·综 述·

可变剪接与疾病的生物信息学研究概况

王科俊[1*], 吕俊杰[1], 冯伟兴[1], 王 鑫[2]

(1. 哈尔滨工程大学 自动化学院, 中国黑龙江 哈尔滨 150001; 2. 剑桥大学 癌症分子研究中心 英国剑桥)

摘 要: 可变剪接是真核基因转录后期的重要调控机制, 它使得同一条蛋白质编码基因能够产生多种转录体, 极大的扩展了遗传信息的应用. 研究发现, 可变剪接与人类疾病有着密切的联系. 错误的剪接会导致疾病, 增加疾病的易感性与病变程度, 甚至直接导致癌变.现对可变剪接调控机制与疾病的生物信息学研究进展进行综述.

关键词: 可变剪接; 疾病; 癌症; 生物信息学

中图分类号: Q752 **文献标识码**: A **文章编号**: 1007-7847(2011)01-0086-09

The Application of Bioinformatics in the Research of Alternative Splicing and Disease

WANG Ke-jun[1], LÜ Jun-jie[1*], FENG Wei-xing[1], WANG Xin[2]

(1. *College of Automation, Harbin Engineering University, Harbin* 150001, *Heilongjiang, China*;
2. *Cancer Research Center, Cambridge University, Cambridge, England*)

Abstract: Alternative splicing is an important mechanism in regulating eukaryotic gene expression during post-transcriptional processing as it generates numerous transcripts from a single protein-coding gene, which largely increases the use of genetic information. Researches showed that alternative splicing was highly relevant to human diseases. Wrong splicing patterns could cause disease, contribute to disease severity and susceptibility, and even cause cancer directly. Research progress of bioinformatics in alternative splicing regulation mechanism and disease was summarized.

Key words: alternative splicing; disease; cancer; bioinformatics

(*Life Science Research*, 2011, 15(1): 086~094)

1977 年 Walter Gilbert 在对 *Adenovirus hexon* 基因的研究中发现并提出可变剪接现象[1], 同时, 他表明一个基因的不同编码区可以拼接在一起被剪接下来, 产生功能不同的信使核糖核酸, 这是对可变剪接的最早描述. 随后 Sharp 和 Roberts 发现高等生物基因的编码区被非编码区所分割, 并提出分割的基因结构[2,3]. 这一发现推翻了传统生物学关于“一种基因对应一种蛋白质”的观点, 接下来 1981 年, 第一次在哺乳动物编码荷尔蒙降血钙素的基因中发现可变剪接现象[4], 20 世纪 80 年代初, 在编码免疫球蛋白的基因中发现可变剪接[4,5]. 此后, 可变剪接被发现广泛存在于真核生物中[6]. 大量关于可变剪接的实验性文章不断发表出来. 很长一段时间, 对可变剪接机制的研究多停留在对单个基因的可变剪接现象和机制的研究上, 而缺少更为系统更为综

收稿日期: 2010-09-01; **修回日期**: 2010-12-01

基金项目: 国家自然科学基金资助项目(61071174); 国家 863 计划项目(2008AA01Z148)

作者简介: 王科俊(1962 -), 男, 黑龙江哈尔滨人, 哈尔滨工程大学自动化学院教授, 博士生导师, 主要从事模式识别、多模态生物特征识别、生物信息学研究, E-mail: wangkejun@hrbeu.edu.cn; 吕俊杰(1982-), 女, 黑龙江齐齐哈尔人, 博士研究生, 主要从事生物信息学研究.

合的分析. 究竟什么是影响基因产生可变剪接的因素，这些基因在产生可变剪接时有着怎样的机制或规律，不同的可变剪接的产物对蛋白质的功能有什么样的影响，等等诸多问题单纯的靠逐个基因地进行实验是无法解决的. 近 20 年来，应用计算方法对可变剪接现象进行研究开始引起人们的关注. 我国对可变剪接的研究起步比较晚，近几年才在个别单位开展，目前开展这方面研究的研究机构主要有：中科院上海生命科学研究院、清华大学生物信息研究所智能技术与系统国家重点实验室、国防科学技术大学、北京大学、上海交通大学、中科院华大基因研究中心、中科院生物物理所、天津大学、内蒙古工业大学等.

可变剪接受时间和空间的限制，在不同的组织中，在相同组织的不同细胞中，在同一组织的不同发育阶段，对病理过程的不同反应等等过程中均会产生不同的剪接变体. 研究发现[7]：94%以上的人类基因存在可变剪接，平均每个人类的基因有多于 5 个转录物变体，其中多达 50%的致病突变会影响剪接；可变剪接的异常改变使得基因在转录后期产生异常的剪接变体，编码出异常的蛋白质，从而导致人类遗传疾病，甚至癌变. 目前，对可变剪接与疾病的相关性已在单基因与单癌症水平上展开了研究[7]. 对可变剪接机制与疾病相关性的深入研究，可以为人类遗传疾病与癌症的临床诊断、预测，治疗方案的制定以及相关的生物制药提供理论上的指导.

1　可变剪接的基本类型

可变剪接的类型主要有 7 种(如图 1 所示)，分别为内含子保留（图 a，intron retention)、可变 3′剪接位点（图 b，alternative 3′ splice site)、可变 5′剪接位点（图 c，alternative 5′ splice site)、外显子跳过(图 d，exon skipping)、互斥可变外显子（图 e，mutually exclusive exons)、可变初始外显子(图 f，alternative initial exon)和可变终末外显子(图 g，alternative last exon). 表 1 为 Christopher Burge 实验室基于 RNA-seq 的最新实验数据，预测出人类基因里各种模式的可变剪接类型所占的比例[8].

