑战。基于基因组测序的变异检测能力大大超过了对变异危害性的认知能力,特别是对于非编码区变异的解释。目前存在诸多变异被判定为临床意义未明的变异(variants of uncertain significance, VUS),因此,临床遗传病诊断工作需要更丰富的基础研究 with 病例数据的支持^[59]。

没有使用 minigene 或患者活检材料进行必要的功能验证的非经典剪接位点的突变,仍然是 VUS 变

异。转录组测序(RNA-seq)作为一种补充诊断工具具有验证和识别外显子和深层内含子区域剪接突变的能力^[60]。对大数据样本的血液进行分析的结果表明,31%归类为VUS的突变可通过RNA-seq及分析预测工具补充的数据将结果更改为可能致病或可能良性^[60-61]。研究证实,RNA-seq可提高WES或WGS对单基因疾病的诊断率,同时在全基因组范围内识别异常的剪接异构体,因此RNA-seq可以作为WES或WGS的一种补充诊断工具^[58]。相信随着基因测序数据的不断积累,以及生物信息分析算法的逐步精确,多组学联合运用将为临床提供更加全面的基因诊断,并揭示更多剪接突变相关的致病机制。

5.2 治疗

pre-mRNA剪接与单基因遗传病相关的研究可为治疗剪接突变所致疾病的药物的开发提供基础,目前正进行着许多基于不同理论的治疗试验,这类研究多集中在纠正异常剪接或更有效的治疗策略上。这里以单基因遗传病SMA为例,它是一种严重

的进行性神经肌肉病,由SMN1基因突变导致。该基因在人体中有一个高度同源拷贝SMN2基因,其与SMN1基因仅有5个碱基的差别,两者同源性>99.9%,编码相同的SMN蛋白。其中最重要的是SMN2基因外显子7的c.840 C>T突变,该突变使得SMN2基因ESE功能受到抑制,导致SMN2基因外显子7跳跃,使得转录本产生截短的不稳定蛋白。

已存在二种针对SMN2基因剪接的治疗策略,其根本目标是在中枢神经系统和其他组织细胞中增加功能性SMN蛋白的表达量^[62]。第一种是反义寡核苷酸(antisense oligonucleotides, ASO),它是人工合成的一类单链DNA,可与SMN2 pre-mRNA上要被去除的外显子序列杂交,其中第一个被美国食品和药物管理局(FDA)批准用于治疗SMA的ASO代表药物是诺西那生钠注射液(Spinraza™/Nusinersen),它可与SMN2内含子7的内含子剪接沉默位点(ISS-N1)结合,并通过抑制其他剪接因子的作用来纠正SMN2外显子7中的错误剪接,促使SMN2的mRNA转录本中包含外显子7,这一机制提高了SMN2基因功能性mRNA的转录水平,并使其翻译出全功能SMN蛋白^[63-64]。诺西那生钠注射液的优点在于:诺西那生序列与ISS-N1高度互补且脱靶风险低;经鞘内注射给药,药物可直达因中枢SMN蛋白水平不足而导致的运动神经元退行性变的部位;临床上接受鞘内注射Nusinesen的SMA患者显示出运动功能的改善,不需要呼吸辅助等。而ASO的不足之处是给药方式为侵入性操作。因为疾病的特异性,目前针对剪接突变导致的单基因遗传病的临床治疗药物较少。第二种是有效的小分子化合物,目前在经高通量筛选技术筛选出的可促进SMN2基因表达完整SMN蛋白的小分子化合物中,已有多数被用于纠正与疾病相关的错误剪接,这些化合物改善了SMA小鼠模型的运动神经元功能,延长了其寿命,并提高了SMA患者中成纤维细胞和运动神经元细胞中的SMN蛋白水平^[65-66],但目前仍在研究阶段。

基因编辑技术具有治疗许多单基因遗传病的临床潜力,能够从基因层面为多种罕见遗传疾病带来治愈的希望。其中CRISPR/Cas9和CRISPR/Cpf1等基因编辑技术的发展使得编辑人类多能干细胞的效率大大提高^[67]。研究表明在SMA的iPSCs中,使用CRISPR/Cpf1可以将SMN2基因原位转化为SMN1基因,提高SMN基因在运动神经元中的表达量,使用

