

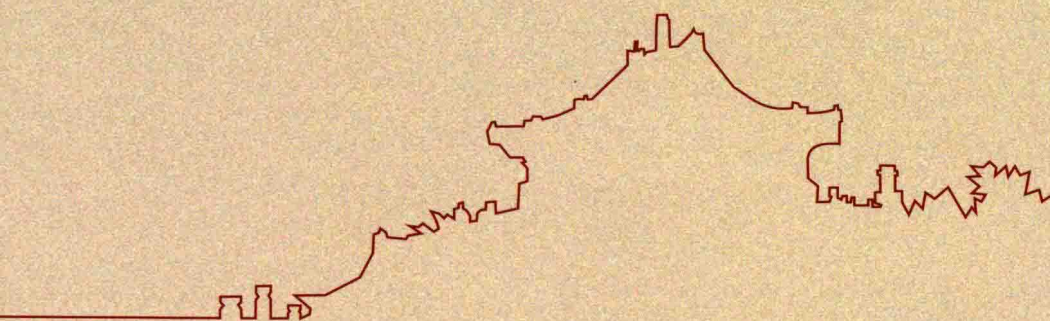
武汉大学优秀博士学位论文文库



抑制SMN2外显子7剪接的 调控蛋白的鉴定和作用机制研究

Screen Novel Splicing Repressors Regulating SMN Exon7 Splicing and
Study the Associated Regulatory Mechanism

肖锐 著



WUHAN UNIVERSITY PRESS

武汉大学出版社

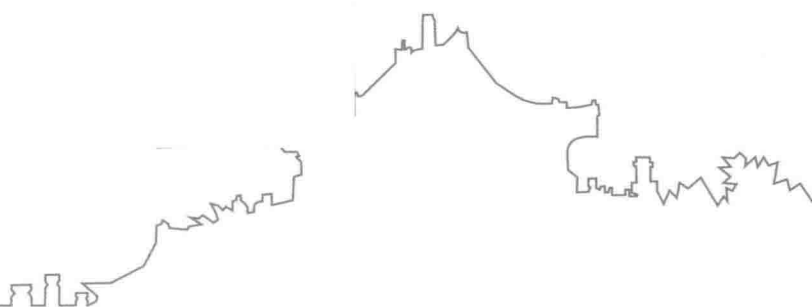
武汉大学优秀博士学位论文文库



抑制 SMN2 外显子 7 剪接的 调控蛋白的鉴定和作用机制研究

Screen Novel Splicing Repressors Regulating SMN Exon7 Splicing and
Study the Associated Regulatory Mechanism

肖锐 著



WUHAN UNIVERSITY PRESS

武汉大学出版社

图书在版编目(CIP)数据

抑制 SMN2 外显子 7 剪接的调控蛋白的鉴定和作用机制研究/肖锐著. —武汉:
武汉大学出版社, 2016.4

武汉大学优秀博士学位论文文库

ISBN 978-7-307-17315-6

I.抑… II.肖… III.肌萎缩—基因—蛋白—调节(生理)—研究 IV.R746.4

中国版本图书馆 CIP 数据核字(2015)第 294975 号

责任编辑:黄汉平

责任校对:汪欣怡

版式设计:马 佳

出版发行:武汉大学出版社 (430072 武昌 珞珈山)

(电子邮件: cbs22@whu.edu.cn 网址: www.wdp.com.cn)

印刷:武汉市洪林印务有限公司

开本: 720×1000 1/16 印张: 8.25 字数: 115 千字 插页: 2

版次: 2016 年 4 月第 1 版 2016 年 4 月第 1 次印刷

ISBN 978-7-307-17315-6 定价: 20.00 元

版权所有, 不得翻印; 凡购我社的图书, 如有质量问题, 请与当地图书销售部门联系调换。

总 序

创新是一个民族进步的灵魂，也是中国未来发展的核心驱动力。研究生教育作为教育的最高层次，在培养创新人才中具有决定意义，是国家核心竞争力的重要支撑，是提升国家软实力的重要依托，也是国家综合国力和科学文化水平的重要标志。

