PIBB 生物化学与生物物理进展
Progress in Biochemistry and Biophysics
2007, 34(12): 1255~1259
www.pibb.ac.cn

真核生物基因组转录诱导融合基因的研究进展 *

吕聪颖 [1)**] 赵刚彬 [1)] 吕贯廷 [2)]
([1)] 南阳理工学院计算机科学与技术系，南阳 473004；[2)] 中国科学院北京基因组研究所，北京 101300)

摘要 真核生物的基因组由基因和基因间区组成. 基因转录时，从转录起始点开始到该基因的转录终止点结束，形成独立的转录单元. 然而有少量的文献表明，转录有时会通读基因间区，产生包含上游基因、基因间区和下游基因的融合基因转录本. 融合转录本经基因间剪接而成为有功能的成熟转录本.对真核生物转录诱导融合基因的基因间剪接方式、产生机制和意义进行了综述.

关键词 基因间区，转录，融合转录本
学科分类号 Q7

在真核生物基因组中，基因(特指蛋白编码基因)被不编码的基因间区(intergenic region)所隔开，形成独立的转录单元．基因在转录时，从特定的起始位点(transcription start site, TSS)开始，把编码区转录出来后，在该基因的终止位点(termination site)结束[1,2]．然而，近年来有零星的报道表明[3~7]，在转录过程中，RNA聚合酶Ⅱ(Pol Ⅱ)有时会通读(read through)终止位点和基因间区，在下游基因的转录终止位点结束，产生一个含有上游基因、基因间区和下游基因的融合转录本．剪接体(spliceosome)对其进行剪接时，把基因间区及其附近的位于上下游基因的UTR(untranslated region)甚至部分外显子作为一个大的内含子从转录本中切除，从而形成有功能的融合转录本(图1)，该转录

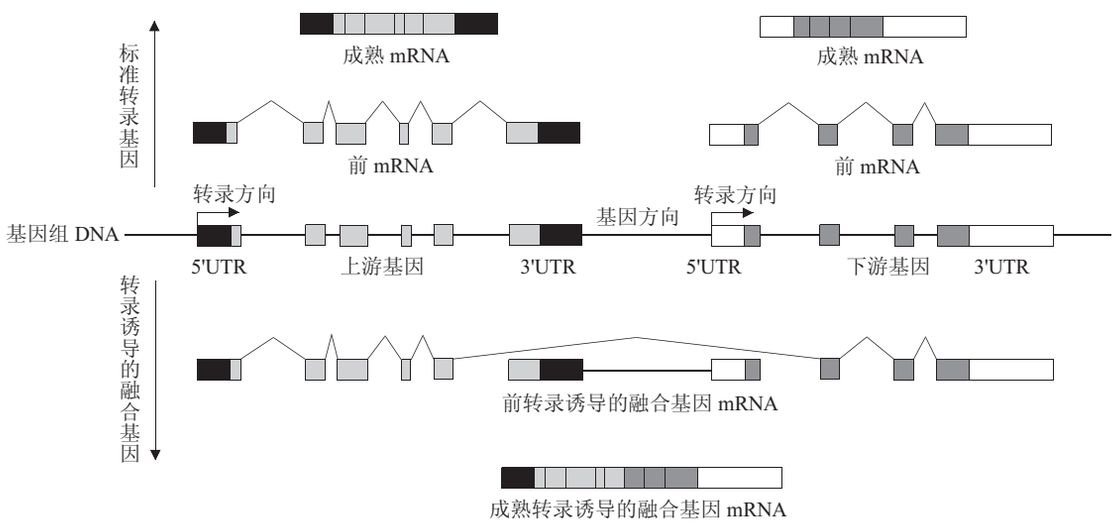

Fig. 1 Specific paradigm of transcription-mediated fusion genes

图1 转录诱导的融合基因模式图

该图上半部分显示在基因组上相邻的两个独立基因在转录时，形成两个独立的转录本并进一步成熟为各自独立的mRNA；下半部分显示这两个相邻的基因还可以在转录时产生一条大的含有两者外显子的融合转录本，剪接体把部分外显子和基因间区作为一个大的内含子从融合转录本中切除，形成成熟的融合基因转录本．