CRISPR/Cas9可以纠正SMN2在SMA iPSCs中的异常剪接,使功能性SMN蛋白的表达量更高,从而改善神经运动功能^[62]。目前CRISPR/Cas9系统的脱靶效应导致其进入临床的进程缓慢,而在最新研究中,研究人员发现了导致CRISPR/Cas9脱靶的分子结构机制,并设计出了新型Cas9蛋白使其脱靶概率降低4 000倍,大幅提升了其安全性,相信这将进一步促进CRISPR/Cas9基因编辑技术进入临床应用^[68-69]。

6 展望

本文综述了剪接突变在单基因遗传病中的致病机制、剪接突变的验证方法及其临床诊断和治疗方法等,为剪接突变致单基因遗传病的相关研究提供了理论基础和研究方向。随着二代测序技术在临床上的大范围应用和生物信息学的快速发展,单基因遗传病的诊断也面临诸多新的问题与挑战,尤其内含子区域的突变受到临床越来越多的关注,但其致病性尚未在大规模试验中得到验证和评估。由于minigene验证效率不高,且作为体外验证试验,无法模拟人体内参与剪接过程的数百种不同的剪接体相关蛋白质对剪接事件的影响,包括对剪接位点的正确识别等,所以针对内含子区域的剪接突变在单基因遗传病中的作用机制难点重重,还需要大量实验和临床数据来研究证明。在未来研究中,以剪接突变的致病机制为方向的基础研究和临床应用研究将有助于相关遗传病的诊断和药物研发,可为单基因遗传病的诊疗提供重要思路。

参考文献 (References)

- [1] HINKLE E R, WIEDNER H J, BLACK A J, et al. RNA processing in skeletal muscle biology and disease [J]. *Transcription*, 2019, 10(1): 1-20.
- [2] MONTES M, SANFORD B L, COMISKEY D F, et al. RNA splicing and disease: animal models to therapies [J]. *Trends Genet*, 2019, 35(1): 68-87.
- [3] SCOTTI M M, SWANSON M S. RNA mis-splicing in disease [J]. *Nat Rev Genet*, 2016, 17(1): 19-32.
- [4] LI D, MCINTOSH C S, MASTAGLIA F L, et al. Neurodegenerative diseases: a hotbed for splicing defects and the potential therapies [J]. *Transl Neurodegener*, 2021, 10(1): 16.
- [5] ZHU L Y, ZHU Y R, DAI D J, et al. Epigenetic regulation of alternative splicing [J]. *Am J Cancer Res*, 2018, 8(12): 2346-58.
- [6] ZHAO S. Alternative splicing, RNA-seq and drug discovery [J]. *Drug Discov Today*, 2019, 24(6): 1258-67.
- [7] YANG Q, ZHAO J, ZHANG W, et al. Aberrant alternative splicing in breast cancer [J]. *J Mol Cell Biol*, 2019, 11(10): 920-9.