武汉大学是一所崇尚学术、自由探索、追求卓越的大学。美丽的珞珈山水不仅可以诗意栖居，更可以陶冶性情、激发灵感。更为重要的是，这里名师荟萃、英才云集，一批又一批优秀学人在这里砥砺学术、传播真理、探索新知。一流的教育资源，先进的教育制度，为优秀博士学位论文的产生提供了肥沃的土壤和适宜的气候条件。

致力于建设高水平的研究型大学，武汉大学素来重视研究生培养，是我国首批成立有研究生院的大学之一，不仅为国家培育了一大批高层次拔尖创新人才，而且产出了一大批高水平科研成果。近年来，学校明确将“质量是生命线”和“创新是主旋律”作为指导研究生教育工作的基本方针，在稳定研究生教育规模的同时，不断推进和深化研究生教育教学改革，使学校的研究生教育质量和知名度不断提升。

博士研究生教育位于研究生教育的最顶端，博士研究生也是学校科学研究的重要力量。一大批优秀博士研究生，在他们学术创作最激情的时期，来到珞珈山下、东湖之滨。珞珈山的浑厚，奠定了他们学术研究的坚实基础；东湖水的灵动，激发了他们学术创新的无限灵感。在每一篇优秀博士学位论文的背后，都有博士研究生们刻苦钻研的身影，更有他们的导师的辛勤汗水。年轻的学者们，犹如在海边拾贝，面对知识与真理的浩瀚海洋，他们在导师的循循善

诱下，细心找寻着、收集着一片片靓丽的贝壳，最终把它们连成一串串闪闪夺目的项链。阳光下的汗水，是他们砥砺创新的注脚；面向太阳的远方，是他们奔跑的方向；导师们的悉心指点，则是他们最值得依赖的臂膀！

博士学位论文是博士生学习活动和研究工作的主要成果，也是学校研究生教育质量的凝结，具有很强的学术性、创造性、规范性和专业性。博士学位论文是一个学者特别是年轻学者踏进学术之门的标志，很多博士学位论文开辟了学术领域的新思想、新观念、新视阈和新境界。

据统计，近几年我校博士研究生所发表的高质量论文占全校高水平论文的一半以上。至今，武汉大学已经培育出 18 篇“全国百篇优秀博士学位论文”，还有数十篇论文获“全国百篇优秀博士学位论文提名奖”，数百篇论文被评为“湖北省优秀博士学位论文”。优秀博士结出的累累硕果，无疑应该为我们好好珍藏，装入思想的宝库，供后学者慢慢汲取其养分，吸收其精华。编辑出版优秀博士学位论文文库，即是这一工作的具体表现。这项工作既是一种文化积累，又能助推这批青年学者更快地成长，更可以为后来者提供一种可资借鉴的范式抑或努力的方向，以鼓励他们勤于学习，善于思考，勇于创新，争取产生数量更多、创新性更强的博士学位论文。

武汉大学即将迎来双甲华诞，学校编辑出版该文库，不仅仅是为百廿武大增光添彩，更重要的是，当岁月无声地滑过 120 个春秋，当我们正大踏步地迈向前方时，我们有必要回首来时的路，我们有必要清晰地审视我们走过的每一个脚印。因为，铭记过去，才能开拓未来。武汉大学深厚的历史底蕴，不仅在于珞珈山的一草一木，也不仅仅在于屋檐上那一片片琉璃瓦，更在于珞珈山下的每一位学者和学生。而本文库收录的每一篇优秀博士学位论文，无疑又给珞珈山注入了新鲜的活力。不知不觉地，你看那珞珈山上的树木，仿佛又茂盛了许多！

李晓红

2013 年 10 月于武昌珞珈山

摘 要

脊髓性肌萎缩症(spinal muscular atrophy, SMA)是一种发病率高并导致婴儿死亡的常染色体隐性神经肌肉障碍疾病。导致这种疾病的遗传学原因是纯合地丢失 SMN1 基因。但其同源的 SMN2 有外显子 7 上 6 位的 C 到 T 的突变导致其不剪接,产生截短的且不稳定的 SMN 蛋白,不能补偿全长的 SMN 蛋白的功能。所以,增加 SMN2 外显子 7 的剪接目前被认为是可行、有效的治疗脊髓性肌萎缩症的方法。因此,SMN2 外显子 7 的可变剪接调控的研究变得尤为重要。