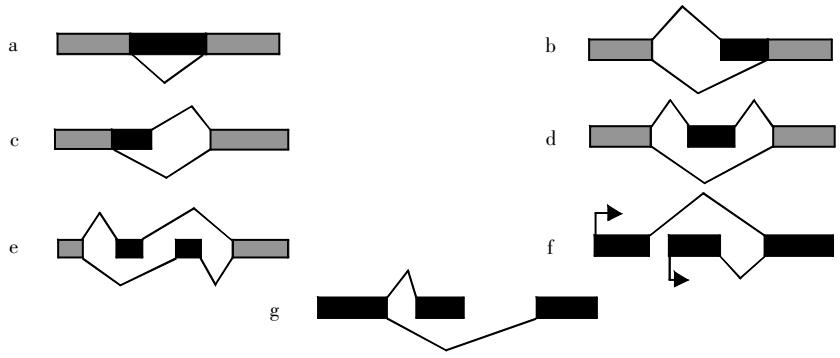

图 1　可变剪接主要类型

黑色方块表示内含子，灰色方块表示外显子；折线表示发生剪接，方块上方的折线表示正常剪接发生的位置，方块下方的折线表示发生可变剪接的位置.

Fig.1　Main patterns of alternative splicing

Black boxes mean exons, gray boxes mean introns; fold lines mean splicing, up-fold lines mean the sites of constitutive splicing, and down-fold lines mean the sites of alternative splicing.

表 1　可变剪接类型及所占比例

Table 1　Patterns and percentage of alternative splicing

Patterns of alternative splicing	Proportion/(%)
Intron retention	1
Alternative 3′ splice site	16
Alternative 5′ splice site	15
Exon skipping	35
Mutually exclusive exons	4
Alternative initial exon	13
Alternative last exo	8

2 可变剪接调控机制

剪接体的核心部分包括一组小核 RNA(small nuclear RNA，snRNA）以及与之结合的蛋白质，它们以严格的程序组装成剪接体(splicesome). snRNA 成员分别为 U1、U2、U4、U5 和 U6，长度在 106(U6)~185(U2)个核苷酸之间. snRNA 与蛋白质结合在一起形成小核核糖核蛋白(small nuclear ribonucleic protein，snRNP)，剪接体复合物中还有大量的其它因子. 如 SR 蛋白质及其相关蛋白质. 这些剪接因子与调节蛋白相互作用来调节剪接，可变剪接过程的调控机制多种多样. 按剪接调控因子结合在 RNA 上的位置及作用方式，可以分为外显子增强子（ESE)、外显子抑制子(ESS)、内含子增强子(ISE)和内含子抑制子(ISS)[9]. 可变剪接的各调控元件，不同的增强子和抑制因子的排列组合共同决定剪接模式的变化. 目前已逐步开展对于这些剪接调控因子的作用机理、作用位置及其对剪接模式的影响等的研究.

近年来对于剪接增强机制的研究主要集中在对 SR 蛋白质的调控机制上. SR 蛋白质对 3′剪接位点的增强通过吸引 U2AF65 结合在信号较弱的多嘧啶区域. 另外一种 3′增强机制则不需要吸引 U2AF65，而是通过 SRm160 和 SRm300 等剪接调控因子桥接 ESE 和剪接体组成元件来实现[10]. 这些调控因子含有 RS(arginine-serine-rich)区域，但没有 RRM(RNA recognition motif)区域，并能和 snRNP、SR 蛋白质之间形成多重作用关系. 内含子增强因子主要通过结合某些反式作用的蛋白质来增强剪接位点的信号.

剪接信号的抑制通常由 hnRNP 蛋白质家族(如 SXL、PTB 和 hnRNPA1 等)的作用来实现.最简单的一种机制是：结合在抑制因子上的蛋白质会直接阻止剪接体的结合. 然而，剪接的抑制经常需要通过结合到多个同类型的抑制因子的蛋白质之间的协调作用来完成. 已发现的有两种协调抑制的形式[11,12]，这两种调控形式或通过结合在高结合强度的抑制因子上的蛋白质的增殖，或通过在 RNA 外将蛋白质形成环形来产生一个“抑制区域”. 在这两种情况下，外显子、剪接位点和增强因子都无法接触剪接因子.

在许多情况下，可变剪接事件可能由更为复杂的机制造成，比如多个增强和抑制因子同时作用. 哺乳动物 *src* *N*1 外显子的神经系统特有的可变剪接是一个典型的多因子调控剪接例子[13]. 在非神经系统的细胞中，*N*1 外显子的剪接取决于4 个侧翼的 CUCUCU 元件，它们和 PTB 协调地进行结合. U2-snRNP 结合在与 PTB 结合位置相反的上游内含子中. 相反，U1snRNP 的结合则不受 PTB 的抑制影响. 在神经系统的细胞中，*N*1 的包含则有赖于下游的 ISE. 该 ISE 和其中的一个 CUCUCU 元件重叠，并与 hnRNPF，hnRNPH 和 KH 类型的剪接调控蛋白质作用. 该 ISE 复合体被 nPTB 蛋白质所吸引，该 nPTB 结合在 CUCUCU 元件，但没有 PTB 的抑制能力强. nPTB 有双重作用，一方面会对抗抑制因子，另一方面却促进 ISE 复合体. nPTB 的表达量有限，因此不足以激活 *N*1 的剪接. *N*1 外显子中的 ESE 结合 SF2/ASF、hnRNPF 和 hnRNPH，而 hnRNPA1 的结合则起到抑制的调控作用. 该剪接机制极为复杂，尚未完全明确.

3 可变剪接与遗传性疾病

可变剪接是一个各种影响因素高度协调的机制，其中任何一个关键因素产生突变，都会改变正常的剪接模式，得到变异的转录体，编码出异常的蛋白质，从而导致疾病甚至癌变. 已有大量与可变剪接相关的疾病记载，但是，因为有一些疾病中的突变不是致病的关键因素，所以记载的数目远远低于实际发生的数量(表 2 给出了现有典型与可变剪接相关疾病的信息). 对可变剪接机制与疾病相关性的深入研究，可以为疾病与癌症提供合理的治疗方案. 下面讨论几种与疾病有密切联系的典型剪接变异.