- [8] OHNO K, TAKEDA J I, MASUDA A. Rules and tools to predict the splicing effects of exonic and intronic mutations [J]. Wiley Interdiscip Rev RNA, 2018, doi: 10.1002/wrna.1451.
- [9] NILSEN T W, GRAVELEY B R. Expansion of the eukaryotic proteome by alternative splicing [J]. Nature, 2010, 463(7280): 457-63.
- [10] XU B, SHI Y, WU Y, et al. Role of RNA secondary structures in regulating Dscam alternative splicing [J]. Biochim Biophys Acta Gene Regul Mech, 2019, 1862(11/12): 194381.
- [11] BERK A J. Discovery of RNA splicing and genes in pieces [J]. Proc Natl Acad Sci USA, 2016, 113(4): 801-5.
- [12] XUE W, MA X K, YANG L. Fast and furious: insights of back splicing regulation during nascent RNA synthesis [J]. Sci China Life Sci, 2021, 64(7): 1050-61.
- [13] WARD A J, COOPER T A. The pathobiology of splicing [J]. J Pathol, 2010, 220(2): 152-63.
- [14] PUCKER B, BROCKINGTON S F. Genome-wide analyses supported by RNA-Seq reveal non-canonical splice sites in plant genomes [J]. BMC Genomics, 2018, 19(1): 980.
- [15] LEWANDOWSKA M A. The missing puzzle piece: splicing mutations [J]. Int J Clin Exp Pathol, 2013, 6(12): 2675-82.
- [16] RAJ B, BLENCOWE B J. Alternative splicing in the mammalian nervous system: recent insights into mechanisms and functional roles [J]. Neuron, 2015, 87(1): 14-27.
- [17] WEISSCHUH N, BUENA-ATIENZA E, WISSINGER B. Splicing mutations in inherited retinal diseases [J]. Prog Retin Eye Res, 2021, 80: 100874.
- [18] WANG Z, XIAO X, VAN NOSTRAND E, et al. General and specific functions of exonic splicing silencers in splicing control [J]. Mol Cell, 2006, 23(1): 61-70.
- [19] KOVÁČOVÁ T, SOUČEK P, HUJOVÁ P, et al. Splicing enhancers at intron-exon borders participate in acceptor splice sites recognition [J]. Int J Mol Sci, 2020, 21(18): 6553.
- [20] ABRAMOWICZ A, GOS M. Correction to: splicing mutations in human genetic disorders: examples, detection, and confirmation [J]. J Appl Genet, 2019, 60(2): 231.
- [21] SINGH R N, SINGH N N. A novel role of U1 snRNP: splice site selection from a distance [J]. Biochim Biophys Acta Gene Regul Mech, 2019, 1862(6): 634-42.
- [22] BARALLE D, BURATTI E. RNA splicing in human disease and in the clinic [J]. Clin Sci, 2017, 131(5): 355-68.
- [23] CAMINSKY N, MUCAKI E J, ROGAN P K. Interpretation of mRNA splicing mutations in genetic disease: review of the literature and guidelines for information-theoretical analysis [J]. F1000Res, 2014, 3: 282.
- [24] RIEDMAYR L M, BÖHM S, MICHALAKIS S, et al. Construction and cloning of minigenes for *in vivo* analysis of potential splice mutations [J]. Bio Protoc, 2018, 8(5): e2760.
- [25] ZHANG S, SAMOCHA K E, RIVAS M A, et al. Base-specific mutational intolerance near splice sites clarifies the role of nonessential splice nucleotides [J]. Genome Res, 2018, 28(7): 968-74.
- [26] LEE M, ROOS P, SHARMA N, et al. Systematic computational identification of variants that activate exonic and intronic cryptic splice sites [J]. Am J Hum Genet, 2017, 100(5): 751-65.
- [27] STINGL K, MAYER A K, LLAVONA P, et al. CDHR1 mutations in retinal dystrophies [J]. Sci Rep, 2017, 7(1): 6992.