本论文通过初步的 RNA 干扰鉴定出了新的 SMN2 外显子 7 剪接的抑制因子 hnRNP U,同时还发现了 hnRNP U 的一项新功能——可变剪接调控因子,同时还发现它既调控 SMN2 外显子 7 剪接,也调控 SMN1 外显子 7 剪接,并不响应 SMN2 外显子 7 上 C 到 T 的突变位点。深入研究后发现重组的 hnRNP U 在体外并不结合 SMN2 外显子 7 和它附近的内含子 RNA,但是在体内结合 SMN2 pre-mRNA。它的作用方式既不依赖于已知的内含子剪接抑制元件,也与该基因的转录速率无关。这暗示 hnRNP U 可能通过未知的方式调控 SMN2 外显子 7 可变剪接。

为了进一步认识这种新的调控机制,我们引入了大规模筛选 RNA 结合蛋白的方法来寻找更多的调控因子,从其他大约 340 个人类 RNA 结合蛋白基因中筛选 SMN2 外显子 7 的剪接抑制蛋白。找寻其中与 hnRNPU 有相互作用的蛋白。我们发现 5 个参与 3' 剪接位点识别的组成型剪接正调控因子,包括 SF1, U2AF65, PUF60, U2AF35 和 CHERP(U2 复合物成员)在 SMN2 外显子 7 可变剪接调控中均扮演抑制因子的角色。我们还发现 hnRNP U 与

U2AF65、U2AF35 和 SF1 相互作用，且这种相互作用依赖于 RNA。该结果暗示 hnRNP U 和这些组成型剪接蛋白可能是通过选择性促进组成型 3' 剪接位点的选择，而间接抑制了可变外显子的选择。因此，当这些组成型剪接蛋白被敲低后，组成型 3' 剪接位点与可变 3' 剪接位点竞争它们共同的 5' 剪接位点的能力减弱了。

为了进一步认识 hnRNP U 是通过哪个 RNA 组成与 U2AF65 等组成型 3' 剪接位点相互作用的，我们采用了 CLIP-seq 方法获得 hnRNP U 在全基因组范围内的结合图谱，发现 hnRNP U 的确不直接结合 SMN2 上关键的外显子 7 及周围 intron 区域，却直接结合 U2 snRNA 的 SmBP-box 区域。因此，hnRNP U 可能通过参与 U2 snRNP 的相互结合而被招募到组成型 3' 剪接位点的附近，与 U2AF65 相互作用。

CLIP-seq 数据显示 hnRNP U 结合在 GU 富集的 RNA 基序上，而且结合位置与一些可变剪接调控事件紧密相关。暗示该蛋白可能是一个广泛的可变剪接调控蛋白。本论文已经获得了该蛋白表达和被 siRNA 沉默的宫颈癌细胞中的转录组，对数据的全面分析有望获得该蛋白调控宫颈癌细胞可变剪接的全貌。

关键词：可变剪接，SMN2 外显子 7，hnRNP U，snRNA

Abstract

Spinal muscular atrophy (SMA) is an autosomal recessive neuromuscular disorder with high incidence, and it is the primary genetic cause of infant mortality. The genetic cause of SMA is a homozygous loss of SMN1. But its homologous gene SMN2 with a C-to-T mutation on exon 7 generates truncated and unstable SMN protein because of the low splicing efficiency of exon 7. Increasing SMN2 exon 7 splicing is considered to be a feasible and effective strategy to therapy SMA. Thus, investigating the mechanism of SMN2 exon 7 splicing regulation is of importance.

Using esiRNA silencing approach to identify SMN2 splicing repressor among heterogeneous nuclear ribonucleoproteins, this study has found the RNA-binding protein hnRNP U as a new RNA splicing factor that represses SMN2 exon 7 splicing, together with the known SMN2 repressor hnRNP A1 and A2. However, unlike hnRNP A1 and A2, hnRNP U repression of SMN2 splicing is not associated with the C-to-T mutation on exon 7, neither with any of the four known ISS elements in the neighboring introns. We further show that the repression is not associated with an altered transcription rate of the SMN2 gene either. These results indicate that hnRNP U may repress splicing of the alternative exon 7 of SMN2 pre-mRNA through a mechanism different from all the known ones.