3.1 点突变

早在 1992 年，Krawczak M 等就在对 101 个造成人类遗传疾病的点突变研究中发现，多达 15%的点突变会破坏 mRNA 的剪接[14]. 该比例在 2005 年提高到 60%[15]. 2009 年 1 月最新公布的人类基因突变数据库（Human Gene Mutation Database，HGMD[16])中有多达 8 532 个人类基因的点突变与遗传病相关，占数据库总数的 10%. 需要注意的是，这个数据是来自于具体的文献记载的实验数据. 实际的破坏剪接的致病点突变所占比例远远高于 10%. 对由于剪接位点的选择变化所导致的疾病研究较为充分的代表是地中海贫血[17]和家族性自律神经失调. 家族性自律神

经失调是一种由 i-kappa-B 蛋白激酶及其相关的复杂蛋白质功能丧失引起的隐性疾病，在德系犹太人中，发病率达 1/3600 [18]. 患儿的和脱髓鞘相关的神经系统发育异常，临床表现为呕吐、病危、步态不稳、对疼痛感减弱或消失. 99.5 %以上的家族性自律神经失调患者的 20 号外显子的 5′剪接位点在 20 号内含子的位置 6 处由 T 突变为 C. 这个点突变打断了与 U1-snRNA 的碱基配对，U1-snRNA 与上游外显子的最后 3 个核苷酸以及下游内含子的前 6 个核苷酸相配对. 大多数 5′剪接位点处形成 7 对与 U1-snRNA 互补的碱基对. 这意味着，将导致 5′剪接位点和 U1-snRNA 有 3 个是错配的. 生物信息学分析表明[19]，这些错配不是随机分布的. 他们或者削弱 5′剪接位点的外显子部分，然后通过增加结合到内含子的部分做补偿，或者削弱 5′剪接位点的内含子部分，然后通过增加结合到外显子的部分做补偿. 在 *IKBKAP* 基因的 20 号外显子中，剪接位点的外显子部分是被削弱的，由于在位置-1 处是 A，与其配对的 T 突变为 C 削弱了 5′剪接位点的内含子部分，导致外显子跳过[19]. 家族性自律神经失调的例子说明了剪接位点突变的复杂性，同时也要联系其他调控元件仔细分析，它进一步表明，关键调控基因的错误剪接会严重影响其它基因.

表 2　互联网上关于遗传疾病的信息
Table 2　Information about genetic diseases on the web

General information about alternative splicing	http: //www.eurasnet.info/
Familial dysautonomia	http://www.familialdysautonomia.org/
Substances that influence alternative splicing	http://www.stamms-lab.net/cpds.htm
Tauopathies	http://www.molgen.ua.ac.be/ADMutations/
Hutchinson-Gilford progeria syndrome	http://www.progeriaresearch.org/
Spinal muscular atrophy	http://www.fsma.org/ http://www.mda.org
Prader-Willi syndrome	http://www.pwsausa.org/ http://www.fpwr.org/
Myotonic dystrophy	http://www.myotonic.com http://www.mda.org
Medium-chain acyl-CoA dehydrogenase (MCAD) deficiency Myotonic dystrophy	http://www.fodsupport.org/mcad_fam.htm
Frontotemporal lobar dementias/-amyotrophic lateral sclerosis	http://www.alsa.org/

3.2　与疾病相关的短重复元件突变

在很多外显子中检测出短重复序列，这些重复序列有助于某些剪接因子的识别[20]. 它们长度的变化还会改变基因的剪接模式，例如：内皮细胞一氧化氮合成酶基因包含一个内含子多态化的 CA 重复区，该重复区与冠状动脉疾病、高同型半胱氨酸血症，在性别特异性剪接模式上的患病几率有着密切的联系[21]; SELEX 实验和功能研究表明[22]: CA 重复元件与 hnRNP L 结合，并且 hnRNP L 活性的激活依赖于 CA 重复元件的长度.基因芯片的研究表明[23]: 内含子序列中丰富的 CA 重复序列可以影响可变剪接，并且其中某些 CA 重复元件的多态性会引起人类疾病. 另一个与疾病相关的短重复元件是 UG，它会引起 *CFTR* 基因的 9 号外显子剪接异常，导致囊性纤维化[23]. Alu 重复元件是人类基因组中最大的重复群体，他们约占基因组序列总数的 10%[24]，Alu 重复元件包含潜在的剪接位点，可以进化成外显子[25]，大约 5%以上的人类可变外显子都来源于 Alu 重复序列[26]. 毫无疑问，Alu 重复元件中的突变会产生异常的 mRNA，从而导致人类疾病，例如 Alport 综合症[27]、先天性白内障[28]、异形面神经综合症[28]、VII 型黏多糖症[29]. 这些例子解释了短重复序列是如何改变可变剪接调控平衡的，尽管他们有时位于内含子中或者看似与 mRNA 无关. 外显子化的 Alu 重复元件表明了如何利用新的重复元件通过进化获得新的功能，不当外显子化的 Alu 重复元件所引起的疾病可以被看作是失败的进化尝试.

3.3　可变剪接产生蛋白质亚型的比例改变

神经退化性疾病是一组描述几种中枢神经系统的疾病，它们的一个共同的病理特征是：细胞内存在异常堆积的 tau 蛋白. 该 tau 蛋白是由位于 17 号染色体上的单个基因编码的，这个基

