

DMD 基因内含子的重复序列与内含子不稳定性 和基因缺失的关联

盛文利, 陈江瑛, 朱良付, 刘焯霖

(中山大学附属第一医院神经科, 广东 广州 510080)

摘 要: 【目的】为了研究 DMD 基因内含子的重复序列与内含子不稳定性及基因缺失的关联, 我们分析了缺失型 DMD 病人的连接片段的断裂点并研究了 DMD 基因内含子结构与内含子不稳定性之间是否存在相关。【方法】多重引物 PCR 法鉴定缺失型 DMD 患者, 分别克隆 46 和 51 号外显子缺失后形成的连接片段, 测定连接片段的序列; 并用计算机分析 DMD 基因可以获得的全部内含子的结构特点。【结果】对 46 号和 51 号外显子缺失后形成的连接片段的序列分析表明, 5' 和 3' 端的断裂点均位于重复顺序, 断裂点附近无小插入、小缺失或点突变。对连接片段的二级结构分析表明, 所有的断裂点均位于单链发夹结构的非匹配区。对 DMD 基因可以获得的全部内含子的结构分析表明, 在 DMD 基因全长内含子中重复序列占大约 34.8%。大量重复序列的存在是许多内含子的重要结构特征。【结论】重复序列是 DMD 基因多数内含子的关键结构, 与内含子的不稳定性相关。在原有 DNA 序列的基础上, 单链发夹结构的形成增加了 DMD 基因的不稳定性, 形成了基因断裂的基础, 单链发夹结构的形成导致两个断裂点的靠近, 最终导致了基因缺失。

关键词: DMD 基因; 连接片段; 内含子; 基因缺失

中图分类号: R746.2

文献标识码: A

文章编号: 1672-3554 (2005)01-0011-05

Relationship Between Repeat Elements and Instability and Gene Deletion of DMD Gene

SHENG Wen-li, CHEN Jiang-ying, ZHU Liang-fu, LIU Zhuo-lin

(Department of Neurology, The First Affiliated Hospital, SUN Yat-sen University, Guangzhou 510080, China)

Abstract: 【Objective】 To study the relationship between repeat elements and instability and gene deletion of DMD gene, we analyzed the breakpoints of junction fragments of deletion-typed DMD patients and the relationship between this gene intron structures and instability. 【Methods】 Deletion-type DMD patients were examined by multiplex polymerase chain reaction (mPCR). The breakpoints of junction fragments with 46 and 51 exon deletions were cloned and sequenced respectively. The structural features of available introns were analyzed by computer. 【Results】 Analysis of sequences of deletion-junction fragments of exon 46 and 51 showed all of the 5' and 3' breakpoints were located in repeat sequences. No small insertion, small deletion, or point mutation was located near the junction point. By analyzing of secondary structure of junction fragments with 46 and 51 exon deletions, it was demonstrated that all of junction fragments breakpoints were located at non-matching regions of single-strand hairpin. Overall 34.8% of total dystrophin intron length is accounted for repeated sequences, a high concentration of

收稿日期: 2004-05-13

基金项目: 国家自然科学基金资助项目 (39700048, 30271378); 广东省自然科学基金资助项目 (980066, 021866, 974151); 广东省科技攻关基金资助项目 (2002C30603); 广东省卫生厅医学基金资助项目 (A2002168)

作者简介: 盛文利 (1964-), 男, 湖北鄂州人, 博士, 副教授. E-mail: wenlisheng@hotmail.com

repetitive elements was found to be a key feature of many dystrophin introns. 【Conclusion】 Repeat elements in many DMD gene introns is the key structure basis and it represents the introns instability. On the foundation of the primary DNA sequences, single-strand hairpin forms, increasing the instability of the DMD gene, forming the base of the breaking of the gene. The formation of the single-strand hairpin results in close of the two breakpoints, then the deletion produces.