* 国家自然科学基金重点项目(60433020)和教育部重点项目(02090).
** 通讯联系人．Tel: 0377-62076312, E-mail: Alin0378@sohu.com
收稿日期：2007-04-09，接受日期： 2007-06-28

本有时还含有来自于基因间区的新外显子．为了和单个基因的剪接有所区别，这种发生在相邻基因间的剪接被称为基因间剪接(intergenic splicing)[3]．这种由于基因的转录而产生融合基因的现象被称为转录诱导的融合基因(transcription-mediated fusion genes)．

通常认为，基因的转录过程受到多重层次的严格调控，当PolⅡ到达转录终止位点时，就会在多种反式作用因子(*trans*-acting factors)与顺式作用元件(*cis*-acting elements)的相互作用下，从模板链上解离下来，转录活动就此终止，因此在转录过程中产生融合基因的频率极小，且表达量非常低，在哺乳动物细胞中很少发生[3, 8]．但根据我们前期对小鼠公共数据库中表达序列的分析结果及Parra等[9]和Akiva等[10]分别对人基因组中转录诱导融合基因的分析来看，一些融合基因的表达具有组织特异性，少数融合基因的表达量甚至高于看家基因，这提示着转录诱导融合基因不但可能是有功能的，且其产生也是一个受调控的过程．

目前发现，转录诱导融合基因在生物界可能是一个广泛存在的现象，已经在多个物种中发现，包括真菌[11]和酵母[12]．在哺乳动物中，转录诱导的融合基因不仅存在于人和小鼠[13]中，在牛[6]的基因组中也有发现．

1 转录诱导融合基因的基因间剪接方式

大多数的转录诱导融合基因的基因间剪接发生在上游基因倒数第二个($n-1$)外显子和下游基因第二个(+2)外显子之间．剪接体把上游基因最后一个外显子5'剪接位点(splicing donor，SD)和下游基因第一个外显子的3'剪接位点(splicing acceptor，SA)之间的区域(包括基因间区)作为一个大的内含子从融合转录本中切除．这种剪接方式通过除去上下游基因的3'UTR和5'UTR而去掉上游基因的翻译终止密码子和下游基因的翻译起始密码子，成为成熟的有功能的融合基因转录本[14]．在发生基因间剪接的融合基因中，约有55%的SD位于最后一个外显子，约80%的SA位于第一个外显子．由于基因间区内存在有潜在的SD和SA，在有些情况下，剪接体把基因间区的某些片段作为外显子与融合基因的其他外显子拼接在一起．在对小鼠基因组转录诱导融合基因的分析中，我们发现含有新外显子的融合基因约占所发现融合基因的11%．

2 转录诱导融合基因的特性

2.1 转录诱导融合基因的表达是一个受到调控的过程且具有独特的表达模式

人肿瘤坏死因子TNF配体家族的2个成员，TNFSF12和TNFSF13所产生的TNFSF12-TNFSF13融合基因编码的蛋白质主要在T细胞和单核细胞前体中发现，而在其他种类的细胞中较少出现；经研究证明，该融合蛋白是一种具有生物活性的受体，主要促进细胞的增殖[5]．人HLA1(HHLA-1)基因和otoconin 90(OC-90)基因所产生的HHLA1-OC90融合基因只发现于畸胎瘤组织中，而不存在于正常细胞中[15]．

2.2 转录诱导的融合基因还能改变来源基因所编码蛋白的特性或其细胞内定位

人泛素结合酶(ubiquitin-conjugating enzyme，UEV)基因编码的蛋白质位于哺乳动物细胞核内，具有DNA修复和促进c-FOS基因转录的功能．Kua基因编码的蛋白质位于内质网膜，具有脂肪酸羟化酶活性. 它们产生的融合基因Kua-UEV编码的蛋白质经试验证明位于内质网膜，而细胞核内未发现它们的融合蛋白[7]．