因存在广泛的可变剪接，16 个外显子中有 8 个经历可变剪接[30]. 人类基因的可变剪接受时间和空间的限制，例如，2 号外显子、3 号外显号子和 10 号外显子具有成人细胞组织特异性，在脑组织的不同区域中剪接是有显著差异的[30]. Tau 蛋白通过微管重复区域与微管结合. 这些微管蛋白结合区中有一个是由可变 10 号外显子编码的，10 号外显子包含会产生有 4 个微管重复区域的蛋白质(4R)，而 10 号外显子跳过则会产生有 3 个微管重复区域的蛋白质(3R). 在成人中这个剪接事件具有物种特异性. 在人类基因中，10 号外显子在成年人细胞中是可变剪接的，而在小鼠中，这个外显子是组成性的. 在这两个物种中，该外显子的作用都是在发育过程中被调节的. Andreadis 对 *Tau* 基因的研究发现[30]：*Tau* 基因中罕见的显性突变会导致额颞叶痴呆症、帕金森病以及 17 号染色体疾病. 大多数的突变会影响 10 号外显子编码微管结合点的部分. *Tau* 的 10 号外显子的突变导致调节元件分解. 该外显子表现为由 4 个增强子和 3 个沉默子组成的可变外显子. 增强子调控区的突变会减弱外显子的作用，反之，沉默子调控区的突变则会增加外显子的作用. 研究发现 10 号外显子的突变改变其正常包含的前体 mRNA 编码 3R 和 4R 重复 *Tau* 亚型的数量，这和 17 号染色体疾病相关[31,32]. 这些数据清楚地表明，剪接突变通过改变 4R 和 3R 亚型的比例导致了神经系统的疾病. 这个例子表明，由可变剪接产生的蛋白质亚型比例变化可引起人类疾病. 进一步说明如何分析基因的变异体有助于理解剪接调控机制. 剪接存在复杂的物种特异性，但可以通过建立动物模型来研究人类病理.

3.4 单核苷酸多态性

高胆固醇是导致动脉粥样硬化的主要因素. 低密度脂蛋白受体(LDLR)将血液中的低密度脂蛋白去除. 低密度脂蛋白受体基因突变是导致高胆固醇血症的主要原因，同时高血脂是引起 LDLR 突变的主要因素[33]. 研究发现[34]：在 LDLR 的 12 号外显子中发现一个单核苷酸多态性(SNP)可以促使这一外显子跳过. 该 SNP 可以促使更年期前的妇女肝脏中 12 号外显子跳过，但是，对更年期后的妇女或者男子并没有该作用. 在更年期前的妇女肝脏组织中该 SNP 和剪接模式都与胆固醇水平相关，在男子中却不存在这种情况. 12 号外显子跳过产生一个截短的 LDLR，该 LDLR 缺乏膜结合和内化必要的跨膜区. 这可能是因为由 12 号外显子跳过产生的异常蛋白严重阻止吸收低密度脂蛋白. 这个实例阐述了由 SNP 引起的 12 号外显子跳过与胆固醇水平有关这一现象. 该 SNP 依赖于性别的原因尚不清楚，但可能是高雌激素水平影响基因的转录水平或者可变剪接[33]. 载脂蛋白 E(*ApoE* 基因)是低密度脂蛋白受体的配体，载脂蛋白 E 的等位状态是导致阿尔茨海默氏病的一个重大因素. 因此，需要研究 LDLR 的 12 号外显子中 SNP 是否与阿尔茨海默氏症有关. 研究发现，在男性中该 SNP 和阿尔茨海默氏病相关，在女性中却不存在这种情况[34]. 这个例子很好地说明了单核苷酸多态性能够影响可变剪接，从而导致人类疾病. 也反映出可变外显子调控机制是由各种因素综合控制的，一个突变依赖于其他因素，在上述实例中依赖于性别和年龄.

3.5 剪接因子的缺失

反式作用因子如果受到剪接的影响，则可能会导致其所调控或参与调控的所有基因的表达发生异常而导致疾病. 剪接体由 5 种 snRNP 和 200 多种蛋白质组成，这些蛋白质包括蛋白激酶、磷酸酶和解旋酶等剪接体所需要的酶，以及相关的 mRNA 输出因子、转录因子和剪接反式作用因子等等[35]. 因此，生成这些蛋白质和 snRNP 的基因如果由于可变剪接被破坏，将会使它们在剪接过程中的调控功能丧失，最终可能导致某种疾病.

Mordes 等对脊髓性肌萎缩（spinal muscular atrophy，SMA)[36]的研究，以及 Winkler 等对色素性视网膜炎(retinitis pigmentosa)[37]的分析表明：这两种遗传疾病都是由于生成 snRNP 的基因发生突变所导致的. SMA 是一种影响运动神经的常染色体隐性紊乱，由于 *SMN*1 基因产物的遗失(即：突变造成该基因所生成的蛋白质异常)而导致. 该基因产物是细胞质中生成核心 snRNP 复合物所需要的，而失去 snRNP 已被证明和疾病有直接联系[38]. 色素性视网膜炎是导致失明的一种最常见的原因，4 000 个人里便有一人患有此疾病[37]. 生成 U4U5U6 三聚 snRNP 的基因中包含色素性视网膜炎的 3 个主要基因——*PRPF*31、*PRPF*8 和 *HPRP*3. 而 U4U5U6 是剪接体的主要组成部分.

上述两个实例是典型的由剪接因子缺失导致剪接异常而产生的疾病，说明剪接因子在剪接调控过程中的重要性，也说明了剪接调控机制的复杂性.

4　可变剪接与癌症

除了上述的人类遗传疾病外，剪接异常也是众多癌症的一个常见特征. 和癌症相关的细胞迁移、细胞生长调控、荷尔蒙响应性、细胞死亡和化疗反应中基因表达变化都可能和可变剪接相关[39~41]. 有证据表明影响致癌基因(*oncogene*)、抑癌基因(*tumour suppressor*)和其它癌症相关的基因剪接的突变在癌症的起始和过程中都有因果关系[41]. 目前，对导致癌症中剪接缺陷的机制理解尚不清楚. 在一些个案研究中发现，顺式作用元件中的遗传和体细胞突变，反式调节因子的成分、浓度、定位以及活性的变异，都会影响剪接位点的识别和作用，从而导致癌变[39,40]. 表 3 给出了典型的由可变剪接异常导致的癌症实例.

我们通过已经得到很好研究的具体实例来阐述癌症中剪接机制的典型改变、致病机理，分析可变剪接对抑癌基因与致癌基因的作用原理，为今后癌症的治疗提供理论上的参考.