Key words: DMD gene; junction fragment; introns; gene deletion

[J SUN Yat-sen Univ(Med Sci),2005,26(1):11-15]

Duchenne 型肌营养不良症 (Duchenne/Becker muscular dystrophy, DMD/BMD) 是 X 连锁隐性遗传的神经肌肉疾病, 该病的致病基因是已知的人类最大的人类基因。该基因在 X 染色体上的长度跨度为 2 500 kb, 占人类基因组的 0.1%, 有 79 个外显子, 外显子的长度只占 DMD 基因长度的 0.6%。该基因的蛋白质产物——抗肌萎缩蛋白, 是血影蛋白系的一员, 该棒状形的蛋白质位于肌细胞膜上^[1,2]。DMD 基因的主要突变类型是基因缺失。约 60% 的 DMD/BMD 基因突变是各种大小的基因外显子缺失, 缺失外显子的频率各不相同^[3-5]。已经确定 DMD 基因外显子缺失存在两个热区, 分别位于该基因的中部和 5' 端。缺失热区的存在表明该基因在缺失热区存在特殊的基因结构, 并使缺失热区的基因结构不稳定而容易断裂。尽管已有的研究都表明, 所有的断裂点均位于内含子, 但对基因缺失的分子机制知之甚少。譬如, 为什么缺失热区的内含子容易断裂, 而非缺失热区内含子却不容易断裂? 内含子基因结构和序列特点与内含子不稳定性的关系如何? 内含子的极不稳定是通过什么方式导致基因断裂的? 据推测, DMD 基因内含子的不稳定性或许与基因缺失有关。现有的研究已表明, 内含子的长度与基因缺失无关。为了揭示 DMD 基因的缺失机制, 系统分析内含子和缺失断裂点的序列特点是很有帮助的^[6,7]。在这项研究中, 我们克隆了缺失型 DMD 患者的连接片段, 测定了断裂点的序列, 分析了 DMD 基因内含子结构与内含子不稳定性的关系。

1 材料和方法

1.1 载体和探针

pBeloBAC11 由 Shizuya 教授 (Division of Biology, California Institute of Technology) 惠赠。DMD cDNA47-4 探针由 Den Dunnen 教授 (Center for

Human and Clinical Genetics, Leiden University Medical Center) 惠赠。其他试剂购自上海生工生物工程有限公司。

1.2 第 46 和 51 号外显子缺失后连接片段的克隆和测序

1.2.1 鉴定 用多重引物 PCR 法鉴定 46 和 51 号外显子缺失的 DMD 病人, 病人均来源于我院门诊。HindⅢ 完全消化病人的基因组 DNA, DNA 片段与 pBeloBAC11 连接, 转化感受态细菌 DH101B, 用氯霉素和蓝白克隆筛选菌落。

1.2.2 筛选 筛选 46 和 51 号外显子缺失后形成的连接片段: 使用地高辛标记 (Roche 公司) 的 DMDcDNA47-4 探针, 与转移并固定在尼龙膜上细菌菌落杂交, 根据化学发光法鉴定并获得第 46 和 51 号外显子缺失后形成的连接片段的序列。

1.2.3 测序 连接片段的测序, 序列特点分析和缺失断裂点的确定: 通过测定 46 号和 51 号外显子缺失后形成的连接片段, 确定缺失断裂点, 并分析连接片段的重复序列和同源序列, 对 5' 和 3' 断裂点两侧各 100 bp 的序列进行二级结构分析。

1.3 内含子结构与内含子不稳定性分析

为了寻找 DMD 基因内含子结构与不稳定性的关系, 我们分析了 DMD 基因所有已经测定序列的内含子。内含子的序列资料从 GeneBank 获得。使用的分析软件可以在下列网站获得: <http://www.hgmp.mrc.ac.uk/Genome Web/Split/Repeat Masker>。

2 结果

2.1 连接片段的断裂点序列分析

第 46 号外显子缺失后, 5' 端断裂点位于 45 号内含子的 AT 富含区内。3' 端断裂点位于 46 号内含子 MER1 (medium reiterated frequency repeats,

MER1)类重复序列内(见图 1)。连接片段有两个 bp 的连接同源序列 ta,局部无小的缺失、插入和碱基的置换。第 51 号外显子缺失后,5'端断裂点位于 50 号内含子的 THE1 (transposon-like human element,THE1)序列内。3'端短裂点位于 51 号内含子 L2 (long interspersed nuclear elements,L2)序列内。连接片段有 3 个 bp 的连接同源序列 cta,局部无小的缺失、插入和碱基的置换。第 46、51 号外显子缺失后连接片段的断裂点的二级结构分析示断裂点均位于单链发夹环的非匹配区。第 46 号和 51 号外显子的测序图及断裂点局部单链二级结构图见文献^[8]。