2.3 转录诱导的融合基因并不一定会产生融合蛋白

大多数的转录诱导融合基因在细胞内都可以产生有功能的蛋白质，比如Kua-UEV、TNFSF12-TNFSF13、CCL14-CCL15、MASK-BP3和NME1-NME2等等．然而真核细胞基因组产生的转录诱导融合基因有时会由于剪接的偏差或者新外显子的出现而导致其开放读码框中出现提前终止密码子(premature termination codons，PTC)，从而使得该融合基因被无义介导的mRNA降解机制(nonsense-mediated mRNA decay，NMD)[16]降解，而无法翻译产生融合蛋白．有时，假基因(pseudogenes)也会与下游基因产生融合基因，由于大部分假基因缺乏必需的翻译调控元件或者终止密码子的提前出现[17]，导致融合基因无法产生融合蛋白．

3 转录诱导融合基因的可能产生机制

3.1 大多数研究认为转录诱导融合基因是一个在转录过程中随机发生的事件

通过对人基因组中可以产生融合基因上下游基因的基因间区(上游基因最后一个外显子3'端至下游基因第一个外显子5'端)的距离进行分析，我们发

现，大多数的融合基因在基因组上距离较近，基因间区距离在20K(20 000 bp)以内的融合基因在人中为298个(总数为627个)．这提示着在基因组上距离越近的单独基因之间越有可能发生转录诱导的融合基因现象．然而人基因组中基因间区在20K以内的独立基因对总共有7 117个，但通过大规模分析公共表达数据所，我们发现基因间区在20K以内可产生融合基因的独立基因对只有294个．这在某种程度上提示，基因间区的距离与融合基因的发生之间不是一个正相关的关系，而且还提示着转录诱导融合基因的发生并不是一个随机的事件．

3.2 转录诱导融合基因的发生与上游基因Poly (A)信号强弱有关

Proudfoot等[2]认为基因的Poly(A)信号越强，其终止转录过程的效率就越高，因此Akiva[10]等推测融合基因的产生与上游基因Poly(A)信号强弱有关，即上游基因Poly(A)信号越弱，其与下游基因之间产生融合基因的可能性越高．然而通过人基因间区的Poly(A)进行分析时，我们发现，约有3 808个基因间区(约占总基因间区总数的15%)没有明显的Poly(A)信号，在发生融合基因的基因间区中，没有Poly(A)信号的有48个，这种现象提示我们，Poly(A)信号的强弱可能与转录诱导融合基因的产生之间也没有直接的正相关关系．

3.3 转录诱导融合基因的发生可能与位于基因间区的一些特殊序列 (motif) 有关

通过对融合基因和非融合基因的基因间区进行分析，我们在融合基因的基因间区筛选到一些非常保守的序列(资料未公开)，这些序列可能通过与某些特殊的蛋白质因子相互作用对基因的转录产生某些未知的影响，进而导致融合基因的发生．

4 转录诱导融合基因的意义

4.1 转录诱导的融合基因在进化上是生物产生新基因的重要步骤

有研究发现，在一个物种多蛋白质复合体或者同一代谢通路中发挥功能的2个独立的蛋白质在另一物种中有时会作为单独的一个蛋白即融合蛋白发挥作用[18]，且蛋白在融合后会发生结构和功能的适应性变化[19]．人琥珀酰Co-A转移酶是由大肠杆菌中乙酰CoA转移酶α和乙酰CoA转移酶β融合而成，酵母的拓扑异构酶Ⅱ是由大肠杆菌促旋酶A和促旋酶B融合而成．而且在对人和大鼠转录诱导融合基因进行分析的过程中，我们常常发现一些能发生融合的2个独立的相邻基因在其他物种中为融合的单独一条基因．因此在某种意义上，转录诱导的融合基因是在进化中产生新基因的一种重要机制[14]．

4.2 转录诱导的融合基因是增加蛋白质多样性的重要机制之一

参与融合的2个独立基因除了可以表达出各自独立的蛋白质外，其产生的融合基因转录本，通过基因间剪接除去上游基因3'端的终止密码子、下游基因5'端的起始密码子及基因间区内随机分布的潜在(cryptic)起始密码子和终止密码子[14]，从而产生有功能的新的含有两基因全部或者部分外显子的成熟转录本，该转录本进一步翻译合成新的蛋白质，它含有上游基因和下游基因的结构域成分．因此，TIFG又是一种产生复杂的多结构域蛋白、增加蛋白质多样性的机制，与选择性剪接(alternative splicing)、选择启动子(alternative promoter)和选择性ploy(A)位点(alternative polyadenylation)等可能具有同样重要的意义．