表 3　可变剪接异常导致的癌症
Table 3　Cancer caused by alternative splicing

Gene	Splice variant	Cancer type
Survivin[42,43]	Survivin 2B with pro-apoptotic properties	Breast carcinoma and late stage or metastatic gastric cancer
FHIT[44,45]	Aberrant transcripts	Gastric, cervical, thyroid and testicular germ-cell tumours
*AIB*1[46]	Isoform lacking exon 3	Breast cancer
VEGF[47]	Isoforms lacking exon 6	Non-small cell lung cancer
Actinin 4[48]	Variant Va	Small cell lung cancer
Cathepsin B[49]	Certain isoforms	Colon cancer
RON[50]	RonΔ165　RonΔ160　RonΔ155	Colorectal carcinoma

4.1　致癌信号

CD44 是一种多功能细胞表面糖蛋白，参与细胞增殖、分化、黏附和迁移等过程[51]. 通过其内部外显子的 10 种变体的组合与包含可以产生多种 CD44 亚型，某些特定的 CD44 剪接变体，特别是那些包含外显子 v5、v6 以及 v7 变体的亚型，在多种肿瘤中都过分表达，已经有研究证实这些变体在肿瘤细胞侵袭和转移过程中起着重要的作用[51,52]. Konig 和 Matter 等的研究表明[53,54]：外显子 v5 包含是由外显子的剪接调控元件——剪接增强子和剪接抑制子共同调节的. 致癌的 RAS 信号转导途径的激活刺激外显子 v5 包含. 由于诱导型剪接不需要蛋白的重新合成，其剪接调控可能是由后转译调节或通过控制剪接因子的细胞内定位实现的. 近来对 STAR（信号转导和 RNA 激活）蛋白质家族的一员 Sam68 的研究表明[55]：对 ERK MAP 酶磷酸化使得其被激活，进而证明是 RAS 诱导突变引起的外显子 v5 包含，从而导致的癌变.

4.2　顺式作用元件的破坏与癌症

与疾病相关的点突变至少有 15%会导致剪接缺陷，这表明顺式调控元件的破坏会导致大量的错误剪接. 已经在许多易发生癌变的基因中发现导致剪接缺陷的胚系突变，例如 *BRCA*1、*BRCA*2、*CDKN*2*A* 和 *APC* 基因[56-61]，而由于体细胞性突变导致的剪接缺陷则很少发生. 胚系序列中的点突变和调节元件突变都会增加癌症的易感性，例如，*BRCA*1 基因的 18 号外显子中的一个遗传性无义突变破坏一个外显子剪接增强子(ESE)及 SR 蛋白质 SF2/ASF 的结合位点，从而结构性 18 号外显子的不当跳过[62]. 抗原递呈细胞（APC）的胚系突变会导致家族性多发性腺癌(FAP). 突变发生在基因的各个位置，特定位置决定疾病的严重程度[60]. 最近的研究发现两个破坏剪接调控元件的突变. 在 4 号内含子的供体位点保守模体 GT 处插入一个 T 引起 4 号外显子跳过，从而导致轻度的家族性多发性腺癌 [60]；7 号内含子受体位点处一个 G 由 A 替代会产生一个隐秘的剪接位点，导致 8 号外显子的起始处有一个单核苷酸缺失；一个单碱基移码导致 APC 缩短，进而引致典型息肉症[57]. 经常在神经纤维瘤的 *NF*1 基因中发现胚系突变和体细胞突变的杂合性缺失. Serra 等对 NF1 体细胞突变做了系统的研究[62]，结果表明大多数点突变会导致剪接缺陷，包括外显子跳跃、可变 5′剪接位点和可变 3′剪接位点的使用. 据我们的了解，这是唯一的体

细胞突变导致剪接位点识别改变的研究.

4.3 反式调节因子的破坏与癌症

许多研究表明在肿瘤中，剪接因子的表达会发生特异性的改变. 已在人类卵巢癌和小鼠的乳腺癌模型发育过程中发现 Tra26、YB-1、SR 蛋白质 SC35 和 ASF/SF2 的高度磷酸化，mRNA 过高表达. 此外，这些改变还与 CD44 剪接模式增加的复杂性有关[63]. 虽然在癌症中很少发现体细胞剪接突变，但是排除了高突变率是剪接缺陷的主要原因这一假设. 一般来说，剪接增强子和沉默子，特别是那些在内含子深处的，保守性不强的，因此，它们的识别以及评价单核苷酸替代对剪接效率的影响，主要视情况应用实验的方法. 目前，很多突变被归类为中性、错义或无义的，进而可能影响剪接的过程. 这使得研究人员致力于研究肿瘤是否有错配修复缺陷及较高的剪接缺陷率. 如果一个剪接因子的表达式或活性在一个癌细胞中发生改变，它很可能会影响其他基因的剪接. 因此，应该根据剪接缺陷为癌症分类. 到目前为止，如何评价剪接因子的表达式或活性改变？这些改变在人类癌症中发生的频率？以及什么基因受到影响？有多少基因受到影响？等等研究还没有开展.

5 研究策略

上面讨论的实例清晰的表明：异常的可变剪接调节是导致人类遗传疾病与癌症的主要来源. 要深入了解疾病的致病机理，研究开发遗传疾病与癌症的治疗方案，以及相应的基因靶向疗法、生物药物设计等，可以在可变剪接层面上展开研究，即对可变剪接进行深入细致的研究，从而找到与疾病相关的异常剪接因素. 对于这方面的研究，主要从顺式调节元件与反式作用因子两方面着手.

由于受计算技术与生物技术的限制，在对顺式作用元件的早期研究上，主要基于剪接位点附近的外显子与内含子中某些特定调节元件的研究上[21,26]，仅仅对单个基因或者某个组织进行研究[11,21]，缺乏在基因组范围的分析，随着生物信息技术的发展，使得在基因组范围内研究可变剪接成为可能. 近期的研究热点转向对各组织的可变剪接调节元件进行综合分析，我们在对 11 对人类组织的可变剪接进行研究中发现[64]：某些剪接在各个组织中都发生，而有些剪接只发生在特定的组织中，即在很多组织中都存在自己特异的剪接模式，或者某些剪接模式在特定的组织中过分的表达，这些具有组织特异性的剪接模式发生突变将导致相应的组织疾病，甚至癌变. 为进一步研究顺式作用元件与疾病的相关性，对顺式作用元件的研究热点将集中在疾病的组织特异性剪接、癌症特异性剪接、物种特异性剪接等方向.