2.2 DMD 基因内含子中的重复序列

在 DMD 基因内含子中发现高密度的重复序列,这是绝大多数内含子的关键结构特征,大约

34.8%的 DMD 基因内含子是由重复序列构成,而 LINE-1 (long interspersed nuclear elements,LINE-1)重复序列是分布在 DMD 基因内含子中的主要序列。令人感兴趣的是,每个内含子的重复序列的总长度与除去内含子的重复序列后剩余的内含子的长度相比较,存在显著的统计学意义 ($P < 0.001$)。

我们还比较分析了我们检测的 42 例缺失型 DMD 患者外显子缺失后上下游内含子的重复序列(表 1),发现了一些有意义的规律。除第 19 号外显子的缺失,其他外显子缺失其相应的上下游的内含子重复序列的比例为 13.7%~52.3%,由于这些内含子均在 DMD 基因的缺失热区,说明在这个范围内的重复序列容易导致外显子缺失。

表 1 中国人外显子缺失热区内含子的重复序列比例

Table 1 The ratio of repeat elements in introns of the Chinese exon-deletion hot-spot (%)

5'-end deletion hot-spot				Central deletion hot-spot			
Deleted Exons	No. of deletion	Upstream introns	Downstream introns	Deleted Exons	No. of deletion	Upstream introns	Downstream introns
PM	3	N	30.2	43	1	28.6	40.3
Pb	2	N	30.2	44	4	40.3	23.6
3	3	33.3	13.7	45	7	23.6	38.3
4	1	13.7	41.1	46	5	38.3	24.2
6	3	21	37.7	47	5	24.2	22.6
12	2	52.3	35.8	48	18	22.6	37.5
17	2	46.9	35.8	49	13	37.5	34.4
19	1	43.6	6.7	50	10	34.4	N
				51	3	N	37.5
				52	3	N	N

N: information unavailable

3 讨论

3.1 DMD 基因内含子不稳定性的结构和序列证据

根据国外和我们对 DMD 基因内含子的重复序列的计算机分析表明,重复序列是 DMD 基因内含子的重要结构特征,在内含子中呈高密度分布。约 34.8% 的 DMD 基因内含子是由重复序列构成,而 LINE-1 重复序列是 DMD 基因内含子分布最广泛的序列。我们对 DMD 基因第 45、46 和 51 号内含子分别进行分析,重复序列分别为 38.3%,

24.2%和 37.5%。在这些内含子中,第 46 号内含子的长度仅有 2 335 bp,但重复序列却占 24.2%。该发现支持我们提出的论点:即 DMD 基因内含子断裂的分子基础只是与分子结构尤其是分子序列相关,而不是与长度相关。因为 34.8%的 DMD 基因内含子是由重复序列构成,这就表明即使断裂点不是发生于某一单一的特定序列,可以预计大约 1/3 的断裂点序列将涉及重复序列。Suminaga 等人^[9]发现在 Alu 和 LINE-1 重复顺序之间的非同源重组可以导致 DMD 基因的缺失,说明重复序列是同源非等位重组的重要靶序列,该作者并且提出

在内含子中存在新的不稳定源的假设与我们提出的论点一致。

我们对 DMD 基因内含子全长的分析还发现两个重要的特点。首先,有 16 个内含子(分别是第 8、14、28、31、35、36、38、39、40、58、68、70、71、72、73、75 号内含子)中并无重复顺序;有 4 个内含子(分别是第 19、32、77、78 号内含子)其重复序列低于 7%,其重复序列所占的比率分别是 6.7%, 6.2%, 6.4%, 4.8%。由于 DMD 基因缺失突变存在两个热点区域,一个在基因中央部位,即 44~51 号外显子(占 70%),另一个在基因的 5' 端,即 2~20 号外显子(占 30%)。所以,这些内含子绝大多数分布于 DMD 基因的 3' 端的非缺失热区(第 68、70、71、72、73、75、77、78 号内含子);或者分布在该基因中央区两侧的非缺失热区(第 28、31、32、35、36、38、39、40、58 号内含子);或者分布在该基因的 5' 端缺失热区(第 8、14、19 号内含子),而这 3 个内含子虽然位于缺失热区但属于断裂频率低发的内含子。复习国内外文献和我们既往对临床确诊的 DMD 患者进行基因检测的结果均发现这些内含子中除第 8、14、19 号内含子(属于断裂低发内含子)外未见断裂点导致相应外显子缺失的报道^[10,11]。这说明 DMD 基因内含子中的重复顺序如果低于 6.7%,发生相应外显子缺失的几率非常小。