hnRNP U has both RNA and protein interaction domains, but has never been reported to associated with alternative splicing before. We hypothesize that hnRNP U may regulate SMN2 splicing through an

interaction with some other known splicing factor. This promoted us to screen the other splicing factors that repress SMN2 splicing through about 340 putative human RNA binding protein genes including all the known splicing factors. We obtained 8 other splicing repressors for SMN2 splicing, strikingly, 5 of them are constitutive splicing factors for the 3' splice site selection, including SF1, U2AF65, PUF60, U2AF35 and CHERP (a component of snRNP). We showed that hnRNP U interacts with U2AF65, U2AF35 and SF1 through an RNA-dependent manner, among which U2AF65 is the strongest one. These results strongly indicate that hnRNP U may interact with these constitutive splicing factors to affect the reorganization and selection of the constitutive 3' splice site, and therefore inhibits the selection of the alternative 3' splice site. Therefore, when this class of proteins is down-regulated, the constitutive splice site becomes less competitive in interacting with their common 5' splice site. As a result, the alternative splice site becomes relatively more competitive.

In order to identify the RNA components through which hnRNP U interacts with U2AF65 and other 3'ss splicing factors, this study used CLIP-seq method to obtain the genome-wide hnRNP U binding sites. It is shown that hnRNP U has no binding site in exon 7 and the adjacent intronic regions, consistent with our data showing that the recombinant hnRNP U does not bind these regions in vitro. However, we found that hnRNP U strongly binds to SmBP-box of U2 snRNA, suggesting that hnRNP U is a constitutive 3' splice site splicing factor. We propose that hnRNP U may be recruited to the 3' splice site through binding to U2 snRNA, where it interacts with U2AF65 and promotes the 3' splice site selection.

CLIP-seq data show that hnRNP U binds to the GU-rich sequence motifs, and the binding is strongly associated with some alternative splicing events, which indicates that hnRNP U may be a general splicing regulator. The transcriptome data have been obtained for the hnRNP U -

expressed and -silenced Hela cells, and we anticipate the global view of the involvement of this protein in regulating of the alternative splicing in Hela cells will be obtained after the data analysis is finished.

Key words: Alternative splicing, SMN2 exon 7, hnRNP U, snRNA

目 录

第 1 章 引言	1
1.1 可变剪接调控的介绍	1
1.1.1 前体 RNA 可变剪接	1
1.1.2 剪接位点的识别和选择	5
1.1.3 帮助剪接位点识别	6
1.1.4 抑制剪接位点识别	8
1.1.5 位置依赖性的剪接调控	11
1.1.6 RNA 结构在可变剪接调控中的作用	12
1.1.7 蛋白因子调控可变剪接	13
1.1.8 转录偶联的可变剪接调控	15
1.1.9 可变剪接和组织特异性	16
1.1.10 组成型剪接调控因子调控可变剪接	19
1.1.11 可变剪接与人类疾病	20
1.1.12 展望	20
1.2 SMA 和 SMN 的介绍	22
1.2.1 SMA 疾病的介绍	22
1.2.2 SMA 疾病相关的 SMN2 外显子 7 可变剪接的 机制	27
1.2.3 SMA 的治疗方面的进展	30
1.3 hnRNP U 的介绍	32
1.4 本课题的研究内容及意义	33
1.4.1 课题研究的内容	33
1.4.2 课题的研究意义	33

第 2 章 材料和方法	35
2.1 材料	35
2.1.1 菌株、质粒及细胞系	35
2.1.2 生化药品及来源	35
2.1.3 细胞培养及操作相关的试剂	36
2.1.4 酶类和 DNA marker, 蛋白 marker	36
2.1.5 抗体及其相关	37
2.1.6 ssRNA 和 siRNA 合成	37
2.1.7 本研究所用引物	38
2.1.8 分析软件和网络资源	41
2.1.9 图像编辑软件	42
2.1.10 仪器	42
2.2 溶液的配制	43
2.2.1 细菌用抗生素的配制	43
2.2.2 蛋白纯化中所用的缓冲液	43
2.2.3 DMEM 细胞培养基的配制	44
2.2.4 细胞专用 PBS 的制备	44
2.2.5 细胞专用胰酶的配制	44
2.2.6 免疫沉淀及免疫印迹相关溶液的配方	44
2.3 方法	45
2.3.1 <i>E. coli</i> RNase III 的制备和纯化	45
2.3.2 人类 RNA 结合蛋白 esiRNA 库的制备	47
2.3.3 细胞培养和转染	54
2.3.4 免疫印迹	56
2.3.5 RT-PCR	56
2.3.6 Trizol 中提取蛋白	58
2.3.7 突变克隆	58
2.3.8 克隆的构建	59
2.3.9 免疫共沉淀	59
2.3.10 RIP-PCR	60
2.3.11 免疫共定位	60