与疾病相关的另一重要机制是反式作用因子，即调节可变剪接的各蛋白质、酶、小核 RNA、hnRNP 等，研究这些调节因子的异常作用与疾病的相关性，首先需要对这些因子的调节机制的进行深入分析，这些调节因子通过 RNA-RNA、RNA-蛋白质、蛋白质-蛋白质等的相互作用来调节可变剪接. 例如：剪接调节因子结合在基因序列的不同位置上，将对剪接产生不同的影响，精确识别出剪接因子结合到基因序列上的位置，对于可变剪接与疾病的治疗均具有重大的意义，以往对这一课题的研究主要集中在对基因序列的分析上，仅考虑序列信息，忽视了结构因素对结合位置的影响，我们将结构因素考虑进来，结合序列信息与结构信息对结合位点进行研究发现，结合了结构因素后可以显著提高结合位点的识别率[65]，这说明，结构因素同样影响可变剪接，结构的改变也会导致异常的剪接，从而导致疾病. 我们考虑的结构信息仅是应用 Vienner 软件中的 Mfold 软件包对最稳定的二级结构进行预测，这是基于最小自由能开发的二级结构预测程序，预测的仅是最稳定的折叠，与生物实际的二级结构存在差异，要进一步研究结构对剪接作用因子与基因序列结合位置的影响，还需要对二级结构进行进一步的研究，这也将是未来的研究热点.

6 结论

可变剪接机制是基因在后转录时期的一项复杂而精密的调控机制，与人类遗传疾病有着密切的关系，更是癌症表达的天然来源. 对可变剪接调控机制与疾病相关性的研究，有助于对致病机理的理解与深入分析；也可以为重大遗传病诊断与治疗提供理论上的指导；为肿瘤的靶向基因疗法提供准确的定位，为开发癌症的治疗方法提供新的思路；更有助于相关疾病的生物制药方案的制定. 目前已有学者提出了在 RNA 剪接层次上治疗人类遗传疾病和癌症的思路[66~68]，如：

Wang Guey-Shin 等提出从顺式或反式作用因子出发改变目标基因的剪接形式，或者在 RNA 和蛋白质层次攻击导致疾病的剪接变体来获得疗效. 但是，目前受到可变剪接调控机制研究的制约，更为深入的实际应用研究尚依赖于该领域基础科学的进一步发展. 对组织特异性、物种特异性、遗传疾病特异性以及肿瘤特异性可变剪接等机制的研究，将会是今后生物信息学的研究热点. 提高正确识别分子的能力，针对剪接作用元件的异常，剪接调节因子的突变，以及剪接因子的结合位点变异等，研发修补措施，阻止异常的剪接发生，将会成为未来人类遗传疾病与癌症的重要治疗方法.

参考文献(References):

[1] GILBERT W. Why genes in pieces?[J]. Nature, 1978, 271(5645): 501.

[2] BERGET S M, MOORE C, SHARP P A. Spliced segments at the 5′ terminus of adenovirus 2 late mRNA[J]. PNAS, 1977, 74(8): 3171-3175.

[3] CHOW L T, ROBERTS J M, LEWIS J B, *et al*. A map of cytoplasmic RNA transcripts from lytic adenovirus type 2, determined by electron microscopy of RNA: DNA hybrids[J]. Cell, 1977, 11(4): 819-836.

[4] LEFF S E, ROSENFELD M G, EVANS R M. Complex transcriptional units: diversity in gene expression by alternative RNA processing[J]. Annual Reviews of Biochemistry, 1986, 55(1): 1091-1117.

[5] MAKI R, ROEDER W, TRAUNECKER A, *et al*. The role of DNA rearrangement and alternative RNA processing in the expression of immunoglobulin delta genes[J]. Cell, 1981, 24(2): 353-365.

[6] BLACK D L. Mechanisms of alternative pre-messenger RNA splicing[J]. Annual Reviews of Biochemistry, 2003, 72(1): 291-336.

[7] VENABLES J P. Unbalanced alternative splicing and its significance in cancer[J]. Bioessays, 2006, 28(4): 378-386.

[8] WANG E T, SANDBERG R, LUO S J, *et al*. Alternative isoform regulation in human tissue transcriptomes[J]. Nature, 2008, 456(7221): 470-476.

[9] GROMAK N, MATLIN A J, COOPER T A, *et al*. Antagonistic regulation of alpha-actinin alternative splicing by CELF proteins and polypyrimidine tract binding protein[J]. RNA, 2003, 9(4): 443-456.

[10] BLENCOWE B J, BAUREN G, ELDRIDGE A G, *et al*. The *SRm*160/300 splicing coactivator subunits[J]. RNA, 2000, 6(1): 111-120.

[11] CARTEGNI L, CHEW S L, KRAINER A R. Listening to silence and understanding nonsense: exonic mutations that affect splicing[J]. Nature Reviews Genetics, 2002, 3(4): 285-298.

[12] WAGNER E J, GARCIA-BLANCO M A. Polypyrimidine tract binding protein antagonizes exon definition[J]. Molecular and Cellular Biology, 2001, 21(10): 3281-3288.

[13] CHOU M Y, UNDERWOOD J G, NIKOLIC J, *et al*. Multisite RNA binding and release of polypyrimidine tract binding protein during the regulation of *c-src* neural-specific splicing[J]. Molecular Cell, 2000, 5(6): 949-957.

[14] KRAWCZAK M, REISS J, COOPER D N. The mutational spectrum of single base-pair substitutions in mRNA splice junctions of human genes: causes and consequences[J]. Human Genetics, 1992, 90(1-2): 41-54.