4.3 转录诱导融合基因的产生可能是一种通过独特的转录干扰方式调节来源基因的组织和发育阶段特异性表达的机制

转录诱导融合基因有时可以通过基因间剪接的方式除去来源基因的大部分外显子，翻译产生全新的蛋白质，改变来源基因编码蛋白的特性或细胞内定位，通过在转录过程中产生融合基因的方式，在某种程度上调节了来源基因自身的表达水平进而影响其功能的发挥，这种特殊的基因表达调控机制被称为转录干扰(transcriptional interference)[4, 20]．

4.4 转录诱导的融合基因可能与某些疾病的发生相关

据文献报道，一些转录诱导融合基因主要在肿瘤等病理组织中表达，而在正常的组织细胞中的表达量却非常低．如人多聚凝集素受体基因(multilectin receptor)DEC-205与C型凝集素基因DCL-1产生的DEC-205/DCL-1融合基因，高表达于Hodgkin和Reed-Strenberg细胞(Hodgkin淋巴瘤组织中的一小群形态独特的恶性细胞)中，而在同一组织的其他细胞中很少出现[8]，HHLA1-OC90融合基因只存在于畸胎瘤组织中，而不存在于正常细胞中，人MDS1-EVI1融合基因主要发生于髓性白血病(myeloid leukemia)细胞中，而在正常的组织中的表达量极低[21]．转录诱导融合基因的产生是一个受到复杂调控的过程，是多种因素共同左右的结果．某些融合基因在病理组织中的表达量较高可能与这

些组织中某种/某类特殊因子的产生或者高表达有关. 转录诱导融合基因的这种独特表达现象提示着它与肿瘤等疾病的发生可能存在着某种联系.

综上所述，转录诱导融合基因的产生可能是一种通过独特的方式对来源基因表达水平进行调控的机制，是在生物进化中产生新基因的重要方式，又是一种产生蛋白质多样性的重要机制，且其产生可能与肿瘤等疾病的发生有关，具有非常重要的理论研究和功能分析的价值. 目前，我们正在对哺乳动物转录诱导融合基因进行全基因组鉴定、组织特异性和发育阶段特异性、产生机制和进化来源等方面进行系统研究，结果会极大地加深我们对融合基因的转录和终止及其在疾病等相关方面的认识.

参 考 文 献

1 Zhao J, Hyman L, Moore C. Formation of mRNA 3' ends in eukaryotes: mechanism, regulation, and interrelationships with other steps in mRNA synthesis. Microbiol Mol Biol Rev, 1999, **63** (2): 405～445

2 Proudfoot N J, Furger A, Dye M J. Integrating mRNA processing with transcription. Cell, 2002, **108** (4): 501～512

3 Communi D, Suarez-Huerta N, Dussossoy D, *et al*. Cotranscription and intergenic splicing of human P2Y11 and SSF1 genes. J Biol Chem, 2001, **276** (19):16561～16566

4 Maeda K, Horikoshi T, Nakashima E, *et al*. MATN and LAPTM are parts of larger transcription units produced by Intergenic splicing: intergenic splicing may be a common phenomenon. DNA Res, 2005, **12** (5): 365～372

5 Pradet-Balade B, Medema J P, Lopez-Fraga M, *et al*. An endogenous hybrid mRNA encodes TWE-PRIL, a functional cell surface TWEAK-APRIL fusion protein. Embo J, 2002, **21** (21): 5711～5720

6 Roux M, Leveziel H, Amarger V. Cotranscription and intergenic splicing of the PPARG and TSEN2 genes in cattle. BMC Genomics, 2006, **7**: 71

7 Thomson T M, Lozano J J, Loukili N, *et al*. Fusion of the human gene for the polyubiquitination coeffector UEV1 with Kua, a newly identified gene. Genome Res, 2000, **10** (11): 1743～1756