2.3.12 CLIP-seq 实验流程及注意事项	61
2.3.13 RNA-seq	72
第3章 结果与讨论	74
3.1 RNAi 筛选发现 HnRNP U 是一个新的抑制 SMN2 可变 外显子 7 剪接的剪接调控因子.....	74
3.2 hnRNP U 沉默同样会增强 SMN1 的剪接, 暗示 hnRNP U 不是通过靶向外显子 7 的 C 到 T 变化所产生的 ESS 而起作用的.....	77
3.3 hnRNP U 抑制 SMN2 的剪接不依赖于 任何已知的 ISS	79
3.4 hnRNP U 可能不是通过影响转录延伸来抑制 SMN 外 显子 7 的剪接的.....	80
3.5 hnRNP U 在体外不与 SMN2 外显子 7 及附近的内含子 RNA 结合, 而在细胞内结合 SMN2 转录本	82
3.6 大规模筛选人类 RNA 结合蛋白发现了另外 8 个新的 抑制 SMN2 外显子 7 剪接的蛋白质, 其中 5 个是与 U2 snRNP 相互作用的蛋白.....	83
3.7 hnRNP U 与 U2AF65/35 通过 RNA 依赖的方式相互 结合, 并协同调控 SMN2 外显子 7 的剪接.....	86
3.8 CLIP 鉴定 hnRNP U 结合的 RNA 靶标	87
3.9 hnRNP U 结合 U2 snRNA, 证明它是 U2 snRNP 的 一个重要组分.....	89
第4章 研究工作总结和展望	92
4.1 工作总结.....	92
4.2 展望.....	93
参考文献	96
缩写索引.....	113
致谢.....	115

第1章 引言

1.1 可变剪接调控的介绍

1.1.1 前体 RNA 可变剪接

在真核生物中,由 mRNA 前体产生成熟的 mRNA 需要经历 5' 加帽,剪接去除内含子(intron)和 3' 加 poly (A)这三个过程。其中剪接去除内含子这一过程有组成型剪接(constitutive splicing)和可变剪接(alternative splicing)两种形式。高等真核生物的基因大多由多个外显子组成。一个基因的外显子以多种不同方式进行拼接的,被称为可变剪接。可变剪接是基因转录后调控的主要形式之一,对于蛋白质的多样性至关重要。目前,估计含多个外显子(exon)的基因中大约 95%都经历了可变剪接过程(Black 2003; Wahl, Will et al., 2009)。在多细胞生物中,可变剪接可产生不同蛋白以行使如细胞生长、分化和死亡等不同的生物学功能。

剪接过程需要剪接体参与完成。剪接体通过其含有的 5 个小核糖核蛋白(small nuclear ribonucleoprotein particles, snRNPs)和大量的辅助因子协同作用来精确识别剪接位点,再通过两步催化反应完成剪接过程。其中剪接体的组装始于 snRNP U1 识别 5' 剪接位点,然后 SF1 (splicing factor 1) 结合分支位点(branch point) (Berglund, Chua et al., 1997), U2AF (U2 auxiliary factor) 结合多嘧啶序列(polypyrimidine tract) 和 3' 剪接位点 (Nelson and Green, 1989; Zamore and Green, 1989)。这一组装过程不依赖于 ATP,从而形成 E 复合物。之后,在 U2 snRNP 替换 SF1 结合分支位点后形成 ATP

依赖性的 A 复合物。随后, A 复合物招募 U4/U6-U5 tri-snRNP 复合物从而形成 B 复合物, 再经过一系列的构象变化成为有催化活性的 C 复合物(图 1-1)。