[15] LOPEZ-BIGAS N, AUDIT B, OUZOUNIS C, *et al*. Are splicing mutations the most frequent cause of hereditary disease?[J]. FEBS Letters, 2005, 579(9): 1900-1903.

[16] STENSON P D, MORT M, BALL E V, *et al*. The human gene mutation database: 2008 update[J]. Genome Medicine, 2009, 1(1): 13.

[17] RADMILOVIC M, ZUKIC B, STANKOVIC B, *et al*. Thalassemia syndromes in Serbia: an update[J]. Hemoglobin, 2010, 34(5): 477-485.

[18] MAAYAN C H, KAPLAN E, SHACHAR S H, *et al*. Incidence of familial dysautonomia in Israel 1977~1981[J]. Clinical Genetics, 1987, 32(2): 106-108.

[19] CARMEL I, TAL S, VIG I, *et al*. Comparative analysis detects dependencies among the 5′ splice-site positions[J]. RNA, 2004, 10(5): 828-840.

[20] HUI J, BINDEREIF A, Alternative pre-mRNA splicing in the human system: unexpected role of repetitive sequences as regulatory elements[J]. Biological Chemistry, 2005, 386(12): 1265-1271.

[21] LAULE M, MEISEL C, PRAUKA I, *et al*. Interaction of CA repeat polymorphism of the endothelial nitric oxide synthase and hyperhomocysteinemia in acute coronary syndromes: evidence of gender-specific differences[J]. Journal of Molecular Medicine, 2003, 81(5): 305-309.

[22] HUNG L H, HEINER M, HUI J, *et al*. Diverse roles of hnRNP L in mammalian mRNA processing: a combined microarray and RNAi analysis[J]. RNA, 2008, 14(2): 284-296.

[23] BURATTI E, DORK T, ZUCCATO E, *et al*. Nuclear factor TDP-43 and SR proteins promote *in vitro* and *in vivo* CFTR exon 9 skipping[J]. The EMBO Journal, 2001, 20(7): 1774-1784.

[24] HASLER J, SAMUELSSON T, STRUB K. Useful 'junk': Alu RNAs in the human transcriptome[J]. Cellular and Molecular Life Sciences, 2007, 64(14): 1793-1800.

[25] LEV-MAOR G, SOREK R, SHOMRON N, *et al*. The birth of an alternatively spliced exon: 3′ splice-site selection in Alu exons[J]. Science, 2003, 300(5623): 1288-1291.

[26] SOREK R, AST G, GRAUR D. Alu-containing exons are alternatively spliced[J]. Genome Research, 2002, 12(7): 1060-1067.

[27] KNEBELMANN B, FORESTIER L, DROUOT L, *et al*. Splice-mediated insertion of an Alu sequence in the *COL4A*3 mRNA causing autosomal recessive Alport syndrome[J]. Human Molecular Genetics, 1995, 4(4): 675-679.

[28] VARON R, GOODING R, STEGLICH C, *et al*. Partial deficiency of the C-terminal-domain phosphatase of RNA polymerase II is associated with congenital cataracts facial dysmorphism neuropathy syndrome[J]. Nature Genetics, 2003, 35(2): 185-189.

[29] VERVOORT R, GITZELMANN R, LISSENS W, *et al*. A mutation (IVS8+0.6kbdelTC) creating a new donor splice site activates a cryptic exon in an Alu-element in intron 8 of the human *beta-glucuronidase* gene[J]. Human Genetics, 1998, 103(6): 686-693.

[30] ANDREADIS A. Tau gene alternative splicing: expression patterns, regulation and modulation of function in normal brain and neurodegenerative diseases[J]. Biochimica et Biophysica acta, 2005, 1739(2-3): 91-103.

[31] GALLO J M, NOBLE W, MARTIN T R. RNA and protein-dependent mechanisms in tauopathies: consequences for therapeutic strategies[J]. Cellular and Molecular Life Sciences, 2007, 64(13): 1701-1714.