其次,有 9 个 DMD 基因内含子其重复序列所占的比率超过 40%(分别是第 4、7、11、13、16、18、25、43、60 号内含子,其重复序列分别是 41.1%、44.5%、52.3%、40.2%、46.9%、43.6%、48.8%、40.3%、40.1%)。这些内含子中除第 25、60 号内含子外,其他内含子均位于缺失热区,且第 4、7、11、13、16、18、43 号内含子国内外的研究资料都表明为断裂点的高聚集区,说明重复顺序的高低与内含子断裂的几率存在一定的相关性。第 25 和 60 号内含子尽管重复序列所占比例较高,但不能完全用上述规律解释,这些内含子中的断裂点导致的相应外显子缺失可能与重复序列形成的局部结构相关,譬如, Nobile 等人^[12]提出 TTTAAA 序列可以导致 DNA 弯曲并导致 DMD 基因内含子断裂支持该假说。

从表 1 也可以看出,除第 19 号外显子的缺

失,其他外显子缺失其相应的上下游的内含子重复序列的比例为 13.7%~52.3%,在这个范围内的重复序列容易导致外显子缺失。外显子缺失不仅与相应内含子的重复序列相关也跟重复序列是否可以形成影响 DNA 稳定的二级结构有关^[13]。

所以我们的结论是,DMD 基因内含子中重复序列的比率高低与内含子的易断裂性具有一定的相关性。如果内含子中重复序列的比例低于 6.7%,该内含子断裂的几率很小。如果内含子中重复序列的比例在 13.7%~52.3%,该内含子断裂的几率明显增多,且在 DMD 基因的两个缺失热区,其相应的内含子中的重复序列的比率较非缺失热区的内含子的比率要高($P < 0.05$)。但也有部分外显子的缺失其相应的内含子中重复序列的比率与外显子的缺失频率不相符,我们认为这种现象可能与断裂点局部的序列是否可以导致局部 DNA 弯曲或形成发夹结构有关,跟重复序列的比率无关。

3.2 重复序列及其二级结构是缺失突变的重要靶序列

迄今为止已经提出几个解释人类基因突变机制的模型。现在普遍接受的是在特定位置的 DNA 的易突变性很大程度上取决于特定位置上及其周围的 DNA 序列环境,比如初始结构和复杂性以及次级结构的形成,后者取决于 DNA 构象^[14]。

同源和非同源重组事件是基因组不稳定的根源。在 Alu 和 LINE-1 重复序列之间的同源重组事件已经在一些人类疾病中报道。在基因突变的两端分别涉及上述两种序列的非同源重组只是在有限的几种疾病突变中报道。然而,在我们的这项研究中,我们发现两个连接片段分别位于不同类型的重复序列。分析第 46 号外显子缺失后形成的连接片段显示 5' 端位于 45 号内含子的 AT 富集区,3' 端位于 MER1 序列。分析第 51 号外显子缺失后形成的连接片段显示 5' 端位于 50 号内含子的 THE1 序列,3' 端位于 L2 序列。这表明并没有唯一的序列导致 DNA 的断裂,在不同重复序列之间的非同源重组导致了 DNA 断裂。对第 7、47、48 号内含子的研究也发现了类似现象^[15,16]。通过分析第 46 和 51 号外显子缺失后形成的连接片段的二级结构表明所有的断裂点均位于单链发夹结构的非匹配区域,这些重复顺序形成的发夹结构容易导

致双链 DNA 断裂。

据推测在单一 DNA 链中发夹环状结构的形成,以及随后在环状区域的非匹配区受酶降解作用可以导致缺失^[7]。我们在断裂点水平的序列特点的发现与这个假设相符,就是缺失是由于 DNA 的双链结构断裂所产生,双链结构的断裂是在内在或外在因素的作用下,或者在正常 DNA 代谢过程中产生的,在修复过程中产生另一个断裂点,结果产生各种大小的 DNA 缺失^[7]。

因此,DMD 基因内含子大量重复序列一级结构的存在是内含子不稳定的基础。单链二级结构取决于序列的一级结构,在序列的一级结构基础上形成发夹环状结构,增加了 DMD 基因的不稳定性,而发夹环状结构的形成,使得断裂位点在空间上彼此靠近,导致缺失的产生。