8 Kato M, Khan S, Gonzalez N, *et al*. Hodgkin's lymphoma cell lines express a fusion protein encoded by intergenically spliced mRNA for the multilectin receptor DEC-205 (CD205) and a novel C-type lectin receptor DCL-1. J Biol Chem, 2003, **278** (36): 34035～34041

9 Parra G, Reymond A, Dabbouseh N, *et al*. Tandem chimerism as a means to increase protein complexity in the human genome. Genome Res, 2006, **16** (1): 37～44

10 Akiva P, Toporik A, Edelheit S, *et al*. Transcription-mediated gene fusion in the human genome. Genome Res, 2006, **16** (1): 30～36

11 Burns D M, Horn V, Paluh J, *et al*. Evolution of the tryptophan synthetase of fungi. Analysis of experimentally fused *Escherichia coli* tryptophan synthetase alpha and beta chains. J Biol Chem, 1990, **265** (4): 2060～2069

12 Kirschner L S, Stratakis C A. Structure of the human ubiquitin fusion gene Uba80 (RPS27a) and one of its pseudogenes. Biochem Biophys Res Commun, 2000, **270** (3): 1106～11010

13 Moore R C, Lee I Y, Silverman G L, *et al*. Ataxia in prion protein (PrP)-deficient mice is associated with upregulation of the novel PrP-like protein doppel. J Mol Biol, 1999, **292** (4): 797～817

14 Long M. A new function evolved from gene fusion. Genome Res, 2000, **10** (11): 1655～1657

15 Kowalski P E, Freeman J D, Mager D L. Intergenic splicing between a HERV-H endogenous retrovirus and two adjacent human genes. Genomics, 1999, **57** (3): 371～379

16 Baker K E, Parker R. Nonsense-mediated mRNA decay: terminating erroneous gene expression. Curr Opin Cell Biol, 2004, **16** (3): 293～299

17 Balakirev E S, Ayala F J. Pseudogenes: are they “junk” or functional DNA?. Annu Rev Genet, 2003, **37**: 123～151

18 Marcotte E M, Pellegrini M, Ng H L, *et al*. Detecting protein function and protein-protein interactions from genome sequences. Science, 1999, **285** (5428): 751～753

19 Lang D, Thoma R, Henn-Sax M, *et al*. Structural evidence for evolution of the beta/alpha barrel scaffold by gene duplication and fusion. Science, 2000, **289** (5484): 1546～1550

20 Petruk S, Sedkov Y, Riley K M, *et al*. Transcription of bxd noncoding RNAs promoted by trithorax represses Ubx in cis by transcriptional interference. Cell, 2006, **127** (6): 1209～1221

21 Fears S, Mathieu C, Zeleznik-Le N, *et al*. Intergenic splicing of MDS1 and EVI1 occurs in normal tissues as well as in myeloid leukemia and produces a new member of the PR domain family. Proc Natl Acad Sci USA, 1996, **93** (4): 1642～1647

Progress of The Transcription-mediated Fusion Genes in Eukaryotic Genomes*

LÜ Cong-Ying[1)**], Zhao Gang-Bin[1)], LÜ Guan-Ting[2)]

([1)] *College of Computer Science and Technology, Nanyang Institute of Technology, Nanyang* 473004,China;

[2)] *Beijing Institute of Genomics, The Chinese Academy of Sciences, Beijing* 101300, China)

Abstract Eukaryotic genomes are composed of individual genes and their intergenic regions. Traditionally, it has been held that the transcription of a gene always begins from the transcription start site and ends at the termination site, the gene is referred as an individual transcription unit. However, some sporadic studies indicate that transcription can sometimes read-through the intergenic region and generates a large fusion transcript which contains the upstream gene, intergenic region and the downstream adjacent gene. The fusion transcript becomes a mature functional transcript after intergenic splicing. The patterns of intergenic splicing, possible generation mechanism and its significance are summarized.

Key words intergenic region, transcription, fusion transcript

---

* This work was supported by a grant from The National Natural Science Foundation of China (60433020) and The Key Project of Chinese Ministry of Education (02090).

**Corresponding author. Tel: 86-377-62076312, E-mail:Alin0378@sohu.com

Received: April 9, 2007 Accepted: June 28, 2007