[32] ANDREADIS A. Misregulation of tau alternative splicing in

neurodegeneration and dementia[J]. Progress in Molecular and Subcellular Biology, 2006, 44(1): 89-107.
[33] ZHU H, TUCKER H M, GREAR K E, *et al.* A common polymorphism decreases low-density lipoprotein receptor exon 12 splicing efficiency and associates with increased cholesterol[J]. Human Molecular Genetics, 2007, 16(14): 1765-1772.
[34] ZOU F, GOPALRAJ R K, LOK J, *et al.* Sex-dependent association of a common low density lipoprotein receptor polymorphism with RNA splicing efficiency in the brain and Alzheimer's disease[J]. Human Molecular Genetics, 2008, 17(7): 929-935.
[35] JURICA M S, MOORE M J. Pre-mRNA splicing: awash in a sea of proteins[J]. Molecular Cell, 2003, 12(1): 5-14.
[36] BRIESE M, ESMAEILI B, SATTELLE D B. Is spinal muscular atrophy the result of defects in motor neuron processes? [J]. Bioessays, 2005, 27(9): 946-957.
[37] MORDES D, LUO X Y, KAR A, *et al.* Pre-mRNA splicing and retinitis pigmentosa [J]. Molecular Vision, 2006, 12(1): 1259-1271.
[38] WINKLER C, EGGERT C, GRADL D, *et al.* Reduced U snRNP assembly causes motor axon degeneration in an animal model for spinal muscular atrophy[J]. Genes & Development, 2005, 19(19): 2320-2330.
[39] VENABLES J P. Unbalanced alternative splicing and its significance in cancer[J]. Bioessays, 2006, 28(4): 378-386.
[40] SKOTHEIM R I, NEES M. Alternative splicing in cancer: noise, functional, or systematic?[J]. The International Journal of Biochemistry & Cell Biology, 2007, 39(7-8): 1432-1449.
[41] SREBROW A, KORNBLIHTT A R. The connection between splicing and cancer[J]. Journal of Cell Science, 2006, 119(13): 2635-2641.
[42] KRIEG A, MAHOTKA C, KRIEG T, *et al.* Expression of different survivin variants in gastric carcinomas: first clues to a role of survivin-2B in tumour progression[J]. British Journal of Cancer, 2002, 86(5): 737-743.
[43] VEGRAN F, BOIDOT R, OUDIN C, *et al.* Distinct expression of survivin splices variants in breast carcinomas[J]. International Journal of Oncology, 2005, 27(4): 1151-1157.
[44] HUIPING C, KRISTJANSDOTTIR S, BERGTHORSSON J T, *et al.* High frequency of LOH, MSI and abnor-mal expression of FHIT in gastric cancer[J]. European Journal of Cancer, 2002, 38(5): 728-735.
[45] KRAGGERUD S M, AMAN P, HOLM R, *et al.* Alterations of the fragile histidine triad gene, FHIT, and its encoded products contribute to testicular germ cell tumorigenesis[J]. Cancer Research, 2002, 62(2): 512-517.
[46] REITER R, WELLSTEIN A, RIEGEL A T. An isoform of the coactivator AIB1 that increases hormone and growth factor sensitivity is overexpressed in breast cancer[J]. The Journal of Biological Chemistry, 2001, 276(43): 39736-39741.
[47] CHEUNG N, WONG M P, YUEN S T, *et al.* Tissue-specific expression pattern of vascular endothelial growth factor isoforms in the malignant transformation of lung and colon[J]. Human Pathology, 1998, 29(9): 910-914.
[48] HONDA K, YAMADA T, SEIKE M, *et al.* Alternative splice variant of actinin-4 in small cell lung cancer[J]. Oncogene, 2004, 23(30): 5257-5262.
[49] KEPPLER D, SLOANE B F. Cathepsin B: multiple enzyme forms from a single gene and their relation to cancer[J]. Enzyme Protein, 1996, 49(1-3): 94-105.
[50] ZHOU Y Q, HE C, CHEN Y Q, *et al.* Altered expression of the RON receptor tyrosine kinase in primary human colorectal adenocarcinomas: generation of different splicing *RON* variants and their oncogenic potential[J]. Oncogene, 2003, 22(2): 186-197.
[51] NAOR D, NEDVETZKI S, GOLAN I, *et al.* CD44 in cancer [J]. Critical Reviews Clinical Laboratory Sciences, 2002, 39(6): 527-579.
[52] FAUSTINO N A, COOPER T A. Pre-mRNA splicing and human disease[J]. Genes & Development, 2003, 17(4): 419-437.
[53] KONIG H, PONTA H, HERRLICH P. Coupling of signal transduction to alternative pre-mRNA splicing by a composite splice regulator[J]. The EMBO Journal, 1998, 17(10): 2904-2913.
[54] MATTER N, MARX M, WEG-REMERS S, *et al.* Heterogeneous ribonucleoprotein A1 is part of an exon-specific splice-silencing complex controlled by oncogenic signaling pathways[J]. The Journal of Biological Chemistry, 2000, 275(45): 35353-35360.
[55] MATTER N, HERRLICH P, KONIG H. Signal-dependent regulation of splicing via phosphorylation of *Sam*68[J]. Nature, 2002, 420(6916): 691-695.
[56] LIU H X, CARTEGNI L, ZHANG M Q, *et al.* A mechanism for exon skipping caused by nonsense or missense mutations in *BRCA*1 and other genes[J]. Nature genetics, 2001, 27(1): 55-58.
[57] CHARAMES G S, CHENG H, GILPIN C A, *et al.* A novel aberrant splice site mutation in the *APC* gene[J]. Journal of medical genetics, 2002, 39(10): 754-757.
[58] AGATA S, DENICOLO A, CHIECO-BIANCHI L, *et al.* The *BRCA*2 sequence variant IVS19t1G->A A leads to an aberrant transcript lacking exon 19[J]. Cancer Genetics and Cytogenetics, 2003, 141(2): 175-176.
[59] RUTTER J L, GOLDSTEIN A M, DAVILA M R, *et al.* *CDKN2A* point mutations D153spl (c.457G>T) and IVS2t1G>T result in aberrant splice products affecting both p16INK4a and p14ARF[J]. Oncogene, 2003, 22(28): 4444-4448.
[60] NEKLASON D W, SOLOMON C H, DALTON A L, *et al.* Intron 4 mutation in *APC* gene results in splice defect and attenuated FAP phenotype[J]. Familial Cancer, 2004, 3(1): 35-40.
[61] TUOHY T M, DONE M W, LEWANDOWSKI M S, *et al.* Large intron 14 rearrangement in *APC* results in splice defect and attenuated FAP[J]. Human Genetics, 2010, 127(3): 359-369.
[62] SERRA E, ARS E, RAVELLA A, *et al.* Somatic NF1 mutational spectrum in benign neurofibromas: mRNA splice defects are common among point mutations[J]. Human Genetics, 2001, 108(5): 416-429.
[63] FISCHER D C, NOACK K, RUNNEBAUM I B, *et al.* Expression of splicing factors in human ovarian cancer[J]. Oncology Reports, 2004, 11(5): 1085-1090.
[64] WANG G S, COOPER T A. Splicing in disease: disruption of the splicing code and the decoding machinery[J]. Genetics, 2007, 8(10): 749-761.
[65] WANG X, WANG G, SHEN C, *et al.* Using RNase sequence specificity to refine the identification of RNA-protein binding regions[J]. BMC Genomics, 2008, 9(1)1: 17
[66] WANG X, WANG K J, RADOVICH M, *et al.* Genome-wide prediction of cis-acting RNA elements regulating tissue-specific pre-mRNA alternative splicing[J]. BMC Genomics, 2009, 10(1): 1-4.
[67] BERASAIN C, GONI S, CASTILLO J, *et al.* Impairment of pre-mRNA splicing in liver disease: Mechanisms and consequences[J]. World Journal of Gastroenterology, 2010, 16(25): 3091-3102.
[68] ZHANG C, FRIAS M A, MELE A, *et al.* Integrative modeling defines the Nova splicing-regulatory network and its combinatorial controls[J]. Science, 2010, 329(5990): 439-443.