参考文献:

- [1] Koenig M, Hoffman EP, Bertelson CJ, *et al.* Complete cloning of the Duchenne muscular dystrophy (DMD) cDNA and preliminary genomic organization of the DMD gene in normal and affected individuals[J]. *Cell*, 1987, 50(3): 509-17.
- [2] Koenig M, Monaco AP, Kunkel LM. The complete sequence of dystrophin predicts a rod-shaped cytoskeletal protein [J]. *Cell*, 1988, 53(2): 219-28.
- [3] Den Dunnen JT, Grootsholten PM, Bakker E, *et al.* Topography of Duchenne muscular dystrophy (DMD) gene: FIGE and cDNA analysis of 194 cases reveal 115 deletions and 13 duplications [J]. *Am J Hum Genet*, 1989, 45(6): 835-47.
- [4] Blonden LAJ, Grootsholten PM, Dunnen JT den, *et al.* 242 breakpoints in the 200-kb deletion prone p20 region of the DMD gene are widely spread [J]. *Genomics*, 1991, 10(3): 631-9.
- [5] 张成,柴建华,刘焯霖,等.抗肌萎缩蛋白基因 DNA 缺失机制的探讨[J]. *中山医科大学学报*, 1996, 17(4): 5-8.
- [6] 盛文利,柴建华,刘焯霖,等.假肥大型肌营养不良症第 50 和 51 号内含子的结构特点及 51 号外显子的缺失机制[J]. *中华医学杂志*, 1996, 76(11): 852-4.
- [7] 盛文利,柴建华,刘焯霖,等.DMD 基因第 50 和第 51 号内含子的克隆和测序[J]. *中山医科大学学报*,

1996, 17(4): 245-50.

- [8] 盛文利,陈江瑛,潘连跃,等.缺失型杜氏肌营养不良症缺失热区内含子断裂点分子结构特点的对比分析[J]. *中华遗传学杂志*, 2003, 20(5): 376-80.
- [9] Suminaga RY, Takeshima K, Yasuda N, *et al.* Non-homologous recombination between Alu and LINE-1 repeats caused a 430-kb deletion in the dystrophin gene: a novel source of genomic instability[J]. *J Hum Genet*, 2000, 45(6): 331-6.
- [10] 潘速跃,张成,盛文利,等.中国人抗肌营养不良蛋白基因缺失的分布特点 [J]. *中华神经科杂志*, 1999, 32(4): 22-3.
- [11] Danieli GA, Mioni F, Muller CR, *et al.* Patterns of deletions of the dystrophin gene in different European populations[J]. *Hum Genet*, 1993, 91(4): 342-6.
- [12] Nobile C, Toffolatti L, Rizzi F, *et al.* Analysis of 22 deletion breakpoints in dystrophin intron 49[J]. *Hum Genet*, 2002, 110(5): 418-21.
- [13] Sironi M, Pozzoli U, Cagliani R, *et al.* Relevance of sequence and structure elements for deletion events in the dystrophin gene major hot-spot [J]. *Hum Genet*, 2003, 112(3): 272-85.
- [14] Krawczak M, Cooper DN. Gene deletions causing human genetic disease: Mechanisms of mutagenesis and the role of the local DNA sequence environment[J]. *Hum Genet*, 1991, 86(5): 425-41.
- [15] McNaughton J, Cockburn DJ, Hughes G, *et al.* Is gene deletion in eukaryotes sequence-dependent? A study of nine deletion junctions and nineteen other deletion breakpoints in intron 7 of the human dystrophin gene[J]. *Gene*, 1998, 222(1): 41-51.
- [16] Toffolatti L, Cardazzo B, Nobile C, *et al.* Investigating the Mechanism of Chromosomal Deletion: Characterization of 39 Deletion Breakpoints in Introns 47 and 48 of the Human Dystrophin Gene [J]. *Genomics*, 2002, 80(5): 523-30.
- [17] Froelich-Ammon SJ, Gale KC, and Osheroff N. Site-specific cleavage of a DNA hairpin by topoisomerase II. DNA secondary structure as a determinant of enzyme recognition/cleavage[J]. *J Biol Chem*, 1994, 269(10): 7719-25.

(编辑 刘清海)