Box 1 | Splicing and spliceosome assembly

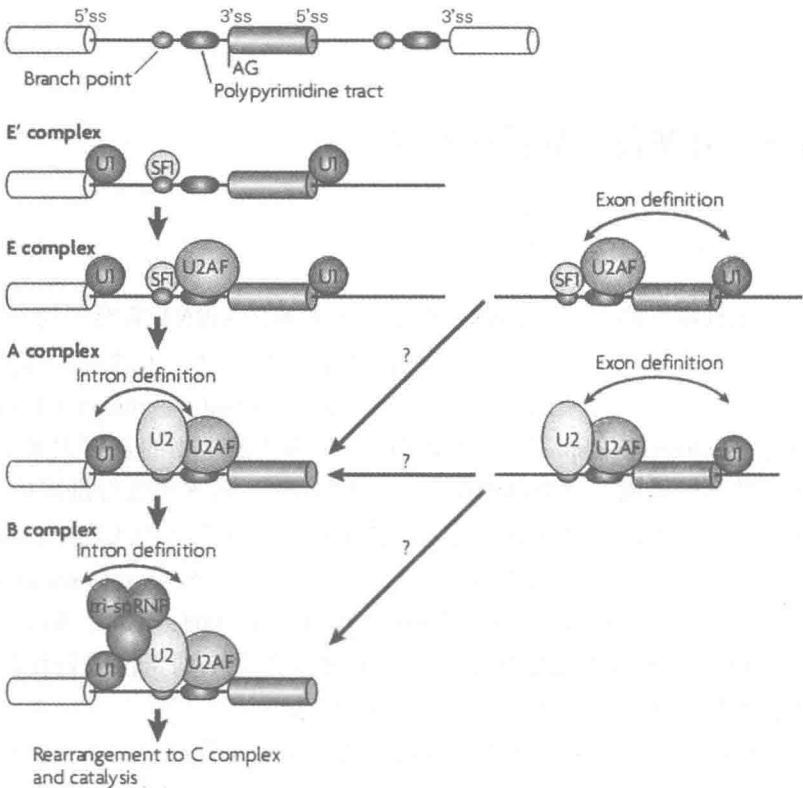


图 1-1 剪接位点的识别和剪接体组装

一般来讲, 这一过程开始于 U1 snRNA 与 5' 剪接位点的配对, SF1 结合分支位点形成 E' 复合物。在 U2AF 异源二聚体(由 U2AF65 和 u2af35 组成)被招募到 3' 剪接位点多嘧啶区和末端 AG 位点后, E' 复合物会转变成 E 复合物。ATP 不依赖的 E 复合物随后又在 U2snRNP 替换 SF1 后转变成 ATP 依赖的剪接体前体 A 复合物。接着, 招募 U4/U6-U5 3snRNP 复合物从而形成含有实施前体 mRNA 剪接所需所有亚基的 B 复合物。随后会发生一系列的构象变化和重构, 包括 U1 和 U4 snRNP 的分离, 最后形成有催化活性的剪接体-C 复合物。

在剪接过程中, 一个可变外显子被剪接或者保留是由其 RNA 序列和蛋白调控因子共同决定的。RNA 顺式作用元件 (cis-regulatory elements) 依据其位置和功能可以分为四种: 外显子剪接增强因子 (exonic splicing enhancers, ESEs)、外显子剪接衰减因子 (exonic splicing silencers, ESSs)、内含子剪接增强因子 (intronic splicing enhancers, ISEs) 和内含子剪接衰减因子 (intronic splicing silencers, ISSs)。通常情况下, SR (Ser-Arg) 蛋白家族 (Tacke and Manley 1999; Graveley, 2000; Long and Cáceres, 2009) 的成员结合 ESE, 而核不均一蛋白 (heterogeneous nuclear RNPs, hnRNPs) 则结合 ESS 和 ISS。其中, hnRNPs 含有一个或多个 RNA 结合结构域和蛋白-蛋白相互作用结构域 (Smith and Valcarcel 2000; Dreyfuss, Kim et al., 2002)。人们现在对 ISE 的了解还不像对其他三种 RNA 顺式作用元件那样清楚, 但也有 hnRNP F, hnRNP H, neurooncological ventral antigen 1 (NOVA1), NOVA2, FOX1 和 FOX2 通过结合 ISE 来促进剪接的报道 (Hui, Hung et al., 2005; Ule, Stefani et al., 2006; Mauger, Lin et al., 2008; Yeo, Coufal et al., 2009)。