- [28] KEEGAN N P, WILTON S D, FLETCHER S. Analysis of pathogenic pseudoexons reveals novel mechanisms driving cryptic splicing [J]. Front Genet, 2021, 12: 806946.
- [29] ROCA X, AKERMAN M, GAUS H, et al. Widespread recognition of 5' splice sites by noncanonical base-pairing to U1 snRNA involving bulged nucleotides [J]. Genes Dev, 2012, 26(10): 1098-109.
- [30] ROCA X, SACHIDANANDAM R, KRAINER A R. Determinants of the inherent strength of human 5' splice sites [J]. RNA, 2005, 11(5): 683-98.
- [31] LIU W, LI X, LIAO S, et al. Activation of the intronic cryptic 5' splice site depends on its distance to the upstream cassette exon [J]. Gene, 2017, 619: 30-6.
- [32] KRAWCZAK M, THOMAS N S, HUNDRIESER B, et al. Single base-pair substitutions in exon-intron junctions of human genes: nature, distribution, and consequences for mRNA splicing [J]. Human mutation, 2007, 28(2): 150-8.
- [33] HABARA Y, TAKESHIMA Y, AWANO H, et al. *In vitro* splicing analysis showed that availability of a cryptic splice site is not a determinant for alternative splicing patterns caused by +1G->A mutations in introns of the dystrophin gene [J]. J Med Genet, 2009, 46(8): 542-7.
- [34] ERKELENZ S, POSCHMANN G, PTOK J, et al. Profiling of cis- and trans-acting factors supporting noncanonical splice site activation [J]. RNA Biol, 2021, 18(1): 118-30.
- [35] MACRAE A J, MAYERLE M, HRABETA-ROBINSON E, et al. Prp8 positioning of U5 snRNA is linked to 5' splice site recognition [J]. RNA, 2018, 24(6): 769-77.
- [36] HARTLEY J N, SIMARD L R, LY V, et al. A homozygous canonical splice acceptor site mutation in PRUNE1 is responsible for a rare childhood neurodegenerative disease in Manitoba Cree families [J]. Am J Med Genet A, 2019, 179(2): 206-18.
- [37] IUSO A, ALHADDAD B, WEIGEL C, et al. A homozygous splice site mutation in SLC25A42, encoding the mitochondrial transporter of coenzyme A, causes metabolic crises and epileptic encephalopathy [J]. JIMD Rep, 2019, 44: 1-7.
- [38] ATTARIAN S, FATEHI F, RAJABALLY Y A, et al. Hereditary neuropathy with liability to pressure palsies [J]. J Neurol, 2020, 267(8): 2198-206.
- [39] KAWARAI T, YAMAZAKI H, MIYAMOTO R, et al. PMP22-related disease: a novel splice site acceptor variant and intrafamilial phenotype variability [J]. Neuromuscul Disord, 2019, 29(6): 422-6.
- [40] FUSCO C, SPAGNOLI C, SALERNO G G, et al. Hereditary neuropathy with liability to pressure palsy (HNPP): report of a family with a new point mutation in PMP22 gene [J]. Ital J Pediatr, 2017, 43(1): 97.
- [41] KAMADA M, KAWARAI T, MIYAMOTO R, et al. Spastic paraplegia type 31: a novel REEP1 splice site donor variant and expansion of the phenotype variability [J]. Parkinsonism Relat Disord, 2018, 46: 79-83.
- [42] KEEGAN N P. Pseudoexons of the DMD Gene [J]. J Neuromuscul Dis, 2020, 7(2): 77-95.
- [43] VAZ-DRAGO R, CUSTÓDIO N, CARMO-FONSECA M. Deep intronic mutations and human disease [J]. Hum Genet, 2017, 136(9): 1093-111.
- [44] MINNEROP M, KURZWELLY D, WAGNER H, et al. Hypomorphic mutations in POLR3A are a frequent cause of sporadic