mRNA 前体的可变剪接形式是多样的 (图 1-2)。

一般来讲, 可变剪接的选择发生在剪接位点识别过程和剪接体组装初期。但是, 最近有几篇报道发现可变剪接的选择可以发生在剪接体组装过程中的不同时期, 甚至可以在两步转酯之间的构象变化这一过程 (House and Lynch, 2006; Sharma, Kohlstaedt et al., 2008)。同时, 已有许多证据表明 RNA 的转录和剪接调控是偶联在一起的 (Batsche, Yaniv et al., 2006; de la Mata and Kornblihtt, 2006; Sims, Millhouse et al., 2007; Lin, Coutinho-Mansfield et al., 2008)。SR 蛋白最近被发现与组蛋白的 3' 尾结合, 在转录的延伸中起非常重要的作用。它与一些在染色体内部和染色体之间相互作用的基因共定位, 而这些基因间的相互作用协同调控了核内基因的表达。因此, SR 蛋白在转录和 RNA 加工中的重要作用为解释它在基因组的稳定性和高等真核生物细胞周期进程中的重要作用提供了解释 (Zhong, Wang et al., 2009)。

基因组的多态性和组织的特异性很大程度上取决于可变剪接过

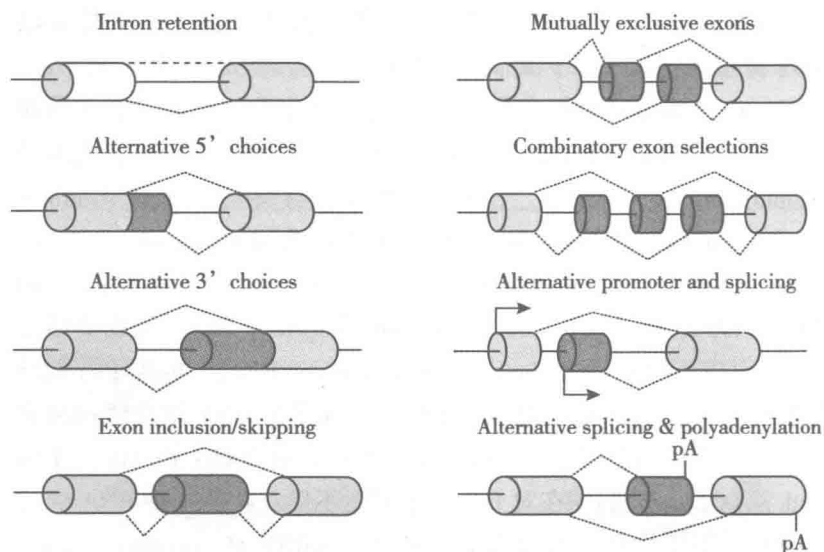


图 1-2 可变剪接 8 种不同的剪接形式

内含子保留、可变 5' 剪接、可变 3' 剪接、外显子剪接或剪接剔除、相互排斥的外显子可变剪接、组合的外显子剪接、可变的启动子剪接和可变的腺苷酸位点。

程 (Graveley, 2001)。了解组织特异性的可变剪接需要知道蛋白-蛋白、蛋白-RNA 和 RNA-RNA 之间的调控网络。组织特异性的可变剪接被认为是由不同的特异性表达和/或广泛性表达的剪接因子的浓度和活性来控制的 (Grabowski and Black, 2001; Park, Parisky et al., 2004; Zhang, Lotti et al., 2008)。目前人们运用高通量的实验方法来揭示更多可变剪接事件 (Blencowe, 2006; Castle, Zhang et al., 2008; Pan, Shai et al., 2008; Wang, Sandberg et al., 2008), 例如, 德国马普研究所 Marie-Laure Yaspo 研究小组 2008 年 7 月在 *Science* 上发表论文, 用 Illumina 高通量测序平台解析了人胚肾和 B 细胞的转录组, 分别获得 8638919 和 7682230 个 27-碱基对的 reads (测序数)。其中 50% 定位到唯一的基因组位点, 其中 80% 是在外显子区, 66% 的多聚腺苷化转录本定位在已知基因。34% 没有被注解的基因组区域, 共发现了 94241 剪接位点, 其中 4096 是新发现的, 同时发现外显子跳过是可变剪接里最丰富的一种 (Sultan,