- and recessive spastic ataxia [J]. *Brain*, 2017, 140(6): 1561-78.
- [45] BOSCHANN F, FISCHER-ZIRNSAK B, WIENKER T F, et al. An intronic splice site alteration in combination with a large deletion affecting VPS13B (COH1) causes Cohen syndrome [J]. *Eur J Med Genet*, 2020, 63(9): 103973.
- [46] WONG J J, AU A Y, RITCHIE W, et al. Intron retention in mRNA: no longer nonsense: known and putative roles of intron retention in normal and disease biology [J]. *Bioessays*, 2016, 38(1): 41-9.
- [47] TRABELSI M, BEUGNET C, DEBURGRAVE N, et al. When a mid-intronic variation of DMD gene creates an ESE site [J]. *Neuromuscul Disord*, 2014, 24(12): 1111-7.
- [48] DE BOER M, VAN LEEUWEN K, HAURI-HOHL M, et al. Activation of cryptic splice sites in three patients with chronic granulomatous disease [J]. *Mol Genet Genomic Med*, 2019, 7(9): e854.
- [49] SOUKARIEH O, GAILDRAT P, HAMIEH M, et al. Exonic splicing mutations are more prevalent than currently estimated and can be predicted by using in silico tools [J]. *PLoS Genet*, 2016, 12(1): e1005756.
- [50] GRIFFIN D A, POZSGAI E R, HELLER K N, et al. Preclinical systemic delivery of adeno-associated α -sarcoglycan gene transfer for limb-girdle muscular dystrophy [J]. *Hum Gene Ther*, 2021, 32(7/8): 390-404.
- [51] ESCOBAR H, KRAUSE A, KEIPER S, et al. Base editing repairs an SGCA mutation in human primary muscle stem cells [J]. *JCI Insight*, 2021, 6(10): e145994.
- [52] ROUZIER C, CHAUSSENOT A, PAQUIS-FLUCKLINGER V. Molecular diagnosis and genetic counseling for spinal muscular atrophy (SMA) [J]. *Arch Pediatr*, 2020, 27(7s): 7s9-7s14.
- [53] QU Y J, BAI J L, CAO Y Y, et al. A rare variant (c.863G>T) in exon 7 of SMN1 disrupts mRNA splicing and is responsible for spinal muscular atrophy [J]. *Eur J Hum Genet*, 2016, 24(6): 864-70.
- [54] PAWLICKA K, KALATHIYA U, ALFARO J. Nonsense-mediated mRNA decay: pathologies and the potential for novel therapeutics [J]. *Cancers*, 2020, 12(3): 765.
- [55] SHEN J J, WORTMANN S B, DE BOER L, et al. The role of clinical response to treatment in determining pathogenicity of genomic variants [J]. *Genet Med*, 2021, 23(3): 581-5.
- [56] BASEI F L, MOURA L A R, KOBARG J. Using the E1A minigene tool to study mRNA splicing changes [J]. *J Vis Exp*, 2021, doi: 10.3791/62181.
- [57] DOHRN M F, GLÖCKLE N, MULAHSANOVIC L, et al. Frequent genes in rare diseases: panel-based next generation sequencing to disclose causal mutations in hereditary neuropathies [J]. *J Neurochem*, 2017, 143(5): 507-22.
- [58] GONORAZKY H D, NAUMENKO S, RAMANI A K, et al. Expanding the boundaries of RNA sequencing as a diagnostic tool for rare mendelian disease [J]. *Am J Hum Genet*, 2019, 104(3): 466-83.
- [59] CUMMINGS B B, MARSHALL J L, TUKIAINEN T, et al. Improving genetic diagnosis in Mendelian disease with transcriptome sequencing [J]. *Sci Transl Med*, 2017, 9(386): eaal5209.
- [60] WAI H A, LORD J, LYON M, et al. Blood RNA analysis can increase clinical diagnostic rate and resolve variants of uncertain significance [J]. *Genet Med*, 2020, 22(6): 1005-14.
- [61] TRUTY R, OUYANG K, ROJAHN S, et al. Spectrum of splicing variants in disease genes and the ability of RNA analysis to reduce uncertainty in clinical interpretation [J]. *Am J Hum Genet*, 2021, 108(4): 696-708.
- [62] KARPE Y, CHEN Z, LI X J. Stem cell models and gene targeting for human motor neuron diseases [J]. *Pharmaceuticals*, 2021, 14(6): 565.
- [63] MESSINA S, SFRAMELI M. New treatments in spinal muscular atrophy: positive results and new challenges [J]. *J Clin Med*, 2020, 9(7): 2222.
- [64] KIRSCHNER J, BUTOIANU N, GOEMANS N, et al. European ad-hoc consensus statement on gene replacement therapy for spinal muscular atrophy [J]. *Eur J Paediatr Neurol*, 2020, 28: 38-43.
- [65] MESSINA S, SFRAMELI M, MAGGI L, et al. Spinal muscular atrophy: state of the art and new therapeutic strategies [J]. *Neurol Sci*, 2021, doi: 10.1007/s10072-021-05258-3.
- [66] HAN F, EBRAHIMI-BAROUGH S, ABOLGHASEMI R, et al. Cell-based therapy for spinal muscular atrophy [J]. *Adv Exp Med Biol*, 2020, 1266: 117-25.
- [67] JANIK E, NIEMCEWICZ M, CEREMUGA M, et al. Various aspects of a gene editing system-CRISPR-Cas9 [J]. *Int J Mol Sc*, 2020, 21(24): 9609.
- [68] BRAVO J P K, LIU M S, HIBSHMAN G N, et al. Publisher correction: structural basis for mismatch surveillance by CRISPR-Cas9 [J]. *Nature*, 2022, 604(7904): E10.
- [69] KIM S, KIM D, CHO S W, et al. Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins [J]. *Genome Res*, 2014, 24(6): 1